



digital



AUTOMATION

22. Leitkongress der Mess- und Automatisierungstechnik

AUTOMATION 2021

Navigating towards resilient Production

29. und 30. Juni 2021

VDI-BERICHTE

Herausgeber:

VDI Wissensforum GmbH

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter www.dnb.de abrufbar.

Bibliographic information published by the Deutsche Nationalbibliothek (German National Library)

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at www.dnb.de.

© VDI Verlag GmbH · Düsseldorf 2021

Alle Rechte vorbehalten, auch das des Nachdruckes, der Wiedergabe (Photokopie, Mikrokopie), der Speicherung in Datenverarbeitungsanlagen und der Übersetzung, auszugsweise oder vollständig.

Der VDI-Bericht, der die Vorträge der Tagung enthält, erscheint als nichtredigierter Manuskriptdruck.

Die einzelnen Beiträge geben die auf persönlichen Erkenntnissen beruhenden Ansichten und Erfahrungen der jeweiligen Vortragenden bzw. Autoren wieder. Printed in Germany.

ISSN 0083-5560

ISBN 978-3-18-092392-5

Inhalt

Vorwort 1
K. D. Bettenhausen, HARTING Führungsstiftung, Espelkamp;
F. Hanisch, Bayer AG, Leverkusen;
U. Jumar, ifak – Institut für Automation und Kommunikation e.V., Magdeburg

► **Fertigungsautomation – Digitaler Zwilling**

A Reference Architecture enabling Sensor Networks based on homogeneous AASs 5
V. Gowtham, A. Willner, Fraunhofer FOKUS, Berlin;
W. Pilar von Pilchau, J. Hähner, Universität Augsburg;
M. Riedl, Institut für Automation und Kommunikation e.V. (ifak), Magdeburg;
N.-S. Koutrakis, J. Polte, E. Uhlmann, Fraunhofer IPK, Berlin;
J. Tayub, V. Frey, Endress+Hauser GmbH + Co. KG, Maulburg

Digitale Zwillinge für Shopfloor-Management und -steuerung 17
A. Sohr, R. Rosen, J. C. Wehrstedt, J. Fischer, F. G. Listl, M. Schellander, Siemens AG, München;
R. Haar, Siemens AG, Köln

Eine Methodik zur Erstellung multi-dimensionaler Modelle des Digitalen Zwillings für automatisierte Produktionssysteme 29
D. Braun, M. Weyrich, Institut für Automatisierungstechnik und Softwaresysteme (IAS), Universität Stuttgart;
W. Schlögl, Siemens AG, Nürnberg

► **Fertigungsautomation – Virtuelle Welten entlang des Lebenszyklusses**

Augmented Reality für hybride Programmierung von Industrierobotern 43
L. Wulff, ICARUS Consulting GmbH, Lüneburg;
M. Brand, T. Schüppstuhl, Institut für Flugzeug-Produktionstechnik, Technische Universität Hamburg

Adaptive VR-Produktionsumgebungen für Evaluations- und Schulungstätigkeiten 55
J. Ulmer, S. Braun, J. Wollert, FH Aachen University of Applied Sciences, Aachen

► **Fertigungsautomation – Verwaltungsschalen in der Anwendung**

Modell einer Pick & Place-Anlage basierend auf Verwaltungsschalen	.65
C. Diedrich, Institut für Automation und Kommunikation e.V. (ifak), Magdeburg; A. Belyaev, T. Schröder, C. Urban, T. Werner, H. Pakala, OvGU Magdeburg	
Creation and Usage of Digital Twins in the Life Cycle of an Automation Solution	.75
D. Göllner, T. Klausmann, M. Kiele-Dunsche, H. Stichweh, Lenze SE, Aerzen	
Flexible Industrie 4.0 Verwaltungsschale für dynamische IIoT-Systeme – Zeitreihen in der Verwaltungsschale	.87
F. Hilbert, S. Soler Perez Olaya, Professur für Prozesskommunikation, Technische Universität Dresden	

► **Fertigungsautomation – Moderne Bereitstellungssysteme von Werkstücken**

Designprinzipien für den Modul- und Serviceentwurf in modularen Logistikanlagen	101
M. Blumenstein, A. Fay, Helmut-Schmidt-Universität, Hamburg; K. Gryczycha, S. Lier, Fachhochschule Südwestfalen, Meschede; C. Kotsch, BASF SE, Ludwigshafen; F. Borgmann, Fraunhofer-Institut für Materialfluss u. Logistik, Dortmund; N. Austermann, BEUMER Maschinenfabrik GmbH & Co. KG, Beckum	
Bereitstellungsmethoden von ungeordneten Werkstücken in den automobilen Produktionsprozess.	117
M. Sarna, J. Weist, A. Tegtmeier, Volkswagen AG, Wolfsburg; A. Lüder, Otto-von-Guericke Universität, Magdeburg	

► **Gebäudeautomation**

BIM im Betrieb durch lebenszyklusübergreifende Verfügbarkeit von Anlagendaten auf Basis von I4.0-Verwaltungsschalen	131
M. Ostermeier, N. Maisch, M. Both, R. Ulmer, J. Müller, Technische Hochschule Köln	

► Industrielle Kommunikation & 5G – Kommunikation

Estimation of 5G Cell Capacity in Industrial Environments 141

L. Underberg, S. Willmann, Institut für Automation und Kommunikation, Magdeburg;
S. Kerschbaum, S. Höme, Siemens AG Digital Industries, Nürnberg

A look inside functional safety over 5G. 155

D. Patel, Ericsson GmbH, Herzogenrath;
A. Hoell, SICK AG;
T. Schildknecht, E. Eberhardt, Schildknecht AG, V. Seelman, FIR, Aachen;
P. Becker, M. Stumpert, Ericsson GmbH, Herzogenrath;

Bridging the Gap in Digital Systems using Automated Model Transformation Concepts 169

A. Karaagac, J. O. Schmitt, D. O. Schulz, ABB Corporate Research, Ladenburg

► Industrielle Kommunikation & 5G – 5G Testbeds

Das Industrial Radio Lab Germany – Vier Standorte, eine Mission 183

N. Franchi, P. Schulz, G. Fettweis, Vodafone Stiftungslehrstuhl Mobile Nachrichtensysteme,
Technische Universität Dresden;
A. Weinand, D. Lindenschmitt, H. Schotten, Lehrstuhl für Funkkommunikation und Navigation,
Technische Universität Kaiserslautern;
N. Bulk, F. Bittner, A. Dekorsy, Arbeitsbereich Nachrichtentechnik, Universität Bremen;
L. Underberg, L. Rauchhaupt, Institut für Automation und Kommunikation, Magdeburg

Plattform zur Untersuchung industrieller Funkkommunikation für die Intralogistik 197

G. Persico, A. Fischer, DEMAG Cranes & Components GmbH, Wetter;
M. Riedl, D. Tran, Institut für Automation und Kommunikation e.V., Magdeburg

► Prozessautomation – Digitalisierung vom Feld bis in die Leitebene

Ethernet-APL für hochverfügbare Sicherheitsanwendungen – Vorteile für den Lebenszyklus von Prozessanlagen.	209
M. Risser, A. Ziegler, P. Salzmann, M. Roser, BASF SE, Ludwigshafen am Rhein	

I4.0-Verwaltungsschalen als Wegbereiter für optimierte Engineering- und Betriebsprozesse in der prozessnahen Gebäudeautomation.	221
M. Krüttgen, J. Müller, Technische Hochschule Köln	

► Prozessautomation – Prozessführung

Adaptive Massenstromregelung einer Pfannen-Gießmaschine mittels virtueller Sensorik . . .	235
P. Ochudlo, S. Stemmler, D. Abel, Institut für Regelungstechnik (IRT) der RWTH Aachen University	

Maximizing Efficiency and Transparency in Batch Processing with Simulation-driven Predictive Decision Support.	249
C. Sonntag, D. Wolff, D. Bleidorn, E. Balling, INOSIM Software GmbH, Dortmund; F. Marx, CSL Behring GmbH, Marburg	

► Prozessautomation – Modulare Produktion

Vergleich und Zusammenführung des Open Process Automation Standard (OPAS) und der VDI/VDE/NAMUR 2658.	261
M. Blumenstein, A. Fay, Helmut-Schmidt-Universität, Hamburg; A. Stutz, M. Maurmaier, Siemens AG, Karlsruhe; M. Barth, Hochschule Pforzheim, Pforzheim	

Rezeptbasierte Orchestrierung einer modularen kontinuierlichen Prozessanlage – Ergebnisse und Erkenntnisse aus dem Forschungsprojekt ENPRO2-ORCA.	275
A. Stutz, M. Maurmaier, Siemens AG, Karlsruhe; M. Eckert, H. Mannheim, H. E. Akar, Merck KGaA, Darmstadt	

Integriertes Vorgehensmodell für die Planung und Realisierung von modularen Plug & Produce Systemen in der Pharmaproduktion	293
A. Jebari, H. Mettler, Exyte Central Europe GmbH, Stuttgart	

► **Prozessautomation – Modellbasiertes Engineering**

Modellgetriebene Entwicklung für service-orientierte Automatisierungssysteme in der Prozessindustrie 305

M. Waßmer, Siemens AG, Karlsruhe, Karlsruher Institut für Technologie, Karlsruhe;
M. Maurmaier, A. Stutz, Siemens AG, Karlsruhe;
V. Pazmino Betancourt, B. Liu, Forschungszentrum Informatik, Karlsruhe;
J. Becker, Karlsruher Institut für Technologie, Karlsruhe, Forschungszentrum Informatik, Karlsruhe

Durchgängig konsistente Darstellung von Verriegelungen im Engineering prozesstechnischer Anlagen 317

F. El Sakka, A. Fay, Helmut-Schmidt-Universität, Hamburg

Testen 4.0 in der Automatisierungstechnik: Agiles modellbasiertes Testen vernetzter Systeme und Komponenten 335

A. Löcklin, M. Müller, M. Weyrich, Institut für Automatisierungstechnik und Softwaresysteme, Universität Stuttgart;
C. Kotsch, K. Krüning, BASF SE, Ludwigshafen am Rhein;
M. Rentschler, Balluff GmbH, Neuhausen auf den Fildern;
C. Ebert, Vector Consulting Services GmbH, Stuttgart

► **Methoden und Synergien – Security**

NAMUR Open Architecture – Security for NOA 353

M. Runde, BASF, Ludwigshafen;
C. Barth, FESTO, Denkendorf;
G. Wartmann, WSC, Hockenheim;
M. Brcic, Endress+Hauser, Reinach, Schweiz

Ontologiebasiertes Security Management 365

C. Siegwart, F. Scherhag M.Sc., ZeMA gGmbH, Saarbrücken;
K. Pervözl, Fraunhofer IAIS, Sankt Augustin,
G. Frey, Universität des Saarlandes, Saarbrücken

Dezentrale Authentifizierung und Autorisierung proaktiver Industrie 4.0-Verwaltungsschalen am Beispiel des IOTA-basierten Industrie-Marktplatzes. 377

A. Dogan, A. Schnakenbeck, M. O. Rögner, A. Fay, Helmut-Schmidt-Universität, Hamburg

ECLASS als semantische Basis eines dezentralen Registers für die Identifikation von Diensten von Industrie 4.0-Komponenten – Distributed Ledger-basierte Infrastruktur für Verwaltungsschalen 391

A. Dogan, A. Fay, Helmut-Schmidt-Universität, Hamburg;

J. A. Nagel, G. Rijo, Neoception GmbH, Mannheim;

A. Belyaev, C. Diedrich, Otto-von-Guericke-Universität, Magdeburg;

C. Block, ECLASS Head Office, Köln;

A. Bondza, Pepperl+Fuchs SE, Mannheim

► **Methoden und Synergien – Verwaltungsschale**

Ein Konzept für Produkt-Verwaltungsschalen zur produktbezogenen Integration von Produktionsdaten für die Qualitätsoptimierung. 405

M. Thies, T. Miny, T. Kleinert, RWTH Aachen University, Aachen

Integration von Module Type Package und Industrie 4.0 Verwaltungsschale 421

S. Grüner, M. Hoernicke, ABB AG Forschungszentrum, Ladenburg;

A. Kehl, C. Barth, M. Freund, M. Hoffmeister, Festo SE & Co. KG, Esslingen;

T. Klausmann, Lenze SE, Aerden

► **Methoden und Synergien – Digitaler Zwilling**

Das AutomationML Komponentenmodell 433

L. Hundt, inpro, Berlin;

M. Wiegand, Festo SE & Co. KG, Esslingen;

A. Lüder, Otto von Guericke Universität, Magdeburg;

T. Meyer, Volkswagen AG, Wolfsburg

Orchestration of Co-Simulations with FMU Integration as a Service. 447

N. Schoch, J. Schlake, S. Grüner, J. Heuschkel, ABB Forschungszentrum, Ladenburg

► Künstliche Intelligenz & Autonome Systeme – Produktionsoptimierung

Assistenzsysteme der nächsten Generation für Produktionssysteme.	461
J. Fischer, R. Rosen, F. G. Listl, A. Sohr, J. C. Wehrstedt, Siemens AG, München	
Automatic Detection of Start and End Times of Batch Phases for the Process Industry	477
S. Merkelbach, L. Urbas, J. Gatter, Professur für Prozessleittechnik/AG Systemverfahrens- technik, Technische Universität Dresden;	
R. Tan, M. Gärtler, R. Gedda, ABB Corporate Research Center Germany;	
F. D. Böhner, Bayer AG Leverkusen	
AIQIIA: General AI Quality Inspector for Industrial Applications	491
D. Díez Álvarez, Mercedes-Benz AG, Stuttgart;	
U. Berger, BTU Cottbus-Senftenberg, Cottbus	
Scaling a Reinforcement Learning Approach for Online Job Shop Scheduling in Flexible Manufacturing Systems to Various Products.	505
S. Baer, Amazon;	
S. Pol, D. Turner, F. Baer, Siemens AG;	
T. Meisen, Universität Wuppertal	
Betriebsparallele Adaption einer Co-Simulation für diskrete Fertigungsanlagen	517
C. Härle, M. Barth, Hochschule Pforzheim;	
A. Fay, Helmut-Schmidt-Universität Hamburg	

► Künstliche Intelligenz & Autonome Systeme – Materialflusssysteme

Lokalisierung von intelligenten Werkstückträgern innerhalb einer vollautomatisierten Fertigungsanlage mittels Bluetooth 5.1	535
B. Schulte, IWN GmbH & Co. KG, Bielefeld;	
H. Fast, H. Flatt, Fraunhofer IOSB-INA, Lemgo	

► Künstliche Intelligenz & Autonome Systeme – Datenanalyse

Compression and Fault Diagnosis of High Frequency Sensor Data	545
I. Amihai, R. Gitzel, A. Boyaci, ABB, Ladenburg	

► Künstliche Intelligenz & Autonome Systeme – Wissensmodellierung

Zustandsüberwachung und Klassifizierung manueller Montageprozesse mit selbstadaptiven Modellen.	557
P. Sehr, N. Moriz, Institut für industrielle Informationstechnik (inIT), Technische Hochschule Ostwestfalen-Lippe, Lemgo	

Semantische Interoperabilität durch Natural Language – Processing als Basis für Self-X-Fähigkeiten von Verwaltungsschalen in semantisch heterogenen Asset-Netzwerken.	571
M. Both, N. Maisch, J. Müller, Technische Hochschule Köln, Köln	

Automatisierte Datenintegration für den Fehleranalyseprozess von Halbleiterbauelementen mithilfe von Ontologien und Graphen	585
S. Kamm, N. Jazdi, M. Weyrich, Institut für Automatisierungstechnik und Softwaresysteme (IAS) der Universität Stuttgart	

► Digitale Geschäftsmodelle

Wie IIoT, Predictive Maintenance und proaktiver Support den Einsatz von Analysatoren und komplexer Sensorik ermöglichen	599
M. Nick, F. Evola, Endress+Hauser Group Services; L. Bondzio, A. Gahr, A. Lemke, S. Schmidt, Endress+Hauser Liquid Analysis	

Zögerliche Digitalisierung der Geschäftsmodelle – Ergebnisse einer repräsentativen Studie	611
G. Sames, Technische Hochschule Mittelhessen, Gießen	

Smart Resilience Services for Industrial Production	621
S. Janzen, N. Öksüz, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Saarbrücken; J. Sporkmann, M. Schlappa, WHU – Otto Beisheim School of Management, Vallendar; J. Gerhard, Feintool System Parts Jena GmbH, Jena; L. Ortjohann, Machine Tools and Production Engineering, RWTH Aachen University, Aachen; P. A. Becker, deZem GmbH, Berlin	

Flexible On-Premises Datenanalyse für KMU	637
G. Steinhagen, M. Banf, fabforce GmbH & Co. KG, Netphen	

Towards an Artificial Intelligence based Approach for Manufacturing Resilience	649
N. Öksüz, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Saarbrücken; S. Bouschery, Institute for Technology and Innovation Management (TIM), RWTH Aachen University; M. Schlappa, J. Sporkmann, WHU - Otto Beisheim School of Management, Vallendar; M. Unterberg, Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen	

► **Posterbeiträge**

Nutzung der BaSys 4.2 Middleware – Lösung in der Umgebung von Industriesteuerungen . . .	665
S. Schäfer, D. Schöttke, T. Kämpfe, K. Ralinovski, Hochschule für Technik und Wirtschaft Berlin; H. Klemppnow, KleRo GmbH Roboterautomation	

Refaktorisierung von Steuerungssoftware cyber-physischer Produktionssysteme – Potentiale und Nutzen	679
E.-M. Neumann, B. Vogel-Heuser, J. Fischer, Lehrstuhl AIS, Technische Universität München; P. Puntel Schmidt, M. Witte, Siemens AG, Nürnberg; N. Langer, Brückner Maschinenbau GmbH & Co. KG, Siegsdorf; L. Becerri, ID-Consult GmbH, München; M. Bengel, teamtechnik Maschinen und Anlagen GmbH, Freiberg am Neckar	

Interoperable Interaktionen semantisch heterogener Industrie 4.0 Komponenten auf Basis einer Capability Ontologie für die Technische Gebäudeausrüstung.	691
N. Maisch, M. Both, J. Müller, Technische Hochschule Köln, Köln	

Auswirkung und Folgen Technischer Schuld in mechatronischen Systemen und daraus resultierende Anforderungen.	705
F. Bi, B. Vogel-Heuser, Technische Universität München AIS, München; H. Dettmering, Prozesswerk, München	

Hygienische Qualitätsparameter für die Prozessautomation – Vom Labor zum Prozess	713
H. Tippe, Endress+Hauser Group Services, Weil am Rhein; A. Gahr, Endress+Hauser Liquid Analysis, Gerlingen; R. Möller, Analytik Jena, Jena	

OPC-UA-Applikationserkennung in Software-defined Networks für Industrie 4.0 725

D. Schulze-Zipper, Institut für Automation und Kommunikation e.V. (ifak), Magdeburg;

W. Mahnke, ascolab GmbH, Kalchreuth;

D. Gerlach, J. Neubauer, SBSK GmbH & Co. KG, Schönebeck

Increased reliability and safety in Coriolis flowmeter operation by monitoring its physical response to process conditions 737

D. Persson, Endress+Hauser Group Services, Reinach, Switzerland;

J. Pohl, Endress+Hauser Flow, Reinach, Switzerland;

A. Rieder, M. Wiesmann, Endress+Hauser Flow, Freising

Selbstorganisation und Co-Management verteilter Automatisierungssysteme mit dem Kommunikationssystem 751

T. Zeun, A. Frotzscher, Fraunhofer Institut für Integrierte Schaltungen IIS, Institutsteil

Entwicklung Adaptiver Systeme EAS, Dresden;

A. Nicklas, Lehrstuhl für Robotik und Systemintelligenz, Technische Universität München;

E. Schmitt, Vodafone Chair Mobile Communications Systems, Technische Universität Dresden;

D. Schulze-Zipper, H. Zipper, Institut für Automation und Kommunikation e.V. (ifak), Magdeburg

Comparison of different spike detection methods in Hall-Héroult cells. 763

R. Kremser, N. Grabowski, D. Tutsch, Bergische Universität Wuppertal, Wuppertal;

R. Düssel, A. Mulder, TRIMET Aluminium SE, Essen

**VDI NACHRICHTEN RECRUITING TAG –
DEUTSCHLANDS FÜHRENDE
KARRIEREMESSE FÜR INGENIEURE.****Ideal für Ihre erfolgreiche Jobsuche:**

- Renommierete Unternehmen
- Direkter Kontakt mit Entscheidern
- Karriereberatung und -vorträge
- Job Board

Wir machen Ingenieurkarrieren. Vor Ort. Und Online.

**VDI nachrichten Recruiting Tag und VDI nachrichten Recruiting Tag Online. Deutschlands führende
Karrieremessen für Ingenieure und IT-Ingenieure.**

Für alle Studierenden der Ingenieurwissenschaften, Absolventen und Young Professionals ein absolutes Muss. Knüpfen Sie Kontakte zu renommierten Unternehmen und sprechen Sie direkt mit Entscheidern aus den Fachabteilungen. Viele Serviceangebote wie Karriereberatung und -vorträge unterstützen Sie bei Ihrem erfolgreichen Einstieg ins Berufsleben.

**Jetzt informieren und kostenfrei teilnehmen: www.ingenieur.de/recruitingtag****Mehr Informationen?**

Silvia Becker, Telefon: +49 211 6188-170

Franziska Opitz, Telefon: +49 211 6188-377

recruiting@vdi-nachrichten.com**VDI nachrichten**
recruiting tag

Vorwort

Die Corona-Pandemie hat uns einmal mehr vor Augen geführt, wie anfällig und verletzlich unsere Gesellschaft und unsere Wirtschaft gegenüber großen Veränderungen sind.

Sobald einfache Modelle und lineare Prognosen nicht mehr greifen, sind unsere kostenoptimierten, weltweit vernetzten Produktionsnetzwerke schnell am Ende ihrer Leistungsfähigkeit.

In unserer Fachdisziplin bezeichnen wir komplexe technische Systeme als resilient, wenn sie auch beim Auftreten innerer und äußerer Ausfälle und Störungen die angeforderten Systemleistungen aufrechterhalten. Hier sind alle Ingenieurdisziplinen und insbesondere die Automation gefragt, denn eines ist offensichtlich: Die Herausforderungen können wir nur gemeinsam mit einem „Blick über den Tellerrand“ meistern.

Unser jährlicher Kongress AUTOMATION ist technologisch geprägt. Das ist Teil des Erfolgsrezepts, welches theoretische Grundlagen, aktuelle Technologieentwicklungen und die Anwendungen in der Produktion – gleichermaßen in der Fertigungs- wie in der Prozessindustrie – vereint. Gerade in Zeiten großer Herausforderungen sind Austausch, Orientierung und Navigation unserer Community besonders notwendig und gefordert. Zugleich gilt es herauszuarbeiten, welche unverzichtbaren Beiträge die Automation ab sofort leisten kann und muss, um die nicht weniger aktuellen und wichtigen Klimaschutzziele zu erreichen.

Wir freuen uns, Sie auf der digitalen AUTOMATION 2021 willkommen zu heißen.

Die Kongressleiter



Dr.-Ing. Kurt D. Bettenhausen,
Vorstand,
HARTING Führungsstiftung,
Espelkamp



Dr.-Ing. Felix Hanisch,
Head of Industrial
Automation, Bayer AG,
Leverkusen



Prof. Dr.-Ing. Ulrich Jumar,
Institutsleiter, ifak – Institut
für Automation und
Kommunikation e. V.
Magdeburg

Programmausschuss

Dipl.-Ing. Heiko Adamczyk, Senior Expert Industrial Security, TÜV SÜD Product Services GmbH, München

Schirin Baer, Area Manager, Amazon, Frankenthal

Prof. Dr.-Ing. Mike Barth, IOS³ – Institut für Smart Systems und Services, Hochschule Pforzheim, Chefredakteur „science“ atp magazin

Prof. Dr.-Ing. Martin Becker, Institut für Gebäude- und Energiesysteme, Hochschule Biberach

Prof. Dr. Joachim Birk, Vice President, Executive Expert of Automation Technology, BASF SE, Ludwigshafen

Prof. Dr.-Ing. Christian Diedrich, Lehrstuhl für Automatisierungstechnik (IFAT), Otto-von-Guericke-Universität

Dr.-Ing. Dagmar Dirzus, Geschäftsführerin VDI/VDE-GMA, VDI e. V., Düsseldorf

Univ.-Prof. Dr.-Ing. Alexander Fay, Leiter des Instituts für Automatisierungstechnik, Helmut-Schmidt-Universität/Universität der Bundeswehr

Dr. Helmut Figalist, Leiter Technologie und Innovation, Industry Automation, Siemens AG, Nürnberg

Univ.-Prof. Dr.-Ing. Georg Frey, Lehrstuhl für Automatisierungs- und Energiesysteme, Universität des Saarlandes, Saarbrücken

Hans Joachim Fröhlich, Director Technology and Portfolio, Endress+Hauser Group Services AG, Reinach, Schweiz

Dr.-Ing. Stefan Gehlen, Geschäftsführer, VMT Vision Machine Technic Bildverarbeitungssysteme GmbH, Mannheim

Dr. Martin Gerlach, Head of Process Analytical Technologies, Engineering & Technology, Bayer AG, Leverkusen

Prof. Dr.-Ing. Iris Gräßler, Fachgruppe Produktentstehung, Universität Paderborn

Prof. Dr.-Ing. Michael Heizmann, Institutsleiter, Institut für Industrielle Informationstechnik – IIIT, Karlsruher Institut für Technologie (KIT)

Dipl.-Ing. Ulrich Hemen, Head of Market Management Industry & Process, WAGO Kontakttechnik GmbH & Co. KG, Minden

Dipl.-Ing. Tim Henrichs, Marketing Manager Europe (Chemical Industry), Center of Excellence Chemicals, Yokogawa Deutschland GmbH, Ratingen

Prof. Dr.-Ing. i.R. Hartmut Hensel, ehemals Fachbereich Automatisierung und Informatik, Hochschule Harz, Wernigerode

Prof. Dr.-Ing. Jörg Kiesbauer, Leiter des Instituts für Mechanik und Mechatronische Systeme (im2s), Hochschule Darmstadt

Dr.-Ing. Niels Kiupel, Production and Technology, Automation/Industry 4.0, Evonik Resource Efficiency GmbH, Marl

Prof. Dr.-Ing. Tobias Kleinert, Lehrstuhl für Prozessleittechnik, RWTH Aachen

Gunther Koschnick, Geschäftsführer Fachverband Automation, ZVEI e. V., Frankfurt/Main

Univ.-Prof. Dr.-Ing. Bernd Kühlenkötter, Lehrstuhl für Produktionssysteme (LPS), Fakultät für Maschinenbau, Ruhr-Universität Bochum

Dr. Christine Maul, Head of Advanced Process Control, Global Process Control Technology, Covestro Deutschland AG, Leverkusen

Prof. Dr.-Ing. habil. Thomas Meurer, Leiter des Lehrstuhls für Regelungstechnik, Christian-Albrechts-Universität zu Kiel

Dipl.-Ing. Martin Müller, Leiter Business Unit I/O and Networks, Phoenix Contact Electronics GmbH, Bad Pyrmont

Dr. Thomas Paulus, Global Executive Officer – CDO, Digital Transformation, KSB SE & Co. KGaA, Frankenthal

Dr. Thorsten Pötter, Chief Digital Officer, SAMSON AG, Frankfurt am Main

Dr.-Ing. Lutz Rauchhaupt, Stellv. Geschäftsfeldleiter IKT & Automation, Senior Engineer Funk in der Automation, ifak e. V., Magdeburg

Dr.-Ing. Eckhard Roos, Leiter Prozessautomation, Festo AG & Co. KG, Esslingen/Neckar

Prof. Dr.-Ing. Martin Ruskowski, Vorstandsvorsitzender der Technologie- Initiative SmartFactory KL e. V., Kaiserslautern

Dipl.-Kfm. Felix Seibl, Geschäftsführer, ZVEI-FB Messtechnik und Prozessautomatisierung, ZVEI e. V., Frankfurt/Main

Prof. Dr.-Ing. habil. Olaf Simanski, Fachgebiet Automatisierungstechnik, Hochschule Wismar

Prof. Dr.-Ing. Leon Urbas, Professur für Prozessleittechnik und AG Systemverfahrenstechnik, Technische Universität Dresden

Nils Weber, Geschäftsführer NAMUR e. V., Leverkusen

Prof. Dr.-Ing. Dr. h. c. Michael Weyrich, Institutsleitung, Institut für Automatisierungstechnik und Software-systeme, Universität Stuttgart

SPONSOREN

Wir danken unseren Sponsoren für die freundliche Unterstützung.



www.pepperl-fuchs.com



www.wago.com

Veranstalter:
VDI Wissensforum GmbH

A Reference Architecture enabling Sensor Networks based on homogeneous AASs

V. Gowtham, Dr.-Ing **A. Willner**, Fraunhofer FOKUS, Berlin;
W. Pilar von Pilchau, Prof. Dr. **Jörg Hähner**, Universität Augsburg;
M. Riedl, ifak e.V., Magdeburg;
N.-S. Koutrakis, Dr.-Ing. **J. Polte**,
Prof. Dr. h. c. Dr.-Ing. **E. Uhlmann**, Fraunhofer IPK, Berlin;
Dr. **J. Tayub**, Dr. **V. Frey**, Endress+Hauser GmbH + Co. KG, Maulburg

Kurzfassung

Sensoren, speziell Sensornetzwerke, sind in Verbindung mit Virtualisierung elementare Bestandteile, wenn man im Kontext der Fertigungs- bzw. Prozessindustrie von einer digitalen Fabrik spricht. Wichtige Konzepte sind hierbei Asset Administration Shells (AAS) und Digital Twins. Unterschiedliche Interpretationen und dadurch resultierende unterschiedliche Implementierungen machen Standardisierungskonzepte unerlässlich. Die Plattform Industrie 4.0 hat den AAS Standard maßgeblich geprägt und resultierte in einer Vielzahl von wissenschaftlicher Literatur und Referenzimplementierungen. Im Kontrast zu bestehenden Architekturen liegt der Fokus, der in diesem Beitrag vorgestellten Architektur, auf homogenen AASs in der Domäne der Sensornetzwerke. Wichtige Aspekte sind die Harmonisierung von unterschiedlichen Virtualisierungskonzepten, die Betrachtung einzelner Komponenten als Blackbox mit Input/Output und die Fähigkeit, von dieser sich mittels Submodels selbst zu beschreiben. Der vorliegende Beitrag präsentiert eine abstrakte Architektur und zeigt wichtige Konzepte anhand einer beispielhaften Implementierung anhand konkreter Komponenten.

1. Motivation - Sensor networks in today's manufacturing and process industry

The goal of Plattform Industrie 4.0¹ (PI4.0) is to drive the formation of an automated factory in cooperation with Cyber Physical Systems (CPS). PI4.0, in other words, aims to close the gap between the contrasting worlds of Operation Technology (OT) and the Information Technology (IT). The field of OT predominantly addresses the field of manufacturing and process industry concerning the operation of manufacturing assets, for example machine tools, and the application of process knowledge. It intends to automate such processes through the means of computing systems keeping in view of safety, reliability and economy. The IT intends to

¹ www.plattform-i40.de

make OT more efficient and transparent by using and adapting concepts of data processing, to enable new business models.

To this end, the concept of a Digital Twin (DT) and an Asset Administration Shell (AAS) aims to create a virtual representation/twin of a physical asset, such that a DT or an AAS can be used by professionals on the IT side to comprehend field level complexities of the machine. In a control loop sequence comprising of "sense-decide-actuate", the sensing stage requires that sensors measuring physical properties are reliable and available. The reliability of a sensor can be established by a calibration which is traceable to a known reference and based on Système International (SI) units. As a result, a measurement uncertainty can be associated with the measured values from that sensor. This is one of the core principles in the field of metrology/measurement science. Establishing reliability of CPSs thus requires integrating metrological principles in the data lifecycle from the physical measurement to the digital representative and the operational decisions in PI4.0.

A Digital Factory does not only consist of one single sensor that has to be digitized but a whole network of such and other important components of the physical process as well. In this paper we will use the term asset for a (physical) component. The heterogeneity and increased complexity of the assets raise the need for the cooperation of different actors using various software approaches for digitalization. To realize a digital occurrence of the underlying process(es) an AAS can be used. There already exists a standardized approach (PI4.0), but the possibility still exists, that implementations differ by the varied development and conceptual approaches need to be merged to one single representation. Our contribution to this field comprises a reference architecture, in particular for sensor networks, that present a guideline of how to design an AAS that combines different underlying digitalization approaches into one digital representation. Important parts of our proposed architecture are the following:

1. **Harmonization:** The use of a standardized approach to present different asset representations in the same way
2. **Black box approach:** The handling of single assets as black boxes, where only an interface definition is required
3. **Sub model registry:** The registrations of a sub model description in the AAS for every asset
4. **Cloud Asset integration:** The handling of routines on the cloud layer, e.g. data analytics or machine learning, as assets and full integration in the AAS

The remainder of the paper is structured as follows. As the terms DT and AAS are highly related, and the lack of clear definitions results in misunderstandings, we present a differentiation in Section 2. Section 3 gives a wide overview of related work that has been done in this field. In Section 4 we present our reference architecture in an abstract manner, so that it is transferable to other applications. In this part we explain the ideas of a sub model registry, harmonization, black box approach and cloud asset integration. In section 5 we show how the implementation of our architecture functions in a concrete scenario. The paper concludes with a short summary and an outlook to future work in Section 6.

2. AAS vs Digital Twin – A Differentiation

There are many discussions about the relationship between a DT to an AAS. In principle the AAS is a DT from perspective of PI4.0. Originated by Grieves in 2011 (summarized in [1]) in collaboration with NASA, the DT is dedicated for the simulation of an asset. Following this intention, the DT provides the behaviour model of a specific task to simulate. The approach is popular for commissioning in both manufacturing and process industries. Here the buzz words are Virtual Commissioning or Offline Configuration. Aim of these approaches is to shorten the real commissioning. Popular are specifications like Functional Mockup Units in combination with 3D models as investigated in [2, 3]. As a result of the Virtual Commissioning, controller programs can be tested at early stages. Following this idea, there is no real online communication with the asset necessary. The Device Description Language [4] supports offline configuration, which is mainly used in the process industry. This description language is an integral part of the FDI concept [5], where virtual manipulation of device parameters in offline mode is also possible using OPC UA interfaces. Here, the DT may also be used at runtime of the industrial process. Offline and online data can be accessed independently of each other.

An AAS is able to offer the same behaviour as described above. The differences are the interfaces, that can be discovered by other AAS instances during runtime of the AAS. By means of this feature, an exchange of information across manufacturers and end users is technically supported. This results in cross-manufacturer and industry-neutral standards for communication, services and semantics (PI4.0). Beyond the pure DT, the AAS is able to represent an asset (products, devices, machines and plants) during its entire life cycle.

The proposed architecture as presented in section 5 uses the approaches of the AAS as well as the DT. The ASS follows the principles of implementation as proposed in [6], so interoperability between different vendors is possible. Only at the highest level specific implementations of the DT for very specific simulations tasks will be used. In case of

standardized interoperability, such implementations may be integrated as sub-model also in an AAS, but this is beyond the scope of recent research activities.

3. Related Work

To place the contribution of the paper into context, we provide a short literature survey to assess the relevance and adaption of AAS and DT.

The convergence of physical and digital worlds is adapted increasingly in manufacturing industries, however to standardize such approaches is challenging. The Edge Computing (EC) and Fog Computing (FC) paradigms have attempted to ease the challenges. A roadmap showing the impact of EC based on Key Performance Indicators (KPI) is available [7, 8]. The required modelling approaches in DT for the transformation have been proposed [9]. The consequence of formalization of DT through EC along with cloud computing, collaborates centralized and decentralized computing [10]. Suitability of smart manufacturing in an open ecosystem were studied [11, 12].

Applicability of DT and its applications open several research opportunities [13]. The importance of data provided by reliable sources contributes to an effective DT [14, 15, 16]. For fields that apply or require an effective data fusion technique, data sources are required to provide data under strict temporal and measurement uncertainties, apart from being reliable [17, 18].

The concept of DT has led to if not made it easier to apply concepts like condition monitoring and machine learning. Condition monitoring for anomaly detection in industrial assets [19] and usage of data analytics to model DTs [20, 21] are some examples of published examples. To forecast performance of industrial assets based on the collected data, DT has also embraced machine learning. [22]. There is an evolving trend of industries seeking Cyber Physical Systems (CPS) as a key enabler for I4.0 [23].

The approach in this paper facilitates the transformation to I4.0 incorporating AAS standard into existing diverse testbeds, in order to harmonize the approaches of generating digital models which can be exchanged electronically and reused.

4. A Reference Architecture for homogeneous AASs

In this section, we present a generalized approach (Figure 1) that was employed in the FAMOUS² project to elevate the level of abstraction of assets belonging to diverse domains. The main motivation is to make physical assets accessible to users using uniform standardized

² www.famous-project.eu

methods, in a digital domain as envisioned by PI4.0. The AAS standard from PI4.0 is a rich set of directives that aims to model aspects of an asset such that they can be embedded in a meta-model. It is with the help of AAS meta-models, that we can represent the diverse assets in the form of exchangeable electronic documents (such as JSON). The mapping of AAS meta-models to a machine-readable format, makes it possible to transform the digital document into elevated information models (for example in the form of OPC UA). In particular, Figure 1 shows a specific approach we chose in the project, to address the requirements of information exchange between assets. Considering the requirements of the project and the diverse set of assets, if an asset could be viewed as black box from an end user perspective, then it is sufficient to model the asset:

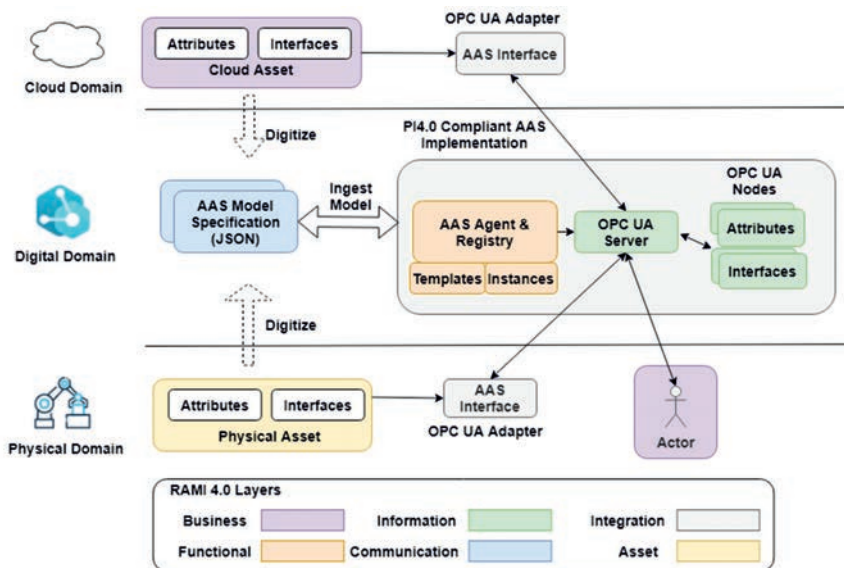


Fig. 1: AAS - A general approach for harmonization

1. To delineate the attributes of the asset in the form of explicit and qualified placeholders, such that the user can infer the exposed attributes either in a standardized protocol such as OPC UA or in the form of exchangeable electronic documents from vendors. This encapsulates the information from the vendor and the assets.
2. The user is pointed to and made aware of available interfaces that can be used to exchange information with the asset using standardized machine to machine communication protocols such as OPC UA.

We found that the physical assets can range from a single sensor component to a large test bed consisting of complex topologies and communication interface.

5. A concrete implementation of the proposed reference architecture

Extending the discussion from the previous chapter, we present our concrete implementation of the proposed reference architecture in the scope of the project.

To offer a better understanding of how to apply the aforementioned principles, we modelled four physical test beds and four business logic applications covering the fields of discrete manufacturing and process industry. In an effort to harmonize the communication approaches, an OPC UA adapter was developed to act as an interface between the AAS and proprietary implementations through the AAS interface.

As a basic classification, the assets could belong to the physical domain, mainly composed of sensors and their associated topological layouts, whereas in the cloud domain, assets are mainly composed of business logic applications. In particular, the propagation of data occurs from sensors and sensor networks encompassing the physical assets, to cloud assets for the purpose of analysis and modelling.

In Figure 2 we present a concrete implementation of the Reference Architecture introduced in Section 4. This architecture is, like the abstract one, aligned to the Reference Architecture Model Industrie 4.0 (RAMI 4.0) which is realized through coloring (find the legends in the bottom of Figure 2). Our architecture can be divided into three logical parts: Testbeds, Edge Node and Cloud.

5.1 Physical Assets Layer (Testbeds)

The lowest layer holds a place for physical assets, and in our implementation, we use four different testbeds. Each testbed is managed by one project partner and is characterized by its own AAS approach. The different approaches are of varying complexity. As a consequence, the same asset type (i.e., a sensor) can be represented in four different ways. Contrary to that, the PI4.0 standard enforces well defined Input/Output interfaces through the meta-model. Our implementation makes use of four different testbeds that are managed with standalone AAS approaches. The use cases covered by the testbeds are originating from the process and the manufacturing industry. The complexity comes from the usage of different sensors and protocols such as: level sensors, acceleration sensors, pressure sensors, density sensors and

speed of sound sensors next to SPI³, I2C⁴, USB⁵, Bluetooth⁶, PROFIBUS and PROFIBUS PA⁷, MODBUS TCP and MODBUS RTU⁸ and Wireless HART⁹.

5.2 Edge Layer

The middle layer which holds the logic of an edge node runs the main AAS logic. Data can be shared with the lower layer via the defined interfaces and following this approach, assets can be viewed as black boxes hiding its complexities. The general AAS system manages the submodels of all registered assets in the system. The decision to locate the main AAS functionality on this layer was met because most of the managed assets are physical ones and the edge node is the nearest reasonable location. If an asset wants to appear in the overall AAS, then it has to register itself here using a PI4.0 conform submodel description. The registry is also the place where other participants can ask for specific submodels or query the entire registry. Besides these management tasks, the AAS platform on the edge layer is responsible for transferring incoming messages to the designated receiver. In our implementation we use OPC UA as the communication protocol for data distribution. The measured values that are acquired from the testbeds are delivered to the edge-level AAS and are subsequently saved in the database. The edge-level AAS simultaneously distributes the data to all subscribed assets from the cloud layer.

5.3 Cloud Layer

The highest layer is described as cloud layer running centralized business logic applications. All the registered submodels from the edge layer are collated on this level into a central reflected instance through the OPC UA server. Utilizing the submodels of the (physical) assets a virtual representation (a DT) of every registered asset is created. These DTs reflect the functionality of the original asset and the interfaces that are defined in the description. A specific cloud asset that is required by the domain of sensor networks is the uncertainty component. One of these constructed assets exists for every physical sensor and is connected

³ <https://patents.google.com/patent/US4816996A/en>

⁴ <https://de.i2c-bus.org/>

⁵ <https://www.usb.org/>

⁶ <https://www.bluetooth.com/>

⁷ <https://www.profibus.com>

⁸ <https://www.modbus.org>

⁹ <https://www.fieldcommgroup.org/hart-specifications>

to its digital representation. When a sensor is registered in the submodel registry, it is paired with a corresponding calibration model. Sensor measurements can be consumed with and without uncertainty information. If an asset requests the uncertainty information, then the Uncertainty Component is commissioned. This component is subscribed to the corresponding sensor and receives every measurement. With this data, the component queries the stored calibration model and can compute the uncertainty related to the measured value. This calculated uncertainty is returned to the virtual sensor and distributed from there to the system. Every sensor has its own calibration model, even if two sensors are of the exact same type and could share one model. This decision was made because sensors are able to drift over time. Using our approach, it is possible to adjust the calibration information of a single sensor. One virtual asset that is using this ability of the system is the so-called Co-Calibration Component. The existence of digital representations of physical assets enables multiple possibilities for the application of Machine Learning (ML) and Data Analytics. A specific application example would be an Anomaly Detection Component that is subscribed to several sensors and uses ML methods to detect drifting or unusual sensor behavior. This knowledge can subsequently be applied to trigger the Co-Calibration Component which in turn adjusts the calibration model of the faulty sensor in the right direction. Other possible ways to utilize these virtual sensor representations are sensor fusion and the creation of complete virtual sensors. This approach can utilize several low accuracy sensors to construct a more accurate value or even combine available physical quantities to derive predictions of additional physical quantities that are not being measured. The derived virtual sensor would also be represented via a DT based on a submodel description. At the same time, the AAS is able to visualize all its managed assets and present them to the user. This can be realized via a REST server that is connected to the virtual representations.

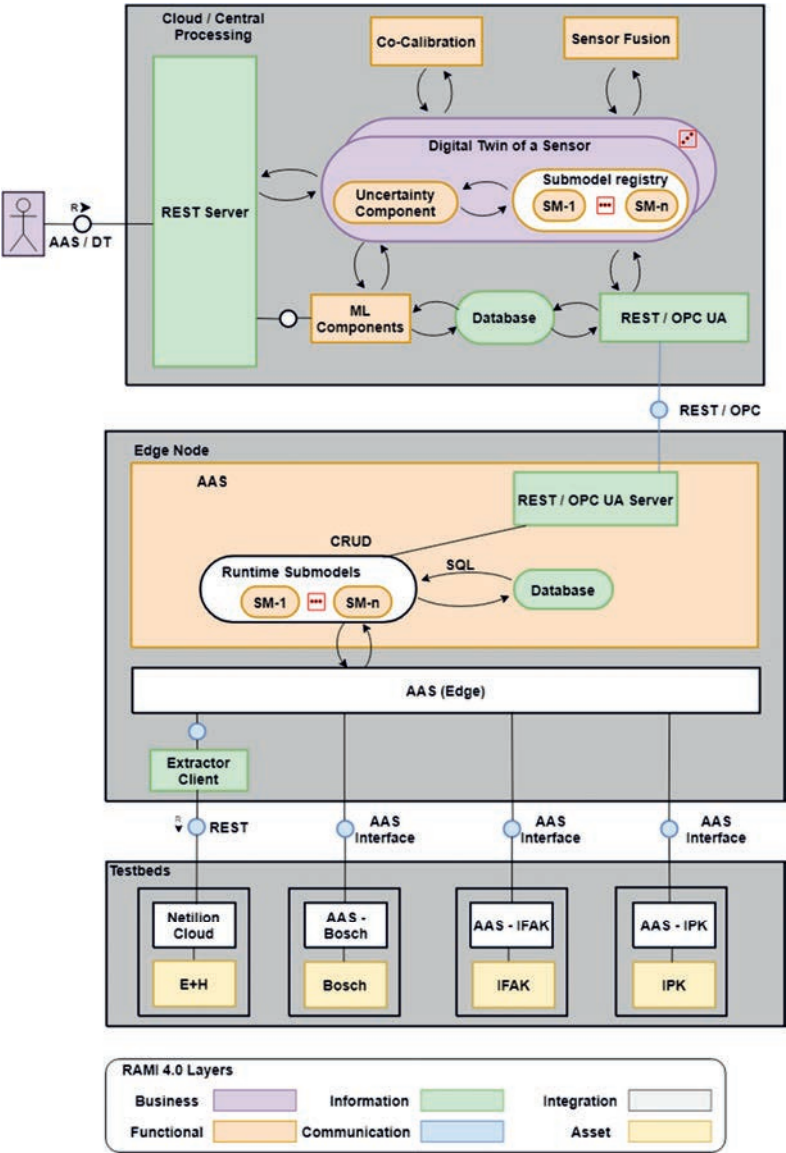


Fig. 2: A concrete implementation/The FAMOUS implementation of the Reference Architecture.

Finally, the received data is persisted utilizing a database on the cloud level for future use.

6. Conclusion and Future work

In this paper we presented a reference architecture for homogeneous ASSs. In the scope of this work, we also offered a clarification of how the concepts of AAS and DT can be differentiated. The proposed architecture intends to homogenize different virtualization concepts (AASs) in an environment of increased of increased complexity and heterogeneity. It encapsulates the physical, the digital and the cloud domain. Based on that approach we described a concrete implementation that brings together different testbeds from the manufacturing and the process industry, that vary in structure and virtualization concepts.

The submodel descriptions can be used to pave the way towards a harmonization for unifying AAS approaches in a heterogeneous environment by elevating the level of abstraction based on the PI4.0 AAS standard.

Our future work will extend the proposed architecture and implement the identified harmonization strategies. Furthermore, the evaluation of interoperable AASs between real world use cases is of interest.

7. Acknowledgements

Research for this paper was financed by the Federal Ministry of Education and Research (BMBF) project FAMOUS (01IS18078). We thank our project partners for their contributions and their collaboration to this research work.

References

- [1] M. Grieves and J. Vickers, "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems," in *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches*, F. Kahlen, S. Flumerfelt and A. Alves, Eds., Cham, Springer International Publishing, 2017, p. 85–113.
- [2] S. Süß, S. Magnus, M. Thron, H. Zipper, U. Odefey, V. Fäßler, A. Strahilov, A. Kłodowski, T. Bär and C. Diedrich, "Test methodology for virtual commissioning based on behaviour simulation of production systems," in *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2016.

- [3] A. Strahilov and others, "Improving the transition and modularity of the virtual commissioning workflow with AutomationML," in *4th AutomationML User Conference*, 2016.
- [4] M. Riedl, R. Simon and M. Thron, *Electronic device description language: EDDL*, Oldenbourg, 2002.
- [5] D. Großmann, M. Braun, B. Danzer and M. Riedl, "FDI-Field Device Integration," *VDE VERLAG GmbH*, 2015.
- [6] H. Zipper, A. Belyaev and C. Diedrich, "Generische Umsetzung von Verwaltungsschalen auf Basis der aktuellen Handlungsempfehlungen der Plattform Industrie 4.0," in *AUTOMATION 2019, Autonomous Systems and 5G in Connected Industries*, 2019.
- [7] D. Georgakopoulos, P. P. Jayaraman, M. Fazia, M. Villari and R. Ranjan, "Internet of Things and Edge Cloud Computing Roadmap for Manufacturing," *IEEE Cloud Computing*, vol. 3, pp. 66-73, 2016.
- [8] M. de Brito, S. Hoque, R. Steinke and A. Willner, "Towards Programmable Fog Nodes in Smart Factories," in *1st International Workshops on Foundations and Applications of Self* Systems (FAS*W)*, Augsburg, 2016.
- [9] Q. Qi, D. Zhao, T. W. Liao and F. Tao, "Modeling of Cyber-Physical Systems and Digital Twin Based on Edge Computing, Fog Computing and Cloud Computing Towards Smart Manufacturing," in *Volume 1: Additive Manufacturing; Bio and Sustainable Manufacturing*, 2018.
- [10] Q. Qi and F. Tao, "A Smart Manufacturing Service System Based on Edge Computing, Fog Computing, and Cloud Computing," *IEEE Access*, vol. 7, pp. 86769-86777, 2019.
- [11] Z. Li, W. M. Wang, G. Liu, L. Liu, J. He and G. Q. Huang, "Toward Open Manufacturing," *Industrial Management & Data Systems*, vol. 118, pp. 303-320, 2018.
- [12] Y. Lu, F. Riddick and N. Ivezic, "The Paradigm Shift in Smart Manufacturing System Architecture," in *IFIP Advances in Information and Communication Technology*, Springer International Publishing, 2016, pp. 767-776.
- [13] Y. Lu, C. Liu, K. I.-K. Wang, H. Huang and X. Xu, "Digital Twin-Driven Smart Manufacturing: Connotation, Reference Model, Applications and Research Issues," *Robotics and Computer-Integrated Manufacturing*, vol. 61, p. 101837, 2020.

- [14] K. Thiyagarajan, S. Kodagoda and N. Ulapane, "Data-driven machine learning approach for predicting volumetric moisture content of concrete using resistance sensor measurements," in *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, 2016.
- [15] Y. Wan, L. Li, J. He, X. Zhang and Q. Wang, "Anshan: Wireless Sensor Networks for Equipment Fault Diagnosis in the Process Industry," in *2008 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, 2008.
- [16] C. Vázquez, A. B. Gonzalo, S. Vargas and J. Montalvo, "Multi-Sensor System Using Plastic Optical Fibers for Intrinsically Safe Level Measurements," *Sensors and Actuators A: Physical*, vol. 116, pp. 22-32, 2004.
- [17] N. Kaempchen and K. Dietmayer, "Data synchronization strategies for multi-sensor fusion," in *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, 2003.
- [18] Y.-W. Hong and A. Scaglione, "A Scalable Synchronization Protocol for Large Scale Sensor Networks and Its Applications," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 1085-1099, 2005.
- [19] G. Shah and A. Tiwari, "Anomaly detection in IIoT," in *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data - CoDS-COMAD '18*, 2018.
- [20] E. Uhlmann, H. Laghmouchi, C. Geisert and E. Hohwieler, "Smart wireless sensor network and configuration of algorithms for condition monitoring applications," *Journal of Machine Engineering*, vol. 17, pp. 45-55, 1 2017.
- [21] E. Uhlmann, A. Laghmouchi, C. Geisert and E. Hohwieler, "Decentralized Data Analytics for Maintenance in Industrie 4.0," *Procedia Manufacturing*, vol. 11, pp. 1120-1126, 2017.
- [22] E. Uhlmann, R. P. Pontes, C. Geisert and E. Hohwieler, "Cluster Identification of Sensor Data for Predictive Maintenance in a Selective Laser Melting Machine Tool," *Procedia Manufacturing*, vol. 24, pp. 60-65, 2018.
- [23] B. Vogel-Heuser, C. Diedrich and M. Broy, "Anforderungen an Cps Aus Sicht Der Automatisierungstechnik / Requirements on Cps From the Viewpoint of Automation," *at - Automatisierungstechnik*, vol. 61, 2013.

Digitale Zwillinge für Shopfloor-Management und -steuerung

Digital Twins for Shopfloor Management and Control

Dr. rer. nat. **Annelie Sohr**, Dipl.-Math. **Roland Rosen**,
Dr. rer. nat. **Jan Christoph Wehrstedt**, Dr.-Ing. **Jan Fischer**,
M. Sc. **Franz Georg Listl**, **Martin Schellander**,
Siemens AG, München;
Dipl.-Ing. **Rainer Haar**, Siemens AG, Köln

Zusammenfassung

Ein effizienter Produktionsbetrieb ist eine wichtige Voraussetzung für die Erreichung der wirtschaftlichen Zielsetzungen in der Fertigung. Zugleich verstärken die Möglichkeiten von Industrie 4.0 den Ruf nach mehr Flexibilität auf dem Shopfloor und verbesserter Resilienz der Produktion gegenüber Schwankungen auch kurzfristiger Natur. Eine kapazitätsgerechte Feinplanung der Fertigungsaufträge auf dem Shopfloor wird dabei immer wichtiger. In diesem Beitrag wird ein Ansatz präsentiert, der auf Basis eines digitalen Zwillings der Fertigung eine situationsgerechte Adaption der Auftragssteuerung ermöglicht, und in einer Werkstattfertigung Informationen und Assistenz für Werker, Schichtleiter und Produktionsverantwortliche liefert. Durch die Verwendung von Simulations- und Optimierungsmethoden werden Mitarbeiter qualifikationsgerecht den Arbeitsplätzen zugeordnet und eine zeitgenaue Einplanung der einzelnen Fertigungsaufträge und ihrer Vorgänge vorgenommen. Über den „Digital Workers Guide“ auf einem mobilen Endgerät bekommt der Werker schließlich konkrete Handlungsanweisungen. Der Ansatz integriert sich in die vorhandene IT und OT der Fertigung. Die Nutzung einer Low-Code-Plattform ermöglicht eine aufwandsarme Realisierung. Ferner gewährleistet ein modularer Aufbau der Lösung die Erweiterbarkeit und Übertragbarkeit auf andere Werke.

Abstract

An efficient production facility is an essential prerequisite for achieving economic objectives in manufacturing. At the same time, the potential of Industry 4.0 reinforces the call for more flexibility on the shop floor and improved resilience in production to fluctuations, even short-term. Therefore, a detailed capacity-based planning of production orders on the shop floor is

becoming increasingly important. In this paper, an approach based on a digital twin of the production is presented which enables a situation-specific adaptation of the production order control. Furthermore, it provides information and assistance for workers, shift supervisors and production managers. Using simulation and optimization methods, workers are assigned to the workplaces according to their qualifications, and the individual production orders and their processes are scheduled precisely in time. Finally, the "Digital Workers Guide" on a mobile device provides the worker with specific instructions for action. The approach is integrated into the existing IT and OT of the production. Thereby, the use of a low-code platform enables a low-cost implementation. Furthermore, a modular structure ensures that the solution can be expanded and transferred to other factories.

1. Einleitung

Ein effizienter Betrieb der Fertigung ist eine wichtige Voraussetzung für die Erreichung der wirtschaftlichen Zielsetzungen bei gleichzeitiger Erfüllung der ökologischen Anforderungen. Einerseits durch die Möglichkeiten von Industrie 4.0 [1], wie z.B. der individualisierten Fertigung, und andererseits durch immer kurzfristigere Auftragszeiten und große Änderungen der Auftragsmenge wird eine ressourcengerechte Feinplanung der Fertigungsaufträge auf dem Shopfloor immer wichtiger. Eine verstärkte Flexibilität und verbesserte Resilienz der Produktion und der Fertigungsabläufe auf dem Shopfloor stehen zunehmend im Fokus. Bei historisch gewachsenen Produktionsanlagen (Bild 1), welche in der Regel nur zum Teil automatisiert und digitalisiert sind, ergibt sich somit die Herausforderung, die Werke fit für Industrie 4.0 zu machen, und dabei die Produktionsstrukturen zu erhalten, um die alten Kundenstrukturen weiterhin bedienen und neue Märkte erschließen zu können. Außerdem ist auch eine Weiterentwicklung der Mitarbeiter von schwer körperlich arbeitenden Facharbeitern zu höher qualifizierten Fachkräften wünschenswert und nötig.

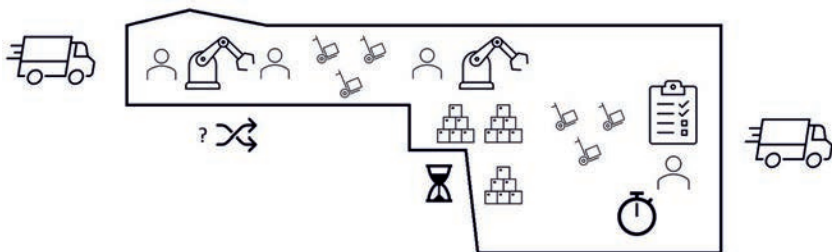


Bild 1: Historisch gewachsenes Werk mit Maschinen, Werkern und geringer, aber zunehmender Automatisierung

Die veränderten Anforderungen treffen auf eine Reihe von Hindernissen: Werkzeuge für eine umfassende Feinplanung fehlen, Arbeiter und Arbeitsplätze sind nicht ausreichend vernetzt und eine übergeordnete Kommunikation auch über digitale Wege findet nicht statt. Zu oft steht eine lokale Optimierung an der einzelnen Maschine einer ganzheitlichen Fokussierung auf die gegebenen Produktionsziele wie der Einhaltung aller Kundentermine im Weg.

Um diese Hindernisse zu überwinden, wird in diesem Beitrag der „Digital Workers Guide“ vorgestellt, dessen Herzstück ein digitaler Zwilling der Anlage bildet [2]. Von der aktuellen Situation auf der Anlage ausgehend wird mit Hilfe des digitalen Zwillings eine Feinplanung durchgeführt und konkrete Handlungsanweisungen für die einzelnen Werker in der Produktion abgeleitet, die über eine personalisierte Softwareapplikation kommuniziert werden.

Hierzu werden in Kapitel 2 die Abläufe in der Fertigung beschrieben, vom Auftragseingang bis zur Produktion auf dem Shopfloor. Kapitel 3 stellt die Rolle der digitalen Zwillinge in der Produktion vor. In Kapitel 4 wird das Konzept des Workers Guide eingeführt, dabei wird sowohl auf die Architektur als auch die Einbettung in den Produktionsablauf eingegangen.

2. Produktionssteuerung: Vom Auftrag zum Produkt

Um im nachfolgenden Kapitel die Rolle des Digitalen Zwillings auf dem Shopfloor einordnen zu können, soll hier zunächst der übliche Ablauf in der Produktion vom Eingang des Kundenauftrags bis zur Auslieferung des fertigen Produktes skizziert werden. Dabei soll zwischen der kapazitätsbezogenen Produktionsplanung, bei der heute ERP-Systeme eine entscheidende Rolle einnehmen, und der unterhalb dieser Planungsebene gelegenen operativen Feinplanung unterschieden werden. Die Feinplanung ist insbesondere in der Werkstattfertigung relevant und wird mit dem hier gezeigten Digitalen Zwilling adressiert.

Die Schritte innerhalb dieser Ebenen und die dabei betrachteten Planungszeiträume können sich von Fertigung zu Fertigung unterscheiden, adressieren aber schließlich doch ähnliche, wiederkehrende Aufgaben. Für eine umfassende Darstellung siehe z.B. [3]. Die folgende Detaillierung und deren graphische Darstellung in Bild 2 ist als beispielhafte Ausprägung zu verstehen, wie sie in einer Werkstattfertigung vorgefunden werden kann.

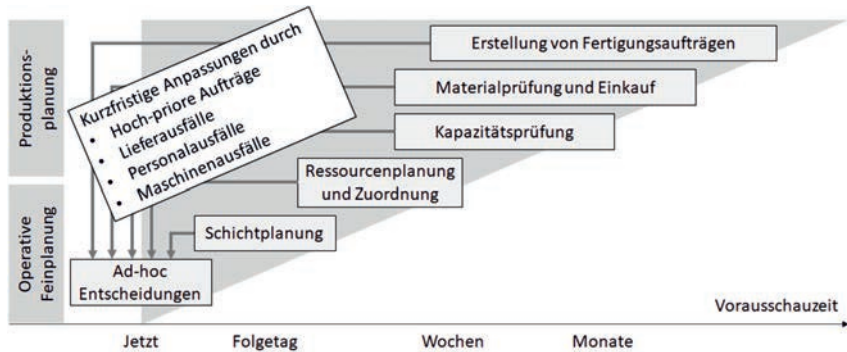


Bild 2: Schritte zur Produktionsplanung und -steuerung

- **Erstellung von Fertigungsaufträgen:** Aus eingegangenen Kundenaufträgen und -terminen, und der Kenntnis zu Fertigungsprozessen, Vorgangszeiten und -abfolgen werden Fertigungsaufträge erstellt und z.B. durch Rückwärtsterminierung in die Produktion eingeplant.
- **Materialprüfung und Einkauf:** Der entstandene Produktionsplan wird in Bezug auf die Materialverfügbarkeit in Lagerbeständen geprüft, Bestellungen werden getätigt und Entscheidungen zu Eigen- oder Fremdfertigung getroffen. Gegebenenfalls wird der Produktionsplan angepasst.
- **Kapazitätsprüfung:** Der Produktionsplan wird mit den bestehenden personellen und maschinellen Kapazitäten abgeglichen und gegebenenfalls angepasst. Bei größeren Abweichungen wird versucht, im Rahmen der gegebenen Möglichkeiten die Kapazitäten zu beeinflussen, z.B. durch Umstellung und Ausweitung der Schichten bzw. durch Einsatz von Zeitarbeitern.
- **Ressourcenplanung und Zuordnung:** Die einzelnen Mitarbeiter werden qualifikationsgerecht und unter Berücksichtigung ihrer Verfügbarkeit den Schichten und Arbeitsplätzen bzw. Arbeitsplatzgruppen zugeordnet.
- **Schichtplanung:** Fertigungsaufträge werden in eine Bearbeitungsreihenfolge gebracht und bestenfalls zeitgenau auf den Maschinen eingeplant und schließlich den verantwortlichen Mitarbeitern zur Bearbeitung übergeben. Besondere Herausforderungen treten da auf, wo Mitarbeiter über die Schicht verteilt an mehreren Arbeitsplätzen zum Einsatz kommen, da hier Abhängigkeiten im Produktionsfluss entstehen und die Wechsel eingeplant und kommuniziert werden müssen.

Trotz dieser Vielzahl an Planungsschritten auf unterschiedlichen Ebenen und mit unterschiedlichen Zeithorizonten treten in der laufenden Produktion Schwierigkeiten auf, sodass der Produktionsfluss ins Stocken gerät und Ad-hoc-Entscheidungen zur Behebung erforderlich sind. Die Ursachen sind zahlreich:

- **Komplexität der operativen Feinplanung:** Die Produktionsplanung erfolgt kapazitätsbezogen. Für die operative Feinplanung müssen dagegen die einzelnen Schritte der Prozessketten und Auftragsnetze aufeinander abgestimmt werden. Verfügbarkeiten und Zuordnungen von Maschinen und Mitarbeitern müssen zeitlich aufgelöst betrachtet werden, um den Produktionsfluss und all seine Abhängigkeiten erfassen und optimieren zu können. Hierfür stehen heutzutage die richtigen Werkzeuge noch nicht zur Verfügung oder sind wie die Simulation oft einem Expertenkreis vorbehalten.
- **Änderungen und Störungen:** Ein hoch-priorer Auftrag kommt herein und wird vorgezogen, eine erwartete Materiallieferung ist nicht eingetroffen, ein Mitarbeiter fällt krankheitsbedingt kurzfristig aus, eine Maschine hat einen Schaden und muss repariert werden. Auf diese Änderungen und Störungen muss schnell und adäquat reagiert werden, was eine Umplanung im laufenden Betrieb erfordert.
- **Lokale Optimierung:** Aufgrund von einerseits fehlenden Möglichkeiten zur Feinplanung und damit verbunden fehlenden Vorgaben für den einzelnen Mitarbeiter sowie auftretenden Störungen im laufenden Betrieb, löst der Mitarbeiter diese Schwierigkeiten über den Weg der "Selbst-Optimierung". Er optimiert lokal die Abläufe an seinem aktuellen Arbeitsplatz, was unter Umständen aus globaler Sicht eher kontraproduktiv sein und den Produktionsfluss behindern kann. Kundenaufträge werden auseinandergerissen, dringende Aufträge lokal zurückgestellt, wartende Aufträge blockieren Maschinen.

Der hier gezeigte Ansatz adressiert genau diese Probleme: Die Komplexität wird durch den Einsatz eines digitalen Zwillings überschaubar und beherrschbar. Eine kurzfristige Reaktion auf Änderungen und Störungen wird so ermöglicht. Die gewonnene Transparenz lenkt den Blick weg von der lokalen "Selbst-Optimierung" auf das große Ganze.

3. Digitale Zwillinge in der Produktion

Digitale Modelle werden beim Design und bei der Auslegung seit jeher verwendet, die verbreitetsten Ansätze sind die Materialflusssimulation und die 3D-Kinematiksimulation [4].

Die 3D-Kinematiksimulation wird eingesetzt, um die Interaktion innerhalb der Fertigungszelle zwischen Maschinen wie Robotern und dem Produkt zu analysieren. Sie wird hauptsächlich in der Phase der Detailplanung und Inbetriebnahme verwendet. Somit ermöglicht sie die

Pfadplanung und Programmierung von Robotern, das Aufsetzen und Validieren von Steuerungskonzepten sowie eine virtuelle Inbetriebnahme.

Materialflusssimulationen werden heute im Rahmen der Produktionssystemplanung häufig eingesetzt, um Fabriklayouts zu überprüfen, Kapazitätsengpässe zu identifizieren oder um Materialpuffer zu dimensionieren [5]. Im Folgenden soll genauer auf einige Details der Materialflusssimulation eingegangen werden.

Ein Materialfluss wird in der Regel mit einem objektorientierten, ereignisdiskreten Simulationsmodell beschrieben, wobei die Objekte Maschinen, Lagerplätzen, Werkern oder Produkten entsprechen. Basierend auf einem vorgegebenen Auftragsmix wird der Durchlauf der Produkte durch die Produktion simuliert. Dabei durchlaufen die Produkte gemäß der für sie festgelegten erforderlichen Produktionsschritte die einzelnen Arbeitsplätze. Für die verschiedenen Arbeitsplätze und Vorgänge werden dafür die Bearbeitungszeiten sowie benötigte Qualifikationen der Mitarbeiter als Simulationsparameter hinterlegt. Zusätzlich werden auch die Mitarbeiter mit ihren Qualifikationen als Objekte angelegt. Somit lässt sich unter Zuordnung der Mitarbeiter auf die Arbeitsplätze der Produktionsfluss simulieren. Engpässe im Produktionssystem lassen sich so bestimmen und geeignete Gegenmaßnahmen definieren.

Bild 3 zeigt ein Simulationsmodell einer Materialflusssimulation einer Fabrik. Man erkennt die verschiedenen Arbeitsplätze sowie mögliche Pfade, die die Produkte durch die Fabrik nehmen können.

Klassischerweise wird die Materialflusssimulation in der Planungsphase verwendet, aber neuere Ansätze [6], [7] schlagen vor, die Simulationsmodelle in betriebsparallele Digitale Zwillinge zu überführen und in Assistenzsysteme für die Betriebsphase zu integrieren. Hierzu müssen die Modelle mit aktuellen Aufträgen und Verfügbarkeiten sowie dem aktuellen Anlagenzustand parametrisiert werden, um Vorhersagen über das Anlagenverhalten treffen und Betriebsstrategien für den konkreten Betrieb optimieren zu können, welche dann von den Operatoren wieder in den Betrieb zurückgeführt werden sollten. Diese Ansätze sind derzeit noch Entwicklungsgegenstand und als individuelle Lösungen teilweise pilotiert und in der Erprobungsphase.

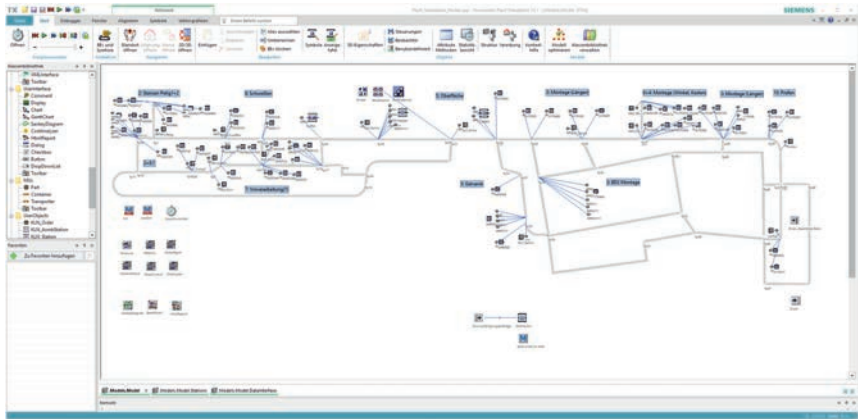


Bild 3: Materialflusssimulation

4. Idee und Konzept Digital Workers Guide

Ein digitaler Zwilling mit einer betriebsparallelen Materialflusssimulation, wie im vorangegangenen Abschnitt skizziert, bildet das Herzstück des Digital Workers Guide. Das Simulationsmodell wird eingebettet und durch eine Applikation orchestriert. Diese adressiert passgenau den Anwendungsfall der operativen Feinplanung und leitet aus den Simulationsergebnissen konkrete Handlungsanweisungen für die Werker ab, welche personalisiert dargestellt werden. Die Simulation wird somit auch jenseits des spezialisierten Expertenkreises nutzbar.

Im Folgenden wird zunächst die funktionale Sicht, dann die Anwendungssicht und schließlich die architektonische Sicht der Lösung beschrieben.

Aus funktionaler Sicht ermöglicht die Simulation ausgehend vom aktuellen Zustand der Anlage, die Auswirkungen operativer Entscheidungen zu prognostizieren. Im Simulationsmodell sind dazu die erforderlichen Parameter zu Produkten, Prozessen und Ressourcen hinterlegt. Der aktuelle Zustand wird aus Daten zur laufenden Produktion, zu anstehenden Fertigungsaufträgen und zur Verfügbarkeit von Werkern und Maschinen aus Quellen wie dem Enterprise-Resource-Planning (ERP) System abgeleitet und ins Modell eingespielt. An operativen Entscheidungen deckt das Modell folgende Punkte ab:

- Die Reihenfolgebildung aller anstehenden Fertigungsaufträge
- Die Wahl der Prozesse und Maschinen, sofern Alternativen existieren
- Die Zuordnung von Werkern zu Maschinen und Vorgängen

Die Simulation berechnet aus den Daten und den getroffenen Entscheidungen detaillierte Produktionspläne mit Anfangs- und Endzeiten für alle Fertigungsaufträge und Prozessschritt-

te. Verschiedene Entscheidungsalternativen können so durchgespielt und volle Transparenz bzgl. des zu erwartenden Produktionsflusses gewährleistet werden.

Andererseits enthält der Digital Workers Guide aber auch Module, die ihrerseits Entscheidungen vorschlagen. Dies kann zum einen durch Einsatz von Heuristiken (siehe z.B. [8] erfolgen, die anschließend mittels Simulation validiert werden. Zum anderen kommen mathematische Optimierungsverfahren zum Einsatz, um eine optimale Zuordnung der Werker auf Basis verfügbarer Qualifikationen zu treffen und Produktionspläne in Bezug auf unterschiedliche Kriterien wie Termintreue und Auslastung zu optimieren. Die mathematischen Optimierungsprobleme werden schnell komplex und können zu langen Rechenzeiten bzw. suboptimalen Lösungen führen [8]. Sinnvolle Abstraktionen sind daher von Nöten, wodurch der Simulation wiederum die wichtige Rolle zukommt, die vorgeschlagenen Entscheidungen zu validieren. Daneben kann in einer Ausbaustufe die Simulation auch noch Unsicherheiten berücksichtigen, die in der Optimierung aus Aufwandsgründen nicht betrachtet werden. Ein weiterer Weg, dem insbesondere im Förderprojekt ASSISTANT [9] nachgegangen wird, ist die simulations-basierte Optimierung, bei der systematisch und automatisiert eine Reihe von Simulationsexperimenten durchgeführt und verglichen werden. Evolutionäre Algorithmen und verschiedene Verfahren aus dem maschinellen Lernen tragen hier bei.

Die Berechnungen von Simulation und Optimierung erfolgen vor Schichtbeginn für die anstehende Schicht. Abweichungen im Verlauf der Schicht führen zu Korrekturen der berechneten Produktionspläne. Bei größeren Änderungen wie dem Hereinkommen eines hochpriorien Auftrags oder dem Ausfall eines Werkers oder einer Maschine muss gegebenenfalls eine komplette Neuberechnung erfolgen.

Abhängig vom Anwender sowie dessen Aufgaben und Anforderungen, werden unterschiedliche Anwendungssichten umgesetzt und nach dem Einloggen bereitgestellt. Der Schichtleiter hat die Aufgabe, die anstehende Schicht im Detail zu planen, zu überwachen und die nötigen operative Entscheidungen zu treffen. In seiner Sicht des Digital Workers Guide hat er die Möglichkeit, an seinem Rechner die aktuellen Daten zu sichten, manuelle Änderungen z.B. zur Verfügbarkeit von Mitarbeitern vorzunehmen, eigene Vorgaben wie feste Zuordnungen von Werken zu Maschinen zu definieren, Ziele wie Termintreue oder Auslastung zu spezifizieren und zu priorisieren, sowie Berechnungen anzustoßen. Deren Ergebnisse werden ihm tabellarisch und graphisch (z.B. als Gantt-Diagramme) präsentiert, so dass er einen Überblick über den zu erwartenden Produktionsfluss mit allen Kundenaufträgen, Maschinen und Werkern bekommt.

Der einzelne Werker bekommt dagegen eine personalisierte Sicht auf seinem mobilen Endgerät, d.h. seinem Tablet, seinem Smartphone oder seiner Smart Watch. Diese Sicht enthält

eine geordnete Liste der anstehenden, dem Mitarbeiter zugeordneten Aufgaben, seien es den Maschinen zugeordnete Produktionsschritte aus geplanten Fertigungsaufträgen oder Transportaufgaben in nachfolgende Versorgungsbereiche (siehe Bild 4). Der Mitarbeiter kann somit jederzeit und von überall sehen, wann, wo und wofür er als nächstes eingesetzt wird. Jede Aufgabe ist mit einem Barcode versehen, den der Werker am Arbeitsplatzterminal seines nächsten Einsatzortes scannen kann, um weiterführende Informationen wie Maschineneinstellungen und Konstruktionsskizzen aus bestehenden Systemen wie dem ERP-System zu erhalten.

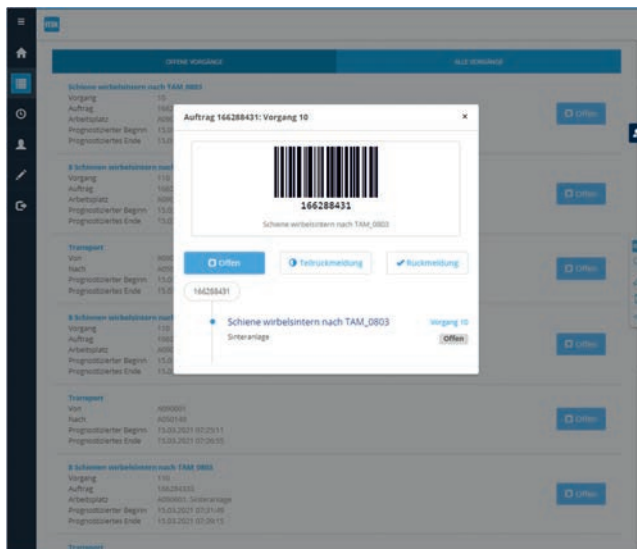


Bild 4: Ansicht für den Werker auf dem mobilen Endgerät

Nach der Bearbeitung des Vorgangs am Arbeitsplatz bzw. Transport des Materials an den Zielort wird dies wahlweise am Arbeitsplatzterminal oder im Digital Workers Guide auf dem mobilen Gerät quittiert. Der Kreis schließt sich, indem der Werker zur nächsten Aufgabe übergeht (siehe Bild 5).



Bild 5: Workflow zum Einsatz des Digital Workers Guide

Der Digital Workers Guide wird als Web-Applikation auf Basis einer Low-Code-Plattform realisiert, was eine aufwandsarme Umsetzung ermöglicht und Standardfunktionen zum Domänenmodell, zum Einlesen und Schreiben von Daten, zum Aufruf von externen Services, zur Benutzerverwaltung und Zugriffsrechten, sowie viele User Interface Bausteine bereits mit sich bringt. Simulations- und Optimierungsrechnungen laufen als Backend-Service auf einem Windows-Rechner (siehe Bild 6). Daten werden aus bestehenden Systemen wie dem ERP-System und weiteren Datenquellen wie z.B. werksspezifischen Qualifikationslisten, Kalendern zur Verfügbarkeit oder Schichtplänen gelesen.

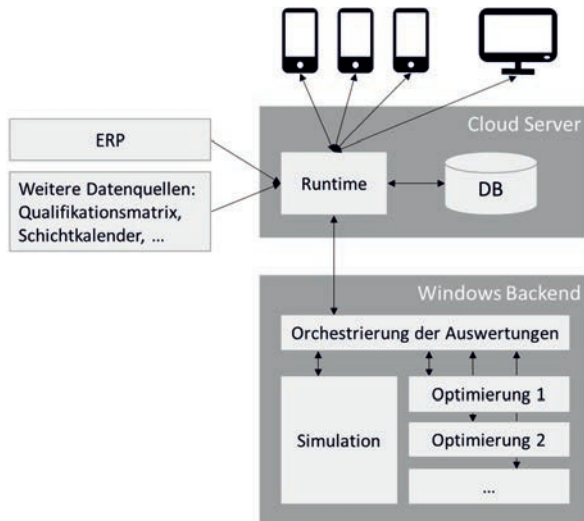


Bild 6: Architektur des Digital Workers Guide

5. Zusammenfassung und Ausblick

Der Digital Workers Guide wird derzeit umgesetzt und prototypisch erprobt. Ähnlich wie im klassischen Lean Management [10] liegt der Nutzen darin, durch einen besseren Einsatz aller Beteiligten, durch höhere Transparenz und durch effizienteren Einsatz von Ressourcen (Arbeitszeit, Kapazität, ...) Verschwendungen in den Bereichen Transport, Bestände, Bewegung, Warte- und Liegezeiten zu vermeiden. Der Verschwendung als einer bedeutenden Ursache für die Minderung des Unternehmenserfolges und für die Demotivation der Mitarbeiter kann mit dem Digital Workers Guide aktiv entgegengewirkt werden.

Erste Rückmeldungen aus dem Pilotbetrieb heben vor allem den Nutzen für den einzelnen Werker auf dem Shopfloor hervor. Er macht die Erfahrung, dass die vom Digital Workers Guide zugewiesenen Aufgaben wie geplant erledigt werden können, da die Voraussetzungen wie z.B. Materialverfügbarkeit bereits geklärt sind. Auch kann er sicher sein, dass die vorgeschlagene Bearbeitungsreihenfolge aus Sicht des Produktionsflusses Sinn macht, so dass er nicht mehr den Weg der „Selbst-Optimierung“ seines Arbeitsplatzes geht. Es wird angeregt, dass mit dem System in Zukunft die Werker auch gezielt an unterschiedlichen Arbeitsplätzen eingesetzt werden, um Qualifikationen auszubauen und Monotonie abzubauen. Generell wird im Digital Workers Guide die Chance gesehen, von einem „Das ist mein Arbeitsplatz“ zu einem „Das ist meine Produktion“ zu kommen.

Bei der Umsetzung wurde und wird auf Modularität und Erweiterbarkeit Wert gelegt. Nach erfolgreicher Einführung des Systems besteht die Möglichkeit, weitere Dienste einzubinden. Als Beispiele seien Lagerhaltung und Materialbereitstellung, Einkauf und Make-or-Buy-Entscheidungen, sowie die Einplanung des Versands genannt. Auch die Übertragbarkeit auf andere Werke mit möglicherweise anderen Anforderungen wird durch die Modularität gewährleistet, da der Ansatz unterstützt, lediglich die relevanten Module anzupassen bzw. auszutauschen.

Der Digitale Workers Guide bietet schließlich aufgrund seiner Erweiterbarkeit auch bislang gering-automatisierten Betrieben einen Einstieg in eine digitale Produktion, welche mit zunehmender Automatisierung der Produktion mitwächst. Neue Komponenten können aufwandsarm im digitalen Zwilling ergänzt und zusätzliche Daten integriert werden. Umgekehrt wird der digitale Zwilling und mit ihm insbesondere das erstellte Simulationsmodell auch zur Evaluation möglicher Erweiterungen wiederverwendet und somit der Transformationsprozess in Richtung einer hochautomatisierten Fertigung unterstützt.

Diese Arbeit wurde von der Europäischen Kommission im Rahmen des Förderprojekts ASSISTANT (GA-Nr.: 101000165) unterstützt.

- [1] BMWi, "Fortschreibung der Anwendungsszenarien der Plattform Industrie 4.0." https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/fortschreibung-anwendungsszenarien.pdf?__blob=publicationFile&v=7 (accessed Apr. 11, 2019).
- [2] R. Rosen, G. von Wichert, G. Lo, and K. D. Bettenhausen, "About the importance of autonomy and digital twins for the future of manufacturing," *IFAC-Pap.*, vol. 48, no. 3, pp. 567–572, 2015.
- [3] G. Schuh, *Produktionsplanung und-steuerung: Grundlagen, Gestaltung und Konzepte*. Springer-Verlag, 2007.
- [4] D. Smirnov, T. Schenk, and J. C. Wehrstedt, "Hierarchical Simulation of Production Systems," in *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, 2018, pp. 875–880.
- [5] J. Fischer, R. Willers, and C. Sinnwell, "Materialflusssimulation für die CTPS Entwicklung," in *Modellbasierter Entwicklungsprozess cybertronischer Systeme*, 2017, pp. 131–138.
- [6] J. C. Wehrstedt, A. Sohr, T. Schenk, R. Rosen, and Y. Zhou, "A Framework for Operator Assist Apps of Automated Systems," *IFAC-Pap.*, vol. 53, no. 2, pp. 10562–10567, 2020, doi: <https://doi.org/10.1016/j.ifacol.2020.12.2805>.
- [7] M. Pinedo, *Scheduling*, vol. 29. Springer, 2012.
- [8] ASSISTANT, "Assistant project – Learning and robust decision support sytem for agile manufacturing environnements." <https://assistant-project.eu/> (accessed Apr. 27, 2021).
- [9] W. Schmeisser and D. Stoeff, *Vom Lean Management zur Digitalisierung*. UVK Verlag, 2018.

Eine Methodik zur Erstellung multi-dimensionaler Modelle des Digitalen Zwillings für automatisierte Produktionssysteme

D. Braun, M.Sc., Prof. Dr.-Ing. Dr. h. c. **Michael Weyrich**,
Institut für Automatisierungstechnik und
Softwaresysteme (IAS), Universität Stuttgart;
Dr.-Ing. **W. Schlögl**, Siemens AG, Nürnberg

Kurzfassung

Der Digitale Zwillinge einer Produktionsanlage kann für viele Use Cases sehr hilfreich sein. Dazu muss die Synchronität des Digitalen Zwillings mit dem realene Produktionssystem sichergestellt werden. Diese herausfordernde Aufgabe kann durch unterschiedliche Methoden unterstützt oder (teil-) automatisiert werden. Ein wichtiger Aspekt, der für Digitale Zwillinge beachtet werden muss, sind die Relationen zwischen den einzelnen Modellen der Dimensionen Mechanik, Elektrik und Software. Daher werden die gängigen Methoden zur Synchronisierung und Erstellung digitaler Modelle aus multiplen Dimensionen vorgestellt und die Relationsbildung betrachtet. Die Ankerpunktmethode als vielversprechendste Methode wird anschließen an einer Beispielsanlage untersucht und deren Grenze aufgezeigt. Basierend auf diesen Ergebnissen wird eine Methodik vorgestellt, die sich mit der Rekonstruktion der digitalen Modelle und deren Relationen beschäftigt. Dabei liegt der Fokus auf bereits existierenden, produktiv eingesetzten Produktionssystemen, sogenannte Brownfield Anlagen, als das häufigste Einsatzfeld.

Abstract

The Digital Twin of a production system can be very helpful for many use cases. This requires to ensure the synchronization of the Digital Twin with the real production system. This challenging task can be supported or (partially) automated by different methods. An important aspect to be considered for Digital Twins are the relations between the individual models of the domains mechanic, electric and software. Therefore, the common methods for synchronization and creation of digital models from various domains are presented considering relation reproduction. The anchor point method as the most promising method is then examined on an exemplary system and its limitation is outlined. Based on these results, a method is presented that deals with the reconstruction of the digital models and their relations.

The focus is on existing production systems already in productive use, so-called brownfield plants, as the most common field of application.

Digitale Zwillinge in der Fertigungsautomatisierung

Automatisierte Produktionssysteme werden heute immer komplexer weshalb die Entwicklung weiter zunehmend computergestützt und modellbasiert stattfindet [1]. Dabei werden über die Engineering Phase einer Anlage verschiedene digitale Modellierung der gesamten Anlage erstellt und Informationen in diese integriert. Dabei werden die verschiedenen Modelle aus den Fachbereichen Mechanik, Elektrik und Software aufeinander aufbauend erstellt. Die Gesamtheit der Modelle sowie deren Beziehungen zu einander formen die Basis für den Digitalen Zwilling der Produktionsanlage [2] und werden bereits heute für verschiedene Systemtests genutzt. Dieser Entwicklungsprozess ist im Idealfall frei von Einschränkungen durch die Umgebung, in der das Produktionssystem arbeiten soll [3]. Dem gegenüber stehen höhere Investitionskosten und Risiken bei der Entwicklung [4]. Es werden sequentiell die verschiedenen Domänen (Mechanik, Elektrik und Software) modelliert und schließlich die Anlage in Betrieb genommen [5]. Dabei werden meist während der Inbetriebnahme kleinere Fehler behoben indem der Aufbau im Vergleich zum Anlagenentwurf modifiziert wird.

Einen Schritt weiter geht die Nutzung eines Digitalen Zwillings, welcher eine Anlage über den kompletten Anlagenlebenszyklus begleitet und diese digital abbildet. Die exakten Bestandteile, Fähigkeiten und Eigenschaften schwanken zwischen den einzelnen Veröffentlichungen abhängig vom geplanten Einsatzzweck. Die Autoren in [2] haben einen Digitalen Zwilling herausgearbeitet, der die Kernpunkte unterschiedlicher Definitionen kombiniert und eine industrietaugliche Architektur definiert. Demnach besteht der Digitale Zwilling im Kern aus den digitalen Modellen und deren Relation zueinander, der sogenannten digitalen Replikation (siehe **Fehler! Verweisquelle konnte nicht gefunden werden.**). Die unterschiedlichen Modelle bilden einzelne mechatronische Komponenten in den unterschiedlichen Dimensionen Mechanik, Elektrik und Software ab, welche über ihre Relationen dimensionsübergreifend verbunden werden aber auch in der topologischen Dimension in Beziehung gesetzt werden.

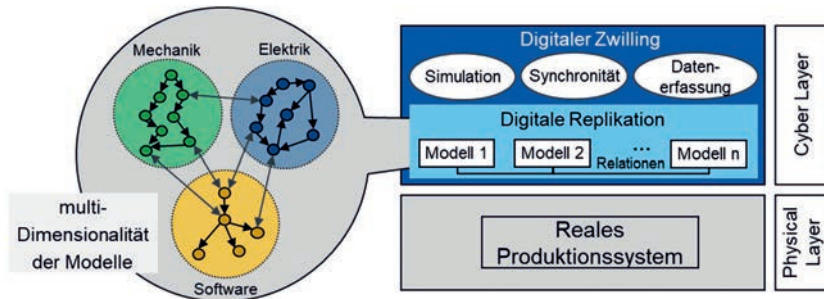


Bild 1: Bestandteile eines Digitalen Zwillings nach [2]

Diese wird durch die Erweiterung um eine Simulationsfähigkeit, einer aktiven Datenerfassung aus und einer Synchronisierung mit dem realen Produktionssystem zum Digitalen Zwilling. Durch diese Fähigkeiten ist ein Digitaler Zwilling nutzbar über den gesamten Lebenszyklus einer Produktionsanlage vom Beginn des Entwurfs über den Betrieb bis hin zum Um- oder Abbau. Die Simulationsfähigkeit ermöglicht den Einsatz eines Digitalen Zwillings für Untersuchungen ohne die reale Produktionsanlage dafür miteinzubeziehen, beispielsweise für eine virtuelle Inbetriebnahme der Anlage oder für Tests von Optimierungs- und Umbaumaßnahmen. Dadurch ergibt sich der signifikante Vorteil gegenüber einer herkömmlichen Rekonfiguration einer Anlage ohne Digitalen Zwilling, dass diese schneller wieder in den Regelbetrieb übergehen kann und damit profitabler ist. Damit eine digitale Replikation einer laufenden Produktionsanlage auch nach einiger Zeit noch dessen Digitaler Zwilling bildet, muss dieser kontinuierlich an Veränderungen in der Realität angepasst werden. Durch die explizite Eigenschaft der Synchronität eines Digitalen Zwillings kann dieser unabhängig der Phase des Anlagenlebenszyklus eingesetzt und seine Vorteile genutzt werden. Dies bedeutet auch, dass für die dynamische Abbildung der Anlage während des Betriebs aktiv alle Daten der Anlage zu erfassen und in den Digitalen Zwilling zu integrieren sind. Zu diesem Zweck muss der Digitale Zwilling über die Fähigkeit der aktiven Datenerfassung verfügen. Ein solcher Digitaler Zwilling kann für eine Vielzahl von Anwendungsfällen vorteilhaft während dem gesamten Anlagenlebenszyklus von Produktionsanlagen verwendet werden. Begonnen mit der Unterstützung durch einen Digitalen Zwilling im Entwurf von Produktionssystemen, über die Produktionsplanung/-steuerung, oder die Produktionsoptimierung. Die verschiedenen Einsatzbereiche profitieren durchweg davon, dass der Digitale Zwilling eine reale Hardware nachbildet und daher unabhängig von dieser in benötigter Geschwindigkeit wirklichkeitsgetreu simuliert werden kann. Eine prominente Anwendung ist hier die virtuelle Inbetriebnahme, die der Realisierung einer Anlage

vorweg gestellt ist und ohne eine wirkliche Montage auskommt. Dadurch lassen sich Fehler beispielsweise in der Steuerungssoftware frühzeitig erkennen, beheben und folglich Zeit und Kosten einsparen [6]. Ebenso lässt sich die virtuelle Inbetriebnahme während der Rekonfiguration gewinnbringend einsetzen und die Umbaudauer direkt sowie indirekt, beispielsweise über eine höhere Wiederanlaufsisicherheit, senken. In [7] wurde gezeigt, dass eine Zeitersparnis von bis zu ca. 60% bei der Rekonfiguration mit Digitalem Zwilling erzielt werden kann. Grundlegend hierfür ist der Aspekt der Synchronität eines Digitalen Zwillings mit der dazugehörigen realen Produktionsanlage. Durch regelmäßige Veränderungen und manuelle Anpassungen der Produktionsanlagen entstehen Inkonsistenzen zwischen den Modellen und dem realen System. Eine Zeitersparnis bietet der Digitale Zwilling daher nur, wenn dieser kontinuierlich und im besten Fall automatisiert an die reale Anlage angepasst wird. Des Weiteren sind aktuell viele ältere Produktionsanlagen im Betrieb für die keine bzw. wenige digitalen Modelle, sondern nur einige papierbasierte Dokumentationen verfügbar sind [8]. Es ist eine große Herausforderung alle Anlagenteile zu identifizieren, zu modellieren und in Beziehung zu setzen. Ein hoher Aufwand für die vollständige Erstellung oder Anpassung der Modelle sowie ein mangelndes, fachbereichsübergreifendes Wissen über das System sind Gründe für eine unvollständige digitale Anpassung. In [7] wurde die sogenannte Ankerpunktmethode verwendet, um diese Synchronisierung des Digitalen Zwillings mit der Produktionsanlage (teilweise) zu automatisieren. Diese und ähnliche Methoden setzen voraus, dass bereits ein Digitaler Zwilling besteht und einige Anforderungen zum Entwurf sowie der Realisierung berücksichtigt wurden. Dies ist speziell für Greenfield Entwicklungen ohne größeren Mehraufwand machbar, führt jedoch zu neuen Herausforderungen, wenn bestehende Produktionssysteme um einen Digitalen Zwilling ergänzt werden sollen.

Die Anpassung von bereits existierenden, produktiv eingesetzten Produktionssysteme, sogenannte Brownfield-Anlagen, ist heutzutage meist der Regelfall [3] und können nicht direkt durch diese Methoden synchronisiert oder modelliert werden. Brownfield Produktionsanlagen sind Produktionssysteme, die vor einigen Monaten bis Jahrzehnte entworfen, gebaut und in Betrieb genommen wurden. Meistens existiert keine oder eine unzureichende Dokumentation der Anlage in ihrer aktuellen Aufbauform, speziell digitale Modellierungen sind kaum verfügbar [8, 9]. Aufgrund der bestehenden Erfahrung und den vorhandenen Ressourcen bieten sie jedoch einen grundlegenden Vorteil gegenüber kompletten Neuentwicklungen. Durch die Wiederverwendung von bestehenden Anlagen(-teilen) und Gebäudeinstallationen sinken die notwendige Investition im Vergleich zur Erstellung eines Produktionssystems im Greenfield [10]. Durch die gesammelte Erfahrung mit bestehenden Anlagenressourcen können einige Fehler-

quellen vermieden und Schwierigkeiten im Produktionsablauf reduziert werden. Dem gegenüber steht die Herausforderung, dass Entwicklungen im Brownfield durch die Bedingungen und Einschränkungen aus der bestehenden Produktion erschwert werden. Aus Kompatibilitätsgründen werden Anpassungen, im Vergleich zu einer Neuentwicklung, technisch eingeschränkt. Außerdem darf der laufende Produktionsbetrieb nicht beziehungsweise möglichst wenig unterbrochen werden, um Gewinneinbußen zu minimieren.

Die Variabilität der Produkte, die diese Produktionssystem herstellen, ist in den letzten Jahren deutlich gestiegen und zeitgleich die Produktionsdauer eines Produkts gesunken [11]. Dieser Trend wird sich voraussichtlich auch in den nächsten Jahren fortsetzen und dazu führen, dass die Rekonfiguration dieser Brownfield-Anlagen häufiger wird. Daher nimmt auch die Relevanz der Unterstützung von Rekonfigurationen weiter zu, um die Anlagenbetriebsdauer im Vergleich zum Anlagenstillstand zu maximieren. Um diese Unterstützung mit einem Digitalen Zwilling liefern zu können, ist es notwendig diesen für Brownfield-Anlagen zu erstellen und anschließen synchron zu halten.

1. Ansätze zur Aktualisierung der Modelle eines realen Assets

In der Literatur finden sich bereits einige Ansätze, die zur Erstellung und Synchronisierung eines Digitalen Zwillings eingesetzt werden können. Dazu gehören bildgebende Verfahren, Scans und Analysen von papierbasierten Dokumentationen, von Kommunikationsnetzwerke oder Konfigurationsdateien oder die Analyse der Steuerungssoftware, wie in der bereits erwähnten Ankerpunktmethode. Jedes dieser Verfahren wurde mit dem Fokus auf einen speziellen Fachbereich entworfen und performt dort besonders gut (siehe **Fehler! Verweisquelle konnte nicht gefunden werden.**). Die bildgebenden Verfahren, wie Photogrammetrie oder spezielle Laserscans, benötigen eine Aufnahme der aktuellen realen Anlage und zielen darauf ab, ein mechanisches Modell der Anlage zu gewinnen [12]. Aufnahmen für eine Photogrammetrie kann mit einer normalen Kamera gemacht werden. Aus mehreren Bildern eines Objekts aus verschiedenen Perspektiven werden dann mittels einer Photogrammetrie-Software die drei dimensionale Form zurückgerechnet. Laserscans benötigen spezielle, kostspielige Scannvorrichtungen die mit einem Laser die Anlage abtasten und diese in einer hochpräzisen Punktwolke darstellen [13, 14]. In beiden Fällen werden nur die mechanischen Aspekte der Anlage abgebildet. Ebenfalls im Bereich der Mechanik jedoch auf einer höheren Abstraktionsebene bewegt sich die Informationsgewinnung über Konfigurationsdaten, speziell von Robotern. Diese beinhalten die Arbeitspositionen und ermöglichen dadurch eine Berechnung des Anlagenlayouts bestehend aus den Robotern und deren Peripherie [15]. Die Analyse der Kommu-

nikation bzw. des Kommunikationsnetzwerk ermöglicht ebenfalls eine Rekonstruktion der mechanischen Topologie mit der Beschränkung auf die kommunikationsfähigen Anlagenkomponenten. Über die Analyse der Teilnehmer in einem Kommunikationsnetzwerk und dem Inhalt der ausgetauschten Nachrichten können jedoch keine exakten Geometrien, sondern nur die topologische Struktur nachgebildet werden. Dafür ermöglicht diese Methode teilweise Rückschlüsse auf die elektrischen Verbindungen und logischen Abläufe zwischen Teilnehmern [16]. Als ausführliche Informationsquelle über die elektrische Anlagenteile eignen sich papierbasierte Dokumentationen, die meist direkt an den Produktionsanlagen hinterlegt und bei Änderungen während des Anlagenlebenszyklus mit einem Rotstift angepasst werden. Diese Dokumentation können eingescannt und mittels Computer Vision als Schaltpläne digitalisiert werden [17]. Basierend auf diesen elektrischen Modellen kann weiter die Signalliste als grundlegende Information für die Steuerungssoftware abgeleitet werden.

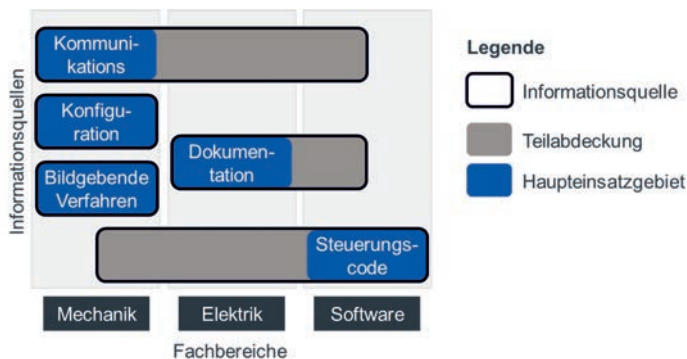


Bild 2: Einordnung der bisher erforschten Methoden zur Erstellung und Synchronisierung eines Digitalen Zwillings in die Fachbereiche des Anlagenengineerings

Der Steuerungscode ist im Vergleich zu den anderen Quellen bereits digital und muss nicht erst erfasst werden. Da dieser im laufenden Betrieb der Produktionsanlage eingesetzt wird und von den beiden anderen Fachbereichen beeinflusst ist, muss dieser stets an die vorhandene Anlage angepasst sein [18]. Daher ist der Steuerungscode eine gute Quelle, um aus den implizite enthaltenden Informationen Rückschlüsse auf das reale Produktionssystem zu ziehen. Eine konkrete Methode ist die Ankerpunktmethode, welche über sogenannte Ankerpunkte in der Steuerungssoftware Änderungen im Produktionssystem rekonstruiert [18]. Der Einsatz dieser einzelnen Methoden ist in unterschiedlichem Ausmaß und Detailgrad auf die Modellsynchronisierung oder Rekonstruktion einzelner Fachbereiche ausgelegt. Teilweise

betrachten sie auch die Verbindung zu einer oder beiden anderen Fachdomänen. Die Verbindungen sind ein wichtiges Element, um Aussagen über die Relationen zwischen den Modellen treffen zu können. Die Ankerpunktmethode bietet die vielversprechendsten Ergebnisse, um diese Relationen zu erfassen.

2. Synchronisierung des Digitalen Zwillings über die Software-Domäne

Die Ankerpunktmethode zur Synchronisierung eines Digitalen Zwillings über den Steuerungscode baut grundsätzlich auf sogenannte Ankerpunkte [18]. Als Ankerpunkt werden markante Elemente in den drei Domänen Mechanik, Elektrik, Software definiert, die in einer festen Beziehung zu Ankerpunkten in einer der anderen Domänen stehen. Über die Anwendung einer Namenskonvention für alle Ankerpunkte, wird eine umfassende Analyse der Relationen zwischen den Domänen ermöglicht. Diese Namenskonvention basiert auf gängigen Standards im Bereich der SPS-Programmierung und Anlagenbenennung muss während der Entwicklung berücksichtigt oder nachträglich manuell nachgezogen werden. Voraussetzung für die Anwendbarkeit der Ankerpunktmethode ist, dass der Digitale Zwilling und der dazugehörige SPS Code von einem vergangenen Zeitpunkt existiert. Findet nun eine Änderung an der Anlage statt, welche aber nicht im Digitalen Zwilling nachgezogen wird, kann diese Methode eingesetzt werden, um den Digitalen Zwilling wieder mit der Anlage zu synchronisieren (siehe **Fehler! Verweisquelle konnte nicht gefunden werden.**). Hierfür wird die der alte und der aktuelle SPS Code auf Änderungen verglichen und die Art der Änderung ausgewertet. Dadurch lassen sich neue und entfernte mechatronische Einheiten identifizieren. Anschließend werden die von einer Änderung betroffenen Ankerpunkte untersucht und deren interdisziplinären Abhängigkeiten analysiert, sodass die resultierenden Änderungen in der Topologie der mechanischen und elektrischen Modelle bestimmt werden können. Abschließend werden die notwendigen Änderungen den relevanten Modellen des Digitalen Zwillings zugeordnet.

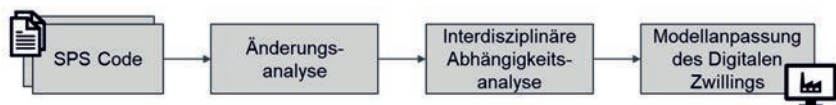


Bild 3: Schritte der Ankerpunktmethode zur Synchronisierung eines Digitalen Zwillings

In [7] konnte gezeigt werden, dass für Anlagen, welche die Anforderungen der Ankerpunktmethode erfüllen, eine deutliche Zeitersparnis für nachfolgende Rekonfigurationen erzielt werden kann. Die Anforderungen der Methode sind, dass bereits der Digitale Zwilling einer vorange-

henden Form der Anlage und der zugehörige SPS Code existiert. Beim Entwurf der Anlagenmodelle und der Steuersoftware muss dabei eine Namenskonvention eingehalten werden, um die benötigten Informationen in den Ankerpunkten zu codieren.

Daher ist diese Methode vor allem für Greenfield-Anlagen geeignet, da auf die Einhaltung dieser Anforderungen mit geringem Aufwand geachtet werden kann. Brownfield-Anlagen besitzen meistens nicht den benötigten Digitalen Zwilling, die vergangenen Versionen des SPS Codes und sind nicht nach der Namenskonvention benannt.

3. Erstellung von Modellen und deren Relationen mit Bezug zur Software

Die Erstellung von digitalen Modellen und deren Relationen wird, in Anlehnung an die Ankerpunktmethodik, ausgehend vom Steuerungscode abgeleitet, da dieser an die Anlage angepasst sein muss. Außerdem wird die SPS durch die beiden vorgelagerten Domänen beeinflusst und bietet daher Informationen über diese. Auf diesen Informationen wird weiter aufgebaut und daraus der Digitale Zwilling einer Brownfield-Anlage erstellt. Dabei müssen die drei Fachbereiche und die draus resultierenden unterschiedlichen Modelle als Dimensionen des Digitalen Zwillings zusammengeführt werden. Dadurch wird der hohe manuelle Aufwand und das notwendige Toolwissen der Ingenieure herabgesetzt. In den vorgestellten Ansätzen aus Kapitel 1 werden lediglich Domänen-spezifische Modelle betrachtet und deren Verknüpfung den Ingenieuren überlassen bzw. explizit über eine Namenskonvention in den Modellen hinterlegt. In dieser neuen Methodik wird ebenfalls vom SPS Code ausgegangen jedoch die Bindung an eine Namenskonvention aufgelöst, indem weitere Informationsquellen in den Analyseprozess mit einbezogen werden, sodass diese Methodik auch auf Brownfield Anlagen anwendbar ist. Die Methodik basiert auf drei Datenquellen:

1. Der aktuelle SPS Code und die SPS Konfiguration
2. Positionsdaten aus der realen Produktionsanlage
3. Prozessdaten der Produktionsanlage

Diese drei Quellen werden mit unterschiedliche Analysemethoden untersucht und auf die Elemente der SPS abgebildet und dadurch ein zusammenhängendes Informationsmodell erstellt (siehe **Fehler! Verweisquelle konnte nicht gefunden werden.**). Die Grundlage bildet die Analyse des SPS Codes und der SPS Konfiguration. Dazu wird der SPS Code in einem herstellerunabhängigem XML Format exportiert und für die Analyse verwendet. Für den Import und die Interpretation der XML Elemente wird eine XML schema definition (XSD) Datei als Metamodel benötigt, welche abhängig vom Hersteller und deren XML Exportstruktur leicht variieren kann. Diese Vorgehen ist angelehnt an die SPS Code Analyse in der Ankerpunktme-

thode [18] und ermöglicht eine herstellerunabhängige Analyse der Elemente, da alle XML Dateien mit den XSD Dateien in eine einheitliche Datenstruktur überführt werden können. Anschließend werden die importierten Elemente unabhängig von deren Namen nach Elementtyp, Abhängigkeit zu anderen Elementen und ähnlichen strukturellen Informationen untersucht und in eine Graph-basierte Informationsstruktur überführt. Die Modelle der unterschiedlichen Fachbereiche werden als Knoten eines Labeled Property Graphs abgebildet und durch deren Relationen als Kanten verbunden. Dabei werden alle Informationen aus den Fachbereichen als Eigenschaften dem entsprechenden Knoten zugeordnet. Diese Informationsdarstellung kann dann aus den weiteren Quellen gespeist werden und um die Eigenschaften und Relationen an der Schnittstelle zur mechanischen und elektrischen Domain ergänzt werden. Dazu wird unter anderem aus einer gemeinsamen Analyse von Positions- und Sensorsignalen eine ungefähre Positionsbestimmung der Sensoren ermittelt. Diese kann anschließend über den beliebigen Signalnamen im SPS Code den Konten übertragen und neue Relationen über die mechanischen Beziehungen ergänzt werden. Über eine separate Analyse der Positions- bzw. Prozessdaten lassen sich die ungefähren Positionen mechatronischer Baugruppen und deren Orientierung abschätzen, zeitliche sowie logische Korrelationen zur Nachbildung des Prozessablaufs abbilden und die Relevanz von Codefragmenten für den Regelbetrieb über die Nutzungshäufigkeit einschätzen. Davon abgesehen können weitere Informationsquellen integriert werden, die aufgrund der Relationsstruktur und der enthaltenen Informationen einzelne Komponenten identifizieren und erweitern können. Die Graph-basierte Informationen können abschließend verarbeitet und in softwarespezifische Modelle des Digitalen Zwillings übertragen werden. Dieser Vorgang kann sofern entsprechende offene Schnittstellen zur Verfügung stehen (teil-)automatisiert vorgenommen werden.

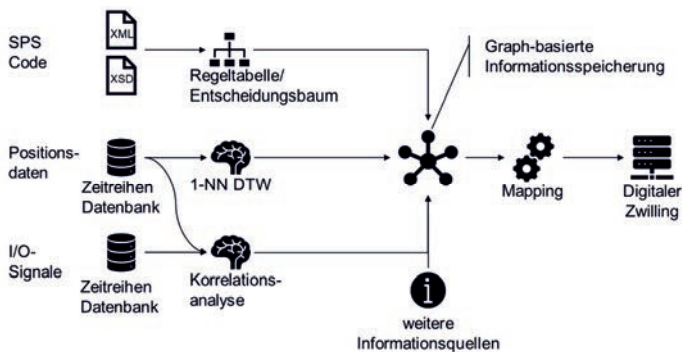


Bild 4: Methodik zur Erstellung multi-dimensional verknüpfter Modelle aus automatisierten Brownfield Anlagen

Diese Methodik wird am Beispiel des intelligenten Lagers der Flexiblen Produktionsanlage in der ARENA2036 (Active Research Environment for the Next Generation of Automobile) erprobt. Die flexible Produktionsanlage besteht aus drei Komponenten, einer Schweißmaschine, einem mobilen Roboter und dem intelligenten Lager. Das Lager stellt verschiedene Rohteile aus Metall in vier unterschiedlichen Werkstückträgern dem Roboter zur Verfügung. Dieser entnimmt die Werkstückträgern aus der unteren Lagerebene, nimmt die Bauteile heraus und stellt die leeren Werkstückträgern zurück in die obere Ebene des Lagers. Der Roboter fügt in mehreren Schritten aus den Metallteilen ein Modellauto und bringt dieses zur Schweißmaschine.

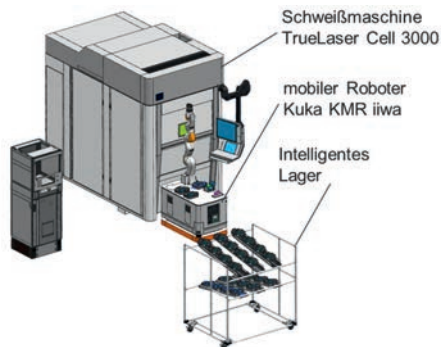


Bild 5: Modell der Flexiblen Produktionsanlage in der ARENA2036

Das intelligente Lager (siehe **Fehler! Verweisquelle konnte nicht gefunden werden.**) besteht aus der Materialausgabe (untere Ebene) und der Rücknahme der leeren Werkstückträger (obere Ebene). In jeder Ebene sind vier Lagerreihen, eine je Werkstückträgertyp, mit jeweils vier Lagerplätzen montiert. Die Werkstückträger auf den einzelnen Lagerplätzen werden durch Lichtschranken detektiert, durch Hubmagnete voneinander separiert und die zugeführten Werkstückträger durch ein RFID System eindeutig identifiziert. Die Lagerverwaltung, die Steuerung des Materialflusses anhand der Lichtschranken und die Kommunikation der Lagerbelegung an die anderen Teilnehmer wird durch eine SPS im Lager übernommen.



Bild 6: intelligentes Lager in der ARENA2036

Mittels der vorgestellten Methode wird die Digitalisierung des Lagers über seine SPS unterstützt. Dazu wird der SPS Code analysiert sowie die Prozessdaten des Lagers über OPC UA erfasst und in einer Zeitreihendatenbank gespeichert. Ein Echtzeit-Lokalisierungssystem in der ARENA2036 liefert die benötigten Positionsdaten für die Nachverfolgung des Lagerprozesses. Die gesammelten Daten und daraus generierten Informationen vereinfachte die Erstellung eines Digitalen Zwillings des Lagers, sodass ein Ingenieur den Lagerprozess softwareseitig, beispielsweise an größere Werkstückträger, anpassen kann.

4. Zusammenfassung und Ausblick

Die Relevanz eines Digitalen Zwillings im Bereich der Rekonfiguration von Produktionsanlagen und deren Software nimmt zu. Um die Rekonfiguration mit einem Digitalen Zwilling unterstützen zu können, muss dieser für Brownfield-Anlagen als häufigste Ausgangssituation erstellt werden. Die bestehenden Methoden liefern hier nur Fragmente und betrachten meist nicht die wichtigen Relationen zwischen den Modellen des Digitalen Zwillings oder erheben einige Voraussetzungen, die von Brownfield Anlagen nicht erfüllt werden können.

Daher wurde eine Methodik vorgestellt mit der die Modellerstellung einer Anlage und deren Relationen ausgehende vom Steuerungscode bewältigt werden kann. Dazu wurden drei Datenquellen vorgestellt und wie diese zur Informationsgewinnung über eine aktuell produktiv eingesetzte Anlage beitragen. Außerdem wurde eine Graph-basierte Speicherung der Informationen vorgeschlagen, in der die unterschiedlichen Informationen anhand ihrer Abhängigkeiten hinzugefügt werden. Dadurch werden die Relationen zwischen den Modellen ebenfalls abgebildet und beispielsweise für des intelligente Lager die SPS-Elemente einer bestimmten Lichtschranke identifiziert sowie deren mechanische Anordnung bestimmt werden.

In der weiteren Forschung wird die Graph-basierte Informationsspeicherung genauer untersucht und speziell die Möglichkeiten von Graph Analysen zur Gewinnung weiterer Informationen betrachtet. Außerdem werden für die einzelnen, eingesetzten Informationsquellen detaillierte Proof-of-Concept erbracht. Hierbei wird zukünftig auch noch untersucht, welche weiteren Informationsquellen in die Methodik integriert werden können und wo die Grenze dieser Integrationsform liegt.

5. Literaturverzeichnis

- [1] R. Haberfellner, O. L. de Weck, and E. Fricke, *Systems engineering: Fundamentals and applications*. Cham: Springer International Publishing, 2019.
- [2] B. Ashtari Talkhestani *et al.*, "An architecture of an Intelligent Digital Twin in a Cyber-Physical Production System," *at - Automatisierungstechnik*, vol. 67, no. 9, pp. 762–782, 2019, doi: 10.1515/auto-2019-0039.
- [3] S. R. Bader, C. Wolff, M. Vössing, and J.-P. Schmidt, "Towards Enabling Cyber-Physical Systems in Brownfield Environments," in *Lecture Notes in Business Information Processing*, vol. 331, *Exploring Service Science: 9th International Conference, IESS 2018, Karlsruhe, Germany, September 19-21, Proceedings*, G. Satzger, L. Patricio, M. Zaki, N. Kühl, and P. Hottum, Eds., Cham: Springer International Publishing, 2018, pp. 165–176.
- [4] T. Lehtonen, J. Pakkanen, M. Lanz, J. Järvenpää, and R. Tuokko, "Brownfield process for developing of product families," in *Proceedings of the 18th International Conference on Engineering Design*, Copenhagen, Denmark, 2011, pp. 248–257.
- [5] A. Strahilov and H. Hämmerle, "Engineering Workflow and Software Tool Chains of Automated Production Systems," in *Multi-Disciplinary Engineering for Cyber-Physical Production Systems: Data Models and Software Solutions for Handling Complex Engineering Projects*, S. Biffl, A. Lüder, and D. Gerhard, Eds., Cham: Springer International Publishing, 2017, pp. 207–234.
- [6] M. Schamp, S. Hoedt, A. Claeys, E. H. Aghezzaf, and J. Cottyn, "Impact of a virtual twin on commissioning time and quality," *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1047–1052, 2018, doi: 10.1016/j.ifacol.2018.08.469.
- [7] B. Ashtari Talkhestani, D. Braun, W. Schloegl, and M. Weyrich, "Qualitative and quantitative evaluation of reconfiguring an automation system using Digital Twin," *Procedia CIRP*, vol. 93, pp. 268–273, 2020, doi: 10.1016/j.procir.2020.03.014.
- [8] M. Strube, "Modellgestützte Modernisierungsplanung industrieller Automatisierungslösungen," Dissertation, Institut für Automatisierungstechnik, Helmut Schmidt Universität, Hamburg, 2014.
- [9] J. Kiefer, "Mechatronikorientierte Planung automatisierter Fertigungszellen im Bereich Karosserierohrbau," Dissertation, Lehrstuhl für Fertigungstechnik/CAM, Universität des Saarlandes, Saarbrücken, 2007.
- [10] A. Vaughn, P. Fernandes, and J. T. Shields, "Manufacturing system design framework manual," pp. 1–64, 2002. [Online]. Available: <http://hdl.handle.net/1721.1/81902>

- [11] S. Biffl, D. Gerhard, and A. Lüder, "Introduction to the Multi-Disciplinary Engineering for Cyber-Physical Production Systems," in *Multi-Disciplinary Engineering for Cyber-Physical Production Systems: Data Models and Software Solutions for Handling Complex Engineering Projects*, S. Biffl, A. Lüder, and D. Gerhard, Eds., Cham: Springer International Publishing, 2017, pp. 1–24.
- [12] F. Buonomici, M. Carfagni, R. Furferi, L. Governi, A. Lapini, and Y. Volpe, "Reverse engineering modeling methods and tools: a survey," *Computer-Aided Design and Applications*, vol. 15, no. 3, pp. 443–464, 2018, doi: 10.1080/16864360.2017.1397894.
- [13] E. Lindskog, J. Vallhagen, and B. Johansson, "Production system redesign using realistic visualisation," *International Journal of Production Research*, vol. 55, no. 3, pp. 858–869, 2017, doi: 10.1080/00207543.2016.1218085.
- [14] G. Erdős, T. Nakano, G. Horváth, Y. Nonaka, and J. Váncza, "Recognition of complex engineering objects from large-scale point clouds," *CIRP Annals*, vol. 64, no. 1, pp. 165–168, 2015, doi: 10.1016/j.cirp.2015.04.026.
- [15] D. Braun, F. Biesinger, N. Jazdi, and M. Weyrich, "A concept for the automated layout generation of an existing production line within the digital twin," *Procedia CIRP*, vol. 97, pp. 302–307, 2020, doi: 10.1016/j.procir.2020.05.242.
- [16] F. Biesinger, D. Meike, B. Kraß, and M. Weyrich, "A digital twin for production planning based on cyber-physical systems: A Case Study for a Cyber-Physical System-Based Creation of a Digital Twin," *Procedia CIRP*, vol. 79, pp. 355–360, 2019, doi: 10.1016/j.procir.2019.02.087.
- [17] G. Koltun, F. Maurer, A. Knoll, E. Trunzer, and B. Vogel-Heuser, "Information Retrieval from Redlined Circuit Diagrams and its Model-Based Representation for Automated Engineering," in *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, Washington D.C., USA, 2018, pp. 3114–3119.
- [18] B. Ashtari Talkhestani, "Methodik zur Synchronisierung der Modelle des Digitalen Zwillings automatisierter Systeme," Dissertation, Institut für Automatisierungstechnik und Softwaresysteme, Universität Stuttgart, Stuttgart, 2020.

Augmented Reality für hybride Programmierung von Industrierobotern

Augmented Reality for hybrid programming of industrial robots

M.Sc. **L. Wulff**, ICARUS Consulting GmbH, Lüneburg;
M.Sc. **M. Brand**, Prof. Dr.-Ing. **T. Schüppstuhl**,
Institut für Flugzeug-Produktionstechnik, Technische Universität
Hamburg

Kurzfassung

Augmented Reality (AR) liefert Methoden, um die Programmierung und Inbetriebnahme von Industrierobotern durch Einblenden digitaler Werkzeuge im dreidimensionalen Arbeitsraum zu unterstützen. Ein besonders hohes Potential für synergetische Effekte liegt dabei in der Kombination der AR mit der klassischen Offlineprogrammierung (OLP). Mithilfe von AR können digitale Modelle und Simulationsbibliotheken der OLP im *Shopfloor* verwendet werden, um Aufgaben wie Evaluation des Zell-Layouts oder Korrekturen an der Position einzelner Pfadpunkte vorzunehmen. In diesem Beitrag präsentieren wir ein System für die hybride Programmierung von Industrierobotern. Um auf den schnelllebigen Markt aktuell verfügbarer AR-Systeme reagieren zu können, haben wir eine erweiterbare Systemarchitektur konzipiert, in der das eingesetzte AR-System als ein austauschbares Plugin integriert ist. Diese Modularisierung erlaubt es durch Auswahl eines geeigneten AR-Systems optimal auf den spezifischen Anwendungsfall zu reagieren und verbessert die Anwendbarkeit der hybriden Roboterprogrammierung.

Abstract

Augmented Reality (AR) offers unique visual and interactive methods to assist workers with programming and the commission of industrial robots. The combination of AR and industrially established offline programming (OLP) bears a highly synergetic potential. Digital models and simulation libraries, commonly used in OLP, can be introduced to a shopfloor worker more feasibly with the help of AR. In turn this can aid with tasks like evaluation of cell layouts and spatial correction of points in robot paths.

In this article we present a system to implement this Hybrid Robot Programming (HRP) and evaluate it in a test scenario. In response to the fast-moving market of available AR systems,

we embedded the system in an expendable architecture concept. Thus, integrating AR systems as interchangeable plugins. This modularization enables us to keep the core functionality of our HRP system while offering the ability of switching the used AR system based on the specific requirements of the treated use case.

1. Einleitung

Mit der Fähigkeit digitale und reale Elemente auf beliebige Weise zu kombinieren, hat die Augmented Reality (AR) das Potential, als innovative Mensch-Maschine-Schnittstelle in Produktionsprozessen eingesetzt zu werden [1]. Ein besonders hohes Maß an Aufmerksamkeit erhält dabei die Anwendung der AR bei der Programmierung von Industrierobotern. Durch gezieltes Einblenden digitaler Elemente kann diese intuitiver gestaltet werden. Vor allem Elemente wie Trajektorie und Roboterkonfiguration erhöhen das Prozessverständnis und erlauben in Kombination mit multimodalen Bedienwerkzeugen eine Programmoptimierung direkt in der realen Anlage [2].

2. Stand der Forschung und Technik

Die Einsetzbarkeit der AR zur Roboterprogrammierung ist Thema verschiedener wissenschaftlicher Arbeiten. Diese umfassen dabei die Nutzung unterschiedlicher Eingabemechanismen wie getrackte Objekte [3], Sprachbefehle [4] oder Gesten [5], um Pfadpunkte eines Roboterprogramms zu erzeugen und durch ein AR-Display zu evaluieren.

Die Programmierung von AR wird dabei typischerweise durch eine Reduktion der Komplexität im Vergleich zu On- und Offlineprogrammierung motiviert. Die entwickelten Applikationen werden daher im Besonderen im KMU Sektor, sowie der flexibleren Programmierung bei Fertigung mit kleinen Losgrößen betrachtet.

Lambrecht et al. präsentieren in diesem Umfeld eine Schnittstelle zum Programmaustausch mit dem industriell etablierten Offlineprogrammiersystem (OLP-System) Tecnomatix von Siemens [5]. Diese erlaubt es, mit AR erzeugte Programme in ein verallgemeinertes Format zu überführen, welches von OLP Systemen importiert werden kann. AR kann damit zu einem gewissen Grad komplementär zur OLP eingesetzt werden.

Diese Kombination von OLP-Systemen mit der AR zeigt das Potential für unterschiedliche synergetische Effekte. Typische OLP-Systeme nutzen modellierte, digitale Anlagenmodelle und proprietäre Simulationsfunktionen, um Roboterprogramme zu entwickeln und zu optimieren. Aufgrund von Abweichungen zwischen digitaler und realer Anlage sind jedoch typischerweise Anpassungen notwendig [6]. Die AR erlaubt es in diesem Fall, die vorhandenen digitalen Mo-

delle direkt in der dreidimensionalen Anlage darzustellen und damit Abweichungen aufzuzeigen. Des Weiteren können geeignete AR-Werkzeuge mit den etablierten Simulationsfunktionen kombiniert werden, um die Fähigkeiten klassischer OLP-Systeme direkt in der realen Anlage einzusetzen. Da dabei direkt auf der Datenbasis des eingesetzten OLP-Systems gearbeitet wird, werden alle Korrekturen automatisch in diese überführt und für anschließende Prozesse wie bspw. der Qualitätskontrolle vorgehalten. Wir bezeichnen diese komplementäre Anwendung von AR und OLP als hybride Roboterprogrammierung (HRP).

3. Hybride Roboterprogrammierung

Vereinfacht verfolgt die HRP die Motivation, die Werkzeuge der klassischen OLP mithilfe der AR auf neuartige Weise in realen Anlagen einzusetzen. Für diesen Beitrag dient die Inbetriebnahme und Optimierung offline erstellter Programme als Anwendungsszenario. Die AR liefert hier Methoden, um Abweichungen zwischen digitaler und realer Anlage aufzuzeigen und zu korrigieren.

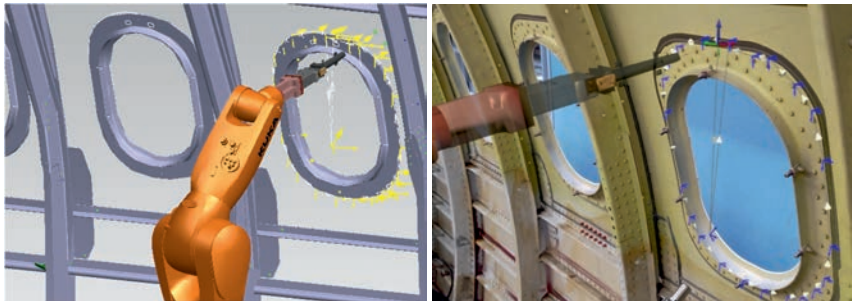


Bild 1: Datenbasis im OLP-System (links); Ansicht im AR-System (rechts)

In Bild 1 ist die ursprüngliche Darstellung im Simulationssystem (links) der Sicht eines Anwenders in der realen Anlage durch das AR-System (rechts) gegenübergestellt. AR ermöglicht es, die in der Realität unsichtbaren Pfadpunkte des Roboters in der realen Anlage visuell wahrzunehmen.

Durch Nutzung der simulativen Funktionen der OLP kann der Anwender die Erreichbarkeit der einzelnen Punkte prüfen, die Konfiguration des Roboters beim Anfahren dieser evaluieren und Korrekturen vornehmen. Dieser direkte Abgleich erlaubt es dem Anwender, die offline erstellten Programme anhand der realen Randbedingungen zu bewerten, sodass etwaige Kollisionen mit nicht digitalisierten Komponenten ohne Risiko erkannt und korrigiert werden können.

Das eingesetzte Medium der AR bietet hier durch seine inhärente Flexibilität zusätzliche designtechnische Freiheitsgrade.

Elementar wird AR anhand von drei Eigenschaften charakterisiert [7]:

1. Kombination von Realität und Virtualität
2. Interaktion in Echtzeit
3. Dreidimensionale Lokalisierung und Registrierung aller Akteure

Da die Umsetzung der AR auf vielfältige Weise möglich ist, kann ein Systemdesigner abhängig von den Anforderungen des Anwendungsszenarios das optimale System auswählen.

Verlangt ein Anwendungsszenario bspw. besonders hohe Anforderungen an Genauigkeit und Stabilität, bietet sich die Verwendung eines statisch montierten Projektors an. Soll dagegen die Analyse von Pfadpunktattributen durch eine AR-gestützte Navigation beschleunigt werden, bietet es sich ggf. an ein mobiles am Kopf angebrachtes AR-System einzusetzen, sodass der Anwender die Hände für andere Tätigkeiten frei hat. Die Möglichkeit das passende AR-System für das spezifische Szenario auszuwählen ist einer der Vorteile des Mediums AR.

Um diesen Vorteil für die HRP zu erhalten, haben wir uns dazu entschieden, unser AR-Bedieninterface in ein übergeordnetes modular aufgebautes System zu integrieren. Das eingesetzte AR-System ist damit ein Plugin, welches je nach Anforderungen des Anwendungsfalls gezielt ausgewählt werden kann. Vor allem im aktuell schnelllebigen Markt der AR-Systeme, kann das entwickelte Bedieninterface auch mit zukünftigen Generationen von AR-Systemen eingesetzt werden.

4. Systemdesign

Eine modulare Systemarchitektur basiert auf der Definition funktionaler Submodule, welche durch eindeutige Schnittstellen miteinander verbunden werden. Den Ansätzen der Clean Architecture folgend, definieren wir fünf funktionale Submodule, welche durch ein Kernmodul orchestriert werden [8]. Ebenso wie das eingesetzte AR-System wird jedes Submodul als Plugin betrachtet und ist damit grundsätzlich austauschbar. Eine Übersicht der einzelnen Module ist in Bild 2 gegeben.

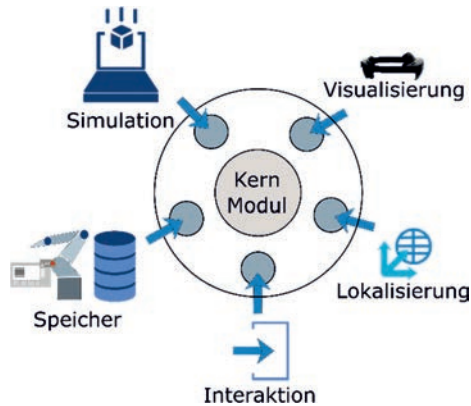


Bild 2: Modulübersicht

Kernmodul

Das Kernmodul steuert als Verwaltungsorgan die weiteren Module, um die Aufgaben der HRP abzubilden. Es interpretiert die Eingaben des Interaktionsmoduls anhand der durch das Lokalisierungsmodul bereitgestellten Position, um die Darstellung im Visualisierungsmodul auf Basis der vom Simulationsmodul bereitgestellten Werkzeuge zu gestalten. Die modifizierten Programme können dann in der Simulationsdatenbasis gespeichert und in die reale Anlage überführt werden.

Visualisierung

Dieses Modul steuert das Rendering von virtuellen Elementen auf dem eingesetzten AR-Display. Die verallgemeinerte Schnittstelle definiert virtuelle Elemente dabei anhand der äußerlichen Gestalt, der räumlichen Position sowie der zeitlichen Veränderung. Diese Abstraktion erlaubt es, sowohl Elemente eines Bedienerinterfaces wie Texte oder Buttons, als auch Zellkomponenten wie Roboter oder Werkzeugteile darzustellen.

Interaktion

Dieses Modul erfasst und verarbeitet Anwendereingaben, um diese dem Kernmodul bereitzustellen. Das Kernmodul interpretiert getätigte Eingaben im Kontext des eingesetzten Eingabesystems. Auf diese Weise können unterschiedliche Eingabesysteme unabhängig voneinander eingesetzt werden.

Lokalisierung

Dieses Modul bestimmt die aktuelle Position des Anwenders in Bezug auf ein definiertes gemeinsames Referenzsystem und stellt diese dem Kernmodul zur weiteren Verwendung bereit.

Speichermodul

Das Speichermodul dient dazu Programm- und Prozessinformationen sowie Metadaten zu verarbeiten und in dedizierten Speichern vorzuhalten. Die Schnittstelle bedient sich dabei standardisierter Formate für Textdaten und Geometrien wie JSON, XML, OBJ und JT.

Simulationsmodul

Das Simulationsmodul liefert Methoden, um Roboterprogramme im Sinne einer OLP zu verarbeiten. Die Schnittstelle ist daher in Form einer API konzipiert, welche Operationen wie die Verschiebung von Pfadpunkten, das Setzen von Attributen oder eine Erreichbarkeitsprüfung in Form von spezialisierten Werkzeugen erlaubt.

Dabei verwendet das Simulationsmodul die Funktionalität eines beliebigen Simulationssystems. Es ist damit möglich unterschiedliche auch proprietäre Simulationssysteme für die Berechnung von Taktzeit oder Kollision einzusetzen.

5. Umsetzung im Forschungsprojekt MiReP

Um die Funktion der konzipierten Systemarchitektur zu evaluieren, wurde diese im Rahmen des Forschungsprojektes MiReP – Mixed Reality Programming in einer Applikation umgesetzt. Im Forschungsprojekt MiReP wird untersucht, inwieweit mithilfe der AR die Integration von Automatisierungslösungen in Produktionsprozessen der Luftfahrtindustrie unterstützt werden kann. Als Beispielszenario wurde die Inbetriebnahme eines offline erstellten Roboterprogramms ausgewählt und an einem Mockup untersucht.



Bild 3: Mockup (links); Wahrnehmung des Anwenders im Mockup (rechts)

Im Mockup, gezeigt in Bild 3 (links), werden eine HoloLens 2, ein Microsoft Controller und ein Kuka KR Sixx Industrieroboter eingesetzt. Als Simulationssystem wird Process Simulate verwendet, welches über einen Netzwerkrechner an das Gesamtsystem angebunden ist. Die Robotersteuerung ist ebenfalls über eine Netzwerkschnittstelle angeschlossen, sodass Programmdateien im Speicher gelesen und geschrieben werden können.

Die Applikation gliedert sich in vier Phasen: „Registrierung und Zellsetup“, „Programmauswahl“, „Modifikation“ und „Upload“.

Registrierung und Zellsetup

Zu Beginn ist es notwendig, das AR-System über ein gemeinsames Koordinatensystem in den dreidimensionalen Arbeitsraum zu registrieren. Dieser Prozess kann entweder manuell oder mithilfe eines eingemessenen Markers ausgeführt werden. Bei manueller Registrierung verwendet der Anwender einen Microsoft Controller, um ein virtuelles Gestell des eingesetzten Roboters über den realen Körper zu legen. Bei Verwendung eines Markers wird die Position des Markers durch das Kamerasystem der HoloLens bestimmt und als Referenz für das gemeinsame Koordinatensystem verwendet. Nach abgeschlossener Registrierung werden zusätzliche Geometrien wie Roboter und Werkzeug aus der Simulationsdatenbasis geladen. In Bild 4 ist links die unvollständige Ausrichtung des Gestells und rechts das digitale Robotermodell sowie die UI bei Initialisierung der Applikation gezeigt.



Bild 4: Unvollständige manuelle Registrierung des AR-Systems (links); Darstellung während der Initialisierung (rechts)

Programmauswahl

In dieser Phase wählt der Anwender das zu bearbeitende Programm direkt aus dem Roboterspeicher aus. Dieses Programm wird dann in das OLP-System importiert und, wie in Bild 5 (links) zu erkennen, im AR-System dargestellt.

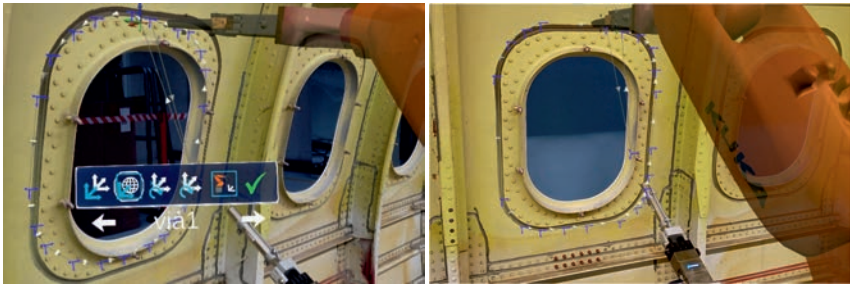


Bild 5: Initiale Ansicht des importierten Programms (links); Aktualisiertes Programm (rechts)

Modifikation

Diese Phase bildet den Kern der Bearbeitung. Der Anwender hat eine Reihe unterschiedlicher Methoden, um das Programm anzupassen, um den in Bild 5 (rechts) dargestellten Zustand zu erreichen.

Pfadpunkte können hinzugefügt oder gelöscht werden. Die Position, Rotation und Attribute wie Geschwindigkeit oder Beschleunigung einzelner Pfadpunkte können angepasst werden. Um die Erreichbarkeit zu evaluieren, kann der Anwender den virtuellen Roboter auf die einzelnen Pfadpunkte springen lassen. Ebenfalls kann eine Simulationssequenz berechnet und wiedergegeben werden.

Das eingesetzte Bedienerinterface ist themenbezogen in vier funktionale Editoren aufgegliedert. Die Editoren sind in einem dreidimensionalen Prisma angeordnet, sodass der Anwender durch Umklappen zwischen diesen Hin- und Herschalten kann.

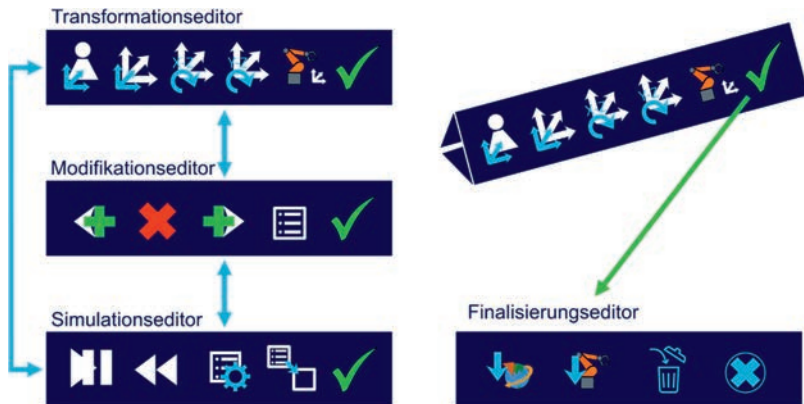


Bild 6: Übersicht der Editoren

Die einzelnen Editoren sind in

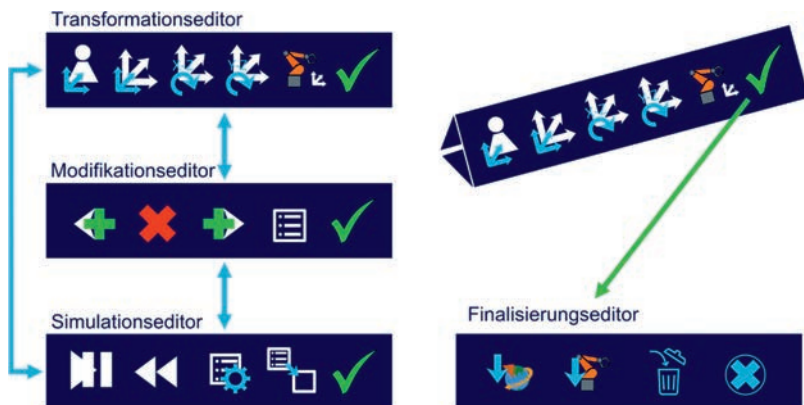


Bild 6 gezeigt. Die hier skizzierte dreidimensionale Anordnung dient dazu die dreidimensionalen Eigenschaften der AR als zusätzliches Gestaltungsmittel eines intuitiven Bedienerinterface zu nutzen.

Upload

Nach Abschluss der Modifikation kann der Anwender das aktualisierte Programm in der Simulationsdatenbasis abspeichern und direkt an den Roboter übertragen. Das Programm kann dann am realen Roboter abgefahren werden.

6. Fazit und Ausblick

Das entwickelte Bedieninterface erlaubt es typische Aufgaben der OLP mithilfe der AR intuitiv in der realen Anlage anzuwenden. Im betrachteten Anwendungsszenario kann ein Anwender beliebige Programme von einem über das Netzwerk angebundenen Industrieroboter laden, mithilfe von AR bearbeiten und weder zurückführen. Durch Nutzung der Simulationsfunktionen des angebundenen Simulationssystems, stellt der Anwender sicher, dass die Modifikationen am Programm valide sind und das Programm vom Roboter abgefahren werden kann.

Am Beispiel des Anwendungsszenario kann die Einsetzbarkeit der HRP bestätigt werden. Die einfache Gestaltung des Anwenderinterfaces erlaubt eine intuitive Bedienung.

Als Limitierung der Umsetzung ist Ortsstabilität der dargestellten Elemente zu nennen. Vor allem bewegte Anlagenkomponenten sind eine Herausforderung für das native HoloLens Tracking.

Für die Zukunft streben wir daher die Integration von Modelltracking-Algorithmen an. Hierbei werden die für das Tracking notwendigen Informationen direkt aus dem CAD-Modell des betrachteten Elementes generiert. Da in der OLP sowohl CAD-Modelle als auch Zell-Layouts mit Positionsinformationen zur Verfügung stehen, sehen wir hier ein hohes Potential. Ein weiteres Ziel ist die Untersuchung von alternativen AR-Systemen wie Tablets. Der Fokus liegt dabei neben einer Performanceanalyse vor allem auf einem ergonomischen Vergleich zur hier eingesetzten HoloLens 2.

Förderung

Die in diesem Beitrag vorgestellten Ergebnisse wurden im Rahmen des Forschungsprojektes MiReP erzielt. Dieses ist durch das Bundesministerium für Wirtschaft und Energie im Umfeld des Forschungsprogramms LuFo V-3 gefördert

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

Literaturangaben

- [1] Ong, S. K., Yuan, M. L. u. Nee, A. Y. C.: Augmented reality applications in manufacturing: a survey. *International Journal of Production Research* 46 (2008) 10, S. 2707–2742
- [2] Wolfgang Vogl: Eine interaktive räumliche Benutzerschnittstelle für die Programmierung von Industrierobotern
- [3] Fang, H. C., Ong, S. K. u. Nee, A.Y.C.: Interactive robot trajectory planning and simulation using Augmented Reality. *Robotics and Computer-Integrated Manufacturing* 28 (2012) 2, S. 227–237
- [4] Rückert, P., Meiners, F. u. Tracht, K.: Augmented Reality for teaching collaborative robots based on a physical simulation. In: Schüppstuhl, T., Tracht, K. u. Franke, J. (Hrsg.): *Tagungsband des 3. Kongresses Montage Handhabung Industrieroboter*. Berlin, Heidelberg: Springer Berlin Heidelberg 2018, S. 41–48
- [5] Lambrecht, J.: Natürlich-räumliche Industrieroboterprogrammierung auf Basis makerloser Gestenerkennung und mobiler Augmented Reality
- [6] Zäh, M. (Hrsg.): *Virtuelle Produktionssystemplanung. Virtuelle Inbetriebnahme und digitale Fabrik* ; Garching, 23. September 2004. *Seminarberichte / Iwb, Institut für Werkzeugmaschinen und Betriebswissenschaften*, Bd. 74. München: Utz 2004
- [7] Azuma, R. T.: A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 6 (1997) 4, S. 355–385
- [8] Martin, R. C.: *Clean architecture. A craftsman's guide to software structure and design*. Robert C. Martin series. Boston: Prentice Hall 2018

Adaptive VR-Produktionsumgebungen für Evaluations- und Schulungstätigkeiten

Adaptive VR production environments for evaluation and training purposes

Jessica Ulmer, Sebastian Braun, Prof. Dr.-Ing. Jörg Wollert,
FH Aachen University of Applied Sciences, Aachen;

Kurzfassung

Industrie 4.0 stellt viele Herausforderungen an produzierende Unternehmen und ihre Beschäftigten. Innovative und effektive Trainingsstrategien sind erforderlich, um mit den sich schnell verändernden Produktionsumgebungen und neuen Fertigungstechnologien Schritt halten zu können. Virtual Reality (VR) bietet neue Möglichkeiten für On-the-Job, On-Demand- und Off-Premise-Schulungen. Diese Arbeit stellt ein neues VR Schulungssystem vor, welches sich flexible an unterschiedliche Trainingsobjekte auf Grundlage von Rezepten und CAD Modellen anpassen lässt. Das Konzept basiert auf gerichteten azyklischen Graphen und einem Levelsystem. Es ermöglicht eine benutzerindividuelle Lerngeschwindigkeit mittels visueller Elemente. Das Konzept wurde für einen mechanischen Anwendungsfall mit Industriekomponenten implementiert und in der Industrie 4.0-Modellfabrik der FH Aachen umgesetzt.

Abstract

Industry 4.0 poses many challenges for manufacturing companies and their employees. Innovative and effective training strategies are needed to keep pace with rapidly changing production environments and new manufacturing technologies. Virtual reality (VR) offers new opportunities for on-the-job, on-demand, and off-premise training. This work presents a new VR training system that can be flexibly adapted to different training objects based on recipes and CAD models. The concept is based on directed acyclic graphs and a level system. It allows a user-individual learning speed by means of visual elements. The concept was implemented for a mechanical use case with industrial components and implemented in the industry 4.0 model factory of the FH Aachen University of Applied Sciences.

1. Einleitung

Produktionsumgebungen und -aufgaben unterliegen einem ständigen Wandel, sodass Mitarbeiter(innen) regelmäßig geschult werden müssen. Je nach Aufgabengebiet erfordern die Schulungen einen hohen Personal- und Zeitaufwand, welcher zu hohen Kosten führt. Virtual Reality (VR) ermöglicht ortsunabhängige Schulungen, welche keine Hardware und keine Präsenz von Facharbeiter(innen) benötigen. [1] In den vergangenen Jahren wurden unterschiedliche VR Lernumgebungen für industrielle Anwendungen entwickelt, integriert und getestet. Typische Anwendungsbereiche sind Bedienungs-, Wartungs- und Notfallschulungen. [2 bis 4]

Um die Vorteile von VR-Schulungen für mehr Unternehmen nutzbar zu machen, müssen Schulungen für einzelne Arbeitsplätze oder für einzelne Produkte bzw. Varianten ermöglicht werden. Daher besteht ein Bedarf an flexiblen VR-Umgebungen, welche von Nicht-Spezialisten intuitiv erstellt bzw. adaptiert werden können. Zudem müssen sich VR-Trainingsanwendungen an das Wissen der Beschäftigten anpassen sowie einen einfachen Einstieg in die VR-Welt ermöglichen.

Um dieses Problem anzugehen, wurde ein adaptives, VR-Training System für die verarbeitende Industrie entwickelt. Es basiert auf drei Säulen:

- Effizienter Aufbau für neue Applikationen: Das VR-Szenario basiert auf bestehenden CAD-Modellen. Die Umgebung passt sich automatisch an Produktvarianten an. Unternehmensspezifische Lernroutinen werden über eine generische Struktur definiert.
- Anpassung der Trainingsschwierigkeit: Verschiedene Spielelemente werden genutzt, um dem Nutzer ein Feedback zu seinem Trainingsstatus zu geben. Ein Levelsystem basierend auf Leistungskennzahlen (KPI) wird genutzt, um die Komplexität der Aufgabe anzupassen. Einführungsszenen unterstützen Benutzer ohne VR-Vorkenntnisse.
- Flexible Aufgabenerfüllung: Arbeitsabläufe werden mittels einer graphenbasierten Struktur definiert. Benutzer können verschiedene Routinen zur Lösung einer Aufgabe selektieren. Die Effizienz einer Routine wird auf Grundlage der Zeit und Qualität bewertet.

2. Stand der Technik

VR-Trainingsanwendungen können verschiedenste industrielle Situationen abbilden, indem sie ihren Benutzern realistische Arbeitsumgebungen, Geräte und Arbeitsaufgaben zur Verfügung stellen. Obwohl VR-Brillen zunehmend kostengünstig verfügbar sind, ist eine weitreichende Nutzung von VR-Trainingsszenarien insbesondere im produzierenden Mittelstand

derzeit kaum möglich. Die Entwicklung von VR-Umgebungen erfordert meist ein hohes technisches Know-How, welches in den Unternehmen nicht oder kaum vorhanden ist. Zudem müssen Änderungen in den Produktionsumgebungen manuell integriert werden. Aus diesen Gründen werden VR Schulung gegenüber realen Schulungen derzeit lediglich bevorzugt, wenn

- eine große Menge an Beschäftigten geschult werden muss,
- Schulungen in Umgebungen stattfinden, welche hohe Schulungskosten beinhalten,
- Situationen nachgestellt werden, welche im Normalbetrieb nicht auftreten und evtl. Gefahren beinhalten. [2, 4 bis 7]

In der Fertigungsindustrie werden VR-Schulungen vorrangig für Montage- und Wartungsschulungen verwendet. [2, 6, 7] Obwohl diverse Studien einen positiven Effekt von VR Schulungen bestätigen [5, 8], bleiben VR Schulungen aufgrund des hohen Erstellungsaufwandes jedoch eine Nischenanwendung.

Roldán et al. [9] ersetzen die manuelle Erstellung von Trainingssequenzen durch die Aufnahme von Expertenaktionen in VR. Die Aufnahmen werden anschließend genutzt, um die Reihenfolge für Schulungen festzulegen und Objekte nacheinander hervorzuheben. Beim vorgestellten Konzept muss allerdings beachtet werden, dass die Testapplikation aus identischen Objekten bestand, welche Löcher und Pins für die Montage beinhalteten. Hierdurch konnten die Montagemöglichkeiten einfach und konstant festgelegt werden, sodass eine flexible Montage für den Experten möglich war. Betrachtet man jedoch komplexe Baugruppen, ist eine Vorgabe der Montagemöglichkeiten jedes Bauteils sehr aufwändig.

Loch et al. [10] entwickelten ein VR System, welches sich an verschiedene Nutzertypen anpassen soll. Ihr Ziel ist es, die VR Schulung in Bezug auf die Darstellung, die Interaktivität und den Inhalt an die Bedürfnisse des Nutzenden anzupassen. Die Bedürfnisse sollen unter anderem mittels Zeitmessung, Eye-Tracking sowie Stimmanalyse analysiert werden. Während in der Arbeit viele Ideen genannt werden, fehlen jedoch noch konkrete Maßnahmen, um die Personalisierung in VR umzusetzen.

Gamification, die Nutzung von Spielelementen in spielfremden Umgebungen [11], bietet die Möglichkeit, Schwierigkeitsgrade individuell anzupassen und den Nutzern ein persönliches Feedback zu geben. Derzeit wird Gamification vorrangig im Bereich der Bildung (Schulen und Universitäten) sowie im Gesundheitsbereich genutzt. Zunehmend kommt Gamification jedoch auch für den industriellen Sektor in Betracht. [12]

3. Konzept der adaptiven VR Trainingsapplikation

Die vorliegende Arbeit basiert auf einem Gamification Framework, welches in Vorarbeiten erstellt wurde. [12] Ziel dieser Arbeit ist es, eine standardisierte VR Trainingsapplikation für manuelle Tätigkeiten zu entwickeln. Hierfür wird die Trainingsapplikation in die Trainingsumgebung sowie dem Auswertungssystem unterteilt (Bild 1).

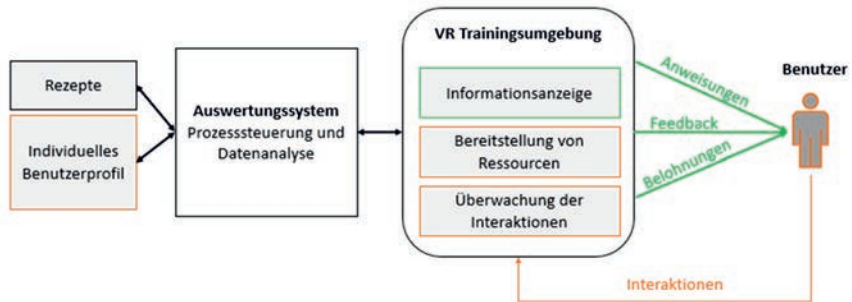


Bild 1: Trennung des VR Trainingsszenarios in das Auswertungssystem und die VR Trainingsumgebung

3.1 Auswertungssystem

Das Auswertungssystem enthält die Logik, um die Trainingssequenzen durchzuführen. Es ist für die Initialisierung sowie die Prozesssteuerung verantwortlich. Die Trainingsaufgaben werden in Rezepten definiert. Jedes Rezept besteht aus Arbeitsschritten, welche die folgenden Informationen beinhalten:

- Benötigte Fähigkeit für die Ausführung,
- Benötigte Teile und Werkzeuge,
- Nachfolgender Schritt bzw. bei mehreren Optionen nachfolgende Schritte.

In der Initialisierung wird die Arbeitsabfolge basierend auf den Rezepten mittels gerichteter, azyklischer Graphen festgelegt. Jeder Einzelschritt stellt einen Knoten im Graphen dar. Hierdurch ist es möglich, flexible Abläufe mit verschiedenen Optionen für den Benutzer umzusetzen. Des Weiteren werden in der Initialisierung die erforderlichen Ressourcen mit den Lagerplätzen bzw. Standorten in der VR Umgebung verknüpft.

Während des Trainingsprozesses sendet das Auswertungssystem die Anweisungen einzeln an die VR Trainingsumgebung. Die Anweisungen enthalten zudem Informationen zum Wissensstand des Benutzers in Form von Leveln. Jeder Fähigkeit wird dabei ein Level

zugewiesen. Basierend auf den Benutzerinteraktionen in VR wird der Graph aktualisiert. Bei einer falschen Aktion wird ein zusätzlicher Knoten zu dem Graphen hinzugefügt, welcher priorisiert bearbeitet werden muss. Somit müssen die Benutzer Fehler korrigieren, bevor sie mit dem Arbeitsprozess fortfahren können.

Während des Trainingsablaufs werden die Benutzerinteraktionen erfasst und in Bezug auf Zeit und Qualität bewertet. Nach Fertigstellung eines Rezeptes wird zudem analysiert, ob der Benutzer in ein höheres Level eingruppiert werden muss. Während des Trainingsprozesses erhält der Benutzer ein regelmäßiges Feedback in Form von Punkteanzeigen. Zudem wird er am Ende über eine eventuelle Levelanpassung informiert.

3.2 VR Trainingsumgebung

Die VR Trainingsumgebung stellt das Trainingsszenario, Interaktionsmöglichkeiten sowie audiovisuelle Unterstützungsmöglichkeiten bereit. Es enthält keine Informationen über die Trainingssequenz und führt keine Interaktionsanalyse durch, da diese Bereiche vom externen Auswertungssystem abgedeckt werden. Daher ist die Trainingsumgebung nicht auf eine spezielle Trainingsaufgabe zugeschnitten, sondern kann für verschiedenste Rezepte genutzt werden.

Um eine automatisierte Anpassung der Trainingsumgebung an die Trainingsaufgabe zu ermöglichen, wurde die VR Umgebung in drei Teile separiert:

- Grundlegendes Layout: Dieser Teil wird für die verschiedenen Anwendungen nicht verändert und kann Objekte wie den VR-Hintergrund, die allgemeine Arbeitsplatzstruktur und die Ausgangsposition des Nutzers enthalten.
- Ressourcen: Dieser Teil wird für jedes Rezept angepasst, indem Ressourcen an vordefinierten Bereichen der VR-Umgebung bereitgestellt werden. Das Laden von Ressourcen kann Modelle von Maschinen, Werkzeugen und einzelnen Gegenständen umfassen, die während des VR-Trainings verwendet werden.
- Instruktions- und Feedbacksystem: Dieser Teil wird für jede Aufgabe entsprechend den Benutzerinteraktionen modifiziert. Er kann das Hervorheben von Objekten oder das Anzeigen von Textnachrichten beinhalten. Als Grundlage dienen CAD Modelle, welche in Unity teilautomatisiert hinterlegt werden.

In der Initialisierungsphase werden die Ressourcen vom Auswertungssystem vorgegeben und von der VR Trainingsumgebung geladen. In der Trainingsphase werden die Instruktionen des Auswertungssystems in audiovisuelle Informationen im VR System umgewandelt und der

Benutzer somit angeleitet. Um die Vorteile von VR zu nutzen, liegt der Fokus auf visuellen Hinweisen zum Trainingsprozess. Die Hinweise hängen hierbei vom individuellen Level des Benutzers ab.

4. Umsetzung

Die VR Umgebung wurde basierend auf der Spieleplattform Unity [14] für eine HTC Vive [15] umgesetzt (Bild 2). Das grundlegende Layout besteht aus dem Hintergrund, dem Handarbeitsplatz sowie dem Papierkorb. Je nach Rezept werden die Boxen, Werkzeuge und Montagevorrichtungen in der Initialisierungsphase als Ressourcen hinzugefügt. In den Boxen sind die virtuellen Bauteile hinterlegt, welche beim Griff in die Box generiert werden. Die Arbeitsfläche, die Fortschrittsanzeige sowie der Monitor werden für jeden Arbeitsschritt aktualisiert.

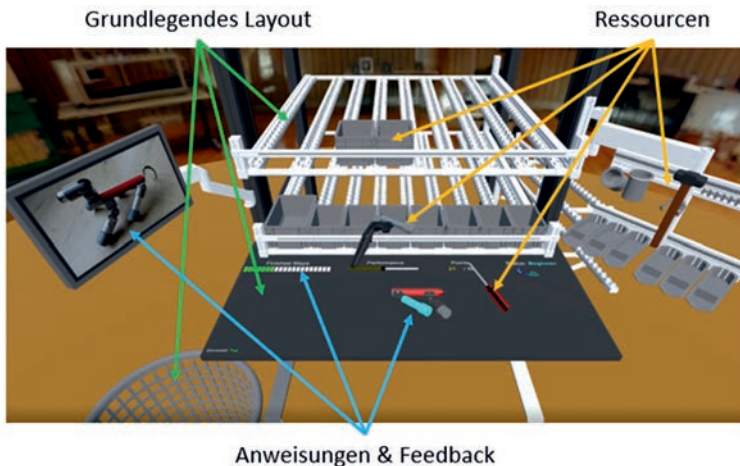


Bild 2: VR Trainingsumgebung mit Beschreibung der Bestandteile nach Kapitel 3.2

Als Anwendungsbeispiel wurden Bauteile der Firma item Industrietechnik GmbH gewählt, welche eine Hundestruktur ergeben. Die Montage der Hundestruktur besteht aus 23 Arbeitsschritten, welche die Benutzung eines Werkzeuges sowie das Drehen der Baugruppe in mehrere Positionen beinhalten.

Während der Montage gibt es drei Hauptkomponenten, welche den Nutzer bei der Durchführung unterstützen: 3D-Ansichten von zu platzierenden Bauteilen; Hervorhebung von Stellen, die mit einem Werkzeug bearbeitet werden müssen; Pick-by-Light System für das Greifen der

benötigten Objekte. Um den Schwierigkeitsgrad benutzerabhängig anpassen zu können, wurde ein Levelsystem in den Hauptkomponenten umgesetzt (Tabelle 1). Es enthält insgesamt 4 Level, wobei Level 0 als Einführungslevel dient, welches dem Benutzer Informationen zur Controllernutzung bietet. Zudem wurde für Level 0 eine Montagesequenz mit reduzierter Komplexität gewählt (Longboard), um den Einstieg in die VR Umgebung zu erleichtern.

Während des Trainings wird die Zeit sowie die Korrektheit jedes Arbeitsschritts als KPI erfasst. Basierend auf diesen Daten wird für jede Fähigkeit, wie beispielsweise Greifen und Montieren, ein Box-Whisker-Plot erstellt, welcher mittels des oberen Whiskers (Maximum) und des unteren Whiskers (Minimum) die Standardwerte für die Ausführung der Arbeitsschritte definiert (Bild 3). Um im Levelsystem aufzusteigen müssen 80% der ausgeführten Aktionen unterhalb des Zeitmaximums liegen sowie 80% der Aktionen korrekt ausgeführt werden.

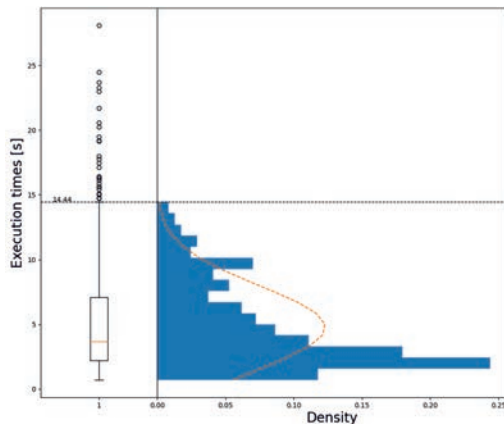
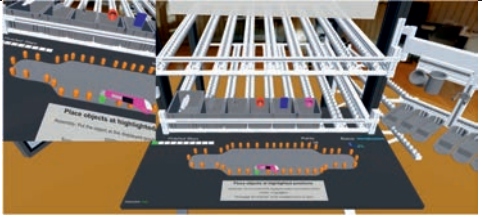
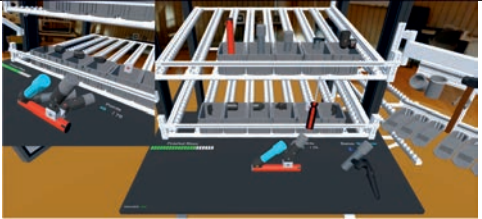
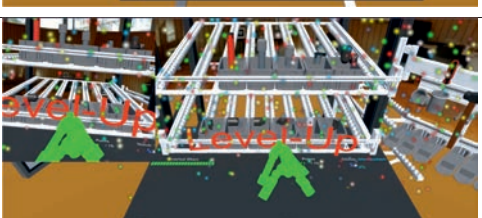


Bild 3: Beispiel eines Box-Whisker-Plots mit Histogramm für Greifaktionen in Level 3

Tabelle 1: Übersicht über das Level- sowie Feedbacksystem; die Bilder bestehen aus einem stationären Kamerabild (Hauptansicht) sowie der Nutzeransicht (oben links)

Level 0	- Textinformationen - 3D Informationen (Werkzeuge & Bauteile) - Pick-by-Light System	
Level 1	- 3D Informationen (Werkzeuge & Bauteile) - Pick-by-Light System	
Level 2	- Transparente 3D Informationen - Hervorhebung von Werkzeugpunkten - Pick-by-Light System	
Level 3	- Durchsichtige 3D Informationen für Bauteile	
Feedback	- Punkteanzeige - Level-Up mit Konfettiregen	

5. Zusammenfassung und Ausblick

Im vorgestellten Projekt wurde eine adaptive VR Umgebung entwickelt, welche sich automatisch an verschiedene Produkte und Varianten anpasst. Die Anpassungen basieren auf CAD Modellen, welche über ein Benutzerinterface in die Entwicklungsumgebung geladen werden. Zudem passt sich die VR Umgebung mittels eines Levelsystems an die Bedürfnisse und die Lerngeschwindigkeit des Nutzers an. Mittels visueller Hinweise wird die Komplexität schrittweise erhöht und dem Nutzer mehr Freiraum gegeben. Die KPIs Zeit und Qualität werden genutzt, um das Unterstützungsniveau zu selektieren.

Das Konzept basiert auf einer Trennung der Auswertungslogik von der VR Umgebung. Das Auswertungssystem enthält das Prozesswissen, um die Benutzerinteraktionen zu validieren sowie den Ablauf zu steuern. Die Trainingsabläufe werden in generischen, wiederverwendbaren Rezepten gespeichert, welche dem Auswertungssystem zur Verfügung stehen. Die VR Umgebung ist modular aufgebaut, sodass sie automatisiert je nach Rezept angepasst werden kann. Sie enthält ein Level- und Punktesystem, welches den Schwierigkeitsgrad schrittweise anpasst und Feedback bereitstellt.

Das vorhandene System wurde genutzt, um weitere Baugruppen als Schulungsmodell zu hinterlegen. Hierbei lässt sich feststellen, dass Standardaufgaben (Greifen, Platzieren, Montieren) einfach und schnell mittels Rezepte und CAD Modellen integriert werden können. Bei Sonderfällen an Interaktionen ist jedoch weiterhin eine Modellierung in VR erforderlich. Des Weiteren wurde deutlich, dass Kleinteile, wie beispielsweise Schrauben und Muttern, nicht in Originalgröße montierbar sind, da die Auflösung der VR Brille zu niedrig und das Handling mit den Controllern schwierig ist. In diesen Fällen ist eine Vergrößerung der Modelle erforderlich.

6. Literaturangaben

- [1] Talbot, J.: Training in Organisations. A Cost-Benefit Analysis. London: Taylor and Francis 2016
- [2] Guo, Z., Zhou, D., Zhou, Q., Zhang, X., Geng, J., Zeng, S., Lv, C. u. Hao, A.: Applications of virtual reality in maintenance during the industrial product lifecycle: A systematic review. Journal of Manufacturing Systems 56 (2020), S. 525–538
- [3] Damiani, L., Demartini, M., Guizzi, G., Revetria, R. u. Tonelli, F.: Augmented and virtual reality applications in industrial systems: A qualitative review towards the industry 4.0 era. IFAC-PapersOnLine 51 (2018) 11, S. 624–630

- [4] Grabowski, A.: Innovative and Comprehensive Support System for Training People Working in Dangerous Conditions. In: Duffy, V. G. (Hrsg.): Digital human modeling and applications in health, safety, ergonomics and risk management. Lecture Notes in Computer Science, Bd. 11581. Cham: Springer 2019, S. 394–405
- [5] Abidi, M. H., Al-Ahmari, A., Ahmad, A., Ameen, W. u. Alkhalefah, H.: Assessment of virtual reality-based manufacturing assembly training system. The International Journal of Advanced Manufacturing Technology 105 (2019) 9, S. 3743–3759
- [6] Pérez, L., Diez, E., Usamentiaga, R. u. García, D. F.: Industrial robot control and operator training using virtual reality interfaces. Computers in Industry 109 (2019), S. 114–120
- [7] Stork, A., Sevilmis, N., Weber, D., Gorecky, D., Stahl, C., Loskyll, M. u. Michel, F.: Enabling virtual assembly training in and beyond the automotive industry. 18th International Conference on Virtual Systems and Multimedia (VSMM), 2012. 2 - 5 Sept. 2012, Milan, Italy ; conference proceedings. Piscataway, NJ: IEEE 2012, S. 347–352
- [8] Numfu, M., Riel, A. u. Noel, F.: Virtual Reality Based Digital Chain for Maintenance Training. Procedia CIRP 84 (2019), S. 1069–1074
- [9] Roldán, J. J., Crespo, E., Martín-Barrio, A., Peña-Tapia, E. u. Barrientos, A.: A training system for Industry 4.0 operators in complex assemblies based on virtual reality and process mining. Robotics and Computer-Integrated Manufacturing 59 (2019), S. 305–316
- [10] Loch, F., Fahimipirehgalin, M., Czerniak, J. N., Mertens, A., Villani, V., Sabattini, L., Fantuzzi, C. u. Vogel-Heuser, B.: An Adaptive Virtual Training System Based on Universal Design. IFAC-PapersOnLine 51 (2019) 34, S. 335–340
- [11] Deterding, S., Tan, D., Khaled, R. u. Nacke, L. E.: Gamification: Toward a Definition. ACM Digital Library. New York, NY: ACM 2011
- [12] Ulmer, J., Braun, S., Cheng, C.-T., Dowe, S. u. Wollert, J.: Human-Centered Gamification Framework for Manufacturing Systems. Procedia CIRP 93 (2020), S. 670–675
- [13] Ulmer, J., Braun, S., Cheng, C.-T., Dowe, S. u. Wollert, J.: Gamified Virtual Reality Training Environment for the Manufacturing Industry. 2020 19th International Conference on Mechatronics - Mechatronika (ME). IEEE 2020 - 2020, S. 1–6
- [14] Unity Technologies: Unity. 2018
- [15] HTC Corporation: HTC Vive Virtual Reality System, 2016.
<https://www.vive.com/eu/product/vive/>, abgerufen am: 24.08.2020

Modell einer Pick & Place-Anlage basierend auf Verwaltungsschalen

Prof. Dr.-Ing. **Ch. Diedrich**, ifak e.V. Magdeburg;
M.Sc. **A. Belyaev**, M.Sc. **T. Schröder**, M.Sc. **C. Urban**,
Dipl.-Ing. **T. Werner**, M.Sc. **H. Pakala**, OvGU Magdeburg

Kurzfassung

Die Digitale Fabrik und Smart Manufacturing benötigen eine durchgängige digitale Repräsentation aller Betriebsmittel. Diese Beschreibung muss lebenszyklusübergreifend ohne Informationsverlust zur Verfügung stehen. Der Digitale Zwilling ist das dafür zurzeit diskutierte Schlagwort. Die Verwaltungsschale (VWS) ist der konkrete Vorschlag der Plattform I4.0 sowie der Industrial Digital Twin Association (IDTA) zur Umsetzung des standardisierten Digitalen Zwillings.

Die Verwaltungsschale ist in ihrer wesentlichen Form definiert ([1], [2]) und kann schrittweise in die industrielle Praxis überführt werden. Die Inhalte der Verwaltungsschale wie die Teilmodelle (eng.: Submodels) und deren Elemente z. B. Properties, Operation, Collections, Events sowie die Identifikationsmöglichkeiten bieten reichhaltige Beschreibungsmittel. Aber wie sollen diese verwendet werden? Wann hat ein Betriebsmittel bzw. Asset eine Verwaltungsschale und wann wird dieses als ein Teilmodell der übergeordneten Verwaltungsschale dargestellt? Wo findet man diese in der VWS enthaltene Informationen heute? Dieser Beitrag berichtet über das Ergebnis einer Arbeitsgruppe aus dem Projekt „Verwaltungsschale vernetzt“, in dem eine kleine fertigungstechnische Anlage mit Verwaltungsschalen modelliert wurde.

Abstract (optional)

The digital factory and smart manufacturing require a consistent digital representation of all equipment. This description must be available across the life cycle without any loss of information. The digital twin is the catchphrase currently being discussed for this. The administration shell is the specific proposal of the I4.0 platform and the Industrial Digital Twin Association (IDTA) to implement the digital twin.

The main form of the administration shell is defined ([1], [2]) and can be gradually transferred to industrial practice. The contents of the administration shell such as the submodel and its elements, e.g. B. properties, operations, collections, events as well as the identification options offer extensive means of description. But how should these be used? When does a device or asset have an administration shell and when does it have which submodel? Where can you find this information today? This article reports on the result of a working group from the "Networked Administration Shell" project, in which a small manufacturing plant with administration shells was modeled.

1. Ausgangspunkte

Die informationstechnische Herausforderung der heutigen Produktionsanlagen besteht in der durchgängigen Datennutzung entlang des Lebenszyklus von der Planung zum operativen Betrieb und der Instandhaltung. Dies gilt für jedes einzelne Gerät und deren Zusammenwirken in Komponenten, Maschinen bis zur gesamten Anlage (Abbildung 1). Besonders kritische Punkte sind Übergänge zwischen Engineeringphasen, z.B. von der Planung zur Inbetriebnahme und in den operativen Betrieb oder bei den Eigentümerübergängen von Produzenten an die Nutzer. Dies liegt vor allem daran, dass bei diesen Übergängen eine zumeist sehr gut aufeinander abgestimmte Werkzeuglandschaft verlassen wird und andere Werkzeuge mit eigenen Datenmodellen zum Einsatz gelangen. Die Daten werden in Form von Dateien oder mittels Schnittstellen der Werkzeuge übertragen. Beispielfaht sind in Abbildung 1 das AutomationML-Format und die Schnittstellen-Standards HTTP und OPC UA dargestellt.

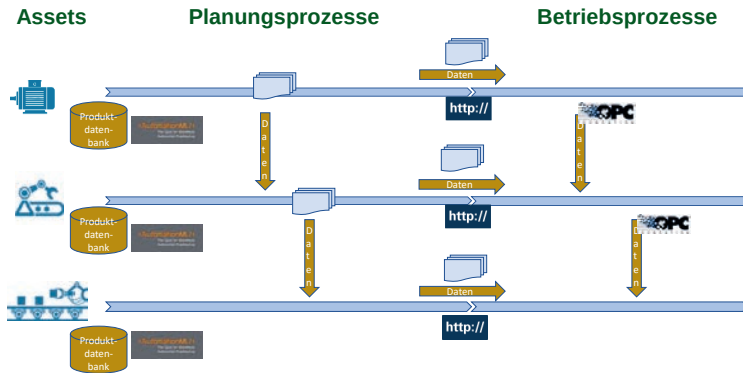


Abbildung 1: Landschaft der Werkzeuge und Datenformate in Lebenszyklusphasen einer Anlage

Bei der Transformation zwischen den Modellen sind Daten- und Informationsverluste zu verzeichnen. Die Frage ist also, wie gerade diese kritischen Stellen besser gemeistert werden können. Die Verwaltungsschale, ein Konzept des Digitalen Zwillinges der Plattform I4.0, ist angetreten, hierfür einen signifikanten Beitrag zu leisten. Doch wie soll dies gelingen?

Für die Beantwortung dieser Frage liegen bisher nur wenige Erfahrungen vor. In dem Projekt „VWS vernetzt“ (vwsvernetzt.de) modelliert eine Arbeitsgruppe eine Pick & Place-Anlage mit Verwaltungsschalen. Die Ergebnisse sind in dem Diskussionspapier der Plattform I4.0 veröffentlicht [4]. Hier werden einige Aspekte daraus dargestellt.

Der Ansatzpunkt und damit die Auswahl, welche Verwaltungsschalen modelliert werden, sind in Abbildung 2 dargestellt. Wie oben beschrieben sind die kritischen Übergabepunkte einmal von der Planung zur Inbetriebnahme und beim Wechsel vom Produzenten zum Kunden. Deshalb wurden Verwaltungsschalen als Ergebnis der Planung und für die Inbetriebnahme modelliert. Hier sei noch einmal erwähnt, dass die Verwaltungsschalen auf einem einheitlichen Meta-Informationsmodell beruhen [1], das sowohl als Grundlage der Datei-Repräsentationsform (passive VWS) als auch für die verschiedenen Schnittstellen (reaktive [2] und proaktive [3] VWS dient.

Am Beispiel eines Motors und einer Arbeitsstation werden explizit die Verwaltungsschalen des jeweiligen Asset-Typs (als Repräsentant der Planung) und der Asset-Instanz (als Repräsentant der Inbetriebnahme) ausgewählt. Zusätzlich werden proaktive Verwaltungsschale modelliert, die eine Fähigkeitsbeschreibung beinhalten, damit über ein Ausschreibungsverfahren [3] eine Auswahl des Assets für den nächsten Produktionsschritt durchgeführt werden kann.

Natürlich ist die Verwaltungsschale für die Asset-Instanz von der seines Asset-Typs abhängig. Deshalb wird sie von diesem abgeleitet. Es ist keine reine Typ-Instanz-Beziehung, wie sie in der Softwaretechnik verwendet wird. Dies liegt daran, dass neue Elemente hinzukommen können, die in der Planung noch nicht benötigt werden (z.B. die Seriennummer) oder auch Änderungen von Datentypen von Merkmalen, wie z.B. die Angabe von einem Bereich möglicher Werte (Datenstruktur Array) beim Typ und einem Einzelwert (Datenstruktur Einzelwert) bei der Instanz.

Zusätzlich werden für die Modellierung von wiederkehrenden Asset-Aspekten wie z.B. Typenschild [8], Dokumentation und technischen Daten [9] als Teilmodelle von der Plattform I4.0 die Teilmodell-Templates bereitgestellt. Sie sind mit den bevorzugten Modellelementen ausgestattet und werden zur Nutzung empfohlen. Ein besonderes Merkmal des Verwaltungsschalen-Metamodells ist die Existenz einer semantischen Referenz für die Modellelemente, deren semantische Definition in einem Vokabular (eng.: Dictionary) definiert ist. Die semantische Definition ist getrennt von der Identifikation der Elemente. Diese Referenz heißt „semanticId“. Der Verweis ist auf sogenannte Dictionaries gerichtet, in denen sich die Beschreibungen befinden. Beispiel für solche Dictionaries sind ECLASS und IEC CDD.

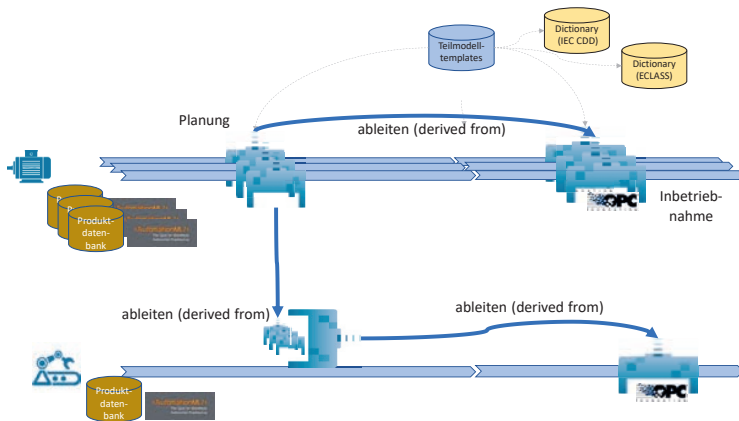


Abbildung 2: Ableitung von Verwaltungsschichten im Lebenszyklus (beispielhaft)

2. Beschreibung der Anlage

Die modellierte Anlage besteht aus drei Bearbeitungsstationen und einer Transporteinheit in Form eines Delta-Roboter. In Abbildung 3 ist die steuerungsrechtliche Struktur der P&P-Station dargestellt. Zum Deltaroboter gehören drei Motoren mit ihren jeweiligen Controllern. Von der P&P-Station können mehrere Produkte bearbeitet werden, hier als Zylinder dargestellt. Die Steuerungsstruktur jedes einzelnen Asset ist klassisch, SPS steuern die jeweilige Station beziehungsweise die Motor-Controller. Jede Steuerung ist mit einer VWS verbunden. Auch jedes Produkt hat seine eigene VWS. Die VWS interagieren untereinander mit der I4.0-Sprache und orchestrieren die Produktion ohne einer zentralen Steuerung. Zusätzlich gibt es die Möglichkeit, in die VWS über eine Browser-Schnittstelle hineinzuschauen.

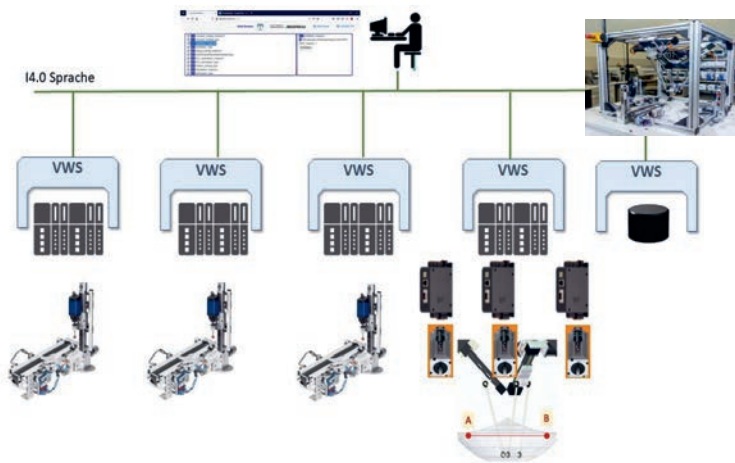


Abbildung 3: Übersicht über die Struktur mit Verwaltungsschichten [4]

3. Beschreibung der VWS ausgewählter Assets

3.1 Auswahl der Assets

Im Kommen dieses Beitrags werden einige Aspekte vorgestellt, die bei der Modellierung von besonderem Interesse waren. Für interessierte Leser sei noch mal auf [4] verwiesen, wo umfangreichere Beschreibungen verfügbar sind. Die Modellierung der Anlage und der Komponenten wurde mit dem Package Explorer, einem open source-Editor für Verwaltungsschalen durchgeführt. Der Package Explorer bietet verschiedene Export-Möglichkeiten, wie z.B. AutomationML, JSON und OPC UA-Nodesets. Auf diese Weise kann die Nutzung der Informationen der Assets in unterschiedlichen Lebenszyklusphasen und Use Cases veranschaulicht werden. Dadurch sind die Ergebnisse auch in ihren Details in einem Web-Server (sogenannter AASX-Server) öffentlich verfügbar.

Neben der Gesamtanlage wurde die Transporteinheit, ein Motor mit Encoder des Deltaroboters und eine Bohrstation als ein Asset mit eigener VWS modelliert (Abbildung 4). Dies ist ein signifikantes Spektrum an Assets, welches die wesentlichen Modellelemente des Verwaltungsschalen-Meta-Modells abdeckt.

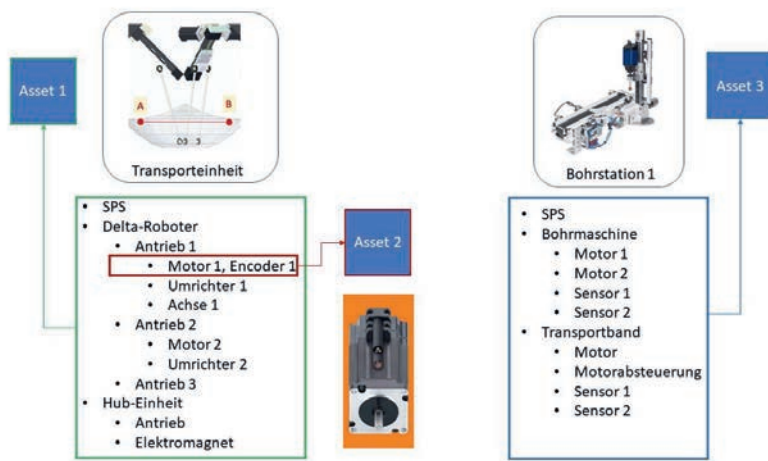


Abbildung 4: Assets, für die ein Verwaltungsschalen erstellt wurden [4]

Alle Verwaltungsschalen haben Teilmodelle für das Typenschild und zur Beschreibung der Dokumentationen. Darauf wird hier nicht weiter eingegangen.

3.2 Die Pick & Place Station als Verbund von Verwaltungsschalen

Maschinen oder Anlagen setzen sich prinzipiell aus einer Reihe von Geräten und Komponenten zusammen. Allen haben entlang ihres Lebenszyklus eine oder mehrere Verwaltungsschalen. Dementsprechend muss sich die Verwaltungsschale der Maschine, hier der Pick & Place-Station, aus den einzelnen Verwaltungsschalen zusammensetzen. Dies entspricht einer Zusammenstellung aller Beschreibungen der Geräte und Komponenten zu einer Gesamtbeschreibung.

Lessons Learned:

Der hier empfohlene Weg ist, eine sogenannte Verbundkomponente zu erstellen, die die Referenzen auf die einzelnen Verwaltungsschalen der Komponenten enthält (Abbildung 5). Dazu wird ein eigenes Teilmodell (eng.: Submodel) angelegt, das alle Referenzen enthält. Dies entspricht einem Inhaltsverzeichnis aller Dokumente, die es gibt. Die Referenzen sind vom Typ „isPartOf“. Dies markiert für eine Maschinenauswertung, dass die referenzierten Verwaltungsschalen zu Assets gehören, die in der Anlage verbaut sind.

Eine weitere Möglichkeit der Beschreibung der Zusammensetzung der Anlage ist es, ein Teilmodell anzulegen, dass die Stückliste (Bill of Material – BoM) aus Sicht der Assets beinhaltet. Beide Varianten ermöglichen es sowohl aus Sicht des Assets, als auch aus Sicht der Verwaltungsschale die Struktur zu erkunden.

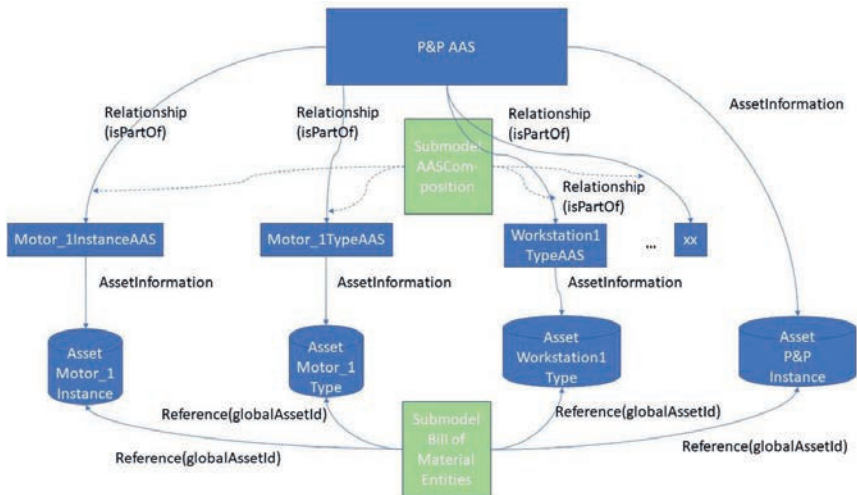


Abbildung 5: Komposition von VWS als Übersicht [4]

Abbildung 6 zeigt die Package Explorer Darstellung des BoM-Teilmodells. Eine semantische Kennzeichnung dieses Teilmodells als BoM ermöglicht es dem PE, auch eine grafische Darstellung der Struktur bereitzustellen. Dies gilt auch für andere standardisierte Teilmodelltypen.

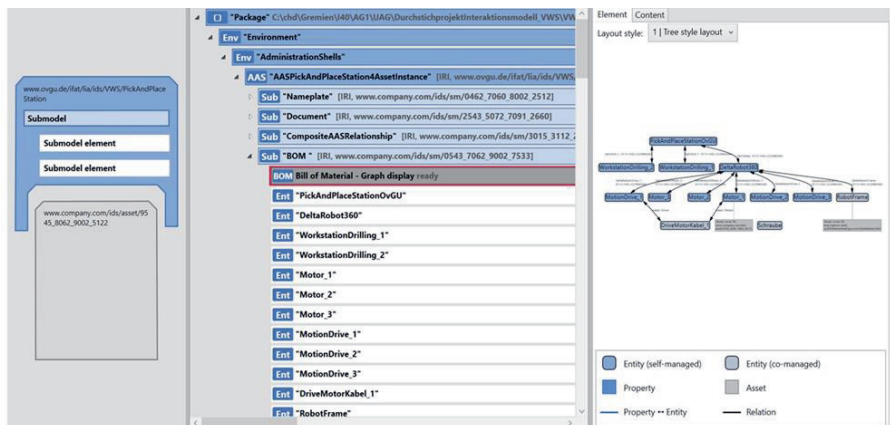


Abbildung 6: Ausschnitt der Verwaltungsschale der Pick&Place Station - Bill of Material (BoM) [4]

3.3 Verwaltungsschale für Planung und Inbetriebnahme – Asset Typ- und Asset-Instanzbeschreibung

Die Transportkomponente beinhaltet drei Schrittmotoren zur Ansteuerung des Delta-Roboters. Alle drei Motoren sind Ausprägungen desselben Produkttyps. Daher wird zunächst eine VWS für diesen Typ modelliert. Abbildung 7 zeigt die Ansicht der VWS des Motortyps im AASX Package Explorer.

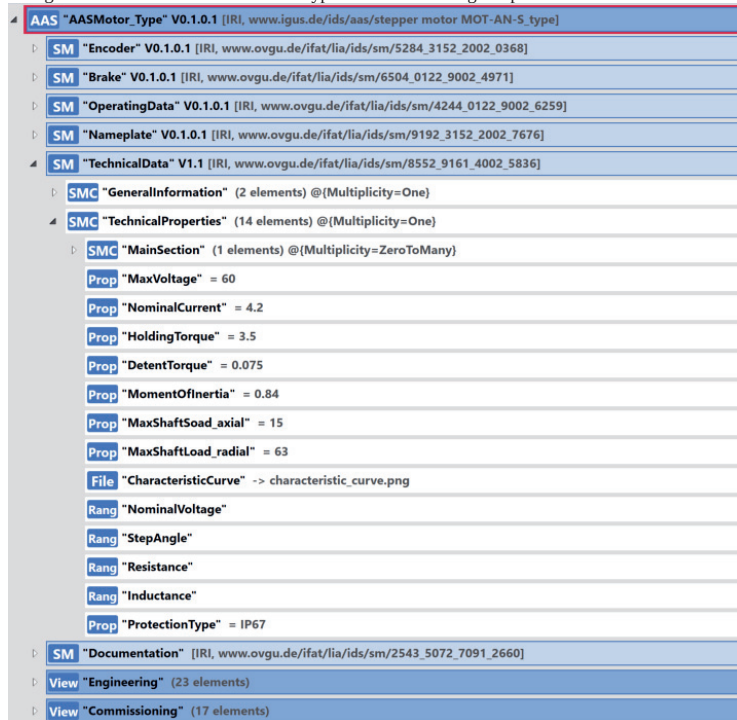


Abbildung 7: Teilmodell TechnicalData mit verschiedenen SubmodelElement-Typen

Angelegt sind ein Asset sowie eine zugehörige VWS. Das Asset ist vom AssetKind Type, da es den Typen des Motors repräsentiert. Die VWS enthält sechs Teilmodelle, deren Wahl sich an der Gliederung der Dokumentationsunterlagen des Motors orientiert. Sämtliche IDs der Teilmodelle fallen unter die Authority www.ovgu.de. Alle Teilmodelle sind hier vom ModelingKind Instance. Die SubmodelElements der Teilmodelle vom Kind Instance sind ebenfalls vom Kind Instance. Die des einen Teilmodells vom Kind Template sind ebenfalls vom Kind Template. Alle SubmodelElements tragen semanticIds, die als GlobalReference vom Typ IRDI oder IRI auf einen Standard wie z. B. ECLASS oder IEC61360-CDD verweisen. Abbildung 7 zeigt außerdem exemplarisch für das Teilmodell TechnicalData, dass verschiedene Subtypen von SubmodelElements Verwendung finden. So enthält dieses Teilmodell neben dem häufig verwendeten Typ Property auch die Typen Range und File, um dem value-Attribut einer Property einen Bereich zuzuweisen bzw. auf eine Datei zu verweisen. Die Property mit dem Bezeichner protection type verwendet zudem eine ValueID, um für deren Ausprägung auf einen standardisierten Wert aus einer ValueList zu verweisen (bspw. IP65, IP67). Einzelne SubmodelElements sind einem der beiden Views „Engineering“ oder „Commissioning“ zugewiesen. Verschiedene Nutzerrollen können damit gezielt auf die für sie als relevant erachteten SubmodelElements zugreifen.

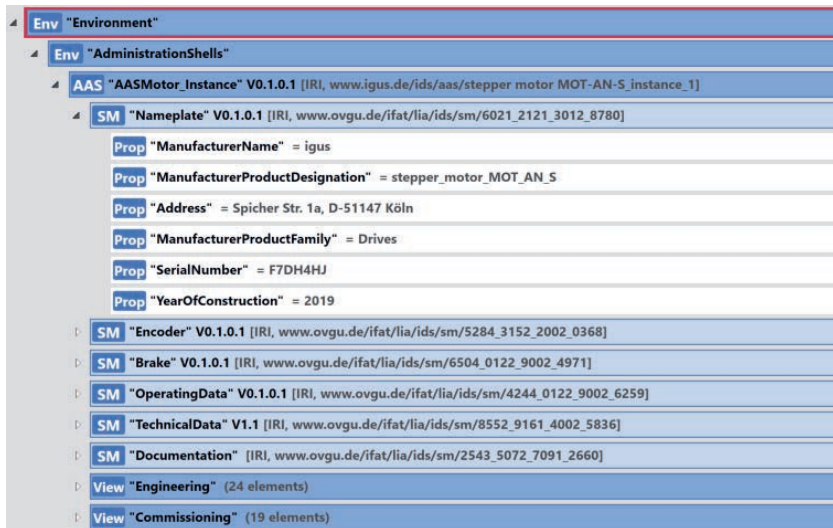


Abbildung 8: Ansicht der VWS einer Motorinstanz

Abbildung 8 zeigt die Ansicht der VWS einer Motorinstanz im AASX Package Explorer. Hiervon liegen insgesamt drei vor. Angelegt sind ein Asset sowie eine zugehörige VWS. Das Asset ist vom *ModelingKind Instance*, da es eine Instanz des Motors repräsentiert. Die VWS enthält nun noch fünf Teilmodelle, da bezüglich des Teilmodells *Nameplate* nun keines vom *ModelingKind Template* mehr benötigt wird. Stattdessen haben nun alle *SubmodelElements* von *Nameplate* konkrete Werte und sind daher in einem einzigen Teilmodell vom *ModelingKind Instance* zusammengefasst.

Auch die Instanz-VWS enthält wieder die beiden Views „Engineering“ und „Commissioning“. Die Elemente der Views verweisen auf *SubmodelElements* der verschiedenen Teilmodelle. Die konkrete Zuordnung der Submodellelemente zu den Views wird je nach Verwendungszweck des Assets und dessen VWS unterschiedlich ausfallen. Das Metamodell selbst kann dafür keine Kriterien bieten. In den übrigen Aspekten unterscheidet sich die Instanz-VWS nicht von der zuvor beschriebenen Typ-VWS.

3.4 Nutzung der Teilmodell-Templates

Teilmodell-Templates sind Schablonen, die für die Modellierung von häufig wiederkehrenden Aspekten der Assets als Teilmodelle die Vorlage für die Erstellung von Teilmodellen mit den dafür benötigten Modellelemente vorgefertigt bereitstellen [4]. Es wird angestrebt für oft wiederkehrend, allgemeingültige Aspekte solche Teilmodell-Templates zur Verfügung zu stellen. Bereits verfügbare Templates sind für Typenschilder und Dokumentationen sowie Zeitreihen und Simulationskomponenten verfügbar. Sie dienen als Kopiervorlage. Die standardisierten Teilmodell-Templates werden sowohl für die Entwickler der VWS als auch für deren Nutzer zugänglich sein. Sie sind so etwas ähnliches wie Profile bei den Feldbussen oder Companion Specifications bei OPC UA.

Lessons Learned (nach [4]):

Das Vorgehen bei der Nutzung der Templates ist nicht so streng, wie man es bei der in der Objektorientierung üblichen Typ/Instanzbeziehung kennt. Einerseits sind optionale Elemente in den Templates, die bei der Nutzung weggelassen werden können und andererseits kann der Nutzer weitere Elemente hinzufügen oder Teile mehrmals instanzieren. Außerdem können sich Hersteller von Assets aus den standardisierten Templates eigene zugeschnittene Templates erstellen und dann herstellerepezifisch umsetzen (Abbildung 9). Außerdem sind so die

Unterschiede für die Beschreibung des Assets-Typs und -Instanz konsistent umsetzbar. Die in den VWS instanziierten Teilmodelle sollen durch die Nutzung der semanticId des zugehörigen Teilmodell-Templates anzeigen, dass sie aus diesem Template entstanden sind. Dies ermöglicht es, den verpflichtenden Teil des Teilmodells bei der Anwendung voraussetzen zu können. Ein Beispiel dafür ist die BoM.

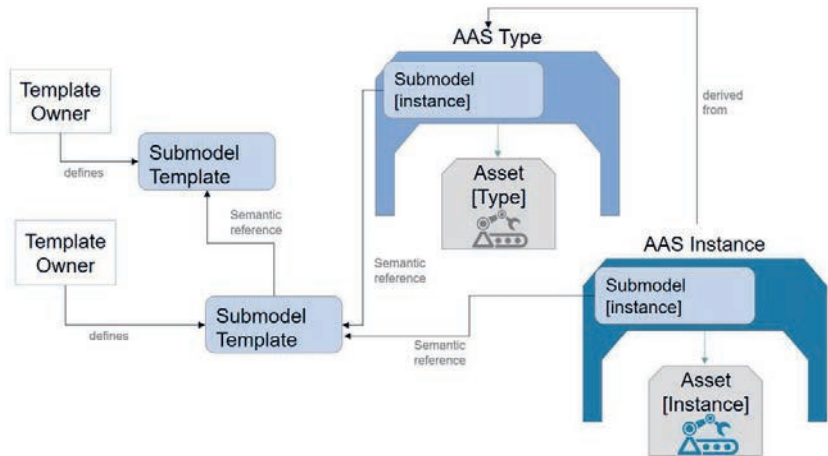


Abbildung 9: Zusammenhang zwischen Teilmodell-Templates (Submodel-Templates) und den Teilmodellen in den VWS

3.5 Fähigkeitsbeschreibung

Ein Capability-Metamodellelement soll die Fähigkeit eines Assets beschreiben, eine bestimmte Funktion zu erfüllen. Die Spezifikation VWS in Detail führt das Capability-Element ein, erklärt aber nicht wie das Element für die Beschreibung von Fähigkeit verwendet werden soll. Im [4] werden zwei Optionen dargestellt, die die Fähigkeiten als Teilmodelle modelliert werden können. Dafür sind die erforderlichen Merkmale in dem Capability-Teilmodell zu definieren. Sie können für jede technische Funktion in Produktbeschreibungsstandards (z. B. ECLASS) zur Verfügung gestellt werden.

Wie in der Abbildung 10 dargestellt, stellt das in diesem Kapitel betrachtete Asset eine Transporteinheit dar. Die Aufgabe bzw. erwartete Wirkung dieses Assets ist es, ein Werkstück innerhalb der P&P-Station von A nach B zu transportieren.

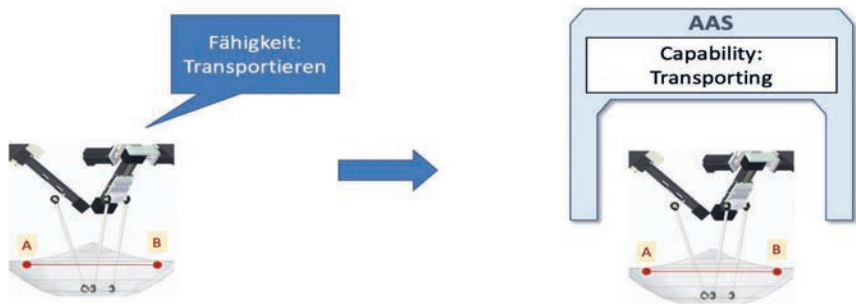


Abbildung 10: Modellierung der Fähigkeit als entsprechendes Teilmodell in der Verwaltungsschale

Lessons Learned (nach [4]):

Der Begriff Fähigkeit (hier englisch „Capability“) wird sehr uneinheitlich verwendet. Dies erschwert die für Verwaltungsschalen beabsichtigte maschinelle Interpretation der Inhalte. In [7] wird beschrieben, dass es ein spezielles Metamodellelement gibt „Capability“, das den Namensstring der Fähigkeit beinhaltet. Zusätzlich verweist dieses Teilmodell mit dem Capability-Element auf das Teilmodell, in dem für die Funktion benötigten Merkmale enthalten sind. Eine andere Nutzungsmöglichkeit besteht darin, dass das Teilmodell mit den funktionsspezifischen Merkmalen den entsprechenden Namensstring beinhaltet und damit für die Fähigkeit steht. Außerdem ist denkbar, die Kommandos die Zustandsübergänge eines ISA-88 Zustandsautomaten (bekannt als PackML-Zustandsmaschine) als Capability zu beschreiben. Das sind jedoch semantische Fragen, die noch weiterer Klärung bedürfen.

5. Zusammenfassung, Aktueller Stand und weitere Schritte

Der digitale Zwilling ermöglicht es, ein durchgängiges Engineering entlang des Lebenszyklus und über die Firmengrenzen hinweg zu gewährleisten. Die Verwaltungsschale der Plattform I4.0 und der Industrial Digital Twin Association (IDTA) bietet konkrete Umsetzungsmöglichkeiten. In diesem Beitrag werden am Beispiel einer fertigungstechnischen Anlage die Nutzungsmöglichkeiten der Verwaltungsschale beispielhaft aufgezeigt. Eine umfangreichere Beschreibung befindet sich in [5]. Die Verwaltungsschalen sind auf einem AASX Server zugreifbar. Zurzeit wird an der online-Kopplung der Verwaltungsschalen mit dem physischen Demonstrator gearbeitet, so dass auch über den AASX-Server die Daten einsehbar sind. Generell sind ein Verwaltungsschalen-Testbed und Demonstratoren verfügbar und Interessenten sind aufgerufen, sich aktiv daran zu beteiligen (nähere Informationen unter <http://vwsvernetzt.de/>).

Acknowledgement

Diese Arbeit wurde finanziert vom Bundesministerium für Wirtschaft und Energie im Rahmen des Projektes „Verwaltungsschale vernetzt“ mit dem FKZ: 01IS17102. Vielen Dank an die Mitarbeiterinnen und Mitarbeiter, die im Projekt zum Gelingen der Referenzmodellierung beigetragen haben. Sie sind in [4] als Autor:innen benannt.

Literatur

- [1] Federal Ministry for Economic Affairs and Energy (BMWi) - Public Relations (publisher). Plattform I4.0 Text and Editing: Details of the Asset Administration Shell Part 1 - The exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0RC01). November 2020.
- [2] Federal Ministry for Economic Affairs and Energy (BMWi) - Public Relations (publisher). Plattform I4.0 Text and Editing: Details of the Asset Administration Shell Part 2 - Interoperability at Runtime – Exchanging Information via Application Programming Interfaces0 (Version 1.0RC01). November 2020.
- [3] Bundesministerium für Wirtschaft und Energie (BMWi) (Herausgeber): I4.0 Sprache. Diskussionspapier Plattform I4.0. April 2018.
- [4] Bundesministerium für Wirtschaft und Energie (BMWi) (Herausgeber). Plattform I4.0 Text and Editing: VWS-Referenzmodellierung – Exemplarische Modellierung einer fertigungstechnischen Anlage mit

- AASX Package Explorer auf Basis des VWS-Metamodells. Diskussionspapier Plattform Industrie 4.0, April 2021.
- [4] Christian Diedrich, Alexander Bieliaiev, Tizian Schröder, Jürgen Bock, Andreas Gössling, Alexander Willner, Rolf Hänisch, Heiko Kozirolek, Andreas Kraft, Florian Pethig, Oliver Niggemann, Johannes Reich, Jens Vialkowitsch, Friedrich Vollmar, Jörg Wende, Thomas Hadlich: Sprache für I4.0-Komponenten. Automation Kongress 2017.
 - [4] VDI/VDE 2193 Blatt 1: Sprache für I4.0-Komponenten. Düsseldorf: VDI, 2019
 - [6] VDI/VDE 2193 Blatt 2: Sprache für I4.0-Komponenten. Interaktionsprotokoll für Ausschreibungsverfahren. Düsseldorf: VDI, 2019
 - [7] Bundesministerium für Wirtschaft und Energie (BMWi) (Herausgeber): Describing Capabilities of Industrie 4.0 Components. Diskussionspapier Plattform I4.0. 2020.
 - [8] Federal Ministry for Economic Affairs and Energy (BMWi) - Public Relations (publisher). Plattform I4.0 Text and Editing: Submodel Templates of the Asset Administration Shell - ZVEI Digital Nameplate for industrial equipment (Version 1.0) . November 2020.
 - [9] Federal Ministry for Economic Affairs and Energy (BMWi) - Public Relations (publisher). Plattform I4.0 Text and Editing: Submodel Templates of the Asset Administration Shell - Generic Frame for Technical Data for Industrial Equipment in Manufacturing (Version 1.1). November 2020.

Creation and Usage of Digital Twins in the Life Cycle of an Automation Solution

M.Sc. **Denis Göllner**, M.Sc. **Tobias Klausmann**,
Dipl.-Ing. **Markus Kiele-Dunsche**, Dr.-Ing. **Heiko Stichweh**,
Lenze SE, Aerzen

Abstract

The concept of the Asset Administration Shell (AAS) has been defined as the Digital Twin (DT) for Industry 4.0. It provides a common data model as well as interfaces and can be used to provide information about any asset. This publication shows the creation of AASs and the integration of their contents into existing tools. It contains a setup using available open-source solutions to add and reuse data throughout the life cycle of a machine that can be applied in various DT use cases.

Deriving the initial DT-data for industrial components from existing databases is the first presented use case. Secondly, the creation of machine-DT from tools in the engineering phase of the machine is discussed. In a third use case this contribution shows the enrichment of implemented DTs with machine data provided by controllers or inverters. The DTs contain process data for machine monitoring as well as the current machine configuration at any time. Reusing the data later in the life cycle for an automated setup of an asset management environment makes the concept of end-to-end solutions tangible.

1. Introduction

Interoperability is one of the main concepts of I4.0. It enables changeable production lines as well as better collaboration across vendors during the engineering phase of a machine life cycle. One of the main technologies needed for reaching interoperability are standardized information models. These models are used to add a self-description to any component of value inside a factory. The emerging solution for an I4.0 information model is the so-called asset administration shell (AAS).

This paper presents the usage of an available open-source implementation of AASs to implement and make use of these information models. The primary goal of this paper is to demonstrate the following aspects:

1. Possibilities of existing AAS implementations.
2. The benefits of using the AAS throughout the life-cycle of components.

Chapter two provides an overview of the relevant terms being used in this paper and describes the context of an automation solution. Afterwards, the implementation is described in chapter three. Finally, an outlook as well as open issues are presented.

2. Background

2.1 Digital Twin

The term “Digital Twin” (DT) is widely used in research as well as in worldwide industries in the last years. It was first mentioned in 2010 in NASA context where it is defined as “an integrated multi-physics, multi-scale, probabilistic simulation of a vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its flying twin” [1]. More definitions were made by various authors, as it is stated in [2]. This paper follows the definition by the Industrial Internet Consortium and the Plattform Industrie 4.0: A DT is a “digital representation, sufficient to meet the requirements of a set of use cases” [3].

2.2 The AAS as the Digital Twin for Industry 4.0

The concept of I4.0 points out that machines and components of different vendors will have to work together in terms of interoperability. To be able to reach full interoperability between machines, new and standardized interfaces are needed [4]. The AAS defines an information metamodel and standardized Application Programming Interfaces (APIs) to enable machine to machine communication as well as easier information sharing between the involved partners in the automation value chain [5]. Each asset, either physical or virtual, may have an AAS that describes the asset and provides the relevant information to anyone who has sufficient rights. To ensure the retrieval of correct information, each AAS as well as each asset shall provide a globally unique id.

The AAS metamodel is designed to provide two kinds of AASs:

1. AAS of an asset type (referred to as “type AAS”): A 1:n connection between AAS and asset instance. Contains information belonging to every asset instance (e.g. data sheet, documentation, ...)
2. AAS of an asset instance (referred to as “instance AAS”): A 1:1 connection between AAS and asset instance. Contains all the information belonging to a specific asset instance. Asset instances exist from the moment they are produced.

The contents inside the AAS are encapsulated into so called submodels, which can be standardized by various working groups [5]. In addition to submodels, views can be defined. Views provide a way to reflect and filter already available information included in the submodels.

To make the contained information machine-readable, semantics are needed. To ensure sufficient semantics, the AAS provides concept descriptions. Existing catalogues (e.g. eCI@ss¹, IEC CDD²) defining properties (by providing unique IDs, units and descriptions) as well as newly defined ones can be integrated in an AAS to enrich the data and achieve a common understanding. Fig. 1 shows the defined structure of an AAS [5]. Following the definition for DT pointed out in section 2.1, the term AAS is used synonymously to digital twin in this paper.



Fig. 1: Contents of an Asset Administration Shell, based on [5]

2.3 The AAS in the machine life cycle

One benefit of the DT concept is the enrichment and reuse of contained information throughout the life cycle. The five life cycle steps for single machine entities described in [6] were adapted in the following way: After developing an initial idea for a machine, the machine has to be designed. This involves the integration of the necessary components, picked by an engineer, into a functional entity. To facilitate the integration step, the engineer can use component information provided in the type AAS structure by the component supplier. The accumulated information that might be necessary in later steps, can be stored in a newly created type AAS

¹ <http://eclasscontent.com/>

² <https://cdd.iec.ch/>

for the specific machine solution. After the engineering is completed, the machine is manufactured.

Accompanying the manufacturing of the machine and its components, the associated instance AASs have to be created. The instance model can be augmented by adding configurations and simulations to conduct initial tests. Finally, the machine is put into production. Live data is gathered and added to the corresponding AAS, allowing "predictive maintenance" algorithms to use the gathered data to further enrich the AAS with its insights. Interactions with the live machines like triggering actions or changing the behavior are possible as well.

2.4 The value chain

Focus of this paper lies on the value chain of a machine, which becomes part of an I4.0 production line. The AAS provides the basis to enable many partners in the value chain to provide integrated solutions. Products, provided by automation solution providers, become part of a machine built by a machine manufacturer. All components have to work together to achieve interoperability. End users and their production line integrators need that interoperability, because their production lines need to be changeable in the future in order to be able to react easily to needed changes.

The variety of the value chain is enriched by providers of Manufacturing Execution Systems (MES) and Enterprise Resource Planning (ERP) solutions, which are used to orchestrate the production process. In the future, given the AAS and standardization concepts are better evolved, there might be more partners in the value chain, because new IT solutions will be much easier to implement. The validation described in the following chapter has the ambition to work along the life cycle of a machine as well as to show the benefits of the AAS throughout the automation value chain.

3. Creation and usage of the Asset Administration Shell in the machine life cycle

This chapter describes implementations for optimized execution of selected use cases. Especially by reusing the provided information across the life cycle of a machine, the benefits of the AAS concept can be presented in more detail. By analyzing the machine life cycle following the methodology from [7], a variety of use cases can be identified, in which an AAS can generate added value. To aid the implementation of the use cases, the necessary basic concepts are presented. The following use cases, put in context in Fig. 2, are introduced in the course of this chapter:

- UC 1. Provision of component information across the value chain
- UC 2. Provision of machine engineering information from engineering tools across the value chain
- UC 3. Provision of process information
- UC 4. Usage of asset information across the value chain

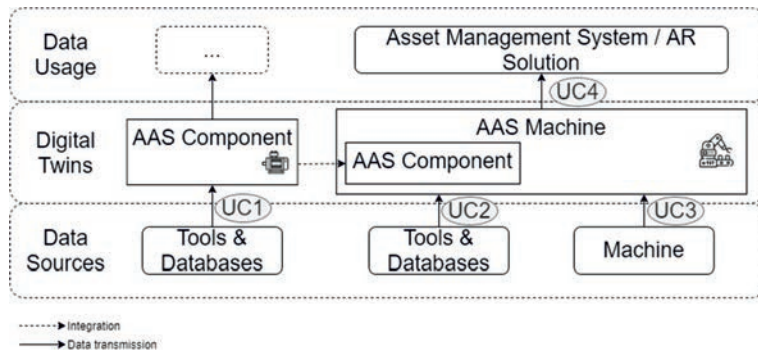


Fig. 2: Reflective overview of the provided use cases³

3.1 Basic concepts

As a basis, a software back-end implementing the AAS meta-model and its concepts is needed. In this paper the BaSyx Java software development kit⁴ ("Basyx Middleware") is used, as it provides a near-full implementation of the AAS meta-model. The middleware provides a Java API for interaction with itself. In combination with an Apache Tomcat⁵ server, BaSyx has the ability to serve AASs via the common http(s) protocol. Based on this, a REST API, using JSON as data exchange format, is available.

Both APIs provide the ability to read data from and write data to the AASs as well as to invoke operations. Based on this technology stack AASs can be implemented.

To enhance the interoperability of the provided information, the recommended submodels published in the Plattform Industrie 4.0 papers "Verwaltungsschale in der Praxis" [8] and "Details of the Asset Administration Shell" [5] are implemented. In the type AASs the following submodels are included:

³ Icons made by Payungkead from <https://www.flaticon.com/>

⁴ <https://projects.eclipse.org/projects/technology.basysx>

⁵ <http://tomcat.apache.org/>

- “Asset Identification”
General asset information, containing basic information, e.g. the assets vendor name, the asset name.
- “Bill of Material (BoM)”
A list of references to assets, used to create a tree structure for the complete machine.
- “TechnicalData”
Contains general descriptive details about the asset, as well as additional necessary information e.g. normative protection classes or certificates.
- “Documentation”
Provides detailed documentation of the asset, e.g. a setup guide.

In the instance AAS the standard submodels “Asset Identification” and “BoM” are implemented. In contrast to the type AASs, these submodels reflect information of a live machine. In addition, the submodels “Operational Data”, a submodel for providing live data of the asset and the connected process, as well as the submodel “Asset Management”, a submodel for managing and reviewing the status of the live asset, are implemented.

One inherent aspect of the AAS is to be machine-readable. Furthermore, it can be distributed among several computer systems and access points. As a result, the AAS is hard to comprehend, yet for creation and inspection of the included machine data, it ought to be understandable by humans. To aid in the understanding of the data, an AAS Viewer is used. The browser-based tool, seen in Fig. 3 and written in Javascript, HTML5 and CSS, provides the data in a human readable form, by querying all the information of an AAS and presenting it. Related data points are shown in a visually compiled form. For interaction purposes data can be changed and AAS operations can be invoked by using the AAS REST interface. The “AAS Browser” is in the process of being released and is intended to become available via the “Plattform Industrie 4.0” open source project⁶.

3.2 Provision of component information across the value chain

In order to enable actors in the value chain to interact with components, information that describes the asset is needed.

- In an engineering process, where mainly the characteristics of a component are important, the type information (see “type AAS” in subchapter 2.2 The AAS as the Digital Twin for Industry 4.0) is used. To identify

⁶ <https://github.com/admin-shell-io/web-aas-client>



Fig. 3: Exemplary view of the AAS Viewer

assets, they have to be distinguishable. To satisfy this, the "Asset Identification" submodel is used. Furthermore, the asset has to be described thoroughly in its characteristics. The submodel "Technical Data" is used for this purpose. In the course of providing information, easy to understand for humans, the "Documentation" submodel is used.

- While newly commissioning or in a maintenance scenario, the instance AAS is usable to identify the component ("Asset Identification" submodel) or to inspect the current component status ("Operational Data" submodel). The submodels structure can be predefined, although a connected asset is needed to contain useful information.

Supplier companies have large amounts of information available already. This information is mainly maintained in internal databases. Parts are already provided in existing data exchange standards like BMEcat or eCl@ss. Using the AAS standard generates benefits by being able to connect the information with asset instance process data in later life cycle phases. To provide the necessary information in the AAS format, it needs to be converted into the defined

structure. The information has to be mapped into the presented submodels introduced in the last chapter, if possible. If sufficient semantics are not maintained yet, the available information has to be enriched to fulfill the AAS requirements.

3.3 Provision of machine engineering information from engineering tools across the value chain

The first step of the machine life cycle is the machine engineering. Information about the machine is generated in the course of this phase. Automation suppliers provide tools aiding the machine builder during these steps. An example is the Lenze Engineering Tool “Easy System Designer (ESD)”. After setting up the machine structure inside the tool, the contained information can be made available for the machine builder to be reused in later life cycle phases. Via a provided REST-API the information about the machine solution can be exported. Providing that information in a standardized way makes it possible to share the data across the value chain and reuse it in specific tools. To fulfill this requirement, the tools data structure is transformed into the AAS structure. Fig. 4 shows the example machine used for the implementation.

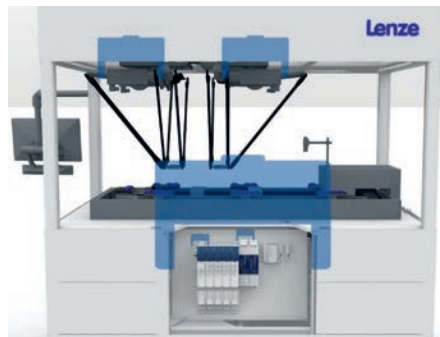


Fig. 4: A sorting cell as demo machine

The main information coming from the tool can be integrated into the machine's AAS. This mainly includes details about the machine topology as well as the components that are part of the machine. The created AAS contains a “BoM” submodel describing a tree structure of all included components, reflecting the machine topology. Furthermore the asset identification, the specific technical information, as well as the specific documentation for the machine are included as submodels.

The machine shown in Fig. 4 consists of four machine modules - two conveyors and two picker arms. To follow the definitions of the Plattform Industrie 4.0, every asset, which may be a physical or logical object, having a value to the organization, needs its own AAS [8]. In terms of a modular machine, an AAS for the machine as well as AASs for the modules of the machine and the components inside these modules are needed. Fig. 5 shows the relations between those AASs. The machine as well as each module contains references to the underlying parts. Following the story across the life cycle of the machine, there is a need of an AAS for every asset type as well as a separate AAS for every asset instance.

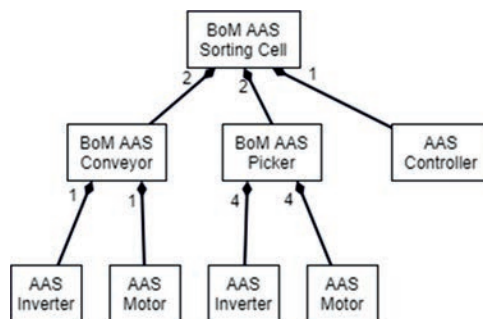


Fig. 5: Topology of the demo machine's nested components

3.4 Provision of live machine information

In the context of production, the "Product, Process and Resource concept" is commonly referred to, for a definition see [9]. For all three, adaptations are possible:

- A product's receipt or characteristics can be altered, which may require changes in the production process or the involved (production) resources.
- A process can be modified to better fit the given production resources and may involve a different set of resources.
- Resources, e.g. production machines, can be replaced, altered or removed.

The AAS has to reflect the changes, as they occur. The changes can be classified into two main aspects:

1. Change of a resource: Within a live production process, the status of the used machines (e.g. the position of a robotic arm in a sorting cell) changes while executing a pick-and-place application. The machine state is reflected by continuously synchronizing the AAS. For the synchronization, a proxy application is used to update the AAS information model based on the connected OPC UA information model. This

application also has the ability to map AAS operations to functions of the asset. As an example, an identification operation is provided that flashes an indication LED in order to identify the asset inside the cabinet.

2. Adaption of available resources: In the context of an evolving production plant, a changed setup of machines or machine components can be reflected, using AAS methodology. For this purpose, the "BoM" submodel is used and updated, containing a list of references to AASs of subcomponents included in the particular component [5]. The "BoM" is updated by adding, replacing or removing references to assets included in the machine. With this, changes in the component hierarchy, as well as in the referenced components, can be reflected.

3.5 Usage of Asset Information across the value chain

One of the main aspects of DTs is to reuse information in later life cycle phases. Setting up an asset management system as a machine end user used to be a mainly manual task. Making use of already available information results in shorter integration and deployment times. When following the AAS concept, relevant information of the machines as well as the included components are available from the engineering phase described earlier (see subchapter 3.2). An "AAS-Importer" was developed to integrate the available information into the "Lenze Asset Management" system. With the help of this tool, the necessary information is converted into the tool-specific information model. With the standardized AAS model as the information source, integration of third party components is easily possible. This enables the asset management to provide use case specific user interfaces upon the information by simultaneously interacting with the AAS.

Providing information in the standardized AAS format enables the development and easy integration of new software solutions. The upcoming solutions can be reused more easily, since the same information can be expected in the AASs of different manufacturers. One example is the augmented reality solution shown in Fig. 6, which is built upon a previously defined AAS-view.

4. Conclusion and future work

This paper illustrates the possibility to implement end-to-end solutions with existing AAS software. By integrating available machine information from engineering tools as well as already existing component information, all relevant information to describe machines can be



Fig. 6: Usage of an AAS View to display information

shared in a single source of truth. Benefits can be achieved by reusing data throughout the life cycle of a machine.

Component information can be made available in the AAS format by converting existing data structures into the relevant submodel structures. Adaptions of the BoM submodel can be used to describe machines consisting of multiple modules and components.

Since the concept is mainly about new interfaces, it is easy to integrate the AAS into existing tools. The implementation of the AR solution by an external supplier made clear, how easy it might be for new and innovative players to create new solutions in the automation industry in the future.

In further studies it is important to run the implementations in larger scale. As a consequence, the distribution of AASs across different locations, needs to be verified as well. The concept of AAS and registries makes it possible to deploy AASs or even single submodels in different locations like on the asset itself, on an edge device or in the cloud.

Additionally, a validation using different AAS implementations at the same time should be done. In the future, there might be AASs implemented in OPC-UA as well as in BaSysx or other tools, which have to be working together.

Acknowledgment

This work is partly funded by the German Federal Ministry of Education and Research (BMBF) through the project BaSys 4.2 (grant no. 01IS19022) as well as by the Ministry of Economic

Affairs, Innovation, Digitalisation and Energy of North Rhine-Westphalia (MWIDE NRW) through the project itsowl-TeDZ (grant no. 1807OW001).

Bibliography

- [1] M. Shafto, M. Conroy, R. Doyle, E. Glaessgen, C. Kemp, J. LeMoigne and L. Wang, "Draft modeling, simulation, information technology & processing roadmap," *Technology Area*, vol. 11, 2010.
- [2] E. Negri, L. Fumagalli and M. Macchi, "A review of the roles of digital twin in cps-based production systems," *Procedia Manufacturing*, vol. 11, p. 939–948, 2017.
- [3] Industrial Internet Consortium, *IIC Vocabulary*, 2019.
- [4] R. Drath, "Industrie 4.0 — eine Einführung," *Open automation*, vol. 3, p. 17–21, 2014.
- [5] Plattform Industrie 4.0, "Details of the Asset Administration Shell Part 1 The exchange of information between partners in the value chain of Industrie 4.0," 2018. [Online]. Available: <https://www.plattform-i40.de/I40/Redaktion/DE/Downloads/Publikation/2018-verwaltungsschale-im-detail.html>.
- [6] C. Wagner, J. Grothoff, U. Eppler, R. Drath, S. Malakuti, S. Grüner, M. Hoffmeister and P. Zimmermann, "The role of the Industry 4.0 asset administration shell and the digital twin during the life cycle of a plant," in *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2017.
- [7] D. Göllner, R. Bernijazov and M. Kiele-Dunsche, "Vorstellung und Anwendung einer Methodik zur Identifikation und Analyse von Anwendungsfällen als Basis für die Ausgestaltung von I4.0-Verwaltungsschalen," in *VDI-Kongress Automation*, 2019.
- [8] H. Bedenbender, J. D. Bock, B. D. Boss, C. P. D.-I. Diedrich, K. Garrels, A. D. Graf Gattenburg, K. Heidrich, O. Hillermeier, B. Rauscher, M. Sauer, J. Schmidt, T. Werner and P. Zimmermann, "Verwaltungsschale in der Praxis - Version 1.0," Plattform Industrie 4.0, <https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/2020-verwaltungsschale-in-der-praxis.html>, 2020.
- [9] J. Pfrommer, M. Schleipen and J. Beyerer, "PPRS: Production skills and their relation to product, process, and resource," in *2013 IEEE 18th Conference on Emerging Technologies Factory Automation (ETFA)*, 2013.

Flexible Industrie 4.0 Verwaltungsschale für dynamische IIoT-Systeme

Zeitreihen in der Verwaltungsschale

Dr.-Ing. **F. Hilbert**, M.Sc.-Ing. **S. Soler Perez Olaya**,
Professur für Prozesskommunikation, Technische Universität Dresden

Kurzfassung

In zukünftigen Industrie 4.0-Systemen wird die Anzahl der kollaborierenden IIoT-Geräte weiter zunehmen. Damit ein solches Gerät als Asset in einem interoperablem Industrie 4.0-System verwendet werden kann, muss es als Industrie 4.0-Komponente agieren, d.h. es muss eindeutig identifizierbar und kommunikationsfähig sein sowie seine Eigenschaften durch eine Industrie 4.0-Verwaltungsschale beschreiben. Eine solche Verwaltungsschale als lebenszyklusbegleitende digitale Repräsentanz eines Assets beschreibt Merkmale, Daten und Funktionen in fachlich voneinander abgegrenzten Teilmodellen. Dabei kann die Verwaltungsschale entweder durch das Asset selbst bereitgestellt werden oder aber durch ein Gateway bzw. über ein Repository in einer Cloud-Anwendung zur Verfügung gestellt werden. Aufgrund der für IIoT-Geräte üblichen Ressourcenbeschränkungen und den damit einhergehenden stromsparenden bzw. kanalbelegungssparenden Kommunikationsstrategien ist eine permanente Kommunikation mit IIoT-Geräten nicht immer gegeben (z. B. batteriebetriebener Sensorknoten). Ein Ansatz um den Nachrichtenverkehr von der Erreichbarkeit der IIoT-Assets durch einen Verwaltungsdienst zu entkoppelt ist in [1] beschrieben. IIoT-Assets sind im Vergleich zu herkömmlichen Feldgeräten wesentlich dynamischer (z. B. Laufzeit-Daten, Änderungen der Eigenschaften und Funktionalitäten, etc. aufgrund Verschleiß, prozessbedingter oder äußerer Einwirkungen, Software-Updates oder gar Funktionsumstellungen). Daher müssen Industrie 4.0-Verwaltungsschalen für solche IIoT-Assets nicht nur zuverlässig erreichbar extern zur Verfügung gestellt werden, sondern sich auch an Änderungen permanent anpassen.

Dieser Beitrag beschreibt ein Konzept, das die Nutzung von Industrie 4.0 Verwaltungsschalen in dynamischen Industrial IoT Umfeld erlaubt. Dafür werden spezielle Synchronisations-Maßnahmen beschrieben, um die betreffenden inhaltlichen, beschreibenden oder funktionalen Merkmale in den spezifischen Teilmodell-Merkmalen zu aktualisieren.

Schlüsselwörter: IoT, IIoT, Industrie 4.0, Verwaltungsschale, Digitaler Zwilling

1. Einführung

Im Industrie 4.0-Kontext bezeichnet ein Asset ganz allgemein einen materiellen oder immateriellen Vermögenswert. Für viele Anwendungsfälle (z. B. flexible Konfiguration für eine größere Produktvielfalt, vereinfachte Inbetriebnahme und Wartung [2]) müssen Informationen über verschiedene Assets standardisiert ausgetauscht werden. Für diesen Austausch wurde das Konzept der Industrie 4.0-Verwaltungsschale (des Weiteren Schale) entwickelt. Unter den materiellen Assets, die direkt für die industrielle Produktion genutzt werden, finden sich auch immer mehr Industrial IoT-Geräte. Daher müssen zukünftige Verwaltungsschalen auch die besonderen Eigenschaften von Industrial IoT-Assets adressieren können. Um diese Problemstellung zu konkretisieren, wird in diesem Kapitel zunächst das Konzept der Verwaltungsschale betrachtet. Anschließend werden kurz die Eigenschaften verteilter IIoT-Systeme analysiert und daraus Anforderungen an die Verwaltungsschale abgeleitet.

1.1. Konzept der Industrie 4.0-Verwaltungsschale

Die Industrie 4.0-Verwaltungsschale soll den Austausch von Asset-Informationen zwischen den verschiedenen Beteiligten im Lebenszyklus eines Produktionssystems unterstützen. Dafür enthält eine solche Verwaltungsschale grundsätzlich die Beschreibung von Merkmalen, Daten und Funktionen für ein Asset. Damit andere I4.0-Komponenten und weitere Systeme auf die Informationen der Verwaltungsschale zugreifen und diese für eine kooperative Zusammenarbeit korrekt verwenden können, müssen die ausgetauschten Informationen maschineninterpretierbar sein und einen ausreichenden Grad an Formalismus sowie semantische Eindeutigkeit bieten [2,3]. Daher werden die grundsätzliche Struktur und die Inhalte einer Verwaltungsschale vorgegeben [2] und obligatorische Basis- und Pflicht-Merkmale, die immer verfügbar und mit sinnvollen Werten belegt sein müssen, definiert [2]. In einem Header-Bereich der Verwaltungsschale informieren eine Liste von Merkmalen über die Verwendung der Assets (z. B. jeweils eindeutige globale und interne Verwaltungsschalen- und Asset-IDs, Bezeichnungen, Kontext, Sichten). In dem darauffolgenden Body-Bereich werden in maschineninterpretierbarer Form Eigenschaften und Funktionalitäten der Assets bezüglich verschiedener technischer Bereiche und Intentionen in voneinander getrennten unabhängigen Teilmodellen beschrieben. Dabei verfügt jedes Teilmodell über Merkmale, welche auf individuelle Daten und Funktionen referenzieren. Insbesondere können auch einzelne Merkmale andere Verwaltungsschalen sowie weitere I4.0-konforme Entitäten und Informationen außerhalb der eigenen Verwaltungsschale **referenzieren**. Somit wird zum einen eine Vernetzung von Informationen unterschiedlicher Wissens Domänen ermöglicht (z. B. unterschiedliche Standards) und zum anderen kön-

nen auf diese Weise anwendungsspezifische Sichten realisiert werden (ISO/IEC 81346, Digitale Fabrik, etc.) [2] oder durch Referenzen auf ein gemeinsames Modell von Fähigkeitsbeschreibungen Capabilities beschrieben werden [5]. Während für die Strukturen der Teilmodelle in strenges, standardisiertes Format gefordert wird (aufbauend auf IEC 61360) können für die referenzierten Informationen und Funktionen unterschiedliche freie und proprietäre (ggf. Hersteller- oder domänenspezifische) Datenformate verwendet werden. Um der Vielzahl von aktuellen und zukünftigen Anwendungsfällen gerecht zu werden sind der Anzahl und Varianz dieser Teilmodelle keine Grenzen gesetzt. Eine Abfrage-Möglichkeit nach außen bietet der Komponenten-Manager, welcher sich gemeinsam mit den Teilmodellen im Body-Bereich befindet. Bild 1 zeigt den strukturellen Aufbau der Industrie 4.0-Verwaltungsschale (Version 3). Die Merkmale eines Assets lassen sich zudem nach dem Aspekt ihrer Änderungsdynamik unterscheiden. So werden sowohl Informationen über statische Merkmale (z. B. IDs, Katalogdaten, Kennblätter, Handbücher, Auslegungen, Sichten Definitionen) als auch über relativ dynamische Merkmale (z. B. Parameter und Variablen) aber auch über hochgradig dynamische Merkmale (z. B. Laufzeitdaten, Zustandsgrößen, Operationen oder Ereignisse) ausgetauscht [4]. Somit dient eine Schale nicht nur als statische Datenquelle für die I4.0 Merkmale, sondern kann auch dynamische Aktualdaten (Laufzeitinformationen) enthalten. So ist es beispielsweise für Predictive-Maintenance-Anwendungen für Batteriebetriebene Sensorknoten essentiell, Informationen über die aktuelle Akkumutzung bzw. Akkustand zu erhalten. Dafür kann die Schale nach außen hin Laufzeit-Daten vom Asset aufnehmen und abbilden (z. B. Stromstarke, Position, Temperatur etc.).

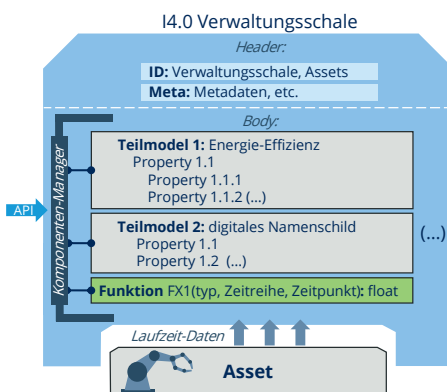


Bild 1: Struktur der I4.0 Verwaltungsschale

Des Weiteren können Verwaltungsschalen, zum einen entsprechend ihrer Rolle in der Wertschöpfungskette und zum anderen bezüglich ihrer Kommunikationsfähigkeit, in unterschiedlichen Ausprägungen zur Verfügung gestellt werden [5]. Zum ersten Aspekt unterscheidet man zwischen Verwaltungsschalen für Asset-Typen und Asset-Instanzen und zum zweiten Aspekt wird unterschieden in passive, aktive und proaktiver Kommunikationsfähigkeit der Verwaltungsschale (in Anlehnung an VDI/VDE GMA FA 7.21). Während eine passive Verwaltungsschale als Datei in XML- oder JSON-Format daherkommt [6], besitzt eine aktive Verwaltungsschale (Reactive AAS) eine API, die z. B. über einen IP-basierten Zugang zugreifbar ist. So kann die Verwaltungsschale als Interaktionsfassade mit anderen Komponenten interagieren und Funktionen zum Lesen und Verändern der Merkmale bereitstellen [7]. Proaktive Verwaltungsschalen dahingegen interagieren aktiv über eine spezielle I4.0-Sprache (VDI/VDE-2193) [8] miteinander und können dezentrale kooperative Aufgaben auf eigener Initiative ausführen [7]. Damit die Verwaltungsschale Asset Laufzeit-Informationen verarbeiten kann, muss sie ein Mindestmaß an Aktivität bieten. Daher beschränken wir uns im Verlauf der weiteren Arbeit auf aktive und proaktive Verwaltungsschalen für Asset-Instanzen.

1.2. Eigenschaften verteilter Industrial IoT-Systeme

Obwohl Industrial IoT-Assets durch eine große Gerätevielfalt geprägt sind, haben sie doch auch Gemeinsamkeiten bezüglich der Beschränkung ihrer Ressourcen. Zumeist unterliegen diese eingeschränkten Geräte (Constrained Devices) technischen Einschränkungen in verschiedenen Aspekten [9]. Die Einschränkungen können beispielsweise die zur Verfügung stehenden Bandbreite, die Stromversorgung oder die Rechenleistung betreffen [10]. Diese Geräte werden nach RFC 7228 [11] in drei Klassen eingeteilt. Geräte der Klasse 0 haben neben anderen Beschränkungen oft nicht die Möglichkeit, selbsttätig und gesichert über das Internet zu kommunizieren und sind daher auf einen Proxy oder ein Gateway angewiesen (z. B. Sensorknoten). Geräte der Klasse 1 sind zwar in ihrer Kommunikation weniger beschränkt, können jedoch im Gegensatz zu Geräten der Klasse 2 nicht direkt über den HTTP-TCP/IP-Stack mit dem Internet kommunizieren, sondern benötigen eigene Protokolle. Zusätzlich sind Industrial IoT-Assets oft nicht direkt mit dem Energienetz verbunden, sondern auf andere, weniger leistungsfähige Energiequellen (z. B. Batterien oder Energy Harvesting) angewiesen [13]. Da die Kommunikationsschnittstelle meist der wesentliche Energieverbraucher solcher Geräte ist, wird hier mit verschiedenen energiesparenden sowie Bandbreitenschonenden Kommunikationskonzepten gearbeitet. [12] unterscheidet hier: (a) jederzeit aktiv (always on) - höchster Energiebedarf, geringste Latenz; (b) komplette Abschaltung der Schnittstelle, wenn nicht be-

nötigt (normally off) - geringster Energiebedarf, höchste Latenz; (c) energiesparender Hibernat-Modus, wenn Schnittstelle nicht benötigt wird (low power) - mittlerer Energiebedarf, erhöhte Latenz. Aufgrund der Rechenleistungs- und Speicher Beschränkungen müssen ggf. komplexe Berechnungen und umfangreiche Speicherbelegungen auf dem Gerät vermieden, bzw. auf ein notwendiges Minimum reduziert werden. Ebenso kann die Erreichbarkeit der Industrial IoT-Asset eingeschränkt sein. Daher muss eine Industrie 4.0-Verwaltungsschale für solche Geräte an einer von außen zuverlässig erreichbaren Stelle positioniert und ausgeführt werden.

1.3. Problemstellung und Lösungsansatz für eine resiliente Verwaltung

Aufgrund der beschriebenen für IIoT-Geräte üblichen Ressourcenbeschränkungen und den damit einhergehenden energie- bzw. kanalbelegungssparenden Kommunikationsstrategien ist eine permanente Kommunikation mit IIoT-Geräten nicht immer gegeben (siehe Störung der Übertragung von Laufzeitdaten in Bild 1). Daher müssen Industrie 4.0-Verwaltungsschalen für solche Industrial IoT-Assets geeignete Kompensationsstrategien zur Verfügung stellen und insbesondere unvollständige oder lückenhafte Laufzeitdaten tolerieren. Während für herkömmliche Feldgeräte die Produktivkommunikation nicht über die Verwaltungsschale realisiert wird, ist diese Vorgehensweise bei Industrial IoT-Assets ohne Feldbus-Schnittstelle durchaus sinnvoll. Angenommen, ein Messwert eines Sensors soll zu vorgegebenen Zeitpunkten bestimmt werden (Aufnahmezeitpunkt) und der Sensor erfasst zu verschiedenen Zeitpunkten einen realen Messwert (Erfassungszeitpunkt) und überträgt diesen. Fallen nun Aufnahmezeitpunkt und Erfassungszeitpunkt nicht aufeinander, so muss der Wert angenähert werden. Liegt der Aufnahmezeitpunkt zwischen zwei Erfassungszeitpunkten, so lässt sich der gesuchte Wert interpolieren. Liegt hingegen der Aufnahmezeitpunkt hinter dem letzten Erfassungszeitpunkt, so lässt sich nur durch Extrapolation ein verwendbarer Messwert annähern. Auf diese Art lassen sich Verzögerungen oder Ausfälle, die durch die Latenzzeit oder Nichterreichbarkeit von IIoT-Geräten hervorgerufen werden, kompensieren. Ein extrapolierte Sensorwert wird aus historischen Zeit-Messwert-Paaren (vergänger Erfassungszeitpunkte) berechnet. Die Berechnung erfolgt anhand einer Extrapolationsvorschrift für einen vorgegebenen Zeitpunkt (Extrapolationszeitpunkt). Während eine Interpolation jederzeit auf der Basis vorhandener historischer Werte auch auf Anwendungsebene realisiert werden kann, so sind bei der Extrapolation verschiedene Randbedingungen zu beachten. Keinesfalls darf ein extrapolierte Wert gekennzeichnet als Realwert übertragen werden. Ebenfalls sollen verschiedene Verwaltungsschalen denselben Wert auf der Basis einer einheitlichen Extrapolationsvorschrift extrapolie-

ren. Dafür muss bereits in der Verwaltungsschale für das betroffene Merkmal angegeben werden, ob und wie es zu extrapolieren ist. Im folgenden Kapitel werden dafür verschiedene Varianten diskutiert und eine Lösungsstrategie für ein robustes System vorgestellt, das die Ressourcenbegrenzung von Industrial IoT-Assets adressiert und die synchronisierte Messwerterfassung von der Latenzzeit und Erreichbarkeit entkoppelt.

2. Resiliente IIoT-Verwaltung in der Verwaltungsschale

Resilienz bedeutet in diesem Kontext, Konzepte bereitzustellen um eine vorübergehende Nichterreichbarkeit von Assets tolerieren zu können. Es sind verschiedene Situationen denkbar, in denen ein Asset nicht in der Lage ist eine Anfrage zu bearbeiten bzw. eine angefragte Information zu liefern. Eine Unterbrechung kann beispielsweise bedingt sein durch Aussetzer bei der Messwerterfassung oder bei der Übertragung erfasster Messwerte. Auch stellen manche Sensoren keine Messwerte bereit, wenn sich die Werte nicht verändern. In all diesen Fällen müssen Eigenschaftswerte extrapoliert werden, um bei der Verwendung von Laufzeitdaten nicht auf die Erreichbarkeit der Assets angewiesen zu sein. Dafür wird im folgenden Abschnitt das Konzept eines Geräteverwaltungsdienstes vorgestellt, der für IIoT-Geräte das Gerätemanagement von der Erreichbarkeit der Geräte entkoppelt. Anschließend werden Varianten und Orte der Extrapolation diskutiert sowie die Persistierung historischer Daten in der Verwaltungsschale betrachtet.

2.1. IIoT-Verwaltungsdienst

Um einen permanent zugreifbaren digitalen Zwilling für gelegentlich offline IIoT-Geräte zur Verfügung zu stellen wurde in [1] ein Geräteverwaltungsdienst vorgestellt, der für IIoT-Geräte das Gerätemanagement von der Erreichbarkeit der Geräte entkoppelt. Für die Synchronisierung der beteiligten IIoT-Geräte puffert der Dienst sowohl Managementbefehle als auch strukturierte Gerätezustände. Durch einen push-basierten Updatemechanismus (Server wartet auf Gerätemeldung) werden die ressourcenlimitierten IIoT-Geräte nicht durch wiederholte Abfragen belasten. Um nicht verfügbare aktuelle Werte abzuschätzen soll der Verwaltungsdienst auch in der Lage sein, selten aktualisierte Informationen entsprechend einem Trend zu extrapolieren (z. B. Batteriestand, Temperaturtrends. Pegeltrends, etc.).

2.2. Varianten der Extrapolation

Um den zeitlichen Verlauf von Messwerte abschätzbar zu machen existieren verschiedene Extrapolations-Algorithmen. So existieren neben den einfachen polynomischen Extrapolationen für gleichförmige Änderungen (Messwerte ändern sich im Lauf mit gleichbleibendem Verhalten) auch Modellbasierte prädiktive Extrapolationen (PID-Regler als Prozessmodell) oder

komplexe Zustandsschätzer (z. B. Model Predictive Control, wo in Kalman-Filter zusätzliche weitere Mess- oder Regelgrößen einfließen). Die Auswahl des verwendeten Extrapolations-Algorithmen hängt vom erwarteten Verhalten des betrachteten Wertes ab. Dieses kann für verschiedene Prozessphasen unterschiedlich sein. Beispielsweise kann für eine Beschleunigungsphase quadratische und für eine Phase gleichförmiger Änderung lineare Extrapolation benutzt werden. Auch lässt sich mit entsprechender Modellierung ein bekannt sprunghaftes Änderungsverhalten abbilden. Wesentlich ist, dass die Extrapolationsvorschrift für einen Wert einheitlich vorgegeben ist.

2.3. Position der Extrapolation

Die Schnittstelle einer aktiven Verwaltungsschale muss permanent erreichbar sein. Bestehende Konzepte der Verwaltungsschale legen nicht fest, an welcher Stelle eine aktive I4.0-Verwaltungsschale ausgeführt wird [9]. Der Ort, wo unvollständige oder lückenhafte Nutzdaten in kontinuierliche Daten umgewandelt werden, lässt sich auf verschiedenen Systemebenen positionieren. Bild 2 zeigt die wesentlichen Varianten.

Die unterschiedlichen Varianten weisen verschiedene Vor- und Nachteile auf. **Variante A** zeigt die Extrapolation auf einem Geräteknoten: In diesem Szenario ist das IoT-Gerät ein intelligentes Gerät, das die Generierung extrapolierter Daten ermöglicht. Vorteile sind die sichere Erreichbarkeit der Daten und die direkte Integration der Funktionalität. Als Nachteil sind die u.U. geringe Rechenleistung und die hohe Komplexität der Interfaces wichtig. Ein mögliches Anwendungsbeispiel für diese Variante sind Smart Meters. Durch die sichere Erreichbarkeit der Daten für das Gerät ist hier eine Extrapolation nicht kommunikationsrelevant.

Variante B beschreibt die Extrapolation auf einem Gateway-Knoten: Ein Gateway, das mehrere Sensoren oder Geräte über verschiedene Wege (BLE, GPIO usw.) verbindet, extrapoliert die Gerätedaten, um Gateway-Funktionen wie Analysefunktionen oder andere geschäftsspezifische Funktionen bereitzustellen. Hier besteht der Vorteil des in der Regel unbegrenzten Vorhandenseins elektrischer Energie. Dadurch wird die Rechenleistung und die Erreichbarkeit nicht mehr ressourcenbedingt begrenzt. Nachteil ist die höhere Komplexität des Gateways und die Erfordernisse für die entsprechende Rechenpower. Ein gängiges Anwendungsszenario ist hier die Integration von IIoT Funktionalitäten in Brownfield Situationen.

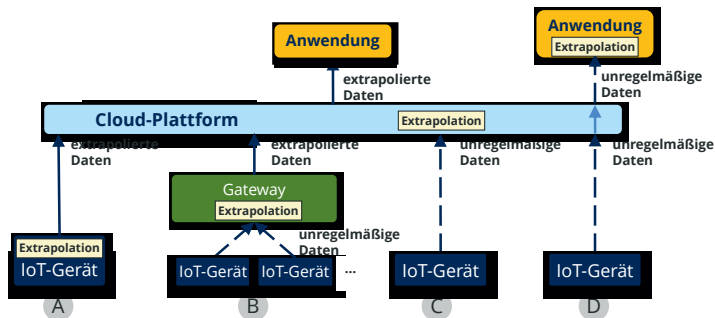


Bild 2: Verschiedene Positionen der Extrapolation

Variante C zeigt die Extrapolation auf einer IoT-Plattform: wie bei einem Gateway-Knoten extrapoliert eine Cloud-IoT-Plattform eingehende Nutzdaten von verschiedenen Protokolladaptern (MQTT, CoAP usw.), für IoT-Lösungen in Richtung Endanwendung (z. B. Datenanalysefunktionen). Der Vorteil einer Extrapolation auf einer IoT-Plattform besteht in der Flexibilität der Gestaltung von verschiedenen Datenverarbeitungsfunktionen und die Skalierbarkeit. Als Nachteil besteht die inhärente Abstraktion der IoT-Plattform, die mit der nicht mehr quellenagnostische Verarbeitung der Daten konfrontiert ist. Typische Anwendungsfall wäre die Normalisierung der Daten bzw. Vorarbeit für die Einspeisung der Daten in KI-Anwendungen.

Variante D Extrapolation auf einem Anwendungsknoten: Eine Anwendung würde Daten extrapolieren, um unabhängig von unterstützten Geräten zu bleiben und sich mehr auf die anwendungsspezifischen Geschäftsfunktionen zu konzentrieren. Die Extrapolation von reinen Nutzdaten auf diese Ebene ergibt wenig Sinn, da bereits andere Strategien, z. B. [11] bekannt sind. Zudem kann es zu Synchronisationsproblemen kommen, wenn miteinander kommunizierende Anwendungen Asset-Merkmale auf unterschiedliche Weise annähern. Unter Berücksichtigung der Eignung der verschiedenen Extrapolationsorte, legen wir uns in der weiteren Arbeit auf die Variante B fest und verorten die Nutzdatenextrapolation auf das Gateway.

2.4. Historische Daten in der Schale

Für das Management von IoT-Geräten ist die Erfassung einiger Merkmale über dem gesamten Lebenszyklus notwendig (z. B. Akku-Ladezyklen oder Werkzeug-Verwendung für Ermittlung der Abnutzung). Diese historische Information bildet eine Basis für Predictive Maintenance Anwendungen. Wie in [14] beschrieben, sollen diese Information das Gerät den ganzen Lebenszyklus begleiten. Die dazugehörenden Daten könnten, wie bei der Verwaltungsschale von Bosch [15], nur in der Laufumgebung der Verwaltungsschale betrachtet werden. Den Nachteil

wäre in Falle eines Laufumgebung-Wechsels, dass diese Daten verloren gehen würden. Es wäre auch möglich diese Daten seitens der Laufumgebung zu speichern, z. B. in Form von Log-Dateien oder XML-Listen. Diese Dateien sollten mit der Verwaltungsschale gekoppelt werden und in Falle eines Laufumgebung-Wechsels mitgezogen werden. Eine andere Variante besteht darin, die Information direkt als Teilmodell in der Verwaltungsschale zu speichern. Unabhängig von der Modalität der Speicherung ist einen Teilmodell-Template zur Interpretation dieser Informationen notwendig. Aus der Notwendigkeit zeitbezogene Informationen zu bearbeiten, folgt das Problem der Repräsentation dieser Informationen. Die Verwaltungsschale [6] hat keinen nativen Stellvertreter für Tupel oder Vektoren in seinem Meta-Modell analog zu Ontologien wie OWL [16]. Dadurch ist es unbedingt notwendig, diese mit einem Teilmodell explizit zu definieren. Insbesondere soll der Teilmodell Zeit-Reihen definieren. Erst dadurch wird es möglich, die während des ganzen Lebenszyklus erfassten Information anzuwenden. Im Folgenden werden zwei Teilmodelle für die Persistenz der historischen Daten sowie für die Beschreibung der Interpolationsoperation vorgestellt.

2.4.1. Teilmodell-Template für historische Information

Um die Repräsentation von Datenreihen, in unseren Fall Zeitreihen, zu ermöglichen, wird auf demselben Prinzip wie in der Ontologien gegriffen. Für die Definition einer verketteten Liste genügt es, ein einzelnes Gliedelement zu definieren. Das Zeitreihenelement (ZRE) hat zwei Properties, Zeitstempel und erfassten Wert sowie ein Referenzelement, das zum nächsten ZRE zeigt. Soll einen ZRE das letzte einer Kette sein, so bleibt die Referenz zum nächsten ZRE undefiniert. Die verkettete Liste die aus der ZRE Reihe entsteht kann nur in die eine Richtung traversiert werden. Sollte die Liste in beide Richtungen traversierbar sein, so wird ein weiteres Referenzelement zur Verknüpfung des vorherigen ZRE benötigt. Für die Extrapolation kann dies von Vorteil sein, da der Algorithmus den letzten ZRE als Startpunkt verwendet.

2.4.2. Teilmodell-Template für Extrapolation historischer Information

Das Teilmodell für Extrapolation historischer Information definiert eine Operation für die polynomische Extrapolation beliebiges Grades eines Wertes. Diese Operation hat als Argument zwei Properties (polynomischer Grad und Extrapolationszeit) und ein Referenzelement (letztes ZRE). Als Ausgabe dienen zwei Properties. Eine bekommt den extrapolierten Wert während die andere einen Statuswert liefert. Über das polynomische Grad wird implizit die Anzahl an involvierten ZREs definiert, z. B. für eine Extrapolation 0-ten Grades reicht mit einen ZRE, für 1. Grades sind zwei ZRE nötig, usw. (siehe Bild 3). Es ist wichtig für die Extrapolation dieselbe Zeitreferenz zu nutzen wie für das Aufnehmen der ZRE. Der Statuswert gibt Auskunft über den Erfolg oder Misserfolg der Extrapolation.

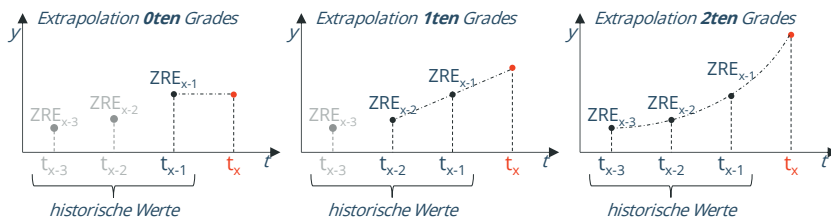


Bild 3: polynomische Extrapolation verschiedenen Grades

Weitere Extrapolationsoperationen sind geplant, z. B. PID-Extrapolation und modellbasierte Extrapolation. Diese Art von Algorithmen benötigen eine anwendungsabhängige große Anzahl an vorhandenen ZREs, um auf einen genauen Extrapolationswert einschwingen zu können. Darüber hinaus wird einen Sollwert parallel zum Istwert benötigt, um diese Operationen durchzuführen. Daher empfiehlt sich, diese Operationen im Rahmen eines Steuerungs-Teilmodells zu definieren. Eine weitere Operation liefert die kumulierte Abweichung von einer Zeitreihe zum Extrapolation mit einen gegebenen Polynomgrad ab einen gegebenen ZRE bis zum neuesten ZRE. Als Input benötigt diese Operation den Grad des Polynoms zum Extrapolieren und die Referenz auf den Anschluss-ZRE. Als Output werden die kumulative Abweichung und einen Statuswert dienen. Im Allgemeinen kann die Operation erfolgreich die Abweichung berechnen, wenn zumindest $N+2$ ZREs bis zum neuesten ZRE vorhanden sind. Diese Operation kann genutzt werden, um die Eignung der verschiedene Polynomgrade zu messen. Die geringste Abweichung sollte den Haupttrend der Zeitreihe bezeichnen. Genauso wie bei der Extrapolationsoperation, liefert der Statuswert Auskunft über den Erfolg oder Misserfolg der Operation.

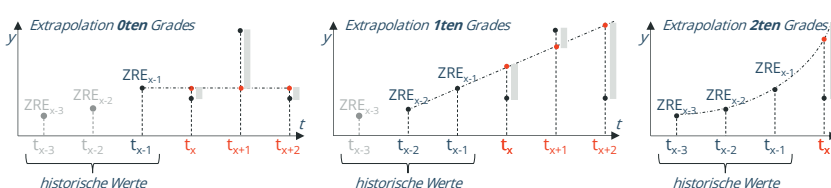


Bild 4: kumulative Abweichung der verschiedenen Grade

2.4.3. Kennzeichnung historischer bzw. extrapolierter Werte

Wie bereits beschrieben, besteht die Notwendigkeit extrapolierte Werte von gemessenen Werten abzugrenzen. In den Eigenschaften der Merkmale sind verschiedene Felder vorgesehen, um mehrere gleichrangige Alternativen oder Detailinformationen abzubilden [7]. Hier lassen

sich Aktualwert, extrapolierter Wert und Extrapolationsvorschrift unterbringen. Ebenso sind hier Referenzen auf eine Extrapolations-Capability der (eigenen oder entfernten) Verwaltungsschale möglich. Der Qualifier „extrapoliert“ deutet darauf hin, dass der Wert über eine Extrapolationsoperation berechnet wurde, d. h. er kann unter Umständen von dem tatsächlichen Wert abweichen. Als Gegenstück dazu dient der Qualifier „gemessen“, der nicht nur tatsächlich gemessene Werte gekennzeichnet, sondern auch auf die Möglichkeit hinweist, extrahierte Werte zu ermitteln. Auf der Seite der historischen Information kann ein Qualifier „hat Historie“ auf ein Merkmal angewendet, mitteilen, dass eine Historie zu diesem Merkmal vorhanden ist. Daraus kann eine unterschiedliche Darstellung folgen, auch die Möglichkeit Extrapolationsoperationen für diese Property anzubieten.

3. Implementierung

Die Implementierung des Verwaltungsdienstes basiert auf dem Eclipse Ditto Projekt [17], welches das Digital Twin Softwarepattern umsetzt. Der Quellcode des Projektes ist öffentlich zugänglich, und die Verwendung ist frei gestattet. Ditto verfolgt einen agnostischen Ansatz, weshalb so wenig Annahmen wie nötig über ein Digital Twin getroffen werden. Ein Digital Twin besteht aus einer ID und mehreren Attributen und Features. Zusätzlich kann ein Digital Twin optional eine PolicyID und eine Definition haben. Der Verwaltungsdienst wurde prototypisch als ED-Microservice-Cluster umgesetzt und durch Hinzufügen weiterer Services zur Orchestrierungskonfigurationsdatei leicht erweiterbar. Die enthaltenen Microservices, ein Microservice für die Extrapolation und ein Microservice für die IIoT-Geräteverwaltung, sind dabei lose über WebSockets gekoppelt. Der Microservice für die Extrapolation sammelt zusätzlich noch Netzwerkänderungen und validiert Managementanweisungen. Der zweite Microservice kümmert sich um An- und Abmeldungen der Assets, Protokollübersetzung, Daten-Plausibilitätsprüfung und der Weiterleitung von Managementanweisungen. Für die Beschreibung der Verwaltungsschale wurde Eclipse Vorto [22] als Domain Specific Language verwendet. Dafür wurde Vorto mittels Konvention erweitert, sodass extrapolierfähige Features gekennzeichnet werden können sowie Extrapolationsfunktionen angegeben werden können. Der Verwaltungsdienst stellt somit für Anwendungen extrahierte Werte zur Verfügung, die mittels der vom Merkmal vorgegebenen Extrapolationsfunktion und einer Historie von im Verwaltungsdienst gespeicherter Werte abgeschätzt werden. Die historischen Werte werden in der prototypischen Realisierung dateibasiert abgelegt. Nicht-operationskritischen Assetnachrichten werden über eine UDP Schnittstelle ausgetauscht. Operationskritische Nachrichten, Alarmsignale und Steueranweisungen werden über TCP ausgetauscht. Bild 5 zeigt die Grundstruktur des Verwaltungsdienstes.

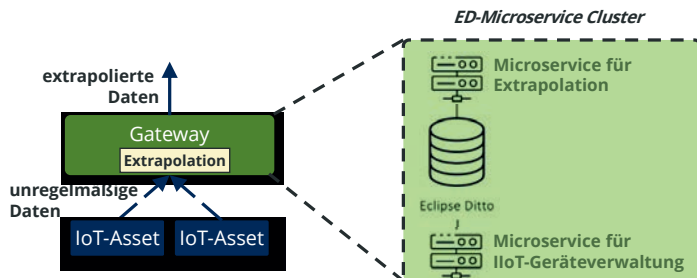


Bild 5: Struktur des IIoT-Verwaltungsdiensstes

4. Zusammenfassung

In diesem Beitrag wurde ein Konzept für eine resiliente IIoT-Geräteverwaltung in der Industrie 4.0-Verwaltungsschale vorgestellt. Resilient bedeutet in diesem Kontext die Möglichkeit vorzusehen, dass Assets eventuell auch nicht erreichbar sind und daher Strategien bereitzustellen, um Eigenschaftswerte zu extrapolieren, um nicht auf Asset-Erreichbarkeit angewiesen zu sein. Aktive Verwaltungsschalen unterstützen die Ausführung von Algorithmen. Allerdings müssen die geeigneten Algorithmen gefunden und innerhalb der Verwaltungsschale vereinbart werden. Für diese Vereinbarung wurde in diesem Beitrag eine allgemeine Nomenklatur für unterschiedliche Extrapolations-Algorithmen und ihre beispielhafte Ausprägung eingeführt. Anhand eines Prototyps wurde grob die Implementierung skizziert.

5. Ausblick

Als erweiterte Verwendung dieses Ansatzes ist es denkbar, geänderte Anforderungen an das Asset zu propagieren, um eine automatische Verteilung von Funktionen zu ermöglichen. Dies ist ein weiterer Schritt hin zur flexiblen Produktionsanlage der Industrie 4.0 unter Einbeziehung von IIoT-Assets. Neben der Extrapolation sind auch weitere algorithmische Vorverarbeitung der Assetinformationen denkbar. So ließe sich beispielsweise auch eine Normalisierung von Werten oder eine Plausibilitätsprüfung zur Eliminierung von Messfehlern oder Messungenauigkeiten im selben Rahmen durchführen. Dazu werden Sensorwerte auch bei erreichbaren Assets parallel extrapoliert und mit den zum gleichen Zeitpunkt gehörenden Messwert verglichen. Bei Überschreitung einer Abweichung ist dann der extrapolierte Wert wahrscheinlicher. Somit wirken sich einzelne Messfehler nicht zu stark aus.

Literaturverzeichnis

- [1] Hilbert, F. et al, 2018, Ein servicebasierter Ansatz für das Management von eingeschränkt erreichbaren IIoT-Geräten, Automation 2019, Baden Baden
- [2] Adolphs, P. et al (2016). Struktur der Verwaltungsschale: Fortentwicklung des Referenzmodells für die Industrie 4.0-Komponente. von BMWi Öffentlichkeitsarbeit. Berlin, Apr, 23-24.
- [3] IEC White Paper: Semantic Interoperability: Challenges in The Digital Transformation Age. Geneva (2019), <https://bit.ly/3dJoN79>. (Abruf am 3.5.2021).
- [4] Henning, K., Wolfgang, W., & Johannes, H. (2013). Recommendations for implementing the strategic initiative INDUSTRIE 4.0. *Final report of the Industrie, 4*, 82.
- [5] BMWi: Plattform Industrie 4.0: Verwaltungsschale in der konkreten Praxis - Wie definiere ich Teilmodelle, bei spielhafte Teilmodelle und Interaktion zwischen Verwaltungsschalen. Berlin 2019.
- [6] Plattform Industrie 4.0, ZVEI, Details of the Asset Administration Shell - Part 1 - The exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0)
- [7] Belyaev, A., & Diedrich, C. (2019). Aktive Verwaltungsschale von I4.0-Komponenten, Erscheinungsformen von Verwaltungsschalen. *Leitkongress der Mess-und Automatisierungstechnik*. Magdeburg, Conference Paper, July 2019
- [8] Diedrich, C. et al (2019) Sprache für I4. 0-Komponenten-Semantik der Interaktionen von I4. 0-Komponenten.
- [9] Çorak, B. H., Okay, F. Y., Güzel, M., Murt, Ş., & Ozdemir, S. (2018, June). Comparative analysis of IoT communication protocols. In 2018 International symposium on networks, computers and communications (ISNCC) (pp. 1-6). IEEE.
- [10] Kayal, P., & Perros, H. (2017, March). A comparison of IoT application layer protocols through a smart parking implementation. In 2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN) (pp. 331-336). IEEE.
- [11] Cuenca, Á., Salt, J., Albertos, P., & Casanova, V. (2006, November). Algebraic design of multi-rate control systems for environments with limited random delays. In IECON 2006-32nd Annual Conference on IEEE Industrial Electronics (pp. 448-453). IEEE.
- [12] Bormann, C., Ersue, M., & Keranen, A. (2014). Terminology for constrained-node networks. Internet Engineering Task Force (IETF): Fremont, CA, USA, 2070-1721.

- [13] Van Kranenburg, R., & Bassi, A. (2012). IoT challenges. Communications in Mobile Computing, 1(1), 1-5.
- [14] Boss, B. et al (2020). Digital twin and asset administration shell concepts and application in the industrial internet and industrie 4.0. Plattform Industrie, 4., <https://www.iiconsortium.org/pdf/Digital-Twin-and-Asset-Administration-Shell-Concepts-and-Application-Joint-Whitepaper.pdf> (Abruf am 3.5.2021).
- [15] Bedenbender, H., Bock, J., & Boss, B. (2019). Verwaltungsschale in der Praxis. Bundesministerium für Wirtschaft und Energie.
- [16] Drummond, N. et al (2006), Sequences in Protégé OWL, in ' 9th International Protege Conference'.
- [17] o.a. Eclipse Ditto. <https://www.eclipse.org/ditto/1.5/>, 2020. (Abruf am 3.5.2021).
- [18] o.a. Vorto. <https://github.com/eclipse/vorto/blob/master/docs/vortolang-1.0.md>, 2020. (Abruf am 3.5.2021).

Designprinzipien für den Modul- und Serviceentwurf in modularen Logistikanlagen

Design Principles for the module and service design in modular logistics facilities

M. Blumenstein, Prof. Dr.-Ing. **A. Fay**,
Helmut-Schmidt-Universität, Hamburg;

K. Gryczycha, Prof. Dr.-Ing. **S. Lier**,
Fachhochschule Südwestfalen, Meschede;

C. Kotsch, BASF SE, Ludwigshafen;

F. Borgmann, Fraunhofer-Institut für Materialfluss u. Logistik, Dortmund;

N. Austermann, BEUMER Maschinenfabrik GmbH & Co. KG, Beckum

Kurzfassung

Das Potenzial modularer Produktionsanlagen kann nur dann vollständig ausgeschöpft werden, wenn auch die zugehörige produktionsnahe Logistik modular gestaltet ist. In diesem Beitrag werden Designprinzipien für die Modularisierung von Logistikanlagen aufgezeigt. Zudem wird ein serviceorientierter Automatisierungsansatz für Logistikmodule vorgestellt, welcher zwischen zwei Varianten der Serviceausführung unterscheidet – einer schnellen, zyklischen Ausführung sowie einer flexibel parametrierbaren Ausführung.

Abstract

Modular production facilities only raise their potential if the associated production-related logistic system is also designed in a modular way. This paper presents design principles for the modularization of logistic systems. In addition, we introduce a service-oriented automation approach for logistic modules, which distinguishes between two variants of service execution - a fast cyclic execution and a flexibly parameterizable execution.

1 Motivation und Einleitung

Die Prozessindustrie sieht sich Trends hin zu einer starken Volatilität der Märkte sowie einer hohen Produktindividualität gegenübergestellt. Diese Trends verlangen Produktionsanlagen sowohl hinsichtlich ihrer hardwareseitigen als auch softwareseitigen Gestaltung ein hohes Maß an Flexibilität und Wandelbarkeit ab. Ein wichtiger Wandlungsbefähiger ist dabei die Modularisierung der Anlagen [1]. In diesem Zusammenhang wurden in den vergangenen Jahren im Rahmen verschiedener Forschungsprojekte [2, 3] und Standardisierungsanstrengungen

[4,5] Konzepte zur Modularisierung von verfahrenstechnischen Produktionsanlagen entwickelt. Die ver- und entsorgende produktionsnahe Logistik dieser Anlagen ist allerdings nach wie vor monolithisch und unflexibel gestaltet. Dies resultiert u.a. in langen Planungszeiten und einer schlechten Wiederverwendbarkeit von Logistikmaschinen. Um das Potenzial der modularen Produktion vollständig auszunutzen, besteht der Wunsch auch die produktionsnahe Logistik modular und damit flexibel und wandelbar zu gestalten [1].

Vor diesem Hintergrund wurde im NAMUR-Arbeitskreis 4.19 „Produktionsnahe Logistik“ die NAMUR-Empfehlung NE 171 [6] entwickelt, welche die Anforderungen an die Automatisierung modularer produktionsnaher Logistikanlagen aufzeigt. Aufbauend darauf stellt [7] einen möglichen Automatisierungsansatz anhand des in [6] eingeführten Praxisbeispiels vor. Es zeigt sich, dass das aus der modularen Produktion bekannte Module Type Package (MTP)-Konzept [4] grundsätzlich auch auf den Bereich der produktionsnahen Logistik anwendbar ist.

Auf Basis dieser Arbeiten wurde das Projekt „Modulare Produktionslogistik (MoProLog)“ [8] initiiert. Dieses hat zum Ziel, Anlagenkonzepte für modulare Logistikanlagen zu entwickeln und diese anhand einer modularen Abfüll- und Verpackungsanlage zu validieren. Hierbei werden neben Material- und Energieschnittstellen für Logistikmodule insb. auch deren informationstechnische Schnittstellen und deren Integration in ein modulares Gesamtsystem betrachtet [9]. Als erste Ergebnisse dieses Projekts zeigt der vorliegende Beitrag Design-Prinzipien für den Modul- und Serviceentwurf modularer Logistikanlagen auf. Diese sollen als Rahmenwerk für die Gestaltung künftiger Logistikmodule dienen. Kapitel 2 grenzt dazu den Anwendungsbeereich der produktionsnahen Logistik und des MoProLog-Projekts ab. In Kapitel 3 wird ein Strukturierungsansatz für modulare Logistikanlagen vorgestellt. Kapitel 4 identifiziert aufbauend auf einer Analyse von Verpackungsprozessen Einflussfaktoren auf die Modularisierung dieser Prozesse. Diese Faktoren resultieren in Designprinzipien für Logistikmodule, die in Kapitel 5 vorgestellt werden. Anschließend zeigt Kapitel 6 einen serviceorientierten Automatisierungsansatz für Logistikmodule auf. Den Abschluss des Beitrags bilden eine Zusammenfassung der vorgestellten Ergebnisse und ein Ausblick auf die nächsten Schritte in Kapitel 7.

2 Abgrenzung der produktionsnahen Logistik

Gemäß [10] bilden der Einkauf, die Produktion und der Vertrieb die Hauptprozesse entlang der Supply Chain eines Unternehmens. Mit jedem dieser Bereiche sind logistische Aufgaben verbunden. Entsprechend lassen sich die Logistikprozesse eines Unternehmens in die Beschaffungs-, Produktions- und Distributionslogistik unterteilen. Bei der produktionsnahen Logistik handelt es sich um den Teil der Produktionslogistik, der die unmittelbare Ver- und Entsorgung von Produktionsprozessen sicherstellt [11]. Im Rahmen des MoProLog-Projekts und

somit auch im Rahmen dieses Beitrags werden aus dem Bereich der produktionsnahen Logistik im Speziellen die Verpackungsprozesse in der Prozessindustrie betrachtet. Ziel dieser Prozesse ist es, Produktströme (Gase, Schütt- oder Flüssiggüter) in individuelle, eindeutig identifizierbare Einheiten zu verpacken und auf diese Weise zu diskretisieren. Diese Einheiten werden in [6] als Logistische Leistungsobjekte (LOe) bezeichnet. LOe sind dabei „Zwischen- und Endprodukte, die über den Einsatz von Verpackungen, Gebinden sowie Transportbehältern in diskretisierte sowie identifizierbare Einheiten gegliedert sind“ [6]. Dieser Beitrag betrachtet die Modularisierung derartiger Verpackungsanlagen zunächst aus struktureller dann aus automatisierungstechnischer Perspektive.

3 Strukturebenen modularer Logistikanlagen

Der Begriff Modularisierung beschreibt im Allgemeinen die Unterteilung eines Systems in einzelne voneinander abgegrenzte Einheiten – die Module. Diese Module können erneut modularisiert und somit in Sub-Module zerlegt werden. Diese Sub-Module können wiederum mehrfach modularisiert werden. Durch die wiederholte Anwendung des Modularisierungsgedankens wird ein System in immer kleinere hierarchisch angeordnete Einheiten zerlegt, es entstehen Strukturebenen. Als erster Schritt zur Modularisierung produktionsnaher Logistikanlagen werden im Folgenden mögliche Strukturebenen für diese Anlage identifiziert und definiert.

Als Grundlage hierfür werden zunächst bestehende Ansätze zur Strukturierung modularer Anlagen betrachtet. In [12] werden fünf räumlich abgegrenzte Strukturebenen für den Bereich der diskreten Fertigung definiert. Das vorgestellte Modell umfasst neben den Betriebsmitteln auch die Unternehmensorganisation und die Raum- und Gebäudestruktur. Im Umfeld modularer Prozessanlagen beschreibt die VDI 2776 [5] vier Strukturebenen, die im Einklang mit dem MTP-Konzept stehen. PackML [13] stellt anhand einer Form Fill Seal (FFS) Maschine ein beispielhaftes Hierarchiemodell für den Bereich der produktionsnahen Logistik vor. Dabei werden vier Strukturebenen vorgesehen, jedoch keine konkreten Definitionen für diese bereitgestellt. In der NE 171 [6] wird lediglich die Ebene der Logistikanlage konkret definiert, die sich aus Modulen und Sub-Modulen zusammensetzen kann. Diese Module werden allerdings nicht genauer spezifiziert. Es zeigt sich, dass noch kein vollständiges Modell für die Strukturierung modularer Logistikanlagen besteht. Aufgrund dessen werden im Folgenden basierend auf den vorgestellten Ansätzen drei Strukturebenen für modulare Logistikanlagen definiert.

Die Ebene mit dem höchsten Aggregationsniveau ist das **Modulare Logistische System (MLS)**. Dabei handelt es sich um ein technisches Gesamtsystem, das logistische Funktionen entlang einer ununterbrochenen Prozesskette an einem Standort verrichtet, um eine umfassende logistische Aufgabe zu erfüllen. Ein Beispiel ist eine Anlage zur Sackabfüllung.

Ein MLS besteht aus mindestens einer **Modularen Logistischen Einheit (MLE)**. Eine MLE ist ein im Anlagenkontext austauschbarer Bestandteil eines MLS. Sie beinhaltet sämtliche notwendige Hardware und Software, um eine logistische Funktion weitestgehend unabhängig von anderen MLEen auszuführen. Eine MLE besitzt eine eigene Steuerung und kann mit einem übergeordnetem Orchestrierungssystem sowie mit anderen MLEen kommunizieren. Ein Beispiel für eine MLE ist eine Form Fill Seal (FFS) Maschine zur Abfüllung von Säcken.

Eine MLE kann **Modulare Logistische Komponenten (MLKen)** enthalten. Eine MLK ist dabei ein im Anlagenkontext austauschbarer Bestandteil einer MLE. Sie ist immer genau einer MLE zugeordnet und ergänzt diese hardware- und ggfs. softwareseitig. Eine MLK kann eine eigene Steuerung besitzen und kann ausschließlich mit der ihr übergeordneten MLE kommunizieren. Die Anbindung an ein Orchestrierungssystem sowie an andere MLEen erfolgt über die übergeordneten MLE. Ein Beispiel ist die Einheit zur Materialeinfüllung einer FFS-Maschine.

Mit den beschriebenen Strukturebenen wird ein erster Ansatz zur Strukturierung modularer produktionsnaher Logistikanlagen zur Diskussion gestellt. In künftigen Forschungsarbeiten und praktischen Anwendungen gilt es, diesen Ansatz zu verfeinern und zu validieren. Der Fokus dieses Beitrags liegt auf der automatisierungstechnischen Gestaltung von MLEen.

4 Analyse von Verpackungsprozessen

Um eine Modularisierung gemäß der definierten Strukturebenen vornehmen zu können, wurde im Rahmen des MoProLog-Projekts eine Analyse von Verpackungsprozessen durchgeführt.

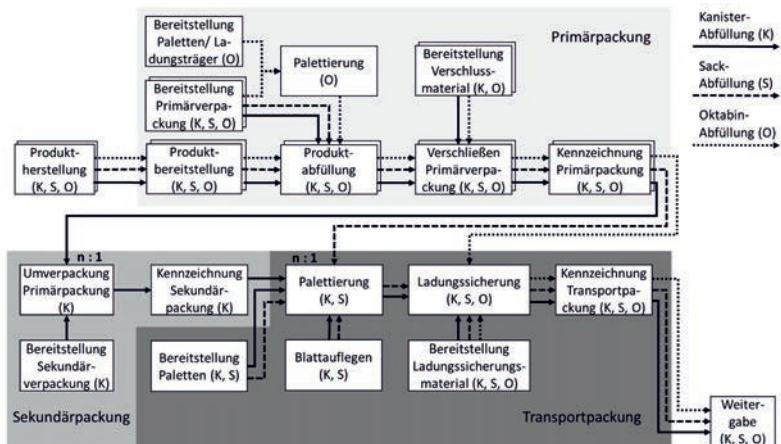


Bild 1: Analyse von Verpackungsprozessen anhand eines generischen Prozessbildes (in Anlehnung an [9])

Diese Analyse erfolgte zum einen auf Basis heutiger Logistikmaschinen sowie realer Anlagenstrukturen. Zum anderen wurden typische Prozessschritte von Verpackungslinien identifiziert und in einem generischen Prozessbild zusammengefasst. Diese Schritte können danach klassifiziert werden, ob sie an der Erstellung von Primär-, Sekundär- oder Transportpackungen (gemäß [14]) beteiligt sind. Bild 1 stellt das generische Prozessbild dar und zeigt das Ergebnis einer Analyse dreier Verpackungsprozesse für Kanister, Säcke und Oktabins auf. Es wird deutlich, dass für die drei Prozesse teilweise verschiedene Prozessschritte notwendig sind. Diese sind mit „K“ für die Kanisterabfüllung, „S“ für die Sackabfüllung und „O“ für die Oktabinabfüllung gekennzeichnet. Aus diesen Analysen lassen sich Erkenntnisse ableiten, die Einfluss auf die modulare Gestaltung von Logistikanlagen haben. Diese werden im Folgenden anhand der drei gezeigten Verpackungslinien für Kanister, Säcke und Oktabins betrachtet.

a) Linienpezifische Prozessschritte: Es existieren Prozessschritte, die spezifische Funktionalitäten für eine bestimmte Verpackungslinie bereitstellen. Dies kann zum einen darin begründet liegen, dass ein Prozessschritt nur in einer Linie benötigt wird. So ist z.B. die Umverpackung Primärpackung im dargestellten Fall nur im Rahmen der Kanisterabfüllung notwendig. Zum anderen können verschiedene Ausführungsvarianten eines Prozessschrittes für unterschiedliche Linien notwendig sein. So erfolgt im dargestellten Fall z.B. die Produktabfüllung für alle drei Linien auf unterschiedliche Art und Weise (Bild 1, doppelte Kästchen).

b) Linienübergreifende Prozessschritte: Es existieren Prozessschritte, die von mehreren Linien verwendet werden können (z.B. Ladungssicherung). Dies bedingt allerdings, dass der grundlegende Funktionsablauf dieser Schritte in verschiedenen Linien identisch sein muss. Eine unterschiedliche Parametrierung der Funktionalität zur Anpassung auf verschiedene Linien ist dabei möglich. In bestehenden Logistikanlagen bringen z.B. Stretchhaubenmaschinen, die zur Ladungssicherung eingesetzt werden, bereits heute eine entsprechende Flexibilität mit, um in verschiedenen Linien eingesetzt zu werden.

c) Logistikmaschinen mit einer spezifischen Aufgabe: Bei der Analyse heutiger realer Logistikanlagen zeigte sich, dass Logistikmaschinen existieren, die innerhalb einer Linie nur für eine sehr spezifische Aufgabe einsetzbar sind. Das bedeutet, sie können nur einen bestimmten Prozessschritt oder mehrere im generischen Prozessbild direkt aufeinanderfolgende Prozessschritte umsetzen. Ein Beispiel ist eine Form Fill Seal (FFS)-Maschine, die die Prozessschritte der Sackbereitstellung, -abfüllung und -verschließung umsetzt.

d) Logistikmaschinen mit mehreren Aufgaben: Es existieren Logistikmaschinen, die innerhalb einer Linie mehrere Aufgaben ausführen können, d.h. mehrere im generischen Prozessbild nicht direkt aufeinanderfolgende Prozessschritte umsetzen. So ist z.B. eine Stretchhaubenmaschine im Kontext einer Oktabinabfüllung in der Lage, den Schritt der Ladungssiche-

nung zu übernehmen. Zudem kann sie dazu genutzt werden, um Folien in die Oktabsins einzulegen, und ist somit am Prozessschritt der Bereitstellung der Primärverpackung beteiligt.

e) Verhältnis der aufgenommenen zu den abgegebenen LOe: Bei jedem Prozessschritt wird eine bestimmte Anzahl an LOe aufgenommen, bearbeitet und eine bestimmte Anzahl an LOe wieder ausgegeben. Das Verhältnis zwischen aufgenommenen und abgegebenen LOe ist dabei in den meisten Fällen 1:1. An den Übergängen zwischen Primär-, Sekundär- und Transportpackung tritt hingegen meist ein $n:1$ -Verhältnis auf (siehe Bild 1, „ $n:1$ “), da dort mehrere LOe zu größeren Gebinden zusammengepackt werden können. So kann ein Palettieren-Prozessschritt z.B. 50 Säcke aufnehmen und in Form einer Palette wieder ausgeben (Verhältnis 50:1). Prozessschritte mit einem solchen $n:1$ -Verhältnis können in der Regel nur für eine Aufgabe in einer Linie eingesetzt werden.

f) Unterschiedliche Taktfrequenzen: Heutige Logistikmaschinen für unterschiedliche Prozessschritte können z.T. unterschiedliche Taktfrequenzen realisieren. So kann beispielsweise die FFS-Maschine BEUMER fillpac® FFS [15] bis zu 2600 Säcke pro Stunde abfüllen. Der Lagenpalettierer BEUMER paletpac® [16] palettiert bis zu 5000 Säcke pro Stunde, was bei z.B. 50 Säcken pro Palette 100 Paletten pro Stunde entspricht. Die Stretchhaubenmaschine BEUMER stretch hood® [17] kann eine Ladungssicherung an bis zu 120 Paletten pro Stunde anbringen. Wird aus diesen Maschinen eine Sackabfülllinie aufgebaut, so können der Palettierer sowie die Stretchhaubenmaschine nicht ausgelastet werden. Um dem entgegenzuwirken, bestehen zwei Ansätze. Zum einen können im Sinne eines Numbering-Up zwei FFS-Maschinen in die Anlage eingebunden werden. Somit wird der Palettierer mit 5200 Säcke pro Stunde beliefert und ist ausgelastet. Um die Stretchhaubenmaschine auszulasten, besteht zum anderen die Möglichkeit, diese Maschine in mehreren Linien und/oder für mehrere Aufgaben innerhalb der gleichen Linie einzusetzen. Um dies zu realisieren, muss die Stretchhaubenmaschine allerdings derart flexibel gestaltet sein, dass sie mit unterschiedlichen ankommenden LOe umgehen kann.

g) Transporte zwischen den Prozessschritten: Die in Bild 1 dargestellten Pfeile zeigen zunächst zeitliche Relationen zwischen Prozessschritten auf, sie können jedoch auch Transporte von LOe repräsentieren. Dabei ist zu beachten, dass Transporte nicht immer möglich und sinnvoll sind. Bei einer Sackabfüllung sind z.B. die Schritte des Sack-Bereitstellens, -Abfüllens und -Verschließens sehr eng verbunden und gehen ineinander über, sodass keine Transporte zwischen ihnen möglich sind. Diese Schritte müssen folglich in einer Logistikmaschine (FFS-Maschine) als integrierte Funktion realisiert werden, eine Verteilung auf mehrere Maschinen ist nicht möglich. Je mehr Prozessschritte in der Art in einer Maschine gekapselt werden, umso spezifischer und unflexibler einsetzbar wird deren Funktion.

Die aufgezeigten Erkenntnissen a) bis g) dienen als Grundlage für die Modularisierung von Logistikanlagen. Darauf basierend werden im Folgenden Design-Prinzipien für den Entwurf von MLEen (siehe Kapitel 5) sowie deren Services (siehe Kapitel 6) abgeleitet.

5 Entwurf Modularer Logistischer Einheiten (MLEen)

Ziel des MLE-Designs ist es, die zuvor untersuchten Prozessschritte durch verschiedene, konstruktiv voneinander getrennte Einheiten umzusetzen. Dabei muss zunächst definiert werden, wie viele und welche Prozessschritte in der gleichen MLE gekapselt werden sollen. Die bisherigen Ergebnisse im MoProLog-Projekt haben gezeigt, dass die heutige Unterteilung von Verpackungslinien in verschiedene Logistikmaschinen eine gute Grundlage für den MLE-Schnitt bildet. Aus den in Kapitel 4 gewonnenen Erkenntnissen lassen sich zudem Charakteristika für MLEen im Anlagenkontext ableiten. Diese werden im Folgenden betrachtet.

5.1 MLE-Charakteristika

Singleline-MLEen vs. Multiline-MLEen

In Kapitel 4 wird aufgezeigt, dass Prozessschritte existieren, die nur in einer Verpackungslinie relevant sind (Punkt a). MLEen, die diese Prozessschritte realisieren, sind folglich genau einer Linie zugeordnet. Dies ist u.a. dann sinnvoll, wenn die von der MLE aufgenommenen zu den abgegebenen LOe ein Verhältnis von n:1 aufweisen (Punkt e). Weiterhin ist die Verwendung einer MLE für nur eine Linie dann sinnvoll, wenn diese MLE im Anlagenkontext eine starke Auslastung erfährt (Punkt f). Sind zwischen Prozessschritten keine sinnvollen Transporte umsetzbar, resultiert das darin, dass mehrere Prozessschritte durch eine MLE umgesetzt werden. Die Funktionalität dieser MLE ist dann sehr spezifisch und kann zumeist nur in einer Linie eingesetzt werden (Punkt g). MLEen, die diese Charakteristik erfüllen, werden im Folgenden als **Singleline-MLEen** bezeichnet.

Aus Kapitel 4 geht weiterhin hervor, dass Prozessschritte existieren, die in mehreren Linien verwendet werden können (Punkt b). MLEen, die diese Prozessschritte realisieren, können somit von verschiedenen Linien genutzt werden. Sie müssen eine gewisse Flexibilität in ihrer Funktionalität mitbringen, um den Anforderungen unterschiedlicher Linien zu genügen. Eine Mehrfachverwendung von MLEen in verschiedenen Linien ist insbesondere dann sinnvoll, wenn eine MLE durch ihre Aufgaben in einer Linie nicht ausgelastet ist (Punkt f). MLEen, die diese Charakteristik erfüllen, werden im Folgenden als **Multiline-MLEen** bezeichnet.

Singlepurpose-MLEen vs. Multipurpose-MLEen

In Kapitel 4 wird beschrieben, dass MLEen existieren, die eine spezifische Aufgabe innerhalb einer Verpackungslinie übernehmen (Punkt c). U.a. für MLEen, die ein n:1-Verhältnis von aufgenommenen zu abgegebenen LOe aufweisen, ist dies (neben ihrer Singleline-Charakteristik)

sinnvoll (Punkt e). Weiterhin sollte eine MLE für nur eine Aufgabe verwendet werden, wenn sie im Anlagenkontext eine starke Auslastung erfährt (Punkt f). In diesem Fall ist es sogar möglich, dass im Sinne eines Numbering-Up mehrere MLEen für eine einzige Aufgabe innerhalb der Linie eingesetzt werden. Sind zwischen Prozessschritten keine sinnvollen Transporte umsetzbar, resultiert das darin, dass mehrere Prozessschritte durch eine MLE realisiert werden. Diese MLE führt dann eine sehr spezifische Aufgabe aus (Punkt g). MLEen, die diese Charakteristik erfüllen, werden im Folgenden als **Singlepurpose-MLEen** bezeichnet.

Kapitel 4 zeigt auf, dass MLEen existieren, die mehrere unterschiedliche Aufgaben innerhalb einer Verpackungslinie übernehmen können (Punkt d). Sie müssen dazu eine gewisse Flexibilität in ihrer Funktionalität mitbringen, um den Anforderungen der Aufgaben zu genügen, für die sie eingesetzt werden sollen. Eine Mehrfachverwendung von MLEen für verschiedene Aufgaben ist insb. dann sinnvoll, wenn eine MLE durch eine Aufgabe nicht ausgelastet ist (Punkt f). MLEen, die diese Charakteristik erfüllen, werden als **Multipurpose-MLEen** bezeichnet.

5.2 MLE-Arten

Durch die Kombination der zuvor beschriebenen Charakteristika ergeben sich vier mögliche MLE-Arten, die in Tabelle 1 anhand je eines praktischen Beispiels aufgezeigt werden.

Tabelle 1: Beispiele für die identifizierten Arten von Modulare Logistischen Einheiten

	Singleline	Multiline
Singlepurpose	Eine Form Fill Seal (FFS)-MLE kann die Bereitstellung, Abfüllung und das Verschließen von Säcken übernehmen. Diese sehr spezifische Funktionalität kann nur am Anfang einer Sackabfülllinie eingesetzt werden. Ein Einsatz in verschiedenen Linien oder für verschiedene Aufgaben innerhalb derselben Linie ist nicht möglich.	Der Prozessschritt zur Bereitstellung von Paletten wird gemäß Bild 1 für Oktabins (Primärpackung), für Säcke und für Kanister (Transportpackung) benötigt. Eine MLE, die diesen Prozessschritt bereitstellt, kann somit in allen drei betrachteten Linien eingesetzt werden. Innerhalb einer Linie übernimmt sie jedoch jeweils nur eine Aufgabe (das Bereitstellen von Paletten).
Multipurpose	In einer Linie zur Oktabinabfüllung wird eine Stretchhauben-MLE zur Ladungssicherung verwendet. Zudem übernimmt sie das Einlegen einer Folie in die Oktabins und ist somit am Prozessschritt der Bereitstellung Primärverpackung beteiligt. Diese MLE übernimmt somit zwei Aufgaben innerhalb einer Linie. In diesem Beispiel existiert nur eine Oktabinabfüllung. Somit wird die Stretchhauben-MLE nur in dieser einen Linie eingesetzt.	Wie im Multipurpose-Singleline-Beispiel übernimmt eine Stretchhauben-MLE das Ladungssichern und das Folie-Einlegen in einer Linie zur Oktabinabfüllung. Zudem existiert eine Sackabfülllinie. Die Stretchhaubenmaschine ist derart flexibel gestaltet, dass sie den Prozessschritt der Ladungssicherung zusätzlich auch für die Sackabfülllinie übernehmen kann. Sie wird somit in zwei verschiedene Linien verwendet.

Auf Grundlage dieser Beispiele kann geschlossen werden, dass alle vier Kombinationen sinnvoll sind. Zudem zeigt sich am Beispiel der Stretchhauben-MLE, dass diese in verschiedenen Anlagenkontexten verschiedene MLE-Arten annehmen kann. Im Speziellen sollte sie die Voraussetzungen mitbringen, um als Singleline- oder als Multiline-MLE eingesetzt zu werden.

Aus den gewonnenen Erkenntnissen in Bezug auf die MLE-Arten wird im Folgenden ein Automatisierungsansatz für modulare produktionsnahe Logistikanlagen abgeleitet.

6 Entwurf von Services modularer Logistikanlagen

Gemäß NE 171 [6] soll das Abrufen und Überwachen von logistischen Operationen innerhalb eines MLS durch parametrierbare Services erfolgen. Daher wird im Folgenden ein service-orientiertes Automatisierungskonzept für MLEen basierend auf der VDI/VDE/NAMUR 2658-4 [18] entwickelt. Im ersten Schritt ist dabei festzulegen, wie feingranular MLE-Funktionalitäten in Services untergliedert werden sollen. Als Grundlage hierfür dienen Erkenntnisse zum Service-Schnitt aus dem Bereich der modularen Produktion (z.B. aus [19]). Zudem fand ein Austausch mit dem ENPRO 2.0-Projekt ORCA [2] statt. Im Ergebnis zeigt sich, dass es im

Bereich der Verpackungslinien in den meisten Fällen sinnvoll ist, einen (Haupt-) Service pro MLE nach außen bereitzustellen. Intern kann die MLE zudem weitere Sub-Services, z.B. zur Anbindung von MLKen, beinhalten.

Die in Kapitel 5.2 vorgestellten Arten von MLEen stellen unterschiedliche Anforderungen an die Arbeitsweise von Logistikserves sowie deren Interaktion mit einem übergeordneten Orchestrierungssystem. So ist eine Singleline-Singlepurpose-MLE dazu konzipiert, eine ganz spezifische Funktionalität in einer einzigen Linie umzusetzen. Die Herausforderung besteht darin, möglichst schnell viele LOe in gleicher Weise zu bearbeiten. Die anderen drei MLE-Arten müssen hingegen in mehreren Linien und/oder für verschiedene Aufgaben einsetzbar sein. Die Herausforderung besteht darin, verschiedene LOe flexibel nach ihren individuellen Auftragsdaten zu bearbeiten. Es resultiert daher die Notwendigkeit, zwei verschiedene Varianten der Serviceausführung für Logistikserves zu spezifizieren. Diese werden im Folgenden als Cyclic Execution Service (CES) und Single Execution Service (SES) bezeichnet.

6.1 Cyclic Execution Service (CES)

Ein Cyclic Execution Service (CES) ist für Singleline-Singlepurpose-MLEen vorgesehen. Er ist dafür konzipiert, im Execute-Zustand zyklisch eine spezifische Abfolge von Tätigkeiten an mehreren LOe des gleichen Auftrags in gleicher Weise durchzuführen. Währenddessen ändert sich die Parametrierung des Service nicht. Dies ermöglicht kürzeste Taktzeiten und einen verringerten Kommunikationsaufwand zu einem übergeordneten Orchestrierungssystem.

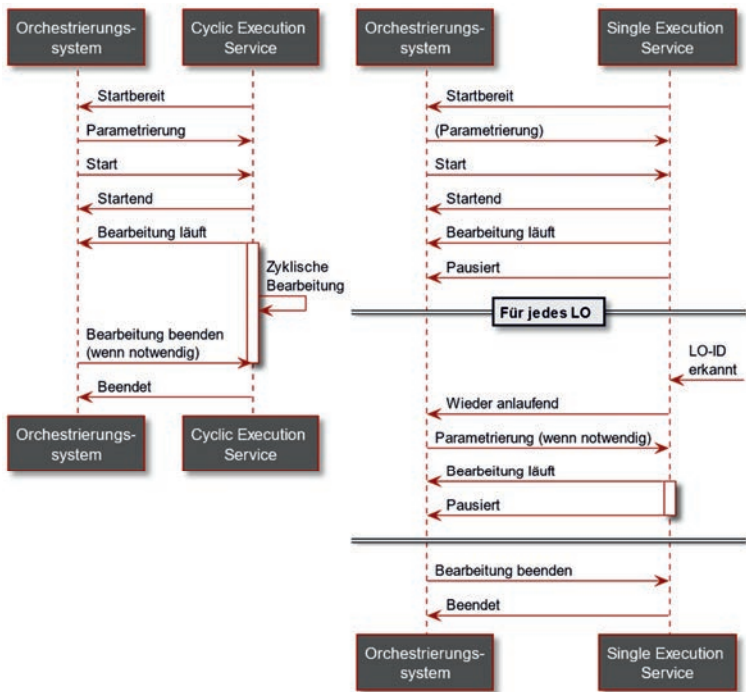


Bild 2: Interaktion eines Cyclic Execution Service (links) und eines Single Execution Service (rechts) mit einem übergeordneten Orchestrierungssystem

Die Interaktion eines CES mit einem solchen Orchestrierungssystem ist in Bild 2 (links) veranschaulicht. Nachdem der Service seine Startbereitschaft signalisiert hat, wird er vom Orchestrierungssystem parametrierung. Charakteristisch für einen CES ist dabei, dass die Serviceparametrierung einmal zu Beginn der Serviceausführung vorgenommen wird und sich bis zum Beenden des Service nicht mehr ändert. Nach der Parametrierung kann der CES gestartet werden und wechselt nach einer Startsequenz in den Bearbeitungszustand. Dort erfolgt eine zyklische Bearbeitung von LOe. Der Bearbeitungszustand kann verlassen werden, wenn der Service vom Orchestrierungssystem einen Befehl zum Beenden erhält oder eine zuvor definierte Anzahl an LOe bearbeitet wurde. In ersterem Fall handelt es sich um einen nicht-selbstbeendenden (nsb) Service, in letzterem Fall um einen selbstbeendenden Service (sb).

6.2 Single Execution Service (SES)

Ein Single Execution Service (SES) ist für alle MLEen mit einer Multipurpose- und/oder Multi-line-Charakteristik vorgesehen. Er ist dazu konzipiert, LOe auf verschiedene Art und Weise gemäß ihrer individuellen Auftragsdaten zu bearbeiten. Die Service-Parametrierung wird dabei LO-spezifisch angepasst. Somit bietet ein SES höchste Flexibilität bei der Bearbeitung von LOe verschiedener Aufträge bzw. verschiedener Bearbeitungszustände.

Bild 2 (rechts) zeigt vereinfacht die Interaktion eines SES mit einem übergeordneten Orchestrierungssystem. Nachdem der Service seine Startbereitschaft signalisiert hat, erfolgt seine Parametrierung. Da zu diesem Zeitpunkt noch nicht feststeht, welche LOe in welcher Reihenfolge beim SES eintreffen, können an dieser Stelle nur LO-unabhängige Parameter (z.B. den konstruktiven Aufbau betreffende Konfigurationsparameter) eingestellt werden. Zudem ist es möglich, LO-spezifische Parametersätze an den SES zu übergeben, die genutzt werden, wenn ein entsprechendes LO eintrifft. Anschließend wird der Service gestartet. Nach seiner Startsequenz wechselt er über den Bearbeitungszustand direkt in einen Pause-Zustand. Dieser signalisiert, dass der SES aktiv ist und auf eintreffende LOe wartet. Nach dem Eintreffen eines LO wird dieses identifiziert und der Service läuft wieder an. Dafür ist eine Serviceparametrierung entsprechend des eingetroffenen LO notwendig. Es wird geprüft, ob ein passender Parametersatz MLE-intern vorliegt. Ist dies gegeben, werden die entsprechenden Parameter angewendet. Ist kein passender Parametersatz verfügbar, kann der SES die erkannte ID des LO an das Orchestrierungssystem übermitteln und wird von diesem parametrierung. Anschließend geht der Service in seinen Bearbeitungszustand über und bearbeitet das eingetroffene LO. Nach Abschluss der Bearbeitung wechselt der SES zurück in den Pause-Zustand und wartet auf das nächste LO. Sollen keine weiteren LOe mehr durch den Service bearbeitet werden, kann dieser durch einen Beenden-Befehl aus dem Orchestrierungssystem beendet werden. SES sind immer nicht-selbstbeendend (nsb), da im Vorfeld nicht feststeht, wie viele LOe während der Serviceausführung bearbeitet werden sollen.

6.3 Realisierung des Servicekonzepts

In [6] und [7] wurde herausgestellt, dass das MTP-Konzept eine gute Grundlage zur Realisierung der Automatisierungssoftware von MLEen bildet. Im Zuge der Arbeiten im MoProLog-Projekt wurde ermittelt, dass das vorgestellte Konzept von CES und SES mit MTP-Mitteln nach VDI/VDE/NAMUR 2658-4 [18] umsetzbar ist. Die konkrete Realisierung des Servicekonzepts sowie notwendiger Schnittstellen für Logistikservices auf Basis des MTP-Ansatzes soll in einer weiterführenden Veröffentlichung aufgezeigt werden.

7 Zusammenfassung und Ausblick

In diesem Beitrag werden Erkenntnisse des MoProLog-Projekts in Bezug auf den Modul- und Service-Entwurf modularer produktionsnaher Logistikanlage präsentiert. Als Grundlage dazu werden drei Strukturebenen eingeführt – das Modulare Logistische System (MLS), die Modulare Logistische Einheit (MLE) sowie die Modulare Logistische Komponente (MLK). Basierend auf einer Analyse von Verpackungsprozessen sowie heutigen Logistikanlagen werden anschließend Design-Prinzipien für die automatisierungstechnische Gestaltung von MLEen dargestellt. Dabei zeigt sich aus funktionaler Perspektive, dass eine MLE die Charakteristika Singleline oder Multiline sowie Singlepurpose oder Multipurpose aufweisen kann. Durch Kombination dieser Charakteristika ergeben sich vier MLE-Arten. Je nach Art werden unterschiedliche Anforderungen an die automatisierungstechnische Gestaltung einer MLE gestellt. Um diese zu erfüllen, werden zwei Varianten der Serviceausführung voneinander abgegrenzt – der auf kurze Taktzeiten fokussierte Cyclic Execution Service (CES) und der auf Flexibilität fokussierte Single Execution Service (SES). Die vorgestellten MLE- sowie Service-Arten können Herstellern als Rahmenwerk für die Ausgestaltung künftiger modularer produktionsnaher Logistikanlagen dienen.

Im nächsten Schritt gilt es, die Ausgestaltung von CES und SES auf Basis der Konzepte der VDI/VDE/NAMUR 2658-4 [18] zu untersuchen sowie in einer beispielhaften Implementierung umzusetzen. Zudem sind entsprechende Orchestrierungsansätze für Modulare Logistische Systeme zu entwickeln.

Literatur

- [1] Lier, S., Wörsdörfer, D. u. Grünwald, M.: Wandlungsfähige Produktionskonzepte: Flexibel, Mobil, Dezentral, Modular, Beschleunigt. *Chemie Ingenieur Technik* 87 (2015) 9, S. 1147–1158
- [2] DECHEMA: ENPRO 2.0 - Effiziente Orchestrierung modularer Anlagen (ORCA). http://enpro-initiative.de/ENPRO+2_0/ORCA.html, abgerufen am: 15.11.2020
- [3] CORDIS: Final Report Summary - F³ FACTORY (Flexible, Fast and Future Production Processes), 2021. <https://cordis.europa.eu/project/id/228867/reporting/de>, abgerufen am: 26.03.2021
- [4] VDI/VDE/NAMUR 2658. *Automatisierungstechnisches Engineering modularer Anlagen in der Prozessindustrie*. https://www.vdi.de/suche?tx_solr%5Bq%5D=VDI%2FVDE%2FNAMUR+2658, abgerufen am: 11.11.2020
- [5] VDI 2776:2020-11. *Verfahrenstechnische Anlagen - Modulare Anlagen - Grundlagen und Planung modularer Anlagen*. <https://www.vdi.de/richtlinien/details/vdi-2776-blatt-1-verfahrenstechnische-anlagen-modulare-anlagen-grundlagen-und-planung-modularer-anlagen>, abgerufen am: 26.03.2021
- [6] NAMUR NE 171:2020-12. *Anwendung des modularen Anlagenkonzeptes in der produktionsnahen Logistik*
- [7] Cordes, S., Busert, T., Fay, A., Kessler, S. u. Schick, A.: NAMUR-MTP für Plug-&-Operate in produktionsnaher Logistik. Anforderungen an eine modulare Automatisierung. *atp magazin* (2020) 01-02, S. 86–93
- [8] DECHEMA: ENPRO 2.0 - Modulare Produktionslogistik (MoProLog), 2020. http://enpro-initiative.de/ENPRO+2_0/MoProLog.html, abgerufen am: 30.04.2021
- [9] Weigel, O., Lier, S., Judel, O., Gryczycha, K., Kocak, I. u. Pannok, M.: Modulare Produktionslogistik in der Prozessindustrie – Herausforderungen und Systematisierungen. In: 10. ProcessNet-Jahrestagung und 34. DECHEMA-Jahrestagung der Biotechnologen 2020
- [10] DIN u. VDI: Deutsche Normungsroadmap Logistik, 2015. <https://www.din.de/resource/blob/110246/2e29c7b22627d9d22e5f9cc7196019a2/logistik-roadmap-data.pdf>, abgerufen am: 17.03.2021
- [11] Lenz, H., Kessler, S., Rytir, C., Schick, A. u. Schneidehan, M.: Plug-and-operate für die produktionsnahe Logistik. *atp edition* 59 (2017) 01-02, S. 9
- [12] Wiendahl, H.-P.: *Planung modularer Fabriken. Vorgehen und Beispiele aus der Praxis*. München, Wien: Hanser 2005

- [13] *ISA-TR88.00.02 (PackML). Machine and unit states: an implementation example of ISA-88*
- [14] DIN 55405:2014-12. *Verpackung - Terminologie - Begriffe*
- [15] BEUMER Group: BEUMER fillpac® FFS: Form, Fill, Seal | BEUMER Group, 2021.
<https://www.beumergroup.com/de/pd/abfuelltechnik/fillpac-ffs-anlage/>, abgerufen am: 11.04.2021
- [16] BEUMER Group: BEUMER paletpac® | BEUMER Group, 2020.
<https://www.beumergroup.com/de/pd/palettiertechnik/paletpac-palettierer/>, abgerufen am: 12.04.2021
- [17] BEUMER Group: BEUMER stretch hood® | BEUMER Group, 2020.
<https://www.beumergroup.com/de/i/beumer-stretch-hood-29966/>, abgerufen am: 04.05.2021
- [18] VDI/VDE/NAMUR 2658-4:2021 (geplant). *Automatisierungstechnisches Engineering modularer Anlagen in der Prozessindustrie Modellierung von Moduldiensten*
- [19] Stutz, A., Maurmaier, M. u. Unser, M.: Services Design Principles zur Unterstützung des Modulbauers bei der Ermittlung möglicher Modulservices gemäß VDI/VDE/NAMUR 2658. In: AUTOMATION 2019, S. 883–893

Bereitstellungsmethoden von ungeordneten Werkstücken in den automobilen Produktionsprozess

M.Eng. **M. Sarna**, Dr. Ing. **J. Weist**, Dr.-Ing. **A. Tegtmeier**,
Volkswagen AG, Wolfsburg;
apl. Prof. Dr.-Ing. habil. **A. Lüder**, Otto-von-Guericke Universität,
Magdeburg

Kurzfassung

Die geordnete Werkstückzuführung aus *Standardladungsträgern* (LT) in den Produktionsprozess, auch *Griff in die Kiste* (GIDK) oder engl. Bin Picking genannt, entsteht als Symbiose zwischen Werkstück, Industrieroboter, Effektor und Bildverarbeitungssystem. Unabhängig davon ob die Werkstücke ungeordnet als Schüttgut oder teilgeordnet in Reihen und Stapeln vorliegen. Der GIDK stellt ein mechatronisches System im Sinne der VDI 2206 dar [1]. Diese stellen Betreiber und Entwickler auch heute noch vor eine umfangreiche Herausforderung. Die Komplexität, der Engineering-Aufwand im Sinne der Planung, Vorbereitung und Inbetriebnahme, und weiter die meist ausbleibende Wirtschaftlichkeit wie auch abnehmende Verfügbarkeit der Anlagen sind die führenden Ursachen dafür. Viele Anlagen entsprechen hinsichtlich Taktzeit und Störanfälligkeit nicht den industriellen Anforderungen und zeichnen sich durch hohen Projektierungs- und Betriebsaufwand aus.

Die wirtschaftlich, ergonomisch und demographisch getriebene Notwendigkeit zur Automatisierung macht es dennoch erforderlich die geordnete Werkstückzuführung aus Standardbehältern in den Produktionsprozess planerisch und praktisch besser zu verstehen. Dieses Verständnis kann auf den funktionalen Anforderungen an den GIDK aufgebaut werden. Zielstellung dieses Beitrages ist es daher eine Klassifikation der funktionalen Anforderungen an das Handling von Werkstücken auf der Basis ihrer Bereitstellungssituation abzuleiten und dadurch die Anwendbarkeit bestehender Konzepte der Werkstückzuführung bewertbar zu machen.

Abstract

The orderly workpiece supply out of standard boxes into the manufacturing process, also known as bin picking, is formed mainly by the symbiosis of workpiece, industrial robot, end-of-arm tooling and image processing system. Thereby the workpieces are unordered as bulk goods or partially arranged in rows and stacks.

Operators and developers are still challenged with the bin picking task. The main reasons for this are the complexity, the engineering effort in terms of planning, preparation and commissioning, and the mostly lacking profitability as well as decreasing availability of the systems. Many systems do not meet industrial requirements in terms of cycle times and susceptibility to failure and are characterized by high project planning and operating costs.

The economically, ergonomically and demographically driven need for automation makes it nevertheless necessary to understand better the orderly workpiece supply from standard containers in the production process in terms of planning, commissioning and production. This understanding can be built on the functional requirements for bin picking. The aim of this article is therefore to derive a classification of the functional requirements for the handling of workpieces on the basis of their availability situation and thereby make it possible to assess the applicability of existing concepts of workpiece supply.

1. Einleitung

Der *Griff in die Kiste* (GIDK) stellt eine wichtige meist ökonomisch induzierte Herausforderung für den industriellen Anlagenbau im Automobilbau dar. Bisher ist die Anzahl der technisch und ökonomisch erfolgreichen Lösungen beschränkt, was zu einer zunehmend ablehnenden Haltung gegenüber GIDK Systemen führt. Dies ist insoweit unberechtigt, da GIDK Systeme keine Lösungen für beliebige Herausforderungen in der Werkstücklogistik bieten können, sondern zielgerichtet auf lösbare Anwendungen konzentriert werden sollten.

Es besteht jedoch derzeit nur unzureichend die Möglichkeit zu bewerten, in welchen Anwendungsfällen ein GIDK System sinnvoll einsetzbar ist. Um dies anzugehen wird in diesem Beitrag eine Klassifikation der funktionalen Anforderungen an das Handling von Werkstücken auf der Basis ihrer Bereitstellungssituation abgeleitet und dadurch die Anwendbarkeit bestehender Systeme bewertbar gemacht.

Zu diesem Zweck werden in Abschnitt 2 der Begriff der Ordnungszustände und in Abschnitt 3 der aktuelle Stand der Technik hinsichtlich Werkstückzuführung vorgestellt. Diesem folgen die Untersuchung der Anlieferungszustände von Werkstücken und der Werkstückentnahme in den Abschnitten 4 und 5. Auf dieser Basis werden dann ein allgemeiner Prozess zur Werkstückzuführung und bestehende technische Lösungskonzepte für diesen Prozess vorgestellt die in einer Bewertung kulminieren. Mit einer Zusammenfassung endet dieser Beitrag.

2. Grundlage Ordnungszustände nach VDI 2860

Ordnen ist das Herstellen eines bekannten oder geplanten Ordnungszustands [2] [3]. Die Ordnungszustände von Werkstücken können, wie in der VDI 2860 definiert, beschrieben werden.

Die räumliche Anordnung eines geometrisch bestimmten Körpers, und somit des Körperkoordinatensystems, ist durch seine Orientierung und Position in kartesischen Bezugssystemen definiert [2]. Dazu wird der *Ordnungszustand* (OZ) definiert, der sich aus *Orientierungsgrad* (OG) und *Positionierungsgrad* (PG) zusammensetzt. Mithilfe des OZ, siehe Tabelle 1, können Werkstückzustände vollständig beschrieben werden.

Tabelle 1: Ordnungszustand OZ nach [2]

Ordnungszustand OZ = (OG ; PG)			
Orientierungsgrad OG		Positionierungsgrad PG	
Orientierungsgrad OG als Kennzahl für die Anzahl der definierten rotatorischen Freiheitsgrade eines Körpers		Positionierungsgrad PG als Kennzahl für die Anzahl der definierten translatorischen Freiheitsgrade eines Körpers	
OG		PG	
3	Orientierung des Körpers in allen Rotationsachsen bestimmt	3	körpereigenes Koordinatensystem befindet sich an einem definierten Punkt
2	Orientierung des Körpers in zwei Rotationsachsen bestimmt	2	körpereigenes Koordinatensystem beliebig auf einer Bahn angeordnet
1	Orientierung des Körpers in einer Rotationsachse bestimmt	1	körpereigenes Koordinatensystem beliebig auf einer Ebene angeordnet
0	Orientierung des Körpers in allen Rotationsachsen unbestimmt	0	Körper beliebig im Raum angeordnet

Dadurch ergibt sich die Zuordnung OZ (3 ; 3) als geordnet, OZ (0 ; 0) als ungeordnet und alle weiteren OZ als teilgeordnet.

Die vorgestellte Beschreibung der Ordnungszustände kann für alle Werkstücke verschiedener Gewerke und Branchen angewendet werden. Der Ausgangszustand der Werkstücke im *Ladungsträger* (LT) ist in Ruhe.

3. Stand der Technik der Werkstückzuführung

Die geordnete Werkstückzuführung in den Karosseriebauprozess nach aktuellem Stand der Technik wird in Tabelle 2 vorgestellt.

Die menschliche Arbeitskraft kommt aus betriebswirtschaftlichen und ergonomischen Gründen zum Einsatz, wenn die Prozessanforderungen (Zuverlässigkeit [4], Wirtschaftlichkeit, Engineering-Aufwand) keine technisch und betriebswirtschaftlich Alternative bieten.

Zuführsysteme, siehe Tabelle 2, genauer Ordnungseinrichtungen, z.B. Vibrationswendelförderer, ordnen Werkstücke durch Vereinzeln für die automatisierte Zuführung. Zuführsysteme sind immer als werkstückspezifisch und durch mechanische Sonderanfertigungen nicht als Universalgerät zu betrachten [5]. Durch Aufwand und Kosten für Entwicklung werkstückspezifischer Handhabungsgräte sind diese heute wie früher nicht wirtschaftlich [6] [7] und eignen sich daher für Normteile, also Komponenten die langfristig unverändert bleiben. Ein von Schmid [8] vorgestelltes branchenübergreifendes und repräsentatives Werkstückspektrum für Förderer zeigt Kleinteile. Die Werkstückmasse der Kleinteile liegt zwischen 1,3 g und 13,9 g. Die Summe der Werkstückabmessungen Länge, Breite und Höhe liegt zwischen 41 mm bis 116,7 mm. Dieses Spektrum unterscheidet sich deutlich von im Karosseriebau üblichen Komponenten in Masse und Abmessung. Zumal sind Werkstückbeschädigung durch Verformung nicht bei allen Werkstücken auszuschließen. Weitere Ursachen dafür sind u.a. Lärmbelastung durch Körperschall in Abhängigkeit der Größe der abstrahlenden Fläche [9], erhöhte Komplexität der Ordnungseinrichtung durch fehlende Werkstücksymmetrien [10], mögliche Taktzeitüberschreitung durch ordnungswahrscheinlichkeitsbedingte Förderleistung bei vollständig asymmetrischen Werkstücken [5] sowie technologiebedingte Skalierung von Werkstückgröße zu Förderergröße, was wiederum auf Zugänglichkeit und Platzbedarf wirkt.

Tabelle 2: Übersicht Werkstückzuführung (Grauabstufung von dunkel zu hell mit zunehmender Grundlagenbedeutung für GIDK)

Mensch	Zuführsysteme	Industrieroboter		
Ein- u. Mehrmaschinen- Bedienung	allgemein Handhabungs-einrichtungen zur automatischen Teilbereitstellung z.B. Förderer [3]	Entnahme	Entnahme mit Sensorik	
Direkt- und Takt- entkoppelt		Regalentnahme	Sensor	3D-Vision
Manipulator z.B. Balancer	Normteile geeignet	Programmiert, Parametrisiert [11]	Entnahme mit Sensorik aus SLT ggf. mit Anwesenheitsprüfung / Lagekorrektur	teilgeordnete OZ
sonstige techn. Hilfsmittel				ungeordnete OZ

Industrieroboter kommen durch höhere Flexibilität als technische Alternative zu Ordnungseinrichtungen in Frage. Neben der programmierten oder parametrisierten Entnahme aus bekannten PG und OG oder errechneten OZ, welche kaum Anwendungsfälle im Karosseriebau kennt,

wurde durch logistische und ergonomische Belange die automatisierte Entnahme aus *Sonderladungsträger* SLT bis zur Serienreife getrieben. Die erhöhte Kapitalbindung durch werkstückangepasste (SLT) und der logistische Mehraufwand durch ausbleibende Mehrfachnutzung dieser erhöhen den Bedarf an LT-basierten Lösungen. Heute sind LT-basierte Automatisierungslösungen mit Bildverarbeitungssystemen (3D-Vision-Sensorik) bekannt, die sich weiter durch teilgeordnete oder ungeordnete Werkstückzustände als Eingangsparameter unterscheiden. In der Praxis wird dafür die Bezeichnung Griff in die Kiste (GIDK) geprägt. Alle anderen in Tabelle 2 genannten Industrieroboteranwendungen erfordern OZ (3 ; 3).

Die Anwendbarkeit der bekannten Methoden der Werkstückzuführung hängt maßgeblich von der Bereitstellung und dem Anlieferungszustand der Werkstücke ab.

4. Anlieferungszustände

Eigene Analysen der Autoren im Rahmen von Expertengesprächen wie auch bestehende Literatur wie Sigloch [12] zeigen, dass in der modernen automobilen Karosseriefertigungen eine relativ konstante Verteilung der Bereitstellung der Werkstückkomponenten hinsichtlich der Werkstücktypen identifiziert werden kann.

Tabelle 3: Anordnung der Werkstücke im Ladungsträger

OZ	geordnet	teilgeordnet	ungeordnet
Werkstück-anordnung im LT	bestimmt	unbestimmt	
Häufigkeit im Karosseriebau	≈ 20 %	≈ 50 %	≈ 30 %
Werkstück-ausrichtung durch	Vorrichtungen, SLT	andere Werkstücke, LT	keine
Zustand	einzeln	Stapel, Reihe, Formstapel	Normalverteilung oder Abweichungen der OW werkstückabhängig
Zusatz	frei	verschachtelt	
		unverschachtelt	
		verhaken / verklemmen / frei	
Anwendung	SLT-Entnahme	<Grenzfälle>	LT-Entnahme, GIDK

Logistisch betrachtet werden 22 % der Werkstücke in SLT und 78 % der Werkstücke in LT bereitgestellt. Letzte können weiter unterschieden werden in Schüttgut 32 %, OZ (0 ; 0), und diverse Stapelarten 46 %, ebenfalls OZ (0 ; 0). Durch Stapeln können signifikant höhere Füllgrade im LT erreicht werden. Beim Umgang mit Stapeln ist das eigenformbedingte Werkstückverhalten [3], auch *Ordnungswahrscheinlichkeit* (OW) [7] genannt, neben dem OZ eine nicht zu vernachlässigende Eingangsgröße bei der Zuführung von ungeordneten Werkstücken. Nach der Übersicht möglicher Werkstückzuführungen und deren Verteilung im Karosseriebau folgt die dazugehörige Untersuchung aus Sicht der Ordnungszustände, siehe Tabelle 3. Der Prozesszugang von 48 % der Werkstücke erfolgt taktentkoppelt durch automatisierte Entnahme aus SLT, Rutschen, Taktbänder o.ä. mit OZ (3 ; 3) durch geordnetes Speichern. Diese werden wie die 52 % taktgebunden eingelegten Werkstücke zuvor manuell bestückt.

Die OZ geordnet, teilgeordnet und ungeordnet sind durch logistische und betriebswirtschaftliche Belange im Karosseriefertigungen mit den Häufigkeiten in Tabelle 3 Zeile 3 zu finden. Der Prozesssicherheit dienend, müssen teilgeordnete OZ im LT wie OZ (0 ; 0) behandelt werden. Trotz der augenscheinlichen Anwesenheit von Ordnung durch Stapel, Formstapel und Reihen, teilweise verschachtelten, ist eine programmierte oder parametrisierte Roboterentnahme nicht möglich, vgl. Tabelle 2. Der wesentliche Unterschied zur geordneten Bereitstellung liegt in der Ursache für die Werkstückausrichtung und im Zustand der Werkstücke. Hier liegen die Werkstücke vereinzelt und bestimmt unabhängig von der Art und Anzahl anderer gleichartiger Werkstücke im SLT.

Durch die industriell geforderte Prozesssicherheit müssen auch nicht vorgesehene Ordnungszustände berücksichtigt werden, solange die Werkstücke nicht wie in SLT in fester Position und Orientierung gespeichert sind.

5. Werkstück-Entnahme-Grad

Tabelle 4: Klassifizierung in Werkstück-Entnahme-Grade WEG

WEG	Eingabe OZ (OG ; PG)	Beschreibung	Beispiel
0	(3 ; 3)	bekannter OZ - Entnahme ohne Sensor	RPS-Ablage
1	(3 ; 3)	bekannter OZ nach Einmessen des SLT	Werkstückmatrize
2	$4 < (OG + PG) < 6$	Erkennung der exakten Position im SLT / Werkstück teillokalisiert	SLT
3	$0 < (OG + PG) \leq 4$	teilgeordnete Sonderfälle ohne karosseriebautechnische Anwendung	ohne
4	(0 ; 0)	Werkstücke in gegenseitiger Beeinflussung	Stapel
5	(0 ; 0)	einzelne ungeordnet	Förderband
6	(0 ; 0)	vollständig ungeordnet	GIDK

Mit der Kenntnis der OZ aus Tabelle 1, den möglichen Werkstückzuführungen aus Tabelle 2 und technischen Besonderheiten der Werkstückanordnungen aus Tabelle 3 wird der *Werkstück-Entnahme-Grad* (WEG) definiert. Der WEG gibt in einer von 7 Stufen einen konkreten karosseriebautechnischen Anwendungsfall wieder, siehe Tabelle 4. Die Grade 0, 1 und 2 gehören mit einer signifikanten Anzahl Anwendungen zum Stand der Technik. Der Grad 3 als Sonderfall wird wie Grad 4 analog der teilgeordneten Bereitstellung ebenfalls wie die Grade 5 und 6 mit OZ (0 ; 0) behandelt.

An dieser Stelle muss erwähnt werden, dass die eingangs beschriebene VDI 2860 keinen Genauigkeitsansprüche bezüglich Position und Orientierung macht. Daher ist die Wahl der Bereitstellungsmethode und den damit realisierbaren Prozessgenauigkeiten in Verbindung mit den Übergabeanforderungen maßgeblich.

6. Prozess der Werkstückzuführung aus Sicht der Ordnungszustände

Tabelle 5: Werkstückzuführung aus Sicht der Ordnungszustände

Arbeitsschritt	Werker-tätig-keit	GIDK	SLT-Ent-nahme	GIDK zentra-lisiert	Geordnete Abnahme im Presswerk
Presswerk	OZ (0 ; 0)	OZ (0 ; 0)	OZ (0 ; 0)	OZ (0 ; 0)	OZ (3 ; 3)
Vorbereitung SLT Beladung			manuell OZ (3 ; 3)	WEG 3 bis 6 OZ (3 ; 3)	
Logistik der LT u. SLT	OZ (0 ; 0)	OZ (0 ; 0)	OZ (3 ; 3)	OZ (3 ; 3)	OZ (3 ; 3)
LT-Entnahme	OZ (0 ; 0)	WEG 3 bis 6 OZ (3 ; 3)			
SLT-Entnahme			WEG 2 OZ (3 ; 3)	WEG 2 OZ (3 ; 3)	WEG 2 OZ (3 ; 3)
Übergabe in Vorrichtung	OZ (3 ; 3)	OZ (3 ; 3)	OZ (3 ; 3)	OZ (3 ; 3)	OZ (3 ; 3)

Der Umformprozess im Presswerk zur Herstellung von Formteilen gilt als Eingabeprozess für den Karosseriebau. Da das Umformen im Presswerk und die Weiterverarbeitung im Karosseriebau durch Gewerkspezialisierung und Taylorismus räumlich getrennt sind, ist der Logistikprozess erforderlich. Die geordnete Übergabe, OZ (3 ; 3), der Werkstücke an den Fügeprozess im Karosseriebau, mit der diese Betrachtung endet, erfolgt bei den aufgeführten Fällen, siehe Tabelle 5, bis auf die Werkertätigkeit automatisiert. Tabelle 5 zeigt dabei von oben nach unten die notwendigen Prozessschritte und die durch sie erreichten Ordnungszustände.

Der GIDK initiiert durch das Zusammenspiel von Effektor, Manipulator, Bildverarbeitungssystem und Werkstück die Werkertätigkeit. Dadurch ergeben sich grundsätzlich keine Änderungen für die vorgelagerten Prozessschritte Presswerk und Logistik.

Betrachtet man den signifikanten Anteil der *SLT-Entnahme* mit 22 % und der dafür vorausgehenden manuellen Bestückung der SLT, wurde hier, unter *GIDK zentralisiert*, die Beladung der SLT durch den GIDK betrachtet.

Eine weitere Möglichkeit zur Vermeidung von Ordnungsprozessen, durch manuelle Tätigkeit oder den GIDK, ist die *geordnete Abnahme im Presswerk* bevor die Formteile als werkzeugfallende Werkstücke den OZ (0 ; 0) annehmen. Dieser Ansatz ist durch die Pressenhubzahl

(Taktzeit) und die karosseriebauübliche Werkstückgröße betriebswirtschaftlich und technisch begrenzt.

Zusammenfassend ist der in Tabelle 5 geführte OZ (3 ; 3) als Kostentreiber durch Ordnen selbst und das folgende Aufrechterhalten zu betrachten. Somit ist, ausgenommen diverser logistischer Belange, ein möglichst im Prozessverlauf spätes Herstellen des OZ (3 ; 3) erstrebenswert.

Da der GIDK mit diesem Potential und den eingangs aufgeführten Herausforderungen noch nicht zum Stand der Technik gehört, werden im Folgenden GIDK Konzepte detailliert untersucht.

7. Bekannte Konzepte

Die Unterschiede der Werkstückzuführungen wurden bisher aus logistischer und ordnungstheoretischer Sicht als Grundlage für Tabelle 6 beleuchtet. In Tabelle 6 folgt die Analyse bekannter Konzepte durch Vergleich der Teilfunktionen in Art und Folge. Als bekannte Konzepte sind geführt Labor-, Messe- und Serienanwendungen. Die *Basisfunktion* (BS), siehe Tabelle 6 Konzept 2, bildet den Funktionsumfang aller darauf aufbauenden Konzepte (Konzepte 3 bis 11).

Die Funktion Handhaben, als definiertes Verändern mit OZ (3 ; 3), unterscheidet sich vom Fördern durch Bekanntheit der rotatorischen Freiheitsgrade OG [2].

Diese Darstellung dient u.a. zur Abschätzung der Taktzeit sowie der erreichbaren Positioniergenauigkeit. Durch Analyse der Einzelungenauigkeiten von Werkstück, Handhabungsmittel, Effektor und z.B. Bildverarbeitungssystem können durch mathematische Operationen zu jedem Zeitpunkt vorherrschende Ungenauigkeiten beziffert werden. Versuche zeigen, dass die Verwendung von mechanischen Einweisern, Konzept 3, die höchste Genauigkeit beim geordneten Speichern der verglichenen Konzepte erreicht. Der Kombination mechanischer Einweiser mit anderen Konzepten steht nichts im Weg. Allgemein gilt, je näher ordnungsherstellende Schritte wie Erkennen oder Ausrichten am Teilprozessende angesiedelt sind, desto genauer ist geordnetes Speichern möglich.

Tabelle 6: Übersicht bekannter Konzepte

Konzept		1	2	3	4	5	6	7	8	9	10	11	12
Teilfunktion am Werkstück		Vereinzeln [13]	Basisfunktionen BS [14] [15] [16]	BS mit Einweiser	Aufnehmen mit Formschluss	BS mit Aufnahmeänderung (Aufnahmeä.)	BS mit Ausrichten [17]	BS mit Aufnahmeä. und Ausrichten [18]	BS mit aktiven Ausrichten	BS mit zweiter Erkennung	BS mit zweiter Erkennung im Effektor [19]	BS mit Aufnahmeä. und zweiter Erkennung [20]	Blind
Vereinzeln	Ungeordnetes Speichern	x	x	x	x	x	x	x	x	x	x	x	x
	PG u. OG erkennen	x	x	x	x	x	x	x	x	x	x	x	
	Aufnehmen	x	x	x		x	x	x	x	x	x	x	x
	formschlüssiges Aufnehmen				x								
	Handhaben	x	x	x	x	x	x	x	x	x	x	x	
	Fördern												x
	PG u. OG erkennen										x		
	Aufnahme lösen					x	x	x	x	x		x	x
	Aufnehmen							x					
	PG u. OG erkennen									x		x	x
	Aufnahme lösen							x					
	PG u. OG ausrichten						x	x	x				x
	Aufnehmen					x	x	x	x			x	x
	Handhaben					x	x	x	x			x	x
	Aufnahme lösen	x	x	x	x	x	x	x	x	x	x	x	x
	OZ ausrichten mit Einweiser			x									
	geordnetes Speichern		x	x	x	x	x	x	x	x	x	x	x
	ungeordnetes Speichern	x											
Ordnen													

Die Anlagenkonzepte 1 bis 4 basieren auf der Vermutung, dass zu jeder Zeit ausreichend viele Werkstücke in einer erreichbaren und verwendbaren Orientierung und Position, OW siehe Kapitel 4, vorliegen. Daher sind lediglich die Konzepte die die Aufnahmeänderung, das Umgreifen,

berücksichtigen im industriellen Umfeld diskutabel. Andere Konzepte sind durch in der Praxis regelmäßig ausbleibenden Verfügbarkeit ausschließlich für die SLT-Entnahme geeignet. Der GIDK wird durch Trennung der Funktionen in Vereinzeln, Ordnungszustand herstellen und Bereitstellen beherrschbar.

Mit den vorgebrachten Daten ist es möglich den Handlungsbedarf im Karosseriebau gegen den GIDK neu einzuordnen. Zum GIDK sind zwei Definitionen bekannt:

(1) *„Eine Handhabungseinrichtung lokalisiert in Ergebnis eines Erkennungsvorgangs ein Teil, greift gezielt in einen Vorrat chaotisch angelieferter Teile, entnimmt das Teil und legt es definiert ab.“* [3]

(2) *„Automatisches Entnehmen einzelner Gegenstände aus einer völlig ungeordneten Menge gleichartiger Objekte mit einem Roboter und unter Nutzung spezieller Bildverarbeitungssysteme.“* [21]

Wie Tabelle 6 zeigt, ist nicht in jedem Fall gezieltes Greifen und Handhaben erforderlich, vgl. Tabelle 6 Konzept 12, was im Rahmen einer Definition bereits einen möglichen Lösungsweg impliziert. Auch die Einbeziehung der Bildverarbeitungssysteme lässt andere Sensortechnologien vollständig unbeachtet. Für eine ganzheitliche Lösung muss objektiv die *„Automatisierung der definierten Werkstückzuführung aus Standardladungsträgern von ungeordneten Werkstücken in den Produktionsprozess“* betrachtet werden. Diese Definition setzt keine Technologie oder gerätebasierten Funktionalitäten voraus.

Die erarbeiteten Strukturen aus Tabelle 6 führen in weiteren Forschungsarbeiten zu neuen Bewertungsmethoden. Diese Methoden bauen auf der Genauigkeit der Einzelfunktionen auf und führen zur Anlagenverfügbarkeit. Dadurch werden kostenintensive GIDK-Applikationen bereits in der Entstehungsphase beherrschbar. Wie in Kapitel 5 erläutert bedarf es für diese Bewertungsmethode einen Übergang von qualitativen OZ hin zu quantitativen Genauigkeiten der Einzelfunktionen mit Einbindung der OZ.

8. Zusammenfassung

Der GIDK als mögliche Bereitstellungsmethode von ungeordneten Werkstücken aus Ladungsträgern in den automobilen Produktionsprozess wurde in dieser Arbeit gegen die Werkertätigkeit und Zuführsysteme verglichen. Durch logistische Einordnung und ordnungstheoretische Beleuchtung konnten Eigenschaften extrahiert und Handlungsbedarfe definiert werden. Die Einführung der Werkstück-Entnahme-Grade ermöglicht es zukünftige Anlagenkonzepte gezielt zu hinterfragen und dahingehend besser zu verstehen. Die Ergebnisse dieser Arbeit lassen sich auf GIDK Anwendungen anderer Gewerke und Branchen übertragen. So sind die Inhalte der Tabellen 2 bis 6 allgemein vor dem GIDK-Kontext zu verstehen.

Konzeptschwächen des GIDK ergeben sich durch Überlappung der Grundfunktionen Vereinzeln und Ordnen. Durch konzeptionelle Trennung dieser Funktionen wird der Prozess beherrschbar. Weiterhin bleiben das standardisierte Positionieren, Orientieren und Umgreifen eine konstruktive und betriebswirtschaftliche Herausforderung im Karosseriebau. Bislang können Werkstücke im Karosseriebau wirtschaftlich nur Vereinzelt werden.

Hinweis: Die VDI 2860 wurde 2018 zur Überarbeitung vorübergehend zurückgezogen. Die darin enthaltene Beschreibung der Ordnungszustände ist einmalig und grundlegend für die Handhabungstechnik und deshalb auch Eingangsgröße für diese Arbeit.

Disclaimer

Ergebnisse, Meinungen und Schlüsse dieser Veröffentlichung sind nicht notwendigerweise die der Volkswagen Aktiengesellschaft.

Literaturverzeichnis

- [1] VDI-Richtlinien 2206 (2004): Entwicklungsmethodik für mechatronische Systeme. Düsseldorf: Verein Deutscher Ingenieure.
- [2] VDI-Richtlinien 2860 (1990): Handhabungsfunktionen, Handhabungseinrichtungen – Begriffe Definitionen, Symbole. Düsseldorf: Verein Deutscher Ingenieure. (zurückgezogen 2018)
- [3] Hesse, S. (2016): Grundlagen der Handhabungstechnik.- 4. Auflage: Plauen, Carl Hanser Verlag.
- [4] VDI-Richtlinien 4003 (2007): Zuverlässigkeitsmanagement. Düsseldorf: Verein Deutscher Ingenieure.
- [5] Groh, W. (1957): Das Ordnen von Massenteilen und ihre selbsttätige Zuführung in die Werkzeugmaschine.- Werkstatttechnik und Maschinenbau 47 (1957), S. 402 – 410, Hannover.
- [6] Flexibowl (2020): Ein Flexibles Zuführsysteme. (Online). <https://www.flexibowl.de/>. Zugriff am 13.09.2020, ARS Automation.
- [7] Warnecke, H. J. (1972): Das Ordnungsverhalten von Werkstücken bei automatisierter Handhabung - Zeitschrift für industrielle Fertigung, Nr. 62 (1972), S. 154 - 160, Heidelberg
- [8] Schmid, S. (2006): Automatisierte Ordnungs- und Kommissionierzelle zur hochflexiblen Bereitstellung von Werkstücken in der Montage. (Dissertation). Institut für Industrielle Fertigung und Fabrikbetrieb der Universität Stuttgart.
- [9] Hering, E.; Martin, R.; Stohrer, M. (2012): Physik für Ingenieure.- 11. Auflage.: S. 642, Heidelberg, Springer Verlag.
- [10] Müller, G. (Hrsg.) (1992): Lexikon Technologie Metallverarbeitende Technologie, Ordnungseinrichtung.- 2. Auflage.: S. 422, Haan-Gruiten, Verlag Europa Lehrmittel.
- [11] Halter CNC Automation (2020):. (Online). <https://www.haltercncautomation.de/>. Zugriff am 15.10.2020, HALTER CNC Automation B.V..
- [12] Sigloch, R. (2020): Griff in die Kiste. (Bachelorarbeit). Duale Hochschule Baden-Württemberg Ravensburg Ravensburg, S. 70.
- [13] FANUC (2014): Robot Uses New Bin Picking Gripper & 3D Vision to Bin Pick Non-Ferrous Valves. (Online), <https://www.youtube.com/watch?v=A6lc9j9zH50>, Zugriff am 13.01.2021, FANUC America Corporation.
- [14] Infaimon (2017): InPicker: INFAIMON Bin-Picking 3D - How it works. (Online), <https://www.youtube.com/watch?v=2QE9PfoUucw>, Zugriff am 13.01.2021, INFAIMON visión artificial.

- [15] Pickit3D (2017): Smart automated bin picking with Pickit 3D. (Online). <https://www.youtube.com/watch?v=mySU2teAtmA>, Zugriff am 13.01.2021, Pickit3D.
- [16] SL-Tech (2014): SL-Tech BinPicking. (Online). <https://www.youtube.com/watch?v=LrZ8UjdvyAQ>, Zugriff am 13.01.2021, SL-Tech.
- [17] KUKA – Robots & Automation (2016): Advanced Robotic Bin Picking from Boll Automation and KR QUANTEC. (Online). <https://www.youtube.com/watch?v=ij7zOdiXnsE>, Zugriff am 13.01.2021, KUKA - Robots & Automation.
- [18] Universal Robots (2019): Jenny Wattle meistert ‚Griff in die Kiste‘ mit UR5 Cobots. (Online). https://www.youtube.com/watch?v=DDaitQEqD_w, Zugriff am 13.01.2021, Universal Robots.
- [19] ROBOT- NET (2019): SoSta: A flexible robot cell used to sort and stack sheet metal parts. (Online). <https://www.ipa.fraunhofer.de/de/referenzprojekte/SoSta.html>, Zugriff am 13.01.2021, Fraunhofer Institut für Produktionstechnik und Automatisierung.
- [20] Scape (2014): Scape Technologies - Standardized & Flexible Bin-Picking with UR Robot. (Online). <https://www.youtube.com/watch?v=krWeJVSZFNI>, Zugriff am 13.01.2021, SCAPE Bin-Picking vlog.
- [21] Hesse, S.; Malisa, V. (Hrsg.) (2016): Taschenbuch Robotik Montage Handhabung.- 2. Auflage.: S. 497, München, Carl Hanser Verlag.

BIM im Betrieb durch lebenszyklusübergreifende Verfügbarkeit von Anlagendaten auf Basis von I4.0-Verwaltungsschalen

M. Ostermeier (B.Eng.), **N. Maisch** (B.Eng.), **M. Both** (M.Eng.),
R. Ulmer (B.Eng.), Prof. Dr. **J. Müller**, Technische Hochschule Köln

Kurzfassung

Der Lebenszyklus gebäudetechnischer Anlagen ist von Informationslücken geprägt. Diese Lücken können durch eine gemeinsame informationstechnische Basis geschlossen werden, die durch standardisierte Industrie 4.0 Teilmodelle umgesetzt wird. Dadurch kann ohne manuelle Aufwendungen im Betrieb sowohl auf die Informationen aus der BIM-Planung als auch auf die Live-Betriebsdaten zugegriffen werden. In diesem Beitrag wird gezeigt, wie standardisierte Industrie 4.0-Teilmodelle für Komponenten der Technischen Gebäudeausrüstung in BIM integriert werden können. Anschließend wird das Konzept in einer prototypischen Monitoringanwendung im Betrieb umgesetzt.

1. Einleitung

In der Technischen Gebäudeausrüstung (TGA) fallen während jeder Phase des Lebenszyklus vielfältige Anlagen-Informationen an, z.B. Planungs- und Herstellerinformationen über die Komponenten oder die Anlagendaten während der Betriebsphase. Diese Informationen werden von verschiedenen Teilnehmern des Wertschöpfungsnetzwerks Bau zu unterschiedlichen Zeiten in digitaler Form benötigt. Die Bild der Planungs- und Betriebsdaten in ein standardisiertes Format und dessen durchgängiger Informationsfluss sind Anforderungen sowohl an Plug and Play Szenarien im Betrieb als auch an verschiedene Anwendungen wie beispielsweise dem Energie- (EM) und dem Asset-Monitoring (AM). Um das volle Potenzial der Digitalisierung auszuschöpfen, muss der Zugriff auf die Informationen während jeder Phase des Lebenszyklus automatisiert erfolgen können, da so der Engineeringaufwand reduziert wird. Die lebenszyklus-übergreifende Bereitstellung von Anlagendaten wird in der Gebäudetechnik bisher nicht umgesetzt. Somit ist ein solcher Informationsaustausch in der TGA nicht bruchfrei möglich. Diese Repräsentanten weisen keine semantischen Ausprägungen auf, die zusätzlich zu den Planungsdaten die Verarbeitung von Anlagendaten im Betrieb und somit einen lückenlosen Informationsfluss ermöglichen.

Obwohl die Konzepte Building Information Modeling (BIM) und Industrie 4.0 (I4.0) die gleichen Ziele verfolgen, da hierbei die gleichen Probleme gelöst werden müssen, und Komponenten wie Pumpen in beiden Bereichen eingesetzt werden, wurden bisher keine Kombinationsmöglichkeiten von BIM und I4.0 veröffentlicht. Weiterentwicklungen sind nicht verzahnt, sondern verlaufen parallel. Untersuchungen zeigen jedoch, dass sich I4.0-Modelle zur Beschreibung von Anlagenkomponenten der Gebäudetechnik eignen und in BIM-Lösungen integriert werden können [1]. Ein Weg, die Informationslücken zu schließen, ist somit der Einsatz von standardisierten I4.0-Verwaltungsschalen, die den gesamten Lebenszyklus der Anlage begleiten [2]. Dadurch kann ohne manuelle Aufwendungen für das Engineering sowohl auf die Informationen aus der BIM-Planung als auch auf die Live-Betriebsdaten zugegriffen werden.

In diesem Paper wird eine Kombinationsmöglichkeit der Konzepte von BIM und I4.0 aufgezeigt. Auf Basis eines Prozessmodells zur Bild von Asset-Informationen wurden exemplarisch I4.0-Verwaltungsschalen für raumluftechnische (RLT) Anlagen und Pumpen entwickelt, durch die ein Zugriff auf die Informationen der Assets, einschließlich der Betriebsdaten, möglich ist. In diesem Paper wird gezeigt, wie standardisierte Planungs- und Betriebsdaten in der Gebäudetechnik mithilfe einer I4.0-Verwaltungsschale automatisiert in einem BIM-Tool ausgewertet und dargestellt werden können. Die standardisierten Informationen in Kombination mit der Verwaltungsschale bilden die Basis für Self-X Fähigkeiten der Komponenten.

2. Kombination von BIM und Industrie 4.0

2.1. Betriebsdaten

Durch die Möglichkeit einer Bereitstellung von Asset-Merkmalen in einem standardisierten Repositorium und der Integrierbarkeit in bestehende Planungsprozesse bietet BIM einen Mehrwert bei der Digitalisierung von Planungen von Gebäuden und deren technischer Ausrüstung. Gerade in der TGA ist eine Durchgängigkeit der Informationen notwendig, um beispielsweise Energieverbräuche überwachen oder Instandhaltungsmanagement betreiben zu können. Ebenso wie in der Planung werden für eine automatisierte Auswertung eindeutige und standardisierte semantische Informationen benötigt. Das Merkmal der Leistung ist somit bei allen Komponenten gleich definiert. Dies ermöglicht Mechanismen zur Erkundung der Komponenten und die automatisierte Einbindung dieser in EM oder AM-Anwendungen. Ein händisches Hinzufügen der einzelnen Merkmale in die entsprechenden Anwendungen ist nicht mehr notwendig.

Im Gegensatz zu BIM fokussiert I4.0 neben der Planung auch den Betrieb von Anlagen [3]. Dies wird beispielsweise bei der Entwicklung der Teilmodelle für Pumpen deutlich. Neben

einem Teilmodell für die technischen Informationen des Herstellers und der Planung wurden Teilmodelle entwickelt, welche Informationen aus dem Betrieb und der Instandhaltung beschreiben [4]. Diese Modelle können einen Beitrag leisten, die Informationsdichte von BIM zu erhöhen.

2.2. Industrie 4.0 als generisches Modell für die Digitalisierung der Bauwirtschaft

Ein weiterer Grund für die Kombination ist die generische Anwendbarkeit der I4.0-Konzepte. Während sich die Bild von Assets in BIM auf die Assets der Bauindustrie beschränkt, können durch I4.0 alle Geräte unabhängig ihres Anwendungsbereichs durch eine I4.0-Verwaltungsschale im digitalen Raum abgebildet werden. Ein Pumpenhersteller, der neben der Bauindustrie auch beispielsweise die Prozessindustrie beliefert, kann eine Pumpe nach dem I4.0-Standard modellieren und muss nicht für gleiche Produkte mehrere Standards abdecken. Im Gegensatz hierzu sind dem BIM-Modell natürliche Grenzen in der Erweiterbarkeit gesetzt, da es ausschließlich den Sektor der Bauwirtschaft fokussiert.

Zusätzlich bietet eine Bild von Asset-Merkmalen als I4.0-Merkmale die Basis für automatisierte Interaktionen zwischen Assets. In der VDI 2193 wird definiert, wie die Struktur von Nachrichten zwischen Assets aussieht. Zudem wird beschrieben, wie neben den Merkmalen auch Interaktionsprotokolle zwischen Assets standardisiert werden können. Grundlage der Interaktionen sind stets die semantisch eindeutig beschriebenen Merkmale der Assets. Die Richtlinie legt die Basis dafür, wie Assets durch den Austausch der Inhalte ihrer Teilmodelle interagieren [5]. Zudem bietet die I4.0-Verwaltungsschale die Möglichkeit, neben Merkmalen eines Assets auch dessen Fähigkeiten zu beschreiben und in Form von „Capabilities“ in die digitale Beschreibung abzubilden. Dies bildet die Grundlage für Self-X-Fähigkeiten [6]. Unter Self-X werden Funktionalitäten eines Systems verstanden, welche intrinsische Automatismen zur Erkundung eines Netzwerks, Selbst-Konfiguration, -Diagnose und -Optimierung ermöglichen.

2.3. Entwicklung von Teilmodellen für Pumpen und RLT-Anlagen

Zur Ableitung von Teilmodellen wurde ein generisches Vorgehensmodell entwickelt [4]. Dieses basiert zum einen auf einer prozess- und lebenszyklusorientierten Modellierung von Asset-funktionen, zum anderen auf einer Klassifikation der Assets. Auf Basis des Modells wurden exemplarisch Teilmodelle für Pumpen und RLT-Anlagen entwickelt. Um eine Eingrenzung der zu spezifizierenden Merkmale vornehmen zu können, wurden die vier Use Cases Identification, Configuration, Optimised Operation und Value Based Services entworfen, welche durch entwickelte Teilmodelle abgedeckt werden sollten. Es wurden 15 Teilmodelle entworfen, die für Pumpen und RLT-Anlagen mit Merkmalen befüllt wurden [7]. Im Teilmodell *Design*

beispielsweise werden Herstellerinformationen des Assets beschrieben. Im Teilmodell *Measurements* werden alle Messwerte physikalischer Größen eines Assets gesammelt. Der Use-Case „Value Based Services“ wird unter anderem durch das Teilmodell *Maintenance* abgedeckt. Dort werden allgemeine Informationen über die Instandhaltung der Pumpe unabhängig von der Instandhaltungsstrategie bereitgestellt.

Insgesamt liegen in den Teilmodellen für Pumpen und RLT-Anlagen 1.000 Merkmale vor. Eine genauere Beschreibung der Teilmodelle für Pumpen und des Prozessmodells zum Ableiten der Teilmodelle kann [7] entnommen werden. Diese Teilmodelle fließen über den VDMA in die Normungsarbeit ein und werden sowohl in den ECLASS Standard integriert als auch als OPC UA Companion Specifications abgebildet. Die entwickelten Teilmodelle bilden die Basis für BIM im Betrieb.

3. Prototypische Integration von Verwaltungsschalen in ein offenes BIM-Tool

Die Ergebnisse dieser Arbeit wurden an realen TGA-Anlagen des Labors für Gebäudeautomation und Regelungstechnik der Technischen Hochschule Köln prototypisch umgesetzt, um Synergien zwischen BIM und I4.0 aufzuzeigen, die sich durch eine gemeinsame informationstechnische Basis ergeben. In der Testumgebung stehen unter anderem eine RLT-Anlage und eine dazu gehörige Kühlkreispumpe zur Verfügung, die konventionell mittels Automationsstationen geregelt und gesteuert werden und deren Datenpunkte sich via BACnet in einem gemeinsamen Netzwerk befinden. Bei der Umsetzung wurden die Möglichkeiten von Verwaltungsschalen im Betrieb mit den Vorteilen von BIM in der 3D-Darstellung in einer Plattform für EM und AM kombiniert. Ziel der Plattform ist dabei eine konfigurationsfreie Einbindung vorhandener Assets und ihrer Planungs- und Betriebsdaten. Ein Ergebnis war die Möglichkeit, sich visuell durch die 3D-Planungen navigieren zu können und vorhandene Assets auszuwählen, um deren Anlageninformationen graphisch dargestellt einsehen zu können (Bild 1). Die Betriebsdaten der Assets für die Plattform werden durch implementierte Verwaltungsschalen mit standardisierten Teilmodellen zur Verfügung gestellt.

Der grundlegende Aufbau der entwickelten Anwendung ist in Bild 2 dargestellt. Da das BIM-Modell und die Verwaltungsschalen (AAS) der Assets auf demselben semantischen Standard basieren, kann die Monitoringplattform auf die Informationen aus Planung und Betrieb ohne manuellen Engineeringaufwand zugreifen. Aufgrund der noch nicht finalen Synchronisation des „bSDD“ und ECLASS, sind die entwickelten I4.0-Teilmodelle in der BIM-Methode noch nicht standardisiert anwendbar. Deshalb wurden die Teilmodelle des Use-Cases „Configuration“ im verwendeten BIM-Planungstool als Psets abgebildet. Hierdurch kann die I4.0-Semantik in die BIM-Methode integriert werden.



Bild 1: 3D-Modell der RLT-Anlage des Labors, Screenshot aus entwickelter Plattform

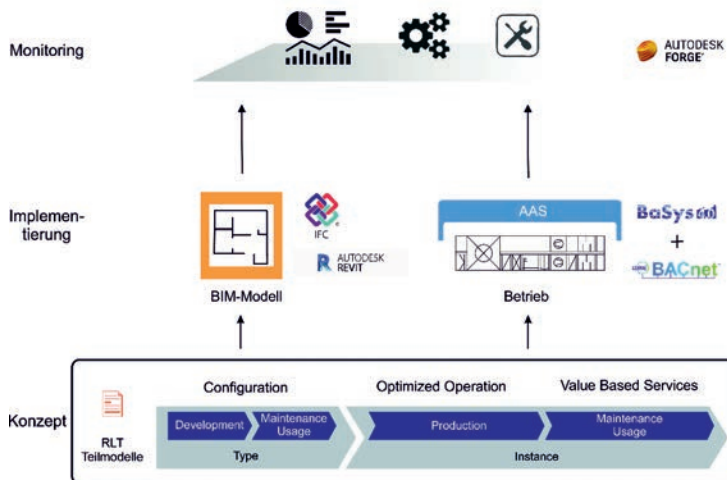


Bild 2: Übersicht der prototypischen Umsetzung mit standardisierten I4.0-Teilmodllen als Basis

3.1. Aufbau einer lauffähigen I4.0-Umgebung

Während beim Erstellen des BIM-Modells eine Bild in Psets nötig war, konnten die Verwaltungsschalen für den Betrieb mithilfe des AASX-Package-Explorers [8] konform zu den

Spezifikationen in [9] als maschinenlesbare Dateien erstellt und auf HTTP-Servern bereitgestellt werden. Hierdurch sind die Informationen der Verwaltungsschalen innerhalb eines Netzwerks erkundbar. Zur Umsetzung wurde auf die Open-Source Referenzimplementierung „Eclipse BaSyx“ zurückgegriffen, welche die Konzepte von I4.0 als Software Development Kit (SDK) umsetzt [10]. Um Zugriff auf alle vorhandenen Verwaltungsschalen im Netzwerk zu haben, wurde eine „BaSyx-Registry“ implementiert. In dieser werden die Deskriptoren und Endpoints aller aktiven Verwaltungsschalen und deren Teilmodellen gesammelt [11]. Wird eine Verwaltungsschale, bspw. einer neu eingebauten Komponente, dem Netzwerk hinzugefügt, muss diese zur Anmeldung lediglich die Netzwerkadresse der Registry kennen. Die Kombination aus Registry und den einzelnen Servern der Verwaltungsschalen ist die Basis für eine lauffähige I4.0-Umgebung. Damit die Inhalte der Verwaltungsschale mit realen Daten gefüllt werden, wurde eine Applikation entwickelt, die Betriebsdaten der Assets der I4.0-Umgebung via BACnet ausliest und in die Verwaltungsschalen schreibt.

Durch die Listung der Deskriptoren der Verwaltungsschalen in der Registry und das Bereitstellen der Verwaltungsschalen auf dazugehörigen Servern kann ein kontrollierter Zugriff auf die aktuellen Daten der Assets ermöglicht und damit die Grundlage für Anwendungen des EM und AM gelegt werden.

3.2. Automatismen im Asset-Monitoring

Die entwickelte Webanwendung integriert verschiedene Programmierschnittstellen der Plattform „Autodesk Forge“, mit der Autodesk die Möglichkeit bietet, auf diverse Bibliotheken zuzugreifen und verschiedene Templates in eigene Applikationen einzubinden [12]. Dadurch konnte das erstellte BIM-Modell, einschließlich der Planungsdaten, in die Anwendung integriert werden. Autodesk Forge wurde um weitere Funktionen erweitert, um die Verwaltungsschalen des Labors integrieren zu können. Die Semantik der spezifizierten Teilmodelle wurde in die Anwendung integriert. Da die semantischen Inhalte des BIM-Modells ebenfalls auf den standardisierten Teilmodellen beruhen, können die Planungsdaten konfigurationsfrei in die Anwendung eingebunden werden. Die Applikation erkennt zudem automatisiert neu angemeldete Assets im Netzwerk und stellt sie auf Basis ihrer Geometriedaten und Koordinaten in dem eingelesenen 3-D-Plan ohne Engineeringaufwand dar. Die implementierte I4.0-Umgebung ermöglicht es zudem, sowohl aktuelle als auch archivierte Betriebsdaten der integrierten Assets konfigurationsfrei einzulesen und auszuwerten. Die Kombination des BIM-Modells und der lauffähigen I4.0-Umgebung ermöglicht dadurch Automatismen im Asset-Monitoring. In der interaktiven Plattform kann man sich durch das 3-D-Modell der Planung bewegen und Assets

auswählen, wodurch sich ein Menü öffnet, in dem man zwischen den verschiedenen Use-Cases wählen kann (Bild 3).

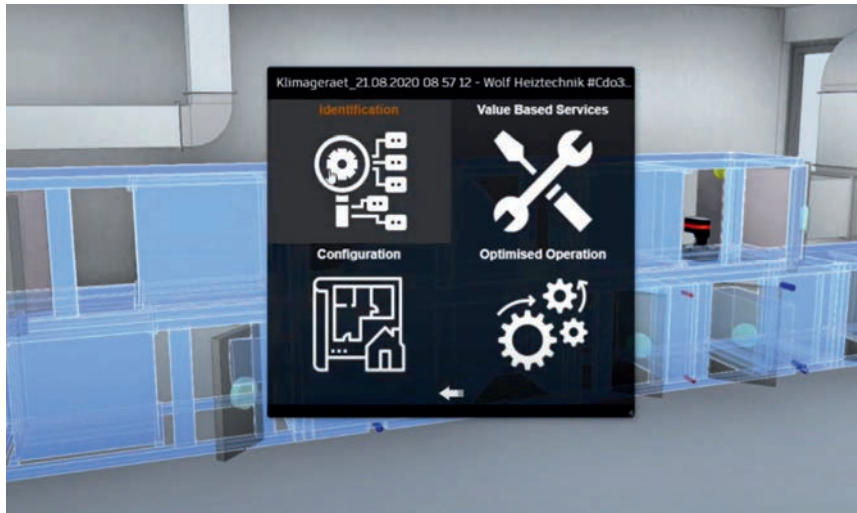


Bild 3: Die vier Use-Cases von Assets umgesetzt als verschiedene Menüs, Screenshot aus entwickelter Plattform

Im Menü zum Use-Case Identification lassen sich Informationen wie die einzigartige Asset-ID oder der Name des Herstellers anzeigen, die im Teilmodell Identification hinterlegt sind. Im Menü zum Use-Case Configuration werden die Informationen der Teilmodelle Design, System Requirements und Implementation dargestellt. Die Menüs zu den Use-Cases Value Based Services und Optimised Operation beinhalten unter anderem mit den Teilmodellen Measurements und Maintenance die tatsächlichen Betriebsdaten und deren zeitliche Verläufe. Bild 4 zeigt beispielhaft den zeitlichen Verlauf der gemessenen Lufttemperaturen der RLT-Anlage.



Bild 4: Beispiel für das Auslesen der verschiedenen Lufttemperaturen der ausgewählten RLT-Anlage, Screenshot aus entwickelter Plattform

Zusätzlich wurde ein Dashboard zum EM entwickelt, in dem verschiedene graphische Auswertungen der Energieverbräuche dargestellt sind. Es kann beispielsweise ein Graph mit dem zeitlich abhängigen Energieverbrauch, die Leistungsaufnahme in Gauges oder durchschnittliche Auslastungen dargestellt werden. Im Dashboard werden dann beispielsweise die Betriebsdaten der ausgewählten Assets kumuliert, was Möglichkeiten bezüglich der Auswertung von Energieverbräuchen bietet. In Bild 5 ist ein Beispiel für das Dashboard zum EM dargestellt, mit dem die Betriebsdaten ausgewertet werden können.



Bild 5: Beispiel für Energiemonitoring der ausgewählten RLT-Anlage, Screenshot aus entwickelter Plattform

4. Fazit

Durch den Einsatz von I4.0-Konzepten in der Gebäudetechnik werden Möglichkeiten eröffnet, die Betriebsdaten von Assets standardisiert abzubilden und in Tools nutzbar zu machen. Fehlende Konzepte für die Einbindung der Betriebsdaten in der Gebäudetechnik können mit der Verwaltungsschale gelöst werden. Durch den Einsatz der Semantik von I4.0 in der Planung von Gebäuden und der Durchführung der Planung mittels der BIM-Methode ist ein lebenszyklusübergreifender Austausch der Informationen ohne Informationsbrüche gewährleistet. BIM wird dabei unter anderem um semantische Merkmale bezüglich des Betriebs erweitert. Im Hinblick auf ein durchgängiges und dynamisches Engineering von Anlagen wird durch die Umsetzung der I4.0-Verwaltungsschale als zentrale digitale Bild eines Assets die Möglichkeit geschaffen, Merkmale während der Planungs- und Betriebsphase einzustellen oder anzupassen. Hierbei stellt die Entwicklung von I4.0-Teilmodellen für Komponenten der TGA eine Schlüsselaufgabe dar. Die Teilmodelle für Pumpen und RLT-Anlagen sind Beispiele dafür, wie das entwickelte Konzept zur prozess- und lebenszyklusorientierten Modellierung von Assetfunktionen als Basis für die Ableitung von Teilmodellen verwendet werden kann. Die entwickelten

Methoden und die prototypische Umsetzung zeigen Möglichkeiten auf, welchen Mehrwert standardisierte I4.0-Verwaltungsschalen in der Bauplanung und im Betrieb von technischen Anlagen im Gebäude bieten. Aktuelle und archivierte Planungs- und Betriebsdaten der Assets stehen Plattformen zum EM und AM automatisiert und konfigurationsfrei zur Verfügung, sobald deren I4.0-Verwaltungsschalen in das lokale Netzwerk eingebunden sind. Ein herstellerunabhängiger Austausch eines Geräts ist demnach ohne Engineeringaufwand möglich, wenn die I4.0-Verwaltungsschalen der Assets auf demselben Standard beruhen.

Literaturangaben

- [1] Both, M. u. Müller, J.: Development of Industry 4.0 models and their applicability for BIM (2019)
- [2] Bundesministerium für Wirtschaft und Energie: Details of the Asset Administration Shell. Part 1 - The exchange of information between partners in the value chain of Industrie 4.0 (Version 2.0) (2019)
- [3] Döbrich, U., Heidel, R., Hoffmeister, M. u. Hankel, M.: Industrie 4.0. Basiswissen RAMI 4.0: Referenzarchitekturmodell mit Industrie 4.0 Komponente. DIN e.V. 2017
- [4] Both, M. u. Müller, J.: Entwicklung einer Industrie 4.0 Verwaltungsschale auf Basis des allgemeinen Geräteprofils für Pumpen. In: Automation 2018
- [5] VDI/VDE 2193 Blatt 1:2020-04. *Sprache für I4.0-Komponenten. Struktur von Nachrichten*
- [6] Describing Capabilities of Industrie 4.0 Components. Whitepaper, Federal Ministry for Economic Affairs and Energy, 2020
- [7] Both, M. u. Müller, J.: Digitization of pumps – Industry 4.0 submodels for liquid and vacuum pumps. 4th International Rotating Equipment Conference (2019)
- [8] AASX Package Explorer. <https://github.com/admin-shell-io/aasx-package-explorer>, abgerufen am: 08.12.2020
- [9] Details of the Asset Administration Shell. Part 1 - The exchange of information between partners in the value chain of Industrie 4.0 (Version 2.0.1), Bundesministerium für Wirtschaft und Energie, Berlin 2020
- [10] Kuhn, T., Schnicke, F. u. Ziesche, C.: In Bewegung. Tutorial, Teil 1: BaSys und Eclipse BaSys – Middleware für die wandelbare Produktion. ix (2019) 11
- [11] Kuhn, T., Schnicke, F. u. Ziesche, C.: Gut verpackt. Tutorial, Teil 2: Verwaltungsschalen und Teilmodelle mit der BaSys-SDK definieren. ix (2019) 12
- [12] Autodesk: Forge - A cloud-based developer platform from Autodesk. <https://forge.autodesk.com/>, abgerufen am: 08.12.2020

Estimation of 5G Cell Capacity in Industrial Environments

Lisa Underberg, Sarah Willmann,
Institut für Automation und Kommunikation, Magdeburg;
Sven Kerschbaum, Stephan Höme,
Siemens AG Digital Industries, Nürnberg

Abstract

The assessment of a hybrid wired-wireless communication network prior to its deployment from an industrial application's perspective is currently challenging due to a lack of proper tools, while it is also requested by industrial end users. Addressing this lack of analytical tools, this paper presents a comprehensive simulations model covering a hybrid network comprised of PROFINET and 5G. The 5G system model is in focus. Its available data rate is calculated based on a signal quality estimation, for which two stochastic channel models are used.

The universal approach to model the 5G system is applied to three example scenarios of an automated guided vehicle (AGV) passing by one or two base stations, respectively. Two scenarios are generic, while a real-world situation is emulated in the third scenario.

The presented approach can be adapted to any wireless or wired communication technology serving an industrial application at choice. Consequently, it is an important step towards the convergence of communication networks for industrial purposes.

1. Introduction

Industrial applications are currently undergoing a paradigm change from mass production of only slightly different goods to the mass production of highly individualized goods. This entails a paradigm change in their communication networks: In contrast to legacy wired communication networks, they have to be increasingly flexible, retro-fittable and support mobile devices. In this context, wireless networks gained relevance during the last years. This is emphasized by increased national and international interest shown by research initiatives and the establishment of frequency bands dedicated to non-public networks. For example, in Germany, there are two frequency bands ranging from 3.7 to 3.8 GHz and from 24.25 to 27.5 GHz reserved for non-public exclusive use.

In future industrial communication, the traditionally deployed Industrial Ethernet networks (IENs) are likely to be combined with a wireless network. By this, a hybrid wired-wireless network is formed, which combines both technologies' advantages as needed. For example, a mobile automated guided vehicle (AGV) can be connected via the wireless network, while

static machines rely on IENs. In the hybrid network, the wireless network is supposed to allow a transparent communication between an IEN's controller and its devices.

In order to be truly transparent, the wireless network has to meet the IEN's communication requirements, i.e., by extension to meet the requirements of the industrial application which is controlled by the IEN. This interdependency is analyzed in [1].

Naturally, there is a need to assess the performance of a network prior to its deployment to determine whether it is suitable to serve a given industrial application. This need is emphasized by, for example, arising joint simulation of plants and their communication network as discussed in [2]. While there are established tools to assess IENs, e.g. Sinetplan [3], and wireless networks, e.g. OMNeT++ [4], separately, the assessment of a hybrid wired-wireless network for the industrial domain remains a challenge. This challenge is addressed in this paper by introducing a comprehensive evaluation concept covering wired and wireless network and the application the overall network is supposed to serve.

The remainder of this paper is structured as follows. Section 2 gives an overview on tools used to assess wired and/or wireless networks. The lack of tools assessing hybrid networks for industrial purposes is revealed. In Section 3, the hybrid network assumed in this paper is shown and the comprehensive simulation approach is described. Here, the 5G cell capacity is estimated. Using this estimation, the maximum capacity for an AGV passing by a base station is evaluated in Section 4, while an example scenario comprising obstacles and two base stations is investigated in Section 5. Section 6 concludes this paper.

2. Evaluation of Communication Networks for Industrial Applications

When assessing the suitability of a communication network for an industrial application, it is crucial to measure relevant parameters indicating the network's performance from this application's point of view. This approach to conduct performance testing from the application's point of view, which parameters to describe (baseline parameters) and which parameters to measure (performance parameters) are introduced in [5, 6, 7].

The hybrid network's topology which is subject to this paper is depicted in Figure 1. Here, the IEN is assumed to be PROFINET (PN) using a line topology, while the wireless network operates in multiple star topologies between base stations (BSs) and user equipment (UEs). The PN devices (PNDs) in the secondary PN networks are connected to their PN controller, which is located in the primary PN network, by a 5G system consisting of UEs, BSs and a 5G core (5GC). The industrial application served by this hybrid network is attached to the PNDs and not explicitly depicted in the figure.

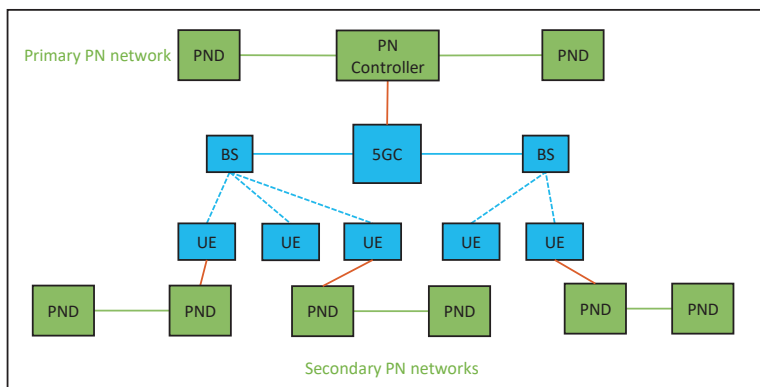


Figure 1: Hybrid wired-wireless network topology

A commonly used event-based simulation tool for communication networks in general is OMNeT++ [4]. Due to its popularity, there are open-source implementations available for various communication protocols, e. g. the INET framework [8]. In principle, the simulation of hybrid networks is possible, as long as the user is able to implement the interfaces, that are typically not included in the open-source implementation, between the different protocols. Also, the user is allowed to implement own protocols or protocols not (yet) covered by open-source software, e. g. PROFINET or certain PROFINET devices.

While OMNeT++ is a strong tool, it is not inherently able to provide a reliable upper or lower bound on the communication behavior due to its event-based approach. And since this is desired in industrial communication networks, tools like Sinetplan [3], which bases on the deterministic network calculus, are deployed. Consequently, the results Sinetplan yields are a reliable bound guaranteeing the communication network's behavior to the industrial application.

However, a wireless network can hardly be covered using an approach with a deterministic calculus. Due to this, an extension to Sinetpan is necessary in order to support a hybrid network topology.

3. Evaluation concept for hybrid wired-wireless communication networks

In accordance with the end user's need to assess a communication network from the industrial application's perspective, the approach proposed in this paper suggests a model for a wireless network, here 5G in particular, that can be integrated to a deterministic network calculus tool like Sinetplan. Therefore, the simulation tool is divided into three parts as depicted in Figure 2.

A deterministic network calculus approach relies on two major inputs: The traffic's arrival curves, which are given by the industrial application modelled in Part C, and the service curves, which are either given by the wired or wireless network. Here, the wireless network's service curves are estimated in Part A and are integrated in Part B's calculation. Part A comprises a signal quality estimation which takes the application's environment as described in Part C into account. Here, stochastic industrial channel models are used, which entails a generalization of a real-world industrial propagation environment. Based on the estimated signal quality, the 5G system's capacity is calculated. This capacity is basis to the service curves. Time behavior is neglected in the present study.

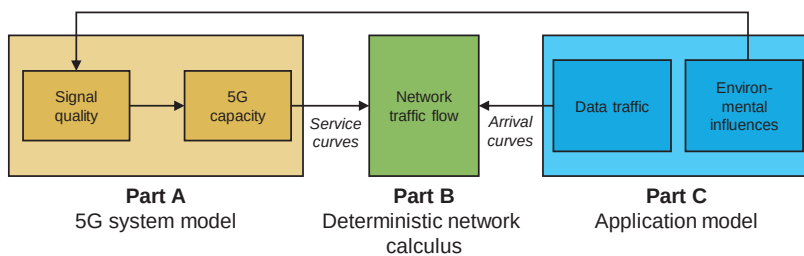


Figure 2: Comprehensive evaluation model for hybrid wired-wireless networks

The results from Part B provide an initial indication whether the overall network is suitable for the application described in Part C prior to the overall network's commissioning.

3.1 Part C: Application model

The application model is comprised of two parts. One part describes the data traffic, which is needed to determine the arrival curves used by the deterministic network calculus. Here, application-related parameters like transmit interval and user data length are provided. An example use case comprising AGVs is described in [9]. The second part describes the environmental influences, which are regarded when estimating the signal quality at the receiver's position. Here, for example, parameters like the spatial extent, trajectory along which a device moves and the position and size of obstacles is described. The environmental influences used in the scenarios in this paper are listed in Section 4.

3.2 Part A: 5G System Model – Signal Quality Estimation

In order to estimate the signal quality at the receiver's position, the overall loss a_{dB} is subtracted from the transmit power. The resulting receiver SNR is estimated taking thermal noise into account.

The overall loss a_{dB} calculated in accordance with Equation (1). The overall loss a_{dB} is the sum of the path loss $a_{\text{PL,dB}}$, i.e. the channel's attenuation caused by the distance between transmitter and receiver, the loss due to fast fading $a_{\text{FF,dB}}$, i.e. the losses caused by multipath propagation, and implementation margin and noise figure $a_{\text{FI,dB}}$.

$$a_{\text{dB}} = a_{\text{PL,dB}} + a_{\text{FF,dB}} + a_{\text{FI,dB}} \quad (1)$$

The path loss $a_{\text{PL,dB}}$ is calculated using two distinct approaches: The eta-power-law and the QuaDRiGa channel model. The fast fading loss $a_{\text{FF,dB}}$ is modelled using a Rician distribution. The implementation margin and noise figure $a_{\text{FI,dB}}$ is estimated as 15 dB. In the following, the eta-power-law and the Rician fading are parametrized in accordance with channel measurements from [10, 11].

Selected path loss model: Eta-power-law

For the path loss $a_{\text{PL,dB}}$, the η -power-law (eta-power-law, EPL), also referred to as one-slope model, is assumed as shown in Equation (2), which is valid for a distance d of at least 1 m.

$$a_{\text{PL,EPL,dB}} = \eta \cdot 10 \cdot \log_{10} \left(\frac{d}{\text{m}} \right) + 20 \cdot \log_{10} \left(\frac{f_0}{\text{GHz}} \right) - G_{\text{Tx}} - G_{\text{Rx}} + 32.44 \quad (2)$$

The value of η has to be selected according to the present environment and the employed carrier frequency. The distance d is given by the application scenario, while the centre frequency f_0 is, in the present case, given by the frequency band n78 from 3.7 to 3.8 GHz and the bandwidth B of the selected component carrier. Antenna gains of transmitting and receiving antenna, G_{Tx} and G_{Rx} , are given by the deployed antennas. Typically, for a scenario with a mobile device, its antenna gain is set to 0 dBi, while the base station's antenna might benefit from a gain ranging from 0 to 18 dBi. In order to obtain a lower bound, 0 dBi is assumed for both antennas.

In order to provide a general impression on values of η , values from channel measurements performed in industrial environments at 5.8 GHz [11] are summarized in Table 1. Note that the values from Table 1 may only be applied to the 3.7 to 3.8 GHz band to a limited extent.

Table 1: Values of η for industrial environments from channel measurements at 5.8 GHz from [11] and for free-space propagation

Environment	η
Line of sight (LOS)	1.76
Obstructed line of sight (OLOS)	1.96
Non line of sight (NLOS)	2.94
Free-space propagation	2

As opposed to an NLOS constellation, where there clearly is no direct line of sight, two cases are distinguished for a line of sight constellation. The LOS case is only valid, if the first Fresnel zone is free from any obstructions. Otherwise, the OLOS case has to be selected.

Selected path loss model: QuaDRiGa

In [12], extensive measurements in industrial environments are described, based on which a model to estimate path loss including large scale fading effects was developed. The path loss according to QuaDRiGa is calculated as shown in Equation (3).

$$a_{PL,Qua,dB} = V_{\mu} + V_{\gamma} \cdot \log_{10}\left(\frac{f_0}{\text{GHz}}\right) + V_{\epsilon} \cdot \log_{10}\left(\frac{d_{2D}}{\text{m}}\right) + X \cdot \left(V_{\sigma} + V_{\delta} \cdot \log_{10}\left(\frac{f_0}{\text{GHz}}\right)\right) \quad (3)$$

The parameters and respective values calculated based on all measurement scenarios are listed in Table 2.

Table 2: QuaDRiGa industrial channel model parametrization according to [12]

Symbol	Parameter	Proposed value LOS	Proposed value NLOS
f_0	center frequency	according to 5G system	
d_{2D}	distance between BS and UE neglecting height	according to scenario	
V_{μ}	Reference value at 1 GHz and 1 m height of transmitter	36.3 dB	29.1 dB
V_{γ}	Coefficient for modelling frequency dependency	19.5 dB	25.4 dB
V_{ϵ}	Coefficient for modelling distance dependency	18.3 dB	24.1 dB
X	Normal distributed random variable with unit variance and zero mean		
V_{σ}	Reference value of standard deviation at 1 GHz and 1 m height of transmitter	1.8	1.15
V_{δ}	Coefficient for frequency deviation dependency	-0.3 dB	3.15 dB

The normally distributed variable X is set to 1.9, which comprises 97.2 % of all measured large scale fading channels.

Selected fast fading model: Rician distribution

The fast fading impact is modelled by a Rician fading channel, which bases on a Rician distribution. For a Rician fading channel, the received amplitude's probability density function (PDF) $f(x; K, \Omega)$ is written depending on the Rice factor K and the total received power Ω as shown in Equation (5.4). $I_0(\cdot)$ denotes the modified Bessel function of first kind with order zero.

$$f(x; K, \Omega) = \frac{2(K+1)}{\Omega} \cdot x \cdot e^{-K - \frac{(K+1)}{\Omega} x^2} \cdot I_0 \left(2 \sqrt{\frac{K(K+1)}{\Omega}} \cdot x \right) \quad (4)$$

The Rician K-factor is assumed as shown in Table 3.

Table 3: Values of K for industrial environments from [11]

Environment	K
Line of sight (LOS); co-polar antenna	0.86
Line of sight (LOS); cross-polar antenna	{0; 0.49}
Non line of sight (NLOS); co-polar antenna	0.27
Non line of sight (NLOS); cross-polar antenna	{0.05; 0.57}

The fast fading loss $a_{FF,dB}$ is interpreted as fast fading margin and set to a value that corresponds to an outage probability p_{out} . To determine the outage probability, that corresponds to a given $a_{FF,dB}$, the Rician PDF is integrated as shown in Equation (5).

$$p_{out} = \int_0^{a_{FF}} f(x; K, \Omega) dx \quad (5)$$

Exemplarily, the outage probability is shown for a fast fading margin of 10, 15 and 20 dB in Table 4. For cross-polar antenna orientation, the pessimistic value is selected.

Table 4: Outage probability for different fast fading margins in LOS and NLOS configurations

Environment	$a_{FF} = 10$ dB	$a_{FF} = 15$ dB	$a_{FF} = 20$ dB
Line of sight (LOS); co-polar antenna	7.8 %	2.5 %	0.8 %
Line of sight (LOS); cross-polar antenna	9.5 %	3.1 %	1.0 %
Non line of sight (NLOS); co-polar antenna	9.3 %	3.0 %	1.0 %
Non line of sight (NLOS); cross-polar antenna	9.5 %	3.1 %	1.0 %

In the following, $a_{FF,dB} = 15$ dB is chosen implying that the estimated signal quality might be lower with a probability of about 3 %.

Estimating the received SNR

The SNR is estimated as shown in Equation (6).

$$SNR = \frac{P_{Tx,dB} - a_{dB}}{k_B \cdot 290 K \cdot B} \quad (6)$$

3.3 Part A: 5G System Model – 5G Capacity Estimation

The 5G capacity is estimated based on the transport block size (TBS) which is calculated in accordance with TS 38.214 [13]. In contrast to the rough estimation as provided in TS 38.306 [14], the TBS takes into account the modulation and coding scheme (MCS), which is selected with respect to the received signal quality. The number of available resource elements (REs) depends on various parameters, e.g. the component carrier's bandwidth and the subcarrier spacing.

3.4 Description of overall procedure

The application model provides the signal quality estimation with information on the environment such as the distance and whether there is a LOS, OLOS or NLOS scenario. This information is regarded when calculating the overall loss in the wireless channel. Using the SNR as estimated based on Equation (6), a suitable MCS is selected. The modulation order and the code rate of the MCS are regarded in the TBS calculation. Using the number of transport blocks, which is given by the component carrier's bandwidth and the subcarrier spacing, and the TBS, the 5G system's data rate is estimated.

In this paper, always the maximum data rate available using all 5G resources at a given position is calculated. This means sharing resources between multiple UEs and the corresponding scheduling mechanisms, which of course reduces the maximal data rate available to an individual UE, are not discussed.

4. Example capacity evaluation for three scenarios

At first, two simple scenarios are investigated as shown in Figure 3. Here, an AGV passes by a base station, which is located centrally in an example industrial hall. In the first scenario, there is always a LOS connection between AGV and BS. In the second scenario, a wall is assumed to be between them rendering an NLOS situation. The minimum distance between AGV and BS is 25 m.

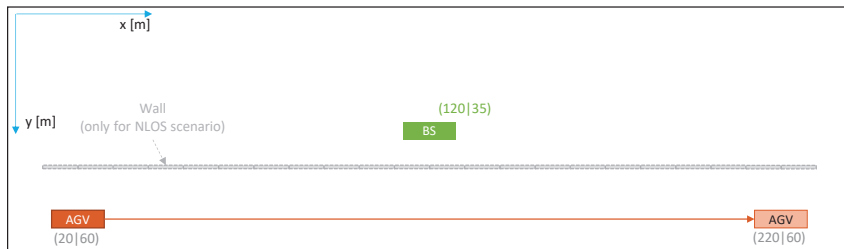


Figure 3: AGV passing by a base station as used in scenario 1 and scenario 2

The 5G system is configured as summarized in Table 5. As an AGV needs only little downlink data [9], the 5G system is configured to allocate as much resources as possible for the uplink communication. Consequently, the resulting uplink data rates are investigated.

Table 5: 5G system configuration used in the scenarios' evaluation

Parameter	Value
Component carrier bandwidth	20 MHz
Centre frequency	3.79 MHz
Subcarrier spacing (SCS)	15 kHz
Base station transmit power	20 dBm
UE transmit power	20 dBm

Scenario 1: AGV passing by with line of sight (no obstacles)

Figure 4 shows the estimated SNR for both EPL and QuaDRiGa and the resulting uplink data rates, respectively. In the LOS scenario, using the EPL results in a higher SNR throughout the entire trajectory. While the SNR continuously changes over the distance, the uplink data rate shows steps. This behavior is caused by changing from one to another MCS when exceeding or falling below a threshold SNR.

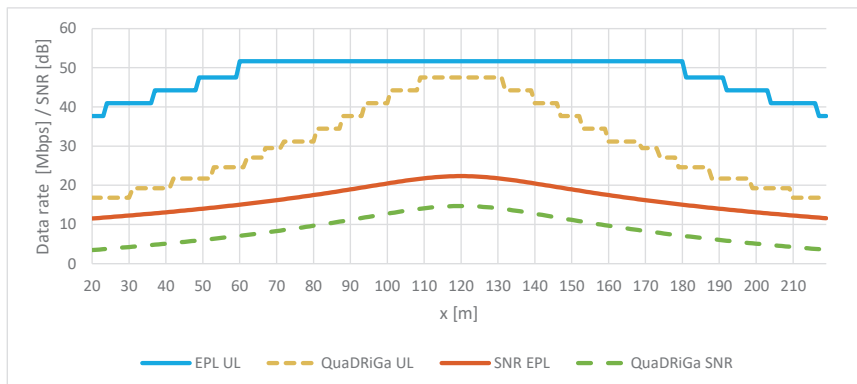


Figure 4: SNR and uplink data rate for scenario 1

Scenario 2: AGV passing by with non line of sight (with obstacle)

In the NLOS scenario, the SNR using the QuaDRiGa channel model is more optimistic than the EPL as shown in Figure 5. As a result, the data rate estimated using QuaDRiGa is higher than for EPL.

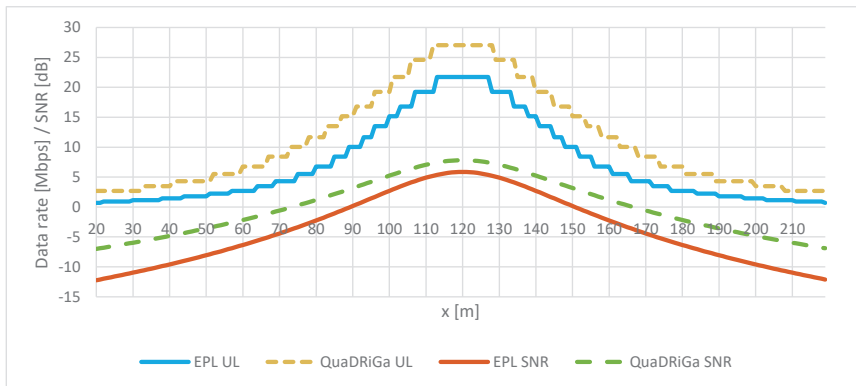


Figure 5: SNR and uplink data rate for scenario 2

Scenario 3: AGV passing by two base stations with obstacles

In this scenario, the AGV's trajectory is equal to scenarios 1 and 2, but there are two base stations and two walls. The walls induce a blocking of the LOS for several positions along the trajectory.

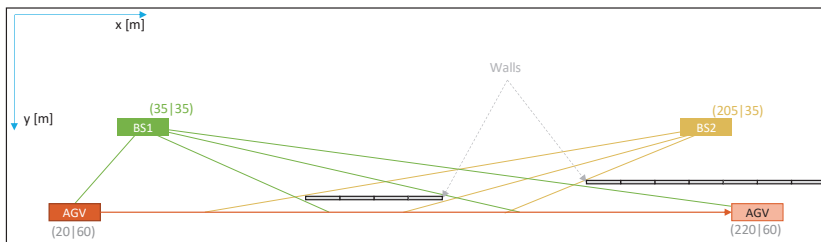


Figure 6: AGV and two base stations with two walls as used in scenario 3

The resulting data rates for both BSs is depicted in Figure 7. Here, the connection to BS1 is lost between about 145 m and 165 m due to an NLOS constellation. The same can be seen for BS2 from about 60 m to 95 m.

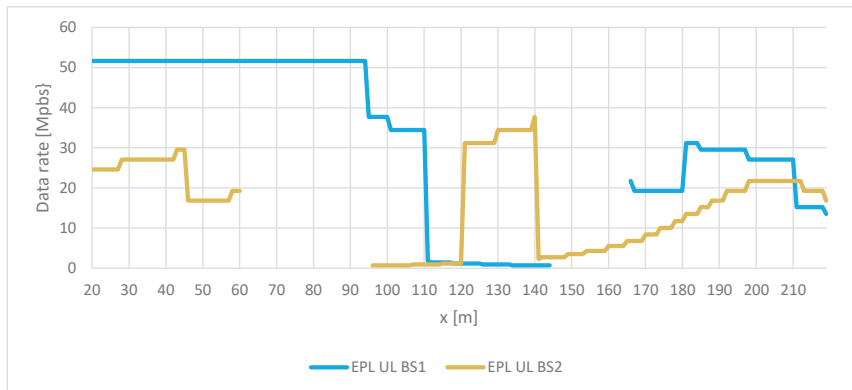


Figure 7: Uplink data rate for both BSs using EPL in scenario 3

Using the QuaDRiGa model as shown in Figure 8, there is no connection loss at all as this channel model is more optimistic than EPL. As state in Section 3.2, the EPL model offers an OLOS setting, while QuaDRiGa only differentiates between LOS and NLOS. This becomes evident when comparing the results e. g. for BS1 between 95 m and 110 m, where the EPL model yields a significant step, when the QuaDRiGa model continues to use the LOS setting.

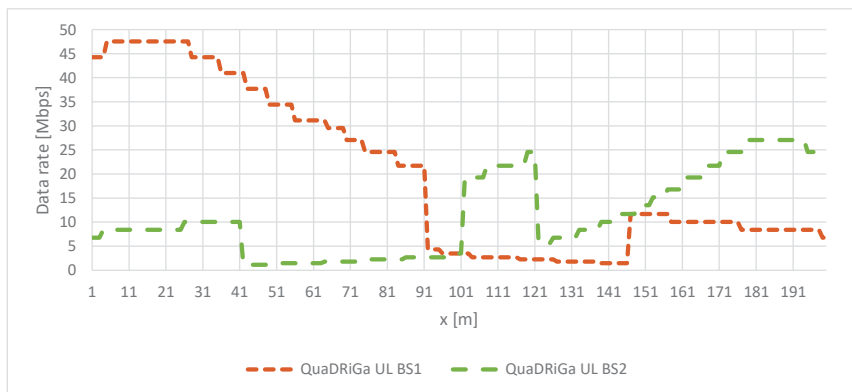


Figure 8: Uplink data rate for both BSs using QuaDRiGa in scenario 3

6. Conclusion and Outlook

This paper presented a comprehensive approach how to simulate a hybrid wired-wireless communication network from an industrial application's point of view. It was argued that the application's perspective is crucial to obtain results indicating an overall networks suitability

prior to its deployment. Addressing the lack of tools yielding these results, a simulation concept comprised of three parts was introduced.

Part A, which estimates the 5G system capacity based on the industrial application's environment, was in focus. In order to draw attention to the selected channel model's impact, two models were described and implemented. Based on the estimated received signal quality, the 5G system's modulation and coding scheme (MCS) was selected and the available data rate using all 5G resources was calculated. Results for both channel models, the eta-power-law and the QuaDRiGa channel model were presented for three example scenarios. Here, two generic scenarios and one real-world emulating scenario of an AGV passing by one or two base stations on a one-dimensional trajectory were selected.

The calculated data rates are the basis to service curves used in a deterministic network calculus approach. The second part needed to define a service curve is the timing behavior, which could be addressed in further work. Also, the signal quality estimation's accuracy could be compared to real-world measurements to validate the 5G capacity model proposed in this paper. These validation measurements could be combined with a comprehensive performance testing in accordance with [6, 7], which also captures the 5G system's time behavior, based on which it could be modelled in a service curve.

References

- [1] L. Underberg, S. Dietrich, R.Kays und G. Fohler, „Towards hybrid wired-wireless networks in industrial applications,” in *IEEE ICPS*, St. Petersburg, Russia, May 2018.
- [2] P. Vukovic, S. Höme und S. Kerschbaum, „Resource-optimized design of communication networks for flexible production plants,” in *Kommunikation in der Automation (KommA)*, Lemgo (online), 2020.
- [3] Siemens AG, „Netzwerkplanung SINETPLAN,” [Online]. Available: [HIER DER LINK....](#)
- [4] OpenSim Ltd., „OMNeT++ - Discrete Event Simulator,” [Online]. Available: <https://omnetpp.org/>.
- [5] VDI/VDE 2185 Blatt 1, „Funkgestützte Kommunikation in der Automatisierungstechnik - Anforderungen und Grundlagen für den Einsatz von Funklösungen für industrielle Mess- und Automatisierungsanwendungen,” 2019.
- [6] VDI/VDE 2185 Blatt 4, „Funkgestützte Kommunikation in der Automatisierungstechnik - Messtechnische Performancebewertung von Funklösungen für industrielle Automatisierungsanwendungen,” 2019.
- [7] 5G-ACIA, „Performance Testing of 5G Systems for Industrial Automation,” White Paper, 2021.
- [8] „INET Framework,” [Online]. Available: <https://inet.omnetpp.org/>. [Zugriff am 10 05 2021].
- [9] S. Willmann, T. Neugebauer und L. Rauchhaupt, „Praktische Aspekte zur 5G Dienstgütebewertung,” in *VDI-Kongress Automation*, Baden-Baden (online), 2020.

- [10] M. Duengen, T. Hansen, R. Croonenbroeck und et. al, Channel measurement campaigns for wireless industrial automation, *Automatisierungstechnik*, vol. 67, no. 1, 2019.
- [11] R. Croonenbroeck, L. Underberg, A. Wulf und R. Kays, „Measurements for the Development of an Enhanced,“ in *IEEE WiMob*, 2017.
- [12] S. Jaeckel, N. Turay, L. Raschkowski, L. Thiele, R. Vuohtoniemi und M. Sonkki, „Industrial Indoor Measurements from 2-6 GHz for the 3GPP-NR and QuaDRiGa Channel Model,“ in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, Honolulu, HI, USA, 2019.
- [13] 3GPP, TS 38.214: Physical layer procedures for data, V16.0.0, 2019.
- [14] 3GPP, TS 38.306: User Equipment (UE) radio access capabilities, V16.0.0, 2020.

A look inside functional safety over 5G

Dhruvin Patel, Ericsson GmbH, Herzogenrath;
Andreas Hoell, SICK AG;
Thomas Schildknecht, **Elena Eberhardt**,
Schildknecht AG, Vasco Seelman, FIR, Aachen;
Paul Becker, **Martin Stumpert**, Ericsson GmbH, Herzogenrath

Kurzfassung

Auf dem Weg zum Aufbau einer flexiblen, intelligenten Fertigung ist die Nutzung drahtloser Kommunikationstechnologien wie 5G ein wichtiger Baustein für die Fertigungsautomatisierung. Hierbei ist auch das Thema der funktionalen Sicherheit zu betrachten, insbesondere da sich hiermit neue Lösungsmöglichkeiten für eine effizientere Produktion ergeben. Im folgende Bericht stellen wir hierzu die untersuchten Schritte zur funktionalen Sicherheit über 5G dar und die neuen Möglichkeiten, welche sich daraus ergeben.

Ausgehend vom etablierten „Black Channel“-Prinzip in heutigen sicheren industriellen Feldbussen wie PROFIsafe und CIP Safety, werden diese Anforderungen für ein industrielles 5G-Campus-Netzwerk betrachtet. Insbesondere im Kontext der Anlagenverfügbarkeit und damit des sinnvollen Einsatzes von 5G im Vergleich zu bestehender industrieller WLAN (IEEE 802.11xx) Kommunikation.

Die Ergebnisse zeigen, dass 5G niedrigere Varianzen in den Zeiten für den Datenaustausch erzielt und deterministisch arbeitet, im Vergleich zur typischen WLAN Kommunikation.

Umgesetzt wurde dies am Beispiel eines fahrerlosen Transportfahrzeuges am 5G Industry Campus Europe. Mittels der aus der Sicherheitsautomation bekannten Stopp-Zeiten konnte der erforderliche Sicherheitsabstand beim 5G-Einsatz aufgezeigt werden. Damit ergibt sich für diese Anwendung, die Möglichkeit einer Effizienzsteigerung im Materialfluss. Der Vortrag endet mit einem technologischen Ausblick auf 5G und der Relevanz für die industrielle Fertigung. Dieser Beitrag ist Teil einer Initiative für eine gemeinsame Sitzung zur Information über Testbeds für industrielle Funksysteme. Die Untersuchungen wurden am 5G-Industry Campus Europe, dem größten industriellen 5G - Forschungsnetzwerk in Europa durchgeführt.

Introduction

Moving forward towards digitization of the production system, more and more sensors and actuators deployed on the shop floor and new manufacturing applications are foreseen by

introduction of new communication technology such as 5G. This also includes emergency stop buttons utilized for the functional safety applications. With adoption of the 5G being foreseen in smart manufacturing, it is important to understand, how to deal with functional safety in such new environment as 5G system becomes important part of industrial communication infrastructure. For functional safety communication today, the industrial wired network is assumed as black channel. The black channel is well established principle for the functional safety related communication independent of the communication technology.

This paper focuses on functional safety communication which is typically realized over wired industrial networks and also, there is a growing trend observed to use wireless technology operating in unlicensed spectrum. Use of 5G based communication as an add on to wired Ethernet based field busses, will increase the flexibility and productivity.

Functional safety and wireless

Safety communication in automation

Industrial communication technology plays an important role in ensuring seamless operation of the functional safety application. To protect people and environment in the production facilities, safety functions are deployed. These safety functions consist of the safety sensors, safety controller and safety actuator. IEC 61508 is a basic functional safety standard which defines a probabilistic failure approach for impact of the safety on device failure and also specifies the contribution of the failure rates of each component.

Safety over wireless

Safety protocols for example CIP Safety or PROFIsafe based upon the standard industrial protocol are realized considering communication technology underneath as black channel for the industrial Ethernet network. It is completely independent from the base transmission channel and works equally well over wired and wireless communication links. A "Black Channel" may comprise several types of transparent network components such as switches, routers, links, and wireless transmission channels. Safety protocols aims at reducing the residual error probability of the data transmission between a safety controller and a safety device to the level required by or better than the relevant standards such as IEC 61508.

Various errors may occur when messages are transferred in complex network topologies, whether due to hardware failures, extraordinary electromagnetic interference, or other influences. As safety systems are real-time systems, process signals must not only be delivered correctly, but also in time. In case of timeliness errors, the safety system will initiate a safe

reaction, e.g. safely stop the movement of a drive. For this purpose, safety devices utilize a watchdog timer that is restarted whenever a new valid safety message arrives. The 1:1 relationship between the Safety controller and a safety device facilitates the detection of misdirected and corrupted message frames.

Response time

Typical safety function response time calculation is based on the input safety device, a safety fieldbus transmission, signal processing in a safety controller, another safety fieldbus transmission, and an output safety device, each with its own statistical cycle times [0]. The maximum time for a safety signal to pass through this chain is called Total Worst-Case Delay Time (WCDT) considering that all parts require their particular maximum cycle times.

5G as preferred black channel

5G is a promising wireless communication technology to support a wide range of industrial automation applications. With release 16 of 5G specification from 3GPP (3rd generation partnership project), new features are introduced to enable support for the smart manufacturing use cases. The 5G-ACIA report [0] provides insight to integration of the 5G with industrial ethernet networks. Furthermore, 5G deployment architecture along with different technical enablers for the reliability and latency are elaborated in the report [3][0].

Concerning functional safety communication over a 5G system, we can assume the 5G system and data network comes under black channel as shown in the **Fehler! Verweisquelle konnte nicht gefunden werden..** It should be noted that data network is not a part of 5G system. For functional safety applications considering 5G system as black channel, it will include end user equipment (UE) connected to safety sensors or safety actuators. The base station is a part of the Next Generation-Radio Access Network (NG-RAN) which supports seamless devices mobility, access control, Quality of Service (QoS) realization and other essential features for radio communication between base station and UE. Considering 5G systems as a black channel from the outside, the internal structure of the 5G system is not visible to functional safety systems. Table 3 highlights the main requirements imposed on the 5G system to support functional safety communication. Along with deterministic bounded latency which is defined based on the response time of the safety function, security and availability are also seen as important requirements. The 5G system capabilities are highlighted which satisfy such requirements. 5G can support functional safety applications if all capabilities listed below are realized on the deployed network. The User plane function (UPF) as part of the core user plane as shown in Fig. 1 below directly interacts with data network. In case of the functional safety application, it

can either interact directly with safety applications or via an industrial ethernet network. Further it is clear that the 5G system being a communication infrastructure operating as black channel does not need to be certified for safety procedure.

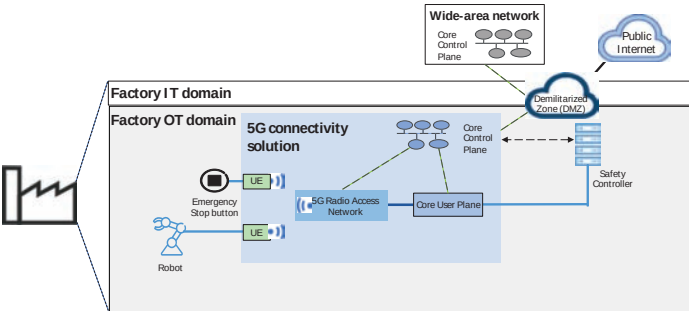


Fig. 1: 5G deployment architecture for the functional safety communication (source:Ericsson)

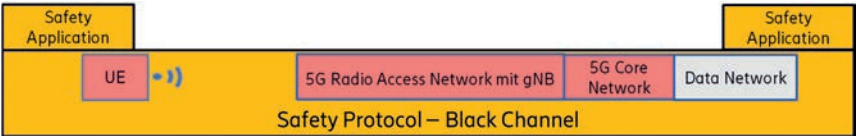


Fig. 2: 5G System as black channel

Table 1: 5G characteristic to support functional safety requirement

Requirements	Characteristic	5G system features/capabilities
Safety function response time	Required Safety function response time depends on the application and is calculated in risk assessment during design phase.	Deterministic latency technical enablers in 5G radio and core network including [Fehler! Verweisquelle konnte nicht gefunden werden.] : - QoS framework - New Radio (NR) toolbox for low latency and low latency variation - Core Network low latency mechanisms
Availability	According to IEC, the ability to be in a state to perform as required. In terms of the communication system, a predictable deterministic performance during the lifetime of the automation system and within the desired coverage.	5G technical enablers for redundancy and radio network design to ensure guaranteed coverage and capacity.
Security, unauthorized safety message modification	Data integrity and authentication between two end nodes	5G support for data integrity of user plane data. 5G authentication mechanisms between UE and network for non-public networks [0].
Integration with industrial ethernet network	Integration of the Industrial network with different communication zones for different purposes (e.g. dedicated zone for functional safety system)	5G support to transport ethernet frames. Network slicing 5G support for VLAN integration and configuration.

Safe crossing use case

Motivation

Transportation times in intralogistics still are one of the most common and critical time components to cause delay, by inefficient fleet planning, unknown obstacles on track and safety behavior of automated and manual guided transportation vehicles (e.g. AGVs, forklifts).

AGV's forklifts are able to run high speed in ideal environment (e.g. modern Warehouses) to move goods from A to B and they are equipped with several safety laser scanners to monitor the environment and to ensure safe and accident-free operation with regard to humans and many other obstacles.

In mixed areas with workers, unknown objects and maybe intersections, AGVs and forklifts must slow down near every intersection. Currently, an AGV decelerates while approaching a blind spot (Intersection) to adapt the required braking distance. This ensures that the AGV can avoid a collision even with objects suddenly entering the monitored safety areas. While this approach ensures safe operation even at blind spots and other critical warehouse black spots, it is inherently inefficient regarding mission time, energy consumption and mechanical wear of the AGV. To solve this task more efficient a safe communication between AGV and his environment (e.g. machine in docking operation, other AGV's, traffic light and gates) is needed.

To get a better understanding of 5G capabilities, the use case safe crossing was realized in an industrial realistic environment in the Demofactory of Aachen as proof of concept.

The goal was to increase operation efficiency due to higher average intralogistics speed, with same level of safety to avoid collision of AGV's with workers or any other obstacle.

To realize this, the AGV's integrated environmental recognition capabilities, as well as additional sensors in infrastructure like intersections to detect objects in critical (for AGV undetectable) areas, are combined. For this proof of concept, a reliable communication via the wireless network is necessary. The system approach was tested by tunneling CIP-Safety, the safety communication protocol for DeviceNet and EtherNET/IP protocols, through a 5G and an IEEE 802.11 WLAN infrastructure, using the black channel principle. The real benefit of the reliable wireless communication is obvious by calculating the application relevant stopping performance of the AGV.

The proof of concept included following components:

- A safety laser scanner in the infrastructure to monitor the intersection (4)
- A safety PLC on the AGV (1)
- A 5G isolated non-public network and
- Communication via 5G, by tunneling of the safety related protocol (Ethernet/IP CIP Safety)

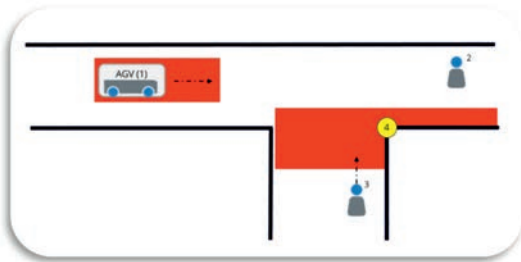


Fig. 3 Intersection with AGV infrastructure component and human operator

Use case description

Fig. 3 shows a typical industrial situation, with AGV and human operators on factory shopfloor performing task independently and simultaneously. Today, based on the safety sensor onboard of the AGV detects the human operator (2) when approached in front of AGV. This results in slowing down by reducing operating speed of AGV (1). The slowing down of the AGV is based on the maximum stopping time of AGV i.e. time to detect the person and stop AGV before collision.

With the introduction of reliable communication infrastructure such as 5G, additional safety sensors can be deployed on the shopfloor which can be further be utilized to gather more situational awareness information e.g. movement of the human operator (3) at the Intersection. The main goal here is to calculate the minimum distance showcasing hazard points at which functional safety application on AGV activate the safe stop of the AGV, resulting being set into safe state. In this scenario minimum distance is very crucial parameter to enable such functional safety loop with safety sensors on shopfloor.

5G isolated non-public network (NPN) deployed on the shopfloor is utilized to connect safety PLC on the AGV (1) with safety laser scanner (4) deployed on the infrastructure to monitor the intersection. Below method to calculate minimum distance is shown

Result: minimum distance comparison

The minimum distance to the hazard zone is calculated accordingly safety standard IEC TR 62998-2:2020 04 and is strongly influenced by the availability of the wireless communication technology. In the case of the WLAN IEEE 802.11 monitoring time was set to 2500 ms to ensure availability of the system. While in the case of the 5G, it was set to 110 milliseconds.

Calculating the minimum distance for the safety zone, we came to a difference of 3,8 m between the two wireless technology 5G (1,6 m) and IEEE 802.11 WLAN (5,4 m). Hence, 5G communication enables the application in industrial environment and more efficiency can be achieved compared to WLAN.

PROFISAFE over 5G test setup

To further investigate the performance of functional safety over wireless infrastructure, laboratory experiments were performed with a 5G network using industrial equipment. This investigation focuses on the performance of the widespread safety protocol PROFINET using PROFIsafe traffic and provides a comparison to legacy technologies like IEEE 802.11 WLAN.

All tests were performed at the 5G-Industry Campus Europe in Aachen¹. This campus network is a research only private network covering several indoor shopfloors of different research institutes and outdoor areas for dedicated testing of industrial use cases. The shopfloors at RWTH WZL institute, Fraunhofer IPT are equipped with heavy machinery and mimic a real-world environments of a manufacturing shopfloors. The test setup considered for this experiment mainly consists of three components: the 5G Communication infrastructure, Schildknecht Dataeagle devices which connect PLC and IO devices via the 5G network and Siemens S7 PLC and IO device running a functional safety PROFIsafe application.

Schildknecht Dataeagle

The Schildknecht Dataeagle is a device which interconnects PLC and IO and enables PROFINET communication between these two devices. The Dataeagle enables conversion of the layer 2 PROFINET frames into layer 3 IP packets. It also performs preprocessing of PROFINET fieldbus telegrams to mimic behavior of fieldbus over wireless connection applied between the two nodes to meet the protocol requirements (e.g. transfer timing requirements, number of retransmissions etc.). Furthermore, it maintains the state of the PROFINET protocol by temporarily storing PROFINET telegrams and performs algorithms to separate hard cable-based timing requirements from the timing over wireless networks. Without such a procedure fieldbus errors will occur causing machine downtime whenever a packet arrives at its destination outside of the PROFINET and PROFIsafe timing window.

¹ <https://5g-industry-campus.com/infrastructure/>

Communication infrastructure

The testbed setup at 5G-Industry Campus Europe was utilized as communication infrastructure for the 5G measurements. The infrastructures on each site of the 5G campus network contain all nodes required to establish a full state-of-the-art 5G network. The 5G non standalone (NSA) network is operating on band B40 (2,3 GHz) as anchor band and uses NR band N78 (3,7 GHz) with a bandwidth of 100 MHz for data transfer. On the mobile end side of the test setup, a Qualcomm X55 Chipset based industrial router was used to establish the wireless network connectivity. The current 5G infrastructure is based on the release 15 of the 3GPP 5G specification. Also, the infrastructure is equipped with 5G RAN and Core Network which is not optimized for the ultra-reliable low latency (URLLC) communication.

The Industrial PC (IPCs) in this setup are enabled with functionality to terminate the L2 tunneling since this cannot be done directly on the UE (user equipment) yet. Being configured as a network bridge, it provides an easy-to-handle mobile network interface to the connected UE and Dataeagle device. To elaborate, 5G system deployment currently supports IP connectivity only. The complete test setup configuration is visualized in the Fig. 5 below. Measurement data can be extracted directly from the Schildknecht Dataeagle by using the Dataeagle integrated USB maintenance and configuration interface, especially a diagnostic function inside Dataeagle is utilized to measure the latency of the communication link.

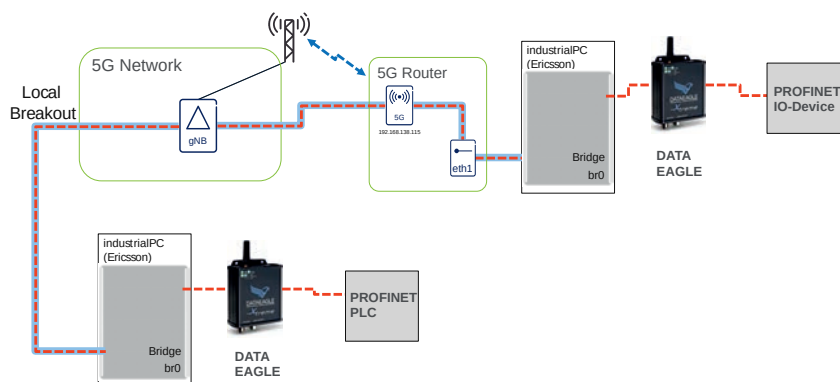


Fig. 4: Test Setup

Test cases

UE radio conditions on the shop floor are observed and it demonstrated excellent quality. The PLC PROFINET cycle time is set to 1 ms, which means that every millisecond a PROFINET packet is generated by the PLC to be transmitted to the I/O device, via the Dataeagle interface. A common Time Division Duplex (TDD) pattern was used for the 5G radio interface with PROFINET cycle time of 1ms. No optimization of the 5G radio access network have been conFig.d, like e.g. prescheduling, which can further optimize the network for ultra-low latency.

Data Exchange time

The main key performance indicator (KPI) considered for the performance investigations was the Data Exchange (DX) time. The DX time can be understood as application exchange time. PROFIsafe define a safety timer (watchdog time). DX time should always be less than the safety timer, if the value is exceeded a safety system will go in safe state. Dataeagle performs data pre-processing and data reduction to ensure PROFINET cycle time. As shown in Fig. 6, for this particular experiment, data exchange time is measured at IO side.

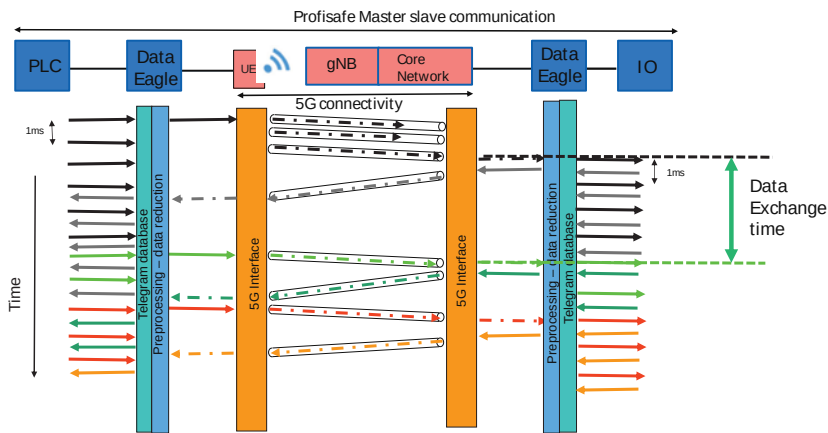


Fig. 5: Sequence flow between different components (Source: Schildknecht AG)

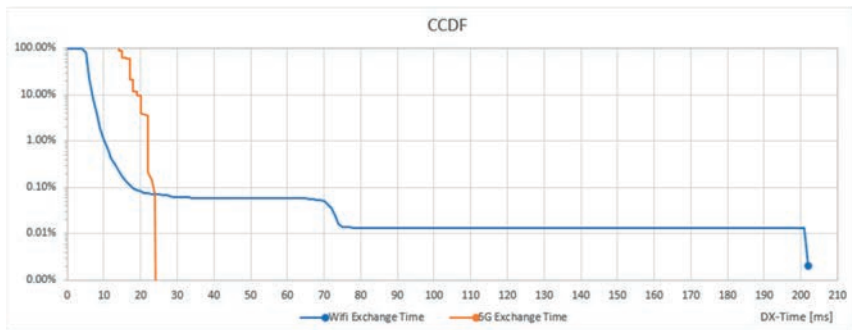


Fig. 6: Data Exchange time observed with 5G and IEEE 802.11 WLAN (Source Ericsson)

Fehler! Verweisquelle konnte nicht gefunden werden. shows the complementary cumulative distribution function of the measured data exchange time for an exemplary set of PROFIsafe data transmitted via different wireless communication systems (5G and WLAN). Looking at the visualization of the 5G transmitted data, the comparatively low variance of the exchange time can be confirmed as expected. The most important transmission characteristic however is the maximum transmission time, which is evaluated by the PROFIsafe protocol to determine the connection status. Even when measured over a long period of time, 5G does not show any outliers in the distribution graph, resulting in a maximum transmission time of 25 ms in the observed dataset compare to IEEE 802.11 WLAN which goes over 200 ms. A comparative measurement with PROFINET operating either over IEEE 802.11 WLAN in the 2.4 GHz band or over the 5G test setup described above, shows that both 5G and IEEE 802.11 WLAN fulfill the requirement of the functional safety communication. The measurement for IEEE 802.11 WLAN was performed in the shielded laboratory without any other interfering radio systems. This can be seen from the fact that 100 % of the telegrams were received. The latency in WLAN is observed lower than 5G due to the fact that no co-existence mechanism where in place. However, in real-manufacturing environment, higher latency in IEEE 802.11 WLAN will be observed as 802.11 WLAN need to co-exist with other license-exempt communication technology.

In practice, all commercially available IEEE 802.11 equipment operates in license-exempt spectrum whereas 5G equipment operates in various licensed bands. In licensed bands, however, no coexistence mechanisms are necessary that limit the performance IEEE 802.11 in license-exempt spectrum. In Europe, Harmonized Standard (HS) EN 300 328 defines the rules

for operating in the 2.4 GHz ISM (Industrial, Scientific, and Medical) band. As this band is accessible to all users and various technologies, different means of coexistence exists. IEEE 802.11 has been designed for operation in such license-exempt spectrum. The coexistence mechanism employed by IEEE 802.11 is Listen-before-talk (LBT) with truncated exponential backoff. According to this mechanism, the radio channel is determined to be unoccupied if no signal above certain thresholds is present. Then, an IEEE 802.11 device waits for a random duration to reduce the likelihood of two or more devices transmitting simultaneously. The latter is denoted a collision that renders both transmissions unreadable and requires them to retransmit. For the measurement results it can be deduced that in practice there is an unpredictable influence on the IEEE 802.11 WLAN latency leading to non-deterministic behavior. This is the starting point for 5G campus networks. Here there is no disturbing external influence and 5G does not have to perform coexistence measures and the latency is predictable and deterministic.

Conclusion

With the digitization of the smart manufacturing ecosystem, new applications are rising with a need for functional safety communication to be realized over wireless technology. 5G is seen as prominent enabler for such type of communications. Considering a rise in the AGV adaption for many advanced manufacturing applications. With 5G, more and more information between AGV, central control system and sensors attached to infrastructure can be gathered in real-time. Further from experimentation, it is shown that 5G fulfils requirement for the functional safety communication. This enables higher efficiency operations in terms of AGV speed and reduction in the cost for computation on AGV. From the use case being described above in the section, it is already observed that 5G technology proves to be reliable enough, parts of the AGV control can be placed into an edge cloud and with environmental sensing from the infrastructure, sensors could be replaced from the individual AGV. This would essentially transform the AGV into a remote-controlled vehicle enabling new use cases. In this case, it would enable the use of emerging technologies, such as machine learning based navigation.

These result in reducing cost of the individual AGVs, increasing scalability and performing easier upgrades. At the same time higher investments in terms of the capital expenditure and operating cost of infrastructure need to be investigated. The challenge here is to analyze the cost of the AGVs which is largely driven by controls, sensors and actuators built into the vehi-

cle. From history point of view, 20 years ago, German Industry Community (VDI) started thinking about adapting wireless communication for manufacturing. After two decades, we have advanced 5G wireless communication with spectrum licenses for industrial segment in Germany.

Reference

- [1] Integration of Industrial Ethernet Networks with 5G, 5G-ACIA, May, 2020
- [2] PROFIsafe System Description – Technology and application – Chapter 1, Profibus & PROFINET international (PI), April 2016
- [3] Dhruvin Patel, Joachim Sachs, "5G E2E Technology to support Verticals URLLC Requirement", NGMN Alliance, February, 2020
- [4] Dhruvin Patel, et. al 5G-SMART deliverable D5.1, "First report on new technological features to be supported by 5G standardization and their implementation impact", May, 2020. <https://5gsmart.eu/deliverables/>
- [5] Security Aspects of 5G for Industrial Networks, 5G-ACIA, May, 2020
- [6] Exposure of 5G Capabilities for Connected Industries and Automation Applications, 5G-ACIA, June 2020
- [7] ETSI EN 300 328 v2.1.1 Wideband transmission systems; Data Transmission equipment operating in the 2.4 GHz ISM band and using wide band modulation techniques; Harmonised Standard covering the essential requirements on article 3.2 of Directive 2014/53/EU

Bridging the Gap in Digital Systems using Automated Model Transformation Concepts

Dr.-Ing. **Abdulkadir Karaagac**, Dr.-Ing. **Johannes O. Schmitt**,
Dr. rer. nat. **Dirk O. Schulz**, ABB Corporate Research, Ladenburg

Kurzfassung

Für ein wettbewerbsfähiges digitales Serviceangebot müssen Daten aus verschiedenen Geräten und Steuerungssystemen zugänglich und verständlich sein. Heutzutage nutzen die von diesen Systemen gesammelten Daten sehr unterschiedlichen Datenmodellen, sodass die Anwendungsentwicklungsprozesse auf die manuelle Integration verschiedener Informationsressourcen mit unterschiedlichen Technologien und Datenmodellen angewiesen sind. Dieser Übersetzungs-/Integrationsaufwand ist ein schwieriger und zeitaufwändiger, welcher das Wissen von Systemexperten und Anwendungingenieuren erfordert. Außerdem ist der Prozess oft fehleranfällig und wiederholt sich für jede neue Anwendung. Um dieses Problem zu lösen, stellt diese Arbeit einen bedarfsbasierten Ansatz vor, bei dem die Anwendungen den Informationsbedarf als *Demand* (Anforderung) gegenüber unbekannten, entkoppelten Informationsquellen beschreibt. Dieser Demand kann dann automatisch von einem zwischengeschalteten Informationsbroker gelöst werden, um die Daten wie angefordert zu liefern. Dieser Ansatz ermöglicht es Web-/Applikationsentwicklern, neue Anwendungen auf einem bestehenden Automatisierungssystem effizient zu realisieren oder bestehende Anwendungen zu verändern, ohne tiefgreifende Kenntnisse über das Automatisierungssystem zu besitzen.

Abstract

A competitive digital service offering requires data from various equipment and control systems to be accessible and understandable. Today, data collected from these systems follow very diverse data models, thus the application development processes rely on the manual integration of different information resources with varying technologies and data models. As this process requires hours of translation and integration effort of system experts and application engineers, it becomes difficult and time-consuming to prepare for data-driven applications of today. Additionally, it is a very costly and error-prone process that repeats for every new application. In order to address this problem, this work presents a demand-based approach where the applications describe the information demand towards unknown, decoupled information sources and then these demands are automatically resolved by an interme-

mediate information broker to deliver the data. This approach will allow non-expert engineers (web/application developers) to realize new applications on an existing automation system efficiently or alter existing applications without in-depth knowledge about the automation system. Therefore, it accelerates application development processes and facilitates the information integration operations.

1. Introduction

With Industrie 4.0, the industrial world is experiencing a strong need to digitalize devices, systems, and processes [1]. Toward this goal, seamless interoperation of devices and software services and the easy integration of data is one of the topics that is currently being addressed by industry. On the field level, emerging standards like OPC UA [2] enable devices to be equipped with a standardized type system (aka companion specifications [3]) and meta-data that allow for a machine-readable understanding of the provided data. From the cloud perspective, concepts like digital twins and Asset Administration Shells [4] are being proposed for the harmonization and standardization of the virtual device representation. Meanwhile, domain-specific standards and exchange formats constitute the fundamentals of engineering-, SCADA- or analytic-tools aiming application development. Although these standards and concepts provide high potential for automated data integration, due to the large diversity of devices and systems and their heterogeneous set of protocols, there is still a certain "semantic gap" between existing technology and the vision of seamless interoperation that needs to be addressed.

Therefore, in this work, we propose an automated scheme for the application of model-transformation mechanisms as part of an information-centric architecture to overcome this semantic gap. The proposed solution considers every model-transformation operation as a black-box or function-block that can be made available by storing these model transformations in a transformation-library (public- or company-wide) and re-used whenever similar models or types must be converted to each other. Having applications and devices that expose their provided model-types (e.g. in OPC UA), it is possible to automatically decide which model-transformation can be applied. Therefore, the proposed solution can automatically resolve the semantic gap between companion specifications provided by devices towards digital twin concepts and towards (view-)models of targeted applications.

In addition, considering the fact that many data integration tasks are to some extent similar and reoccurring over different installations, there is a high potential for automation and re-use of transformation mechanisms for data integration resulting in a reduction of efforts - as well as faster and potential less error-prone execution of data integration tasks. In addition to the

concept description and solution design, this work also includes a proof of concept implementation of the proposed solution and validation of its' potential via a prototype based on a cloud platform and a real use case of a Gearless Mill Drive (GMD).

The remainder of this paper is structured as follows: Section 2 describes the addressed problem in detail, while in Section 3 we briefly introduce the proposed concept and describe the proposed integration solution. After that, Section 4 further elaborates on the functional blocks and key enablers of the proposed solution, which is followed by the description of the Automated Integration Workflow in Section 5. Finally, summary and conclusions are provided in Section 6.

2. Problem Description

Typically, in an application development process, the information consumer (e.g. a dashboard or analytics application) has no or only limited insights to details/context of the information provided by the information source (e.g. control systems, smart devices). In many cases expert knowledge and translation effort is required to fill the existing semantic gap (e.g., for adapting model structure, variable names, or inconsistencies) and integrate the data provided by the information sources with the view models required and known by the application. Unfortunately, these manual steps are typically very costly, error-prone, and repeating for every new application. That means extensive amount of efforts spent on manual configurations and hard coded transformations by System and Model Experts.

For instance, as presented in Fig. 1, CRISP-DM [5], which stands for Cross-Industry Standard Process for Data Mining, envisions the efforts for data integration typically requires a noticeable portion (~25%) of the overall efforts. Some other studies even claim that this process might constitute up to 50-80% during a typical data science project [6].

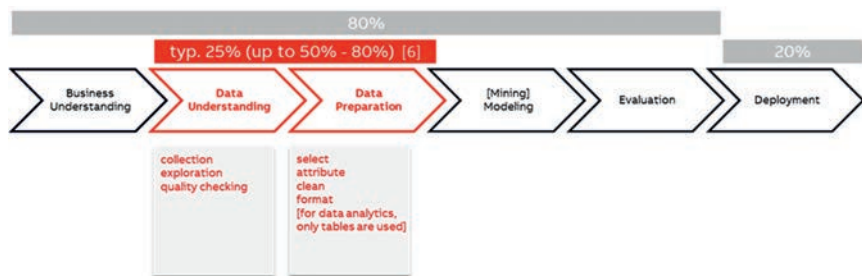


Fig. 1: Efforts for data integration in CRISP-DM [5]

Besides new applications, an in-depth expert knowledge about the automation system would be required for extending or altering existing applications, which makes such a process very complicated, costly and time-consuming.

In a typical workflow of a manually performed system integration process, as presented in Fig. 2 (simplified), there is a need for a multiple integration step that needs to be taken by a number of roles: Application Developer, Model Expert and System Expert. Initially, the Application Developer documents the view/target model of the application, which specifies the requested information from the information resources. Similarly, a System Expert, who must understand all application and system-specific documentation, has to setup the system and conFig. all components in the right way. Additionally, a Model or Domain Expert that understands the context (business) of the use-case must explore the data to interpret the human readable documentation and understand the meaning of the data. Then, this expert uses data integration tool and performs integration steps to prepare the data. This preparation is required to "integrate" the input data with the target application. In this step, the input data is reformatted, cleaned up, filtered in order to suit the needs (i.e. the view model) of the target application. This manual integration effort might slightly divert in different domains or from application to application, but similar tasks are part of the integration processes in the domains like data analytics, solution development or engineering of modular plants.

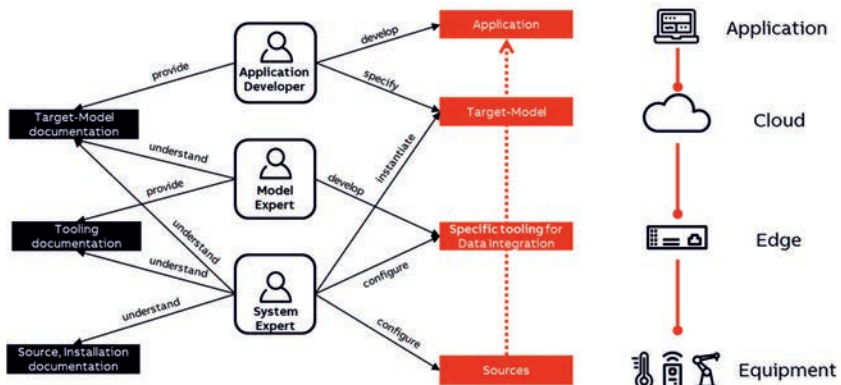


Fig. 2: Typical steps in a manual integration workflow

In this context, although there are emerging concepts (e.g. in OPC UA) in industrial domain, which are trying to simplify these processes to some extend by providing interfaces that al-

allows a client to browse through the set of available information to identify and pull or subscribe to relevant data, these concepts are not addressing the problem of the semantic gap towards the application view models and semantics known to the application developer. Therefore, there is still a clear need for an automated transformation solution that can facilitate the data integration operation and thus speed up the application development and extension process.

3. Concept

The main idea behind the proposed solution is that the whole information provisioning can be self-organizing based on demands formulated by the consuming application that finally knows and is able to define what data is required (and this often late – sometimes even during runtime). From the top-down view of the application, the application should be able to “just to search and use” information. The whole information acquisition, preprocessing, transformation, and delivery should be like a black-box for the application. Fig. 3 provides a simple representation of this demand-driven information provisioning and data integration process.

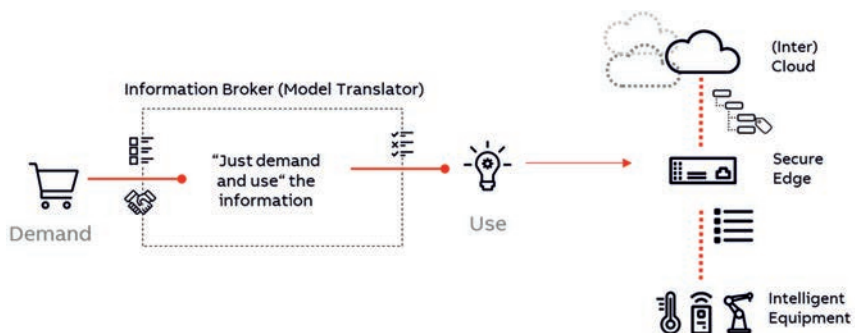


Fig. 3: Demand Driven information provisioning

Typically, the application knows only the data (output) that is required and (optionally) also a pointer to the required Source(s) to be used as Input. But it does not know how to get (from the input) to the output – also no knowledge about existing transformations and what type of extra transformation shall be required. For that reason, the application provides all the relevant knowledge as *Demand*. Within such a *Demand*, it should be possible to identify what information the application is searching for and how it is going to use it. Therefore, a *Demand*

should specify at least the source (e.g. which system or device to consider) and the target/view model that is understood and required by the consuming application.

On the other side, specific function-blocks are used - named *Transformation*. They basically specify input model (to declare that only instances of that model or submodel can be processed), output model (to declare the model of the instances generated by this transformation), optionally Instance Filter (to filter instances by e.g. ID, name, parameter-value) and structural aspects (e.g. following references, child of). Internally, such Transformation holds the transformation steps/tasks that shall be executed (e.g. remove, add, rename, transformation commands with conditions, loops, programmatic concepts with access to models/model-structures. The input/output of a Transformation can be one or multiple information models (incl. references between those or other information models).

The technology used to implement the transformation steps inside of a Transformation is not fixed or limited – it is just required to implement the interfaces of the surrounding function-block for Transformation.

As component between information source (e.g. field equipment) and targeted applications, an Information Broker automatically processes the *Demands* and provide the information as demanded – ready to “use”. Based on the Pointer to potential information sources and the Instance Filter this Information Broker identifies the information sources to retrieve the data from. Then, it translates the information provided from the selected information sources towards the view model requested by the consuming application. In order to enable further re-use, this transformation knowledge can be created and stored as Model Transformation Functions as Reusable Functional Blocks (Black box), that can be used by the information broker as growing system and modeling knowledge. This whole automated information acquisition and data integration process is presented in Fig. 4 (simplified).

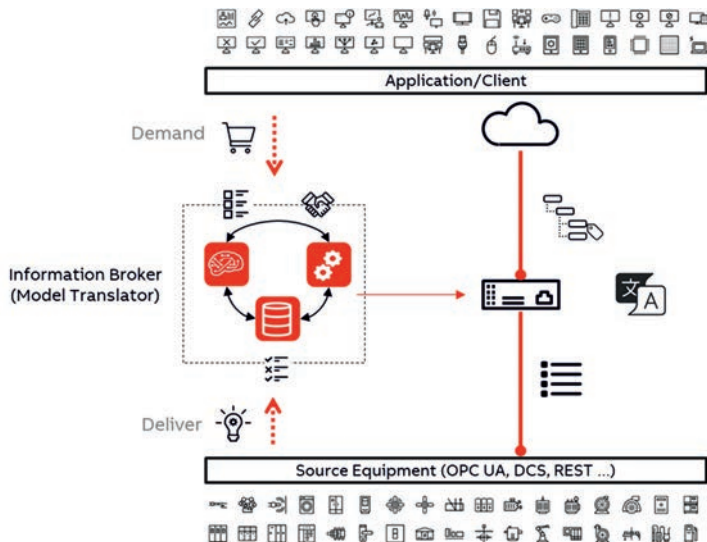


Fig. 4: Automated information acquisition & system integration processes

4. Functional Components

The basic intention of Model-Transformations is to integrate source models delivered by an information *Provider* with target models or view models required by an information *Consumer*. In order to achieve this functionality, there are a number of functional components with varying capabilities and responsibilities. This section provides short description of each of these components.

4.1 Transformation Library (TL)

TL is a kind of library for available Transformations that can be (re-)used for various purposes. The stored Transformations are typically created and verified by Model Experts once and stored in TL for further re-usage. Especially considering the ongoing adoption of the common protocols, standards and applications, it can be assumed that same or similar Transformations have the potential to be used in different contexts and systems. These Transformations may address various translation needs (values, structure, types, format, clean-up, filtering etc.) in order to suit the needs of the Consumer. A Transformation can perform complex operations like changing the whole structure of the model or it can be simple like a unit-

handling or adding, removing or renaming fields etc. It can also be used to decompose a model into smaller subsets and compose subsets into one model.

4.2 Information Broker (IB)

IB is the entity which keeps track of the *Demands* and executes Transformations for *Demands* provided by the Transformation Engine. It actively monitors available *Demands*, Transformations and resolves unresolved/changed *Demands* in case of any changes. Similarly, it also monitors source models that are specified as entry-point by *Demands* and starts Transformations to generate corresponding target models in case of any change. Therefore, this component is responsible for the execution of the selected Transformations between the information producer and the consuming application.

4.3 Transformation Engine (TE)

TE is the entity which resolves the *Demands* by using the appropriate set of transformations retrieved from the TL to transform from Source Model to the Target Model as specified in the *Demand*. In case of a perfect match for the source and target model with an existing Transformation, the TE can directly provide this Transformation to IB as a response for the *Demand*. If not, then the TE may try to resolve the *Demand* by following various approaches:

- Combining or concatenating existing Transformations so that the output of one Transformation is used for one or multiple subsequent Transformations.
- Reversing an existing Transformation: For some transformations, the transformation can be created as bi-directional, so that changes on both sides – source or target model – would be reflected on the corresponding other side. For some others, it would be possible to generate a reverse Transformation ($B-A \rightarrow A-B$) automatically from an existing Transformation: e.g. unit-handling, scaling.
- Inheritance-aware Transformation generation: Also considers the Transformations with a source and target model which has some inheritance with the *Demand* at the hand, then tries to come up with a Transformation which partially or fully resolves the *Demand*.
- Decomposing complex/specific objects into smaller/generic objects: Complex (specific) objects like the device model can be decomposed into (typically more generic) objects.
- Re-use of commonalities along model standards in the same domain. A Transformation addressing a generic type in another standard could be applied to a similar type in the given standard. e.g. unit handling.

If TE achieves to resolve a new *Demand* via one of these approaches, it will add the outcome of this process as a new Transformation that can be used to resolve similar *Demands* in the future. Fig. 5 provides insights about a sample Demand resolution process.

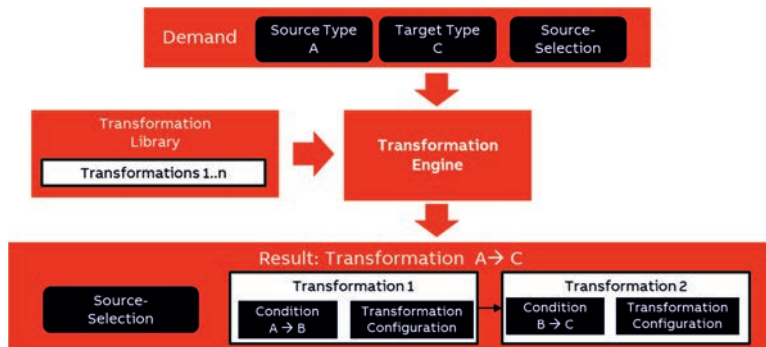


Fig. 5: Transformation Engine – Demand Resolution

4.4 Transformation Management Workplace

This workplace is the user interface that offers the capability to manage, specify and validate Transformations and manage *Demands* and monitor Demand resolution. Typically, this workplace will be used by Model Experts to add, manipulate, and remove model Transformations. In case of Demands that cannot be resolved by the TE and a new Transformation is needed, a ME will be automatically notified and will be requested to enter a new Transformation (from scratch, from an existing Transformation, or by manipulating existing transformations) via this tool. This tool can provide means to realize a wide range of transformation operations: from the simplest transformation tasks to very complex and complicated transformation operations. Similarly, this workplace will be also used by the Application Developers to monitor the status of the created Demands and further follow-up on that. In addition, this workplace offers mechanisms to verify and validate the generated models based on schemas for Source/Target Models (if available).

5. Automated Integration Workflow

By enabling the automatic resolution of the Transformation Demands created by the information Consumer, this solution decouples the development of an application (Consumer) from the corresponding information resources (Producer) and it results in a selective data acquisition, which provides (collect, transform, deliver) only data that is demanded by the Consumer. The Consumer (typically defined by an Application Developer (AD) who has deep

understanding of the application) describes what information (Target Model) is required and specifies the decoupled information sources (Source Model e.g. system/device to consider) – in the form of an Intent or “Demand” for information. This demand definition process is the initial step in the automated integration workflow, which must be performed once per application/solution.

Then this Demand will be processed by an intermediate system, named Information Broker (IB) that resolves a Demand by evaluating available information sources and how these sources can be integrated with the demanded models by means of existing Transformations. If the required Transformations are available (perfect match for the source and target model of an existing Transformation) in the Transformation Library (TL) of the IB, the creation of a Demand is directly triggering the execution of the Transformation and the delivery of the demanded target model. If there is no simple match, then Transformation Engine (TE) will try to automatically try to resolve the Demand by various approaches which are described in Section 4.3. As simplified flow diagram for this Demand resolution process is provided in Fig. 6.

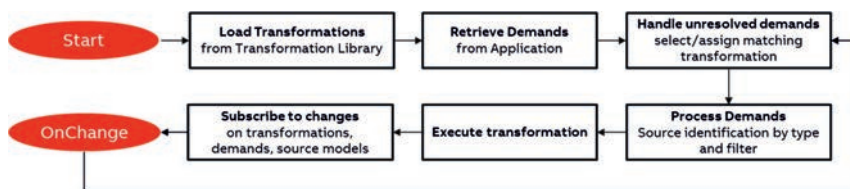


Fig. 6: Demand Processing Flow Diagram

If it is not possible to resolve a Demand fully (due to a missing Transformation), this state is reflected in the Demand and the application or application developer gets informed. In the case of unresolved Demands, a Model Expert (ME) (who has sufficient domain knowledge, understands the context of the use-case and knows input/output models and how to integrate them) will be automatically notified and will be requested to enter a new Transformation using the Transformation Management Workplace, in order to address the Demands .

Thus, data integration can be achieved with no or limited effort from ME, only if there is a need for a new Transformation or extending a partially matched Transformation in order to answer the Demand. New Transformations are stored in the TL and can be re-used further as a black box (independent from what is inside the Transformation and how it was implemented) by the IB for all kinds of Demand resolution tasks in the future. Therefore, this

Transformation generation/creation process is performed only once per required Transformation – e.g. from one model to another.

If the TE can resolve the Demand fully or partially, a verification and validation operation for the resulting models will be performed by the TE by using schemas (e.g. XSD or JSON-Schema). This will allow to identify missing optional/mandatory fields or unwanted fields and to validate the value-type (e.g. int, string, boolean) of the resulting field. Secondly, in case of partial resolution, TE can also calculate a distance factor to the Target and the resolutions with the lowest distance could be used for resolving the Demand or a limited manual effort could be requested from ME. That means partial resolution would still save a lot of resources.

Finally, IB will subscribe to the changes in Demands, sources and Transformations and start continuously monitoring for new Demands or changes in the existing Demands and provide Demand resolution in a continuous manner. Thanks to this capability, the proposed solution also addresses dynamic systems, where the underlying information sources or consumers might change over time.

Fig. 7 provides an overall architecture and workflow of the automated data integration and acquisition process.

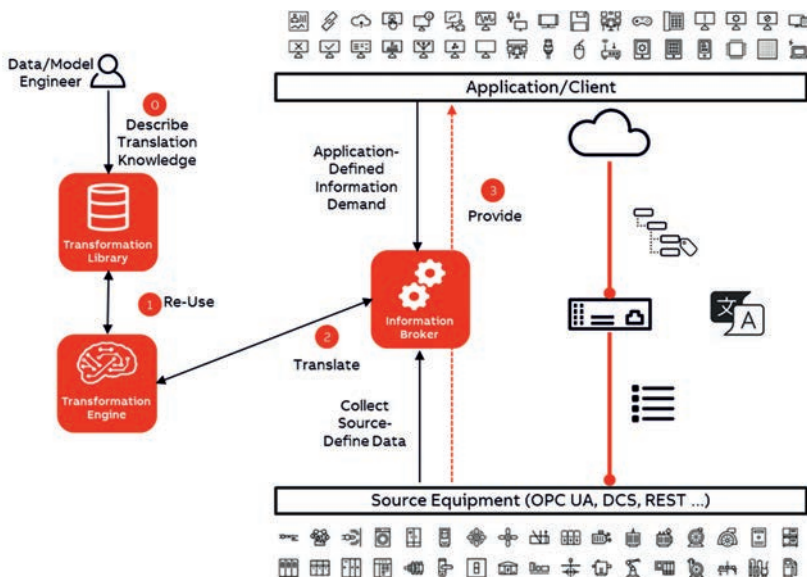


Fig. 7: Automated Data Integration and Information Acquisition Workflow

5. Summary and Conclusion

In this paper, the concept of automatic demand-driven setup of communication path from Consumer to Provider, which decouples information consumption completely from information sources, is described in detail. By automating and simplifying the data integration effort, this solution can reduce inHactegration cost, enable faster application development, and enable added business value for industrial companies. It allows for automated deployment of many/complex information acquisition paths within a large system. In contrast to the traditional model transformation practices, the proposed solution is aiming to create a transformation workflow that simplifies the specification, integration, and re-usage of Transformation along various applications. By automating the Transformation selection and execution, it allows application developers to realize their applications without deep domain knowledge. By reduction of the manual tasks and re-usage of established Transformations, it minimizes efforts and errors during the integration process. Finally, using function-blocks as generic and abstract concept to handle Transformations and meta-data as self-description, it enables maintaining a growing transformation knowledge in the form a Transformation Library.

In addition to concept description and solution design, this work also verified the proposed solution via a use case study with a Gearless Mill Drive (GMD) as proof-of-concept for a large system that consists of many subsystems. This verification study showed that automatic demand-driven integration can be realized for real systems and it can help to improve the integration process for such large industrial systems.

References

- [1] Schlaepfer, R. C., Koch M.: Industry 4.0 Challenges and solutions for the digital transformation and use of exponential technologies. Deloitte AG, 2015.
- [2] OPC Foundation: OPC UA Online Reference. 2021. Available online: <https://reference.opcfoundation.org/#:~:text=The%20OPC%20UA%20Online%20Reference,appropriate%20section%20in%20the%20specification.>
- [3] OPC Foundation: UA Companion Specifications. Available online: <https://opcfoundation.org/about/opc-technologies/opc-ua/ua-companion-specifications/>
- [4] Plattform Industrie 4.0: Specification: Details of the Asset Administration Shell - Part 1. November 2020.
- [5] Data Science Process Alliance: What is CRISP DM?. Available online: <https://www.datascience-pm.com/crisp-dm-2>
- [6] Dodds, L.: Do data scientists spend 80% of their time cleaning data? Turns out, no?. January 31, 2020. Available online: <https://blog.ldodds.com/2020/01/31/do-data-scientists-spend-80-of-their-time-cleaning-data-turns-out-no/>.

Das Industrial Radio Lab Germany

Vier Standorte, eine Mission

Norman Franchi, Philipp Schulz, Gerhard Fettweis,
Vodafone Stiftungslehrstuhl Mobile Nachrichtensysteme,
Technische Universität Dresden;
Andreas Weinand, Daniel Lindenschmitt, Hans Schotten,
Lehrstuhl für Funkkommunikation und Navigation,
Technische Universität Kaiserslautern;
Niklas Bulk, Frank Bittner, Armin Dekorsy,
Arbeitsbereich Nachrichtentechnik, Universität Bremen;
Lisa Underberg, Lutz Rauchhaupt,
Institut für Automation und Kommunikation, Magdeburg

Kurzfassung

Funkkommunikationssystemen wird eine zunehmend wichtige Rolle in professionellen und gewerblichen Anwendungen zugeschrieben, doch gleichzeitig verläuft die Adaption neuer Kommunikationstechnologien zurückhaltend. An dieser Stelle setzt das Industrial Radio Lab Germany (IRLG) an. Als Forschungsinitiative des BMBF werden vom IRLG wissenschaftlich-technische Fragestellungen industrieller Funkkommunikationssysteme erarbeitet, Innovationen entwickelt sowie Unternehmen bei deren Umsetzung unterstützt. Das IRLG fördert in einem ganzheitlichen Ansatz die Adaption von Funkkommunikationssystemen in professionellen und gewerblichen Anwendungen. Ein Beispiel zum Wissenstransfer ist das Merkblatt zu 5G Campusnetzen, welches vom IRLG veröffentlicht wurde.

Das IRLG ist ein Verbund von vier regional verankerten Laboren, den IRLs. Jedes regionale IRL hat eine spezifische Forschungsausprägung. Die IRLs stellen ihre Infrastruktur kooperierenden, kleinen und mittleren Unternehmen zur gemeinsamen Forschung bereit. So erlauben sie den Unternehmen eine effiziente Durchführung von Demonstrator-Entwicklung im Bereich der Funkkommunikation.

In diesem Beitrag wird diese gemeinsame Mission der vier Regionallabore (IRLs) anhand von Beispielen aus aktuellen Arbeiten skizziert. Dazu wird die Robustheit von PROFINET über Bluetooth, die Effiziente SDR-Implementierung eines GFDM Systems, die Resilienz von 5G-Campusnetzen sowie die sichere Kommunikation in Produktion und Internet of Things (IoT) betrachtet.

1. Was ist das IRLG?

Das Industrial Radio Lab Germany (IRLG) untersucht wissenschaftlich-technische Fragestellungen aktueller und zukünftiger sicherer, industrieller Funkkommunikationssysteme. Das IRLG sieht sich hier als Innovations- und Impulsgeber und setzt sich durch die Arbeit mit wirtschaftlich-politischen Gremien für den Wissenstransfer in Richtung einer digitalisierten Gesellschaft ein. Ein Kernpunkt ist dabei auch die unabhängige Unterstützung von Unternehmen bei der Umsetzung ihrer Anwendungen mit industriellen Funkkommunikationssystemen. Dazu gehört auch das Ziel ein Schaufenster durch die Aufbereitung von Forschungsergebnissen zu schaffen. Zudem engagiert sich das IRLG in der Förderung des akademischen Nachwuchses durch den fachlichen Austausch zwischen verschiedenen Forschungseinrichtungen. Die Aufträge des IRLG sind in Abbildung 1 zusammengefasst.

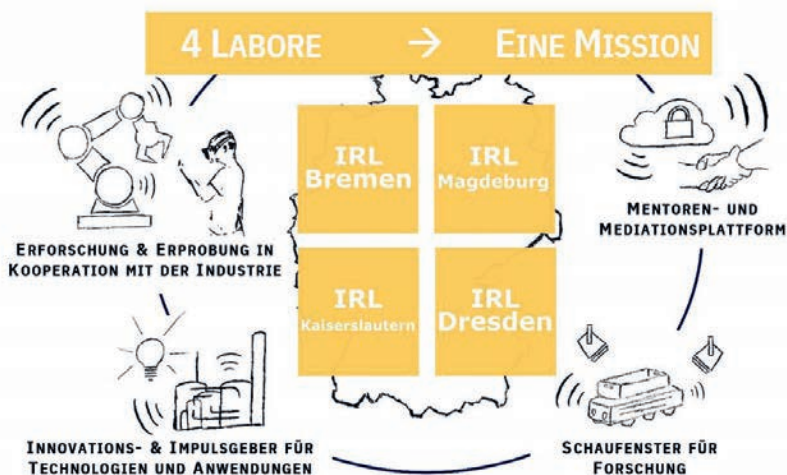


Abbildung 1: Ganzheitlicher Ansatz des IRLG

Das IRLG ist als Verbundprojekt von vier verteilten Forschungs- und Entwicklungslaboren aus dem Bereich Funkkommunikation organisiert. Die IRLs blicken wie in Abschnitt 2 beschrieben auf gemeinsame Erfahrungen im Bereich Industriefunk. Jedes regionale IRL verfügt über ein ausgeprägtes und stetig wachsendes Netzwerk zu kleinen und mittleren Unternehmen, sowie zur Großindustrie. Kooperierende Unternehmen werden in die gemeinsame Forschung einbezogen und bei der Durchführung der Entwicklung von Demonstratoren unterstützt. Hierbei ergänzen sich die unterschiedlichen Forschungsschwerpunkte der IRLs, die in Abschnitt 3 näher erläutert werden.

2. Wo kommen wir her?

Das Konsortium aus der Technischen Universität Dresden, der Universität Bremen, der Technischen Universität Kaiserslautern und dem Institut für Kommunikation und Automation (ifak) in Magdeburg bringt langjährige Erfahrung in der Forschung und Entwicklung von Funkkommunikationstechnologien mit, deren Einsatz in industriellen Anwendungen bei allen Partnern seit bis zu 20 Jahren ein Fokusthema ist. Aus dieser Expertise heraus bildete das Konsortium im Rahmen der BMBF-Initiative zu "Zuverlässige drahtlose Kommunikation für die Industrie" (ZDKI), deren Projekte von 2015 bis 2018 liefen, die Begleitforschung (BZKI). In diesem Projekt wurden die Grundlagen für die Beschreibung von Anforderungsprofilen industrieller Anwendungen gelegt [1]. Ebenso wurden Zuverlässigkeitsaspekte beleuchtet [2]. Gleichzeitig wurde die BZKI zum Ko-Initiator der 5G Alliance for Connected Industries and Automation (5G-ACIA), wodurch die Aktivitäten rund um Anforderungsspezifikation und Zuverlässigkeitsaspekte verstetigt wurden. Basierend auf den Erfahrungen der BZKI konnte das Konzept für Campusnetze entwickelt werden, welches von der Bundesnetzagentur (BNetzA) aufgegriffen und für 3,7 bis 3,8 GHz sowie inzwischen ebenfalls für 24,25 bis 27,5 GHz umgesetzt wurde [3]. Diese erfolgreiche Zusammenarbeit durch die Vernetzung von Industrie, Kommunikationstechnologie und Forschung wird im Rahmen des IRLG fortgesetzt.

3. Womit beschäftigen wir uns in den vier Industrial Radio Labs?

Um das gemeinsame Ziel des IRLG – "Wir schließen die Lücke zwischen Forschung und Anwendung" – zu erreichen, hat jedes regionale IRL eigene Schwerpunkte, sodass insgesamt das breite Spektrum industrieller Anwendungen abgedeckt wird. Tabelle 1 fasst die Schwerpunkte zusammen, aus denen im Folgenden jedes IRL ein schwerpunktübergreifendes Thema herausgreift und seine aktuellen Arbeiten beschreibt.

Tabelle 1: Schwerpunkte der IRLs

 TECHNISCHE UNIVERSITÄT DRESDEN • Resiliente, industrielle Campusnetze • Echtzeitüberwachung d. Funkspektrums • Detektion u. Entschärfung von Störquellen • Co-Design von Kommunikation und industrieller Regelung • Design der Funkschnittstelle für URLLC • Anwendungen & Technologien für das Taktile Internet	 Universität Bremen • Verbindung von Satellitensystemen und terrestrischen Kommunikationssystemen • Künstlicher Intelligenz in Kommunikationssystemen • Co-Design von Kommunikation und industrieller Regelung • Überwachung des Funkspektrums

 • Resiliente Systeme • Koexistenzmanagement • IT & Netzwerksicherheit • Verifikation und Zertifizierung • Verlässlichkeit • 5G Campusnetze. • Künstliche Intelligenz	 • Anwendungen im Wassermanagement • Kabellose Lösungen f. d. Wasserindustrie • Absicherung gegen aktive Umwelteinflüsse • Spezifizierung von Interferenzprofilen • Erhöhung der Robustheit • Koexistenzmanagement

a) IRL Bremen: Effiziente SDR-Implementierung eines GFDM Systems

Um im internationalen Wettbewerb langfristig bestehen zu können, muss die deutsche Industrie Prozesse konsequent digitalisieren. In der Vision einer „Industrie 4.0“ kooperieren Menschen, Roboter und Maschinen eng miteinander. Dazu bedarf es einer durchgängigen Vernetzung aller Systeme und Industrieakteure. Die Erforschung und Entwicklung verlässlicher Funktechnologien für industrielle Anwendungen wurde im BMBF- Forschungsprogramm "Zuverlässige drahtlose Kommunikation in der Industrie" adressiert. Die Umsetzung und Erprobung der erforschten Technologien wurden an der Universität Bremen über Demonstratoren z.B. im Projekt HiFlecs mit eigenständigen Ansätzen gezeigt.

Mit der 5. Generation des Mobilfunks (5G) wurden die Grundlagen geschaffen, die Potenziale für die Industrie 4.0 über standardisierte Ansätze in einem breiten Spektrum von Anwendungen einzusetzen, wie bspw. in den EU-Projekten Fantastic-5G, METIS und iJOIN untersucht wurde. In den BMBF-Programmen "5G: Industrielles Internet" und "5G: Taktiles Internet" wurde diese Einbindung von 5G-Technologien u.a. im Projekt TACNET 4.0 erforscht.

Im Rahmen dieses Projektes wurden zur Erprobung neuer Kommunikationskonzepte Testbeds entwickelt und aufgebaut. Das Ziel dieser Testbeds war es, typische Szenarien aus den Bereichen Produktionsautomatisierung mit echtzeitfähigen Anwendungen sowie mobilen Arbeitsmaschinen abzudecken. Das Testbed Bremen zeigt eine Robotik-Verkehrsregelung zwischen mobilen Robotern. Dabei wird Edge Computing u. a. als Traffic Optimizer eingesetzt. Dieser befasst sich mit dem Management von Autonomous Guided Vehicles (AGVs), um deren reibungslose und kollisionsfreie Bewegung zu ermöglichen, insbesondere in kritischen Situationen wie Kreuzungen und Einmündungen (siehe Abbildung 2). Der Traffic Optimizer steuert die Roboter so, dass das Verhalten der AGVs dynamischer ist und ein Überfahren von Kreuzungen, das Verlassen von Einmündungen und Verfolgungsfahrten ohne Notstopps bewältigt werden.

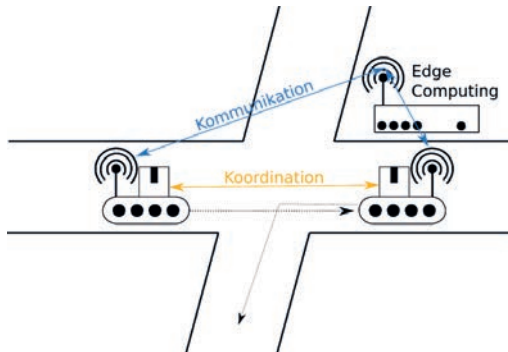


Abbildung 2: Robotik-Verkehrsregulierung

Dabei verwendet das Testbed eine Kommunikationsumgebung welche auf Software Defined Radios (SDR) basiert. Für die Untersuchung verschiedener Kommunikationsaspekte ist eine performante SDR-Implementierung, die sowohl in der Simulation als auch im Feldversuch eingesetzt werden kann, wünschenswert, da sie ihre Flexibilität ohne große Leistungseinbußen beibehält. Darüber hinaus ist eine SDR-Implementierung für das verwendete Edge Computing System von entscheidender Bedeutung, da es aufwendige Rechenoperationen in der Basisstation durchführt, welches zu einer effizienten Nutzung gegebener Hardware führt.

Um die Übertragungseigenschaften zu optimieren, wurde ein besonders leistungsfähiges Modulationsverfahren eingesetzt. Im Gegensatz zum konventionellen OFDM werden bei GFDM, wie es hier genutzt wird, ganze Frames statt einzelner Zeitschlitzte verwendet. Dieser Ansatz führt zu mehr Flexibilität und dadurch zu einer individuellen Anpassung an den Anwendungsfall. Auch die Latenzzeiten können durch eine entsprechende Auslegung des Systems minimiert werden [4].

Diese in TACNET4.0 erarbeiteten Innovationen sollen in einem Testbed in Bremen in Kooperation mit Industriepartnern weiterentwickelt werden. Der Demonstratoraufbau setzt weiter auf die Implementierung mit SDRs und strebt eine 5G-standardnahe Implementierung an, dadurch können Systemkomponenten in einem modifizierbaren 5G-Kommunikationsnetz auf ihre Funktionsweise untersucht werden. Der Kern des Demonstratoraufbaus ist die Robotik-Verkehrsregulierung durch einen Traffic Optimizer. Um einen reibungslosen Verkehrsfluss an Kreuzungen und Einmündungen zu realisieren, ist eine hohe Abstimmungsrate, eine latenzreduzierte, störffeste und drahtlose Kommunikationsumgebung notwendig.

Das sich im Aufbau befindliche Digital Hub Industry (DHI) Bremen wird zukünftiger Standort des IRL Bremen sowie des Demonstratoraufbaus sein. Das DHI gibt kleinen und mittelständischen Unternehmen die Möglichkeit, ihre eigene digitale Transformation voranzutreiben. Hierbei unterstützt das IRL Bremen die Unternehmen durch eine offene und zielgerichtete industriennahe Forschung. Die Unternehmen können dabei auf das Testbed zurückgreifen um für ihre Produkte spezifische Problemstellungen zu identifizieren und in Kooperation mit dem IRL Bremen bearbeiten. Änderungen am System können aufgrund der eingeführten SDR-Schnittstellen, schnell und einfach durchgeführt werden. Der modulare Aufbau des SDR lässt ebenso eine unkomplizierte Erweiterung des Systems durch Industriekomponenten zu. Etwaige Herausforderungen können so schon vor der Implementierung in ein Produktivsystem aufgezeigt und Lösungsansätze unmittelbar kommuniziert werden.

b) IRL Dresden: Resiliente Campus-Netze durch Spectrum Sensing

Ein wichtiger Aspekt des 5G-Mobilfunks ist die Bereitstellung von Frequenzbändern für Campusnetze. Der Einsatz von Campusnetzen gestattet den Aufbau eines exklusiven lokalen Netzes für die industrielle Nutzung. Dieser Ansatz wird neuartige Anwendungen ermöglichen und neue Märkte für Dienstanbieter eröffnen. Für eine ausführliche Erklärung der möglichen Ausprägungen von Campusnetzen sei auf das zugehörige IRLG Merkblatt verwiesen [5].

Essentiell für den Erfolg von Campusnetzen im industriellen Umfeld ist der Aufbau resilienter Netzwerke. Ein vielversprechendes Mittel dafür ist die (Echtzeit-) Überwachung des Spektrums, die Erkennung und Lokalisierung von bewussten und unbewussten Störsendern sowie die Selbstheilung und Wiederherstellung des Kommunikationssystems im Falle eines Ausfalls. Am IRL Dresden liegt ein Forschungsschwerpunkt im Aufbau von resilienten Campusnetzen mithilfe von Sensornetzwerken. Die Ortsinformation des Senders ist beispielsweise für ein effizientes Spektrum-Management oder die Strahlformung von MIMO-Systemen wertvoll.

Im Folgenden wird ein kurzer Überblick über die am IRL Dresden verfolgten Ansätze gegeben. Dabei wird zwischen zwei Informationsquellen (Empfangsleistung, Kanalschätzung) und zwei Lösungsansätzen (modellbasiert, Machine-Learning) unterschieden.

Die Empfangsleistung (received signal strength, RSS) wird als eine Informationsquelle für die Lokalisierung verwendet. Auf diese Weise ist keine Vorkenntnis des übertragenen Signals erforderlich, was auch die Lokalisierung nicht-kooperierender Sender, d.h. beabsichtigter und unbeabsichtigter Störer, ermöglicht. Außerdem werden auf der Kanalschätzung basierende Ansätze für die Senderlokalisierung verfolgt, die Vorwissen über das Sendesignal erfordern. Damit sind die Ansätze für reguläre Nutzer geeignet. Ausgangspunkt ist die Kanalmodellierung unter Berücksichtigung der Position des Senders, des Empfängers und deren Umgebungen.

Dieses Modell wurde für Proof-of-Concept-Simulationen erstellt und veröffentlicht [6]. Aktuelle Arbeiten im IRL Dresden beschäftigen sich mit der Verbesserung der Auflösung der Kanalschätzung, der Analyse verschiedener Lokalisierungsalgorithmen und der Erweiterung des Lokalisierungsalgorithmus durch Trackingfunktionen.

Machine Learning wird als potentielle Schlüsseltechnologie für Lokalisierung unbekannter Sender mithilfe von Sensornetzwerken mit hoher Genauigkeit und gleichzeitig niedriger Komplexität angesehen. Aktuelle Arbeiten am IRL Dresden nutzen einen Deep-Learning-Ansatz für die Lokalisierung in einem industriellen Szenario mithilfe von gesammelten RSS-Daten. Dieser Ansatz erzielt nahezu die Genauigkeit eines optimalen Maximum-Likelihood-Schätzers, erlaubt dabei aber eine deutliche Reduktion der Komplexität.

Modell-basierte Ansätze haben hingegen den Vorteil, dass sie keine großen Mengen an Trainingsdaten benötigen, die möglicherweise nur sehr spezifische Situationen abdecken. Den Ansätzen liegt die Analyse der Empfangsleistungen durch ein Sensornetzwerk unter der Annahme eines gegebenen Pfadverlustmodells zugrunde, welches sich an verschiedene Umgebungsbedingungen anpassen lässt. In aktuellen Arbeiten des IRL Dresdens wurden mehrere Lösungsansätze vorgeschlagen, die die Lokalisierung einzelner (Maximum Likelihood Search, Generalized Least Squares) und mehrerer Quellen (Radio Environment Map Localization, Compressed Sensing) unter dem Einfluss von Abschattungseffekten zu ermöglichen und auf Genauigkeit und Komplexität untersucht [7]. Außerdem wurde der Compressed-Sensing-Ansatz durch einen Kalman-Filter erweitert, um Echtzeit-Tracking von Sendern zu ermöglichen [8]. Weiterhin wird der Einsatz einer durch die Physik inspirierte Partikelsimulation für die Lokalisierung mehrerer unbekannter Sender mit unbekannten Sendeleistungen und unter Abschattungseffekten vorgeschlagen [9]. Durch den iterativen Ansatz, eignet sich auch diese Methode für das Tracking. Auch andere Informationsquellen könnten leicht integriert werden.

Die theoretische Forschung am IRL Dresden wird durch die Kooperation mit Rohde und Schwarz ergänzt. Gemeinsam wurde das in Abbildung 3 illustrierte Konzept einer zukünftigen intelligenten industriellen Fabrik mithilfe von Echtzeiterkennung von Störungen und Jamming vorgestellt. Im Rahmen der Kooperation werden dem IRL Dresden Hochleistungsnetzwerk-scanner durch Rohde und Schwarz bereitgestellt. Dies ermöglicht die praktische Erprobung der theoretischen Ergebnisse des IRL Dresden. Schließlich werden die Arbeiten durch die praktische Erfahrung und industrielle Perspektive von Rohde und Schwarz ergänzt.

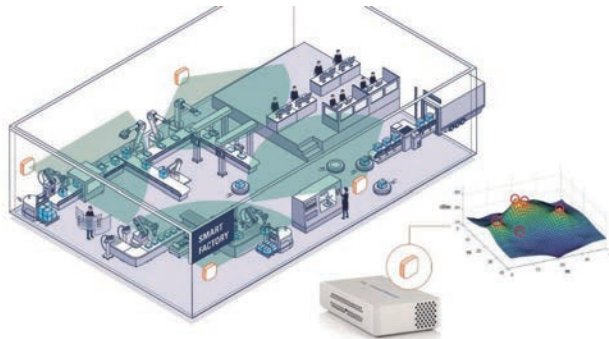


Abbildung 3: Echtzeitmessungen von Störungen und Jamming in einer intelligenten Fabrikhalle (Quelle: Rohde & Schwarz).

c) IRL Magdeburg: Performance von PROFINET über Bluetooth

Trotz der Entwicklung zuverlässiger Funkkommunikationslösungen für industrielle Automatisierungsanwendungen sind Befürchtungen weitverbreitet, dass die Funkübertragung ungewollt oder gezielt gestört werden kann. Entsprechende Diskussionen bei der Überarbeitung der VDI/VDE-Richtlinie 2185 "Funkgestützte Kommunikation in der Automatisierungstechnik" im VDI/VDE-GMA Fachausschusses 5.21 wurden zum Anlass genommen, das Störpotenzial bei industrieller Funkkommunikation im IRL Magdeburg näher zu untersuchen.

In der nachfolgend vorgestellten Testreihe wurden Funkgeräte mit Bluetooth 5, die in der Lage sind PROFINET-I/O-Kommunikation zu übertragen, verwendet. Der Testaufbau beinhaltet die drahtlose Kommunikation zwischen einer PROFINET-I/O-Steuerung und drei PROFINET-I/O-Geräten (siehe Abbildung 4). Mithilfe speziell entwickelter Hardware, den Multifaces, werden über die digitalen Eingänge der PROFINET-Geräte die Dateninhalte der PROFINET-Nachrichten gezielt geändert. Die Änderungen werden nach der PROFINET-Funk-Übertragung an den Ausgängen der PROFINET-Geräte registriert und ausgewertet. Auf diese Weise wird zum einen die Zuverlässigkeit der PROFINET-Kommunikation und zum anderen die Erfüllung der Anwendungsanforderungen bewertet.

Im vorgestellten Testfall wurden je logischer Verbindung ca. 27.500 Nachrichten versandt. Zwischen Minute 20 und 40 wurde ein Störsignal aktiviert. Weitere Bedingungen für diesen Testfall sind nachfolgend aufgelistet:

- Testumgebung: Labor, leitungsgebundene Übertragung unter Verwendung eines Kanalemulators, Funkgeräte befinden sich in geschirmten Boxen
- Anzahl logischer Verbindungen: 4
- Nutzdaten je Verbindung: 16 Bit
- Sendezeitabstand je Verbindung: 131 ms
- PROFINET-Zyklus: 16 ms
- Bluetooth Connection Interval: 7,5 ms
- Abstand zwischen den Funkgeräten: 10 m (über Kanalemulator verbunden)
- Abstand zwischen der Störquelle und einem Bluetooth-Slave-Gerät: 3 m (über Kanalemulator verbunden)
- Störsignal: 802.11g, 20 MHz (mit Vektorsignalgenerator erzeugt)
- Stördauer: kontinuierlich
- Störleistung: 100 mW
- Beobachtungsdauer: 1 Stunde mit 20 Minuten Störung zwischen Minute 20 und 40

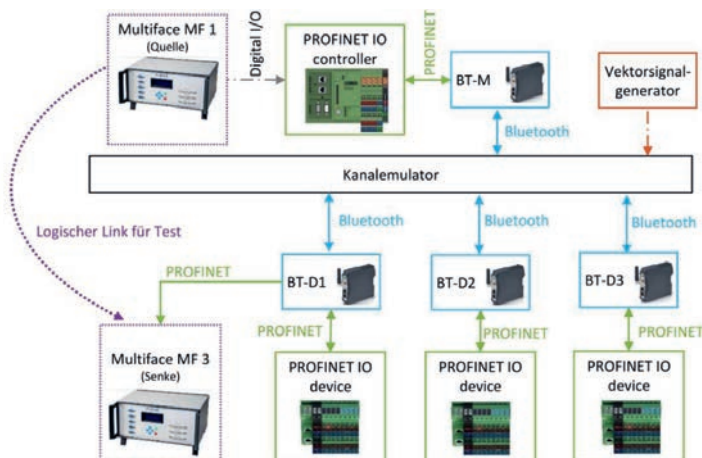


Abbildung 4: Aufbau zur Untersuchung von PROFINET-Kommunikation über Funk

Die Auswirkung auf die gemessene Aktualisierungszeit ist in Abbildung 5 deutlich zu erkennen. Allerdings wurde trotz längerer Aktualisierungszeit die PROFINET-Kommunikation nicht unterbrochen. Das bedeutet, dass in einer geplanten Anwendung, die das ermittelte Zeitverhalten während einer Störung toleriert, die untersuchte Funkkommunikationslösung eingesetzt werden kann.

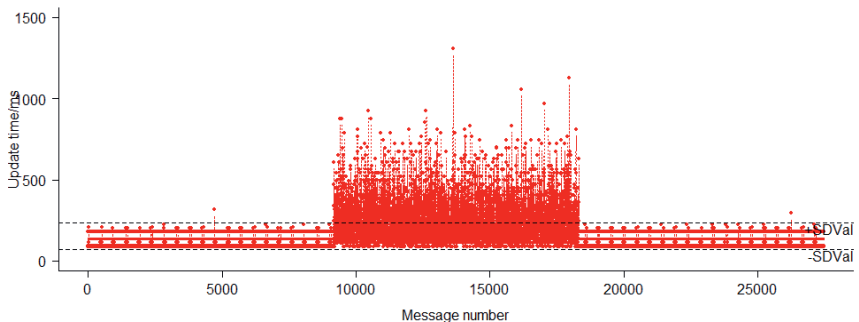


Abbildung 5: Aktualisierungszeit von Bluetooth-Kommunikation ohne und mit Interferenz

Durch Variation der Werte der oben aufgelisteten Parameter (beispielweise des PROFINET-Zyklus) ergeben sich weitere Testfälle, die im Rahmen der IRL Magdeburg Aktivitäten untersucht wurden.

Mit der vorliegenden Testmethodik und -plattform unterstützt das IRL Magdeburg Hersteller von Funklösungen bei der Entwicklung und bei der quantitativen Bewertung ihrer Produkte. Da das Testvorgehen für alle Funklösungen in gleicher Weise und reproduzierbar anwendbar ist, liefern die Ergebnisse eine verlässliche Orientierung für potenzielle Anwender von Funklösungen. Aufgrund der Neutralität des IRLG werden keine Funktechnologien oder Funkprodukte bevorzugt behandelt. Die Forschungspartner können sich auf eine anwendungsbezogene unabhängige Bewertung verlassen.

d) IRL Kaiserslautern: Sichere Kommunikation in Produktion und IoT

Ein aktuelles Thema, mit welchem sich das IRL Kaiserslautern beschäftigt, sind Aspekte zur Informationssicherheit, Datensouveränität und Vertrauenswürdigkeit in Funknetzen im industriellen Kontext. Während die Informationssicherheit insbesondere technische Aspekte, wie z. B. den Einsatz kryptographischer Verfahren umfasst, spielen bei den Themen Datensouveränität und Vertrauenswürdigkeit auch Aspekte der Systemarchitektur bzw. -konfiguration eine wichtige Rolle.

Im Rahmen der Informationssicherheit in industriell eingesetzten Funksystemen sind insbesondere die Aspekte Authentifizierung, Datenvertraulichkeit und -integrität besonders relevant, da die zugrundeliegenden Applikationen bei Kompromittierung dieser Schutzziele große Auswirkungen auf die physikalischen Umgebungen haben können (z. B. Zerstörung physischer Güter oder Gefährdung von Personen).

Die Authentifizierung von Teilnehmern oder Entitäten in einem industriellen Funksystem bildet die Vertrauensbasis für alle weiteren Sicherheits- und Systemfunktionen. Dazu werden z. B.

in Mobilfunksystemen SIM-Karten eingesetzt. In 4G dienten diese zur Gewährleistung der Vertrauensbeziehung zwischen Nutzer und Netzbetreiber. Im Rahmen von 5G ist insbesondere die Ende-zu-Ende Sicherheit zwischen Endgeräten und Anwendungen auf der Backend-Seite in den Fokus gerückt, was durch zusätzliche Sicherheitsfunktionen zwischen Netzbetreibern und Anwendungsanbietern („Verticals“) ermöglicht wurde (zuvor nur als Over-the-Top Lösung optional möglich). Weiterhin wurden die symmetrischen Verfahren des Universal Subscriber Identity Module (USIM) zur Authentifizierung mit der Einführung von 5G durch asymmetrische Verfahren mit höherem Sicherheitsniveau ersetzt (z. B. Verhinderung bekannter Angriffe in 4G) [10].

Im Gebiet der Datenvertraulichkeit bieten nahezu alle etablierten Funksysteme, welche heutzutage im industriellen Kontext eingesetzt werden, die Möglichkeit einer Verschlüsselung um z. B. ein Mithören des Nutzdatenverkehrs zu verhindern. Jedoch sind Verfahren zur Einbringung des Schlüsselmaterials nicht immer praktikabel gelöst. Im Vergleich zu den Mobilfunksystemen bieten beispielsweise Funksysteme für IoT-Anwendungen wie Low Power Wide Area Networks (LPWANs) nicht immer die Möglichkeit dies mit Hilfe von SIM-Karten zu ermöglichen. Häufig müssen hier manuell Schlüsselmaterial und andere Attribute entweder im Endgerät oder in entsprechenden Backend-Systemen hinterlegt werden (z. B. LoRaWAN) oder es besteht keine Möglichkeit diese im laufenden Betrieb im Falle einer Kompromittierung zu widerrufen. Weiterhin zeichnen sich die genannten Anwendungen durch eine große Anzahl von vernetzten Geräten, geringen Paketgrößen und Sendezyklen aus, sodass insbesondere hohe Anforderungen an die Ressourcen- und Energieeffizienz bestehen. Klassische kryptographische Verfahren bieten hier aufgrund des hohen Signalisierungsaufwands keine optimale Lösung zur Erreichung der genannten Ziele. Eine alternative Lösung bieten daher Physical Layer Security (PHYSEC) basierte Verfahren. Dabei werden aus zufallsähnlichen physikalischen Prozessen (z. B. des Funkkanals) kryptografische Schlüssel abgeleitet, welche mit sehr geringem Signalisierungsaufwand eingesetzt werden können.

Am Beispiel von LoRaWAN wurde dazu im IRL Kaiserslautern eine Untersuchung durchgeführt [11]. Hier wurde gezeigt, dass PHYSEC-Verfahren eine Alternative zu konventioneller Kryptographie bieten können. Weitere Untersuchungen, insbesondere zur Optimierung des Signalisierungsaufwands zum Sitzungsschlüsselaustauschs, sowie weiterer Anwendungsszenarien sind Gegenstand aktueller Forschungen des IRL Kaiserslautern.

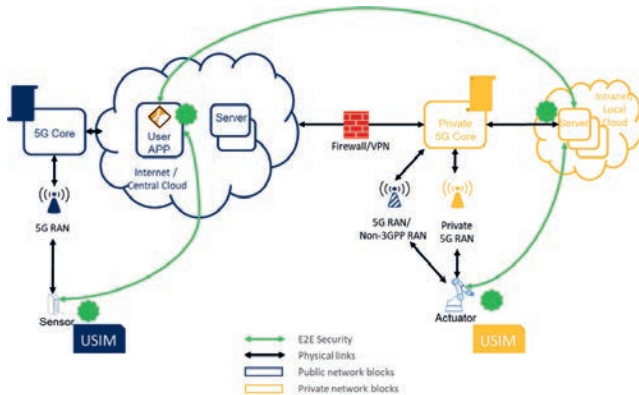


Abbildung 6: Datenhoheit im 5G System in verschiedenen Betreibermodellen

Ein wichtiger Aspekt, welcher eine hohe Relevanz im Bereich industrieller Anwendungen hat, ist die Gewährleistung von Datensouveränität und Vertrauenswürdigkeit. Dies hängt in erster Linie vom gewählten Betreibermodell eines 5G Systems (z. B. Campusnetz, siehe Abbildung 6) ab. Dies bestimmt u. a., inwiefern die Akquirierung von Metadaten aus dem Datenverkehr durch die beteiligten Entitäten möglich ist. Der Betreiber bzw. Home Mobile Network Operator (HMNO) des Kernnetzes ist technisch in der Lage, die anonymisierten Identitäten der Teilnehmer zu entschlüsseln. Daher liegt hier die Datensouveränität nicht vollständig auf der Nutzerseite. Wird der Funksystemnutzer hingegen selbst zum Kernnetzbetreiber (nicht-öffentliches Campusnetz), so kann ein hohes Maß an Datensouveränität gewährleistet werden. Eine Ende-zu-Ende Sicherheit durch Authentifizierung und Verschlüsselung bleibt zum Erhalt der Datensouveränität hingegen in jedem Betreiberszenario notwendig.

4. Was erwartet uns in Zukunft?

Das Industrial Radio Lab Germany vernetzt Industrieanwendungen und Forschung. In den vorgestellten Themenschwerpunkten wird deutlich, wie vielfältig die Forschungsgebiete der unterschiedlichen IRLs sind. Die Forschung innerhalb des gesamten IRLGs streckt sich dabei von Problemstellungen der physikalischen Übertragung bis hin zur Souveränität der Daten sowie von der Implementierung einzelner Forschungsaspekte in Demonstratoren bis hin zur konkreten Umsetzung in nicht-öffentlichen Campusnetzen.

Das Industrial Radio Lab Germany ist die nationale Kollaborationsplattform für wissenschaftlich-technische Fragestellungen im Bereich der industriellen Funkkommunikation. Interessierten Industriepartnern ermöglicht das IRLG einen direkten Kontakt zu den für die jeweilige Herausforderung relevanten IRLs. Akute sowie perspektivische Themen rund um die

industrielle Nutzung drahtloser Kommunikationstechnologien werden somit auf Augenhöhe diskutiert und gelöst. Als Verbund unabhängiger Forschungseinrichtungen ist das IRLG offen und neutral gegenüber allen bestehenden sowie zukünftigen Kommunikationstechnologien.

Hinweis

Das diesem Beitrag zugrundeliegende Vorhaben "Industrial Radio Lab Germany" wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBG) unter den Förderkennzeichen 16KIS1010K, 16KIS1011, 16KIS1012 und 16KIS1013 gefördert.

Verweise

- [1] L. Rauchhaupt, D. Schulze, A. Gnad und M. Krätzig, „Anforderungsprofile im ZDKI,“ 2016.
- [2] S. Willmann, A. Gnad, E. Hintze, M. Krätzig und L. Rauchhaupt, „Aspekte der Zuverlässigkeitsbewertung in ZDKI,“ 2017.
- [3] Bundesnetzagentur, „Regionale und lokale Netze,“ [Online]. Available: <http://www.bundesnetzagentur.de/lokalesbreitband>. [Zugriff am 10 05 2021].
- [4] J. Demel, C. Bockelmann, A. Dekorsy, A. Rode, S. Koslowski und F. K. Jondral, „An optimized GFDM software implementation for future Cloud-RAN and field tests,“ in *GNU Radio Conference 2017*, San Diego, USA, 2017.
- [5] Industrial Radio Lab Germany (IRLG), „Merkblatt: 5G Campus-Netze,“ September 2020. [Online]. Available: <https://industrial-radio-lab.eu/publikationen/#campusnetze>.
- [6] Z. Li, R. Bomfin, I. Bizon Franco de Almeida, A. Nimr, N. Franchi und G. P. Fettweis, „Position-Driven Wireless Multipath Channel Simulator for Square-Shaped Environments,“ in *Proceedings of 1st IEEE International Online Symposium on Joint Communications & Sensing (JC&S 2021)*, Dresden, virtuelle Konferenz, 2021.
- [7] K. Kulkarni, N. Franchi und G. P. Fettweis, „Reliable Real-time Localization and Tracking of Interferers Using Cooperative Spectrum Sensing,“ in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC 2019)*, Marrakeesch, Marokko, 2019.
- [8] K. Kulkarni, N. Franchi und G. P. Fettweis, „Real-time Cooperative Spectrum Sensing: Reliable Localization of Unknown Interferers in Shared Spectrum,“ in *Proceedings of International ITG Conference on Systems, Communications and Coding (SCC 2019)*, Rostock, 2019.
- [9] P. Schulz, N. Franchi und G. P. Fettweis, „RSS-Based Localization of Multiple Unknown Transmitters Through Particle Simulation,“ in *Proceedings of IEEE International Online Symposium on Joint Communications & Sensing (JC&S 2021)*, Dresden, virtuelle Konferenz, 2021.
- [10] A. Koutsos, „The 5G-AKA Authentication Protocol Privacy,“ *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*, pp. 464-479, 2019.
- [11] A. Weinand, A. G. de la Fuente, C. Lipps und M. Karrenbauer, „Physical Layer Security based Key Management for LoRaWAN,“ *Proceedings of 2020 Workshop on Next Generation Networks and Applications (NGNA 2020)*, 2020.

Plattform zur Untersuchung industrieller Funkkommunikation für die Intralogistik

Platform for the investigation of industrial wireless communication for intralogistics

G. Persico, A. Fischer, DEMAG Cranes & Components GmbH, Wetter;
M. Riedl, D. Tran, ifak e.V., Magdeburg

Kurzfassung

Funktechnologien wie 5G und WiFi 6 versprechen zuverlässige Kommunikation für industrielle Steuerungsanwendungen. Ein Problem stellen jedoch ausgedehnte Funktionstests und Leistungsuntersuchungen unter realen Einsatzbedingungen dar. Im Rahmen des Forschungsvorhabens OPTIMUM entstand ein Demonstrator zum Erforschen und Entwickeln zukunftsweisender Intralogistikkonzepte im Kontext vom Smart Manufacturing und Industrie 4.0. Dieser Beitrag stellt das Potenzial des Material Handling Demonstrators vor, der auch als Testbed für die Untersuchungen industrieller Funkkommunikation für die Intralogistik zur Verfügung steht. Das Testbed wird weiter ausgebaut und für künftige Kooperationsprojekte genutzt.

Abstract

Wireless technologies such as 5G and WiFi 6 promise dependable communication for industrial control applications. However, extensive functional and performance tests under real operating conditions are a problem. As part of the OPTIMUM research project, a demonstrator was built for researching and developing future-oriented intralogistics concepts in the context of smart manufacturing and Industry 4.0. This article presents the potential of the material handling demonstrator, which is also available as a testbed for investigating industrial wireless communication for intralogistics. The testbed will be further expanded and used for future co-operation projects.

1. Intralogistik für Industrie 4.0

Anforderungen wie kürzere Inbetriebnahme, schnelle Umrüstzeiten bei Produktwechseln oder agile und temporäre Zusammenarbeit sollen dem Anwender im Zuge der Flexibilisierung der Produktionssysteme durch IIoT oder standardisierte Lösungen im Umfeld von Industrie 4.0 deutliche Kosteneinsparung bringen. Für die Gerätehersteller oder Ausrüster bedeutet dies, eigene Produkte um entsprechende Funktionalitäten zu erweitern. Im ITEA3-Projekt OPTI-

MUM lag aus Sicht der deutschen Projektpartner der Schwerpunkt bei Intralogistikkomponenten und hier insbesondere bei Kranen. Diese sollten zunächst in die Lage versetzt werden, flexibel miteinander, mit anderen Anlagenbestandteilen aber auch mit Mitarbeiter*innen zu kommunizieren und zu kollaborieren. Unter Berücksichtigung aktueller Diskussionen zu Industrie 4.0 konformen Architekturen, Komponenten und Diensten verfolgte OPTIMUM folgende Ziele:

- Verbesserung von Aspekten der verteilten Steuerung,
- Nutzung und Anpassung von Technologien aus dem Umfeld (I)IoT für reale industrielle Anwendungen,
- Verbesserung von Steuerungslogik und Anwendungen durch die Berücksichtigung von Kontext- und Ortsinformationen,
- Modell- und 3D- basiertes Engineering und Anlagenüberwachung

Gegenüber den heute etablierten, zentralen und hierarchisch gegliederten Steuerungs- und Anwendungsprozessen (S95/IEC 62264) erfordern künftige, komponentenbasierte Konzepte und „kollaborative Automatisierung“ verteilte Funktionalitäten. Die in OPTIMUM entwickelten Technologien sollen Maschinen- und Hebezeughersteller, wie z. B. den Projektpartner DEMAG, dazu befähigen, proprietäre Hardwarevarianten der eingesetzten Steuerungstechnik zu reduzieren und gleichzeitig die Modularität der eingesetzten Software zu erhöhen. Kontext- und Positionsinformationen erlauben Assistenzfunktionen, die zu höherer Effizienz, Komfort und funktionaler Sicherheit führen und somit die Chancen neuer innovativer Produkte am Markt verbessern.

Ein häufiger Anwendungsfall, der im Forschungsprojekt zunächst mit kleinen Demonstratoren und anschließend mit realen Kranen demonstriert wurde, ist die Tandemfahrt, bei der mehrere Krane und Laufkatzen zusammenarbeiten müssen. Die Lösung aus Sicht der Steuerung liegt in einem **verteilten Steuerungssystem**, bei dem die einzelnen Kransteuerungen sowohl autark als auch im Verbund arbeiten können. Weiterhin wurde untersucht, wie bestimmte Gebiete einer Werkhalle bei der Kranfahrt ausgeschlossen werden können, um beispielsweise Kollisionen zu verhindern. Da diese Gebiete sich mit der Zeit ändern können, müssen die Pfadplanungen stets dynamisch erfolgen. Die erlaubten Überfahrgebiete lassen sich zur Laufzeit jederzeit neu festlegen. Diese Änderungen dürfen natürlich nur durch autorisiertes Personal durchgeführt werden. Dazu wurden neben der Sicherung der Kommunikation auch Zugriffskonzepte erarbeitet, die auf Schnittstellen aufsetzen, die häufig in IIoT Lösungen zu finden sind [1]. Neben den konzeptionellen Lösungen sollte auch die Standardisierung für die vertikale Integration der industriellen Geräte vorangetrieben werden. So wurde im Rahmen des VDMA eine firmenübergreifende Arbeitsgruppe von Partnern aus dem OPTIMUM-Projekt ins

Leben gerufen. Diese arbeitet an der Erstellung einer OPC UA Companion Specification für Krane.

Die in OPTIMUM erarbeiteten zukunftssträchtigen Konzepte erfordern weitreichende Vernetzung der beteiligten, örtlich verteilten Systemkomponenten. Da diese in der Intralogistik meist beweglich oder mobil sind spielt die Funkkommunikation eine entscheidende Rolle.

2. Herausforderungen für die Funkkommunikation

Um die anlageninternen Geräte für den intelligenten Automatikbetrieb zu nutzen, werden für jedes Gerät Informationen über Situation und Umgebung benötigt. Daher muss jedes Gerät mit einem Netzwerk verbunden sein, um Kontext- und Positionsinformationen, beispielsweise für die angestrebten Assistenzfunktionen, auszutauschen.

Die Netzwerktopologie und das angewandte Protokoll werden je nach Zweck der Systemkonfiguration festgelegt. Beim OPTIMUM-Projekt kamen exemplarisch zwei Vernetzungsarten zum Einsatz, die IIoT-Vernetzung für die Überwachung und die Vernetzung der verteilten Steuerung für das Senden und Empfangen von Daten zwischen den Automatisierungsgeräten. Der konzeptionelle Unterschied zwischen der Kommunikation für verteilte Steuerungen und der IIoT-Vernetzung besteht darin, dass verteilte Steuerungen ein horizontales Netzwerk für die Kommunikation zwischen Automatisierungsgeräten bilden, während das IIoT ein vertikales Netzwerk für die Kommunikation mit überlagerten Geräten bildet. Der funktionale Unterschied besteht darin, dass die IIoT-Vernetzung für **das Senden und Empfangen relativ großer Daten**, wie z. B. Firmware-Updates, verwendet werden soll. Das erfordert eine hohe **Funktionszuverlässigkeit** des Kommunikationssystems, also eine möglichst lange ununterbrochene Klarzeit (siehe auch [2]). Im Fall der verteilten Steuerungen **kommunizieren viele Automatisierungsgeräte mit großer Häufigkeit** untereinander. Das erfordert eine hohe **Verfügbarkeit** des Kommunikationssystems, also eine möglichst geringe Nachrichtenverlustquote.

Je nach Anzahl der Komponenten und den darauf laufenden Anwendungen in einem Automatisierungssystem, muss das **Kommunikationssystem skalierbar** sein. Aufgrund des Zielkonflikts zwischen leistungsfähiger Hardware und Preis sollte ein **möglichst effizientes Kommunikationsprotokoll** zum Einsatz kommen.

In den letzten Jahren wurde die (Cyber-)Security zu einem neuen Hauptziel im industriellen Bereich mit einer Vielzahl von Vorgaben. Dabei konzentriert sich die Laufzeitsicherheit auf die vertrauenswürdige Umgebung, in der eine industrielle Anwendung ohne böswillige Unterbrechung ausgeführt werden kann, während sich die Kommunikations- und Systemsicherheit auf **Vertraulichkeit, Authentifizierung, Autorisierung, Integrität** und in industriellen Anwendungsfällen insbesondere auf **Verfügbarkeit und Verantwortlichkeit** konzentriert.

Für die beschriebenen Vernetzungsarten sind sowohl bestehende Echtzeitprotokolle, die meist auf Schicht 2 des OSI-Referenzmodells operieren, als auch Schicht-3-Modelle zu berücksichtigen. Die Applikation kann unterschiedliche Kommunikationsprotokolle verwenden, z. B. **OPC UA** mit Read/Write/Subscription über TCP, Pub/Sub über UDP oder Pub-Sub über **TSN**.

Es ist naheliegend, dass die Kommunikation in der Intralogistik oft drahtlos erfolgen muss. In Bild 1 ist die Funkkommunikation in einer intralogistischen Anwendung exemplarisch dargestellt. Zu den oben genannten Herausforderungen für die Kommunikation kommen bei der Funkkommunikation noch die **Einflüsse auf die Signalausbreitung** in einer industriellen Umgebung hinzu. Die Bewegung von Funksender und -empfänger ändern laufend die Signalqualität am Empfänger. Selbst die Bewegung der Personen oder anderer, nicht an der Funkkommunikation beteiligten beweglichen Objekte, haben Einfluss auf die Signalausbreitung. Die **Information über die Position** der mobilen Komponenten kann in diesem Zusammenhang hilfreich sein.

Schließlich ist in einem industriellen Umfeld auch das Potenzial für ungewollte und bewusst herbeigeführte Störungen der Funkübertragung zu berücksichtigen.

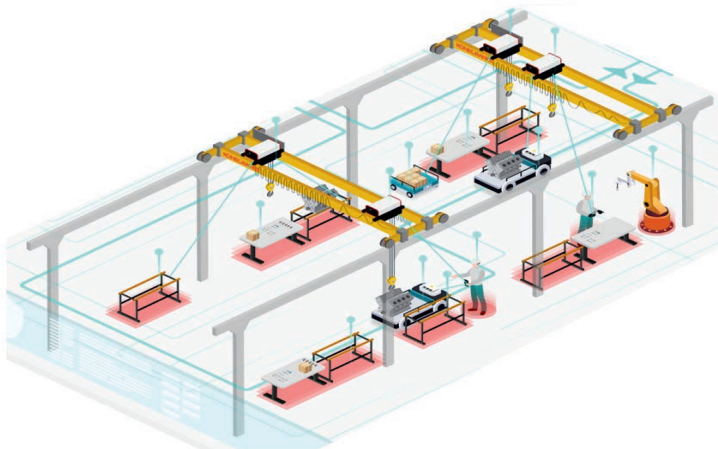


Bild 1: Funkvernetzung einer intralogistischen Anwendung

Funktechnologien wie 5G versprechen eine zuverlässige Vernetzung industrieller Steuerungsanwendungen. Ein Problem stellen jedoch ausgedehnte Funktionstests und Leistungsuntersuchungen unter realen Einsatzbedingungen dar. Diese sind aber erforderlich, um passive

Umgebungseinflüsse (Dämpfung, Mehrwegeausbreitung, Bewegung) und aktive Umgebungseinflüsse (Interferenzen, andere Funkssysteme) auf die Funkkommunikation bewerten zu können. Nur dann kann Vertrauen in die Erfüllung der genannten Herausforderungen geschaffen werden. Reale Produktionsanlagen sind für solche Untersuchungen in der Regel nicht zugänglich, insbesondere nicht bei Kooperationsprojekten mit firmenfremden Partnern. Unter anderem deshalb wurde im Ergebnis des OPTIMUM-Projektes die DEMAG Research Factory mit dem Material Handling Demonstrator (MHD) errichtet.

3. Ziele der Untersuchung industrieller Funkkommunikation

Mit der DEMAG Research Factory wird das strategische Ziel verfolgt, Produkte und Systeme für eine hochvernetzte Produktion im Sinne von Smart Manufacturing und Industrie 4.0 zu entwickeln. Es geht darum nachzuweisen, dass mit der Implementierung der neu entwickelten Konzepte die angestrebte Steigerung der Effizienz der Produktionsabläufe erreicht werden. Hauptziele sind die Realisierung der kollaborativen Konnektivität zwischen verschiedenen Maschinen unterschiedlicher Hersteller sowie die Indoor-Lokalisierung der bodengebundenen und bodenfreien, dynamisch bewegten Maschinen und ihrer Bediener*innen. Darauf aufbauend werden innovative Assistenzfunktionen implementiert.

Ein Aspekt der Forschungs- und Entwicklungsaktivitäten betrifft die 5G-Kommunikation im industriellen Umfeld. Die DEMAG Research Factory bietet mit dem Material Handling Demonstrator eine praxisnahe Plattform zur Untersuchung industrieller Funkkommunikation für die Intralogistik, basierend auf 5G-Technologien. Innovative Assistenzfunktionen für Krane oder fahrerlose Transportfahrzeuge wie "Follow Me", "Go to Machine" und "No Go Areas" erfordern eine zuverlässige Funkkommunikation bei gleichzeitiger Gewährleistung der funktionalen Sicherheit. Sie können auch an Funktionalitäten teilhaben, die das 5G-System bieten kann. Dazu gehört die Synchronisation der kommunizierenden Geräte oder deren Lokalisation in der Fertigungshalle. Weitere Konzepte, die gegenwärtig in der 5G Alliance for Connected Industries and Automation (5G-ACIA) diskutiert werden und deren Praxistauglichkeit noch nachzuweisen sind, betreffen die Integration von OPC UA und 5G sowie die Nutzung von TSN mit 5G [3]. Und schließlich gibt es Überlegungen, Industrie 4.0-Konzepte der Anlagenvirtualisierung auch auf relevante Bestandteile von 5G-Netzwerken anzuwenden [4].

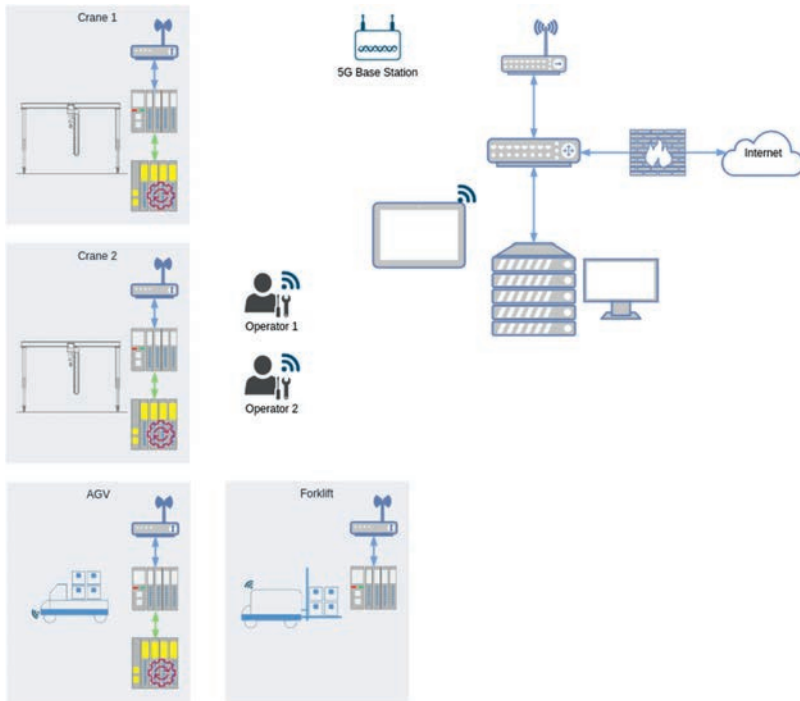


Bild 2: Elemente eines drahtlosen Kommunikationssystems für eine intralogistische Anwendung

Bild 2 zeigt beispielhaft Elemente einer intralogistischen Anwendung, die mit einem 5G-System vernetzt werden. Basierend auf der im OPTIMUM-Projekt entwickelten verteilten Steuerung zwischen Kranen, fahrerlosen Transportsystemen, Gabelstaplern, Bediener*innen usw. können die oben genannten Anforderungen an die Kommunikation evaluiert werden. Es soll aber auch untersucht werden, inwiefern neue Funktionalitäten drahtloser Kommunikationsnetzwerke, wie z. B. die Lokalisierung, zu innovativen Intralogistikkonzepten beitragen können.

Darüber hinaus soll es möglich sein, gezielt Einflüsse auf die 5G-Kommunikation zu untersuchen. Zu diesen Einflüssen gehören beispielsweise:

- verschiedene Datenverkehrsmodelle,
- verschiedene Übertragungsentfernungen mit und ohne Sichtverbindung,
- bewegliche Objekte, die Hindernisse für die Funkkommunikation darstellen,

- verschiedene synchronisierte oder asynchrone Bewegungsszenarien und
- verschiedene 5G Netzwerkkonfigurationen.

Es ist wichtig anzumerken, dass die DEMAG Research Factory eine Plattform für die Evaluierung der fortschreitenden Entwicklung bei den Anwendungen und bei der 5G-Kommunikation ist. Insofern ist der nachfolgend vorgestellte Aufbau eine Momentaufnahme.

4. Aufbau der DEMAG Research Factory

Mit den zentralen Systemen 5G-Netzwerk, Krane, Hebezeuge, FTS in einer Fertigungsumgebung steht mit der DEMAG Research Factory eine Plattform zur Verfügung, um den Nutzen der 5G-Kommunikation zu demonstrieren und den Erfüllungsgrad der Anwendungsanforderungen für 5G-Kommunikation quantitativ zu verifizieren. Bild 3 zeigt einen Blick in die Anlage.



Bild 3: Material Handling Demonstrator in der DEMAG Research Factory

Das nicht-öffentliche 5G-Campus-System in der DEMAG Research Factory ist auf kommerziell verfügbaren Netzwerkkomponenten aufgebaut. Es wird das dafür in Deutschland bereitgestellte Spektrum 3,7 - 3,8 GHz (Frequenzbandbreite 100 MHz) genutzt. Das aktuell installierte 4G-basierte Campusnetz (Release 15) wird gegenwärtig für eine 5G-Stand-alone-Architektur aufgerüstet.

Danach sind funktionale Upgrades geplant, um die Funktionen von 5G-Release 16/17 zu unterstützen. Diese werden eine Verbesserung der Architektur, Funktionalität, Leistung und Zuverlässigkeit der Anwendungen ermöglichen. Das Anwendungsframework Digital Automation Cloud von Nokia unterstützt die verwendeten Industrieprotokolle. Zusätzliche Anwendungen können bei Bedarf eingesetzt werden. Das Testbed wird die Möglichkeit bieten, Feedback und Erfahrungen zu sammeln, wie Parametereinstellungen des 5G-Netzwerks angepasst werden können, um die Leistung des Systems entsprechend den unterschiedlichen Anforderungen und Bedingungen zu optimieren.

Der Material Handling Demonstrator umfasst verschiedene Maschinen, die am Materialfluss in industriellen Prozessen beteiligt sind. In der DEMAG Research Factory sind bereits vorhanden:

- 2 Brückenkrane (manuelle und automatische Betriebsart)
- 4 Laufkatzen, die auf den Kranen fahren
- 1 fahrerloses Transportfahrzeug
- 1 Gabelstapler (manuell betrieben)
- 3 Montagewerkbänke
- 2 Bediengeräte

Für die manuelle Kransteuerung verwenden die Bediener*innen ein drahtloses Bedienpult. Die verschiedenen Maschinen können den Materialfluss in verschiedenen Betriebsarten realisieren, manuell, halbautomatisch und vollautomatisch. Alle Maschinen sind über das 5G-Netz miteinander verbunden. Entsprechend der Anforderungen der zu lösenden Aufgabe, werden verschiedene Kommunikationsbeziehungen zwischen diesen Komponenten etabliert. Die Steuerungsmodule kommunizieren mit anderen Zeitanforderungen und Kommunikationsprotokollen als die Bedienmodule mit den Maschinen. Das genutzte Kommunikationssystem muss in der Lage sein, diese verschiedenen Datenströme zu übertragen und dabei das Funkmedium optimal zu nutzen.

Für die genaue Positionsbestimmung aller beweglichen Komponenten werden gegenwärtig Ultra-Weitband-Lösungen erprobt. Eine künftige Alternative könnte die Lokalisierung mithilfe des 5G-Netzwerkes sein.

Es ist beabsichtigt, die Anzahl der Systemkomponenten verschiedener Hersteller für eine Produktionsumgebung laufend zu erhöhen, den praktischen Nutzen einer weitreichenden Vernetzung und Interoperabilität zwischen Produkten verschiedener Hersteller zu demonstrieren.

Für die genaue quantitative Bewertung der 5G-Kommunikation wird ein Testsystem verwendet, dessen prinzipieller Aufbau in Bild 4 dargestellt ist.

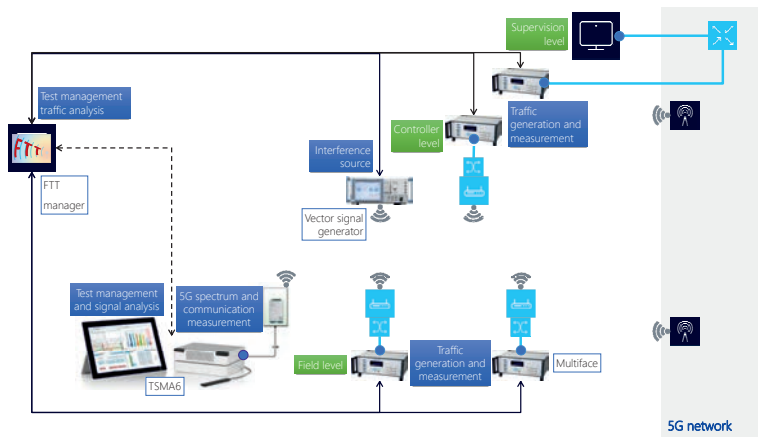


Bild 4: Prinzipieller Testaufbau zur Bewertung der 5G-Kommunikation

Der anwendungsbezogene Datenverkehr wird durch FPGA-basierte Geräte namens Multiface erzeugt [5]. Auf diese Weise kann das in der Steuerungstechnik übliche Zeitverhalten mit hoher Dynamik und hoher Genauigkeit erzeugt werden. Gleichzeitig werden auch Messungen des Zeit- und Fehlerverhaltens des Nachrichtenaustausches mit hoher Auflösung und hoher Genauigkeit durchgeführt. In diesem Aufbau wird der Datenverkehr über Ethernet-Schnittstellen an die Komponenten des 5G-Funksystems übergeben.

Mit Hilfe des FTT-Managers kann der Verkehr in Bezug auf logische Verbindungen, Nutzdateninhalt, Nutzdatenlänge, Datenverkehrsart und Sendezeitabstand frei konfiguriert werden. Die wichtigsten ermittelten Leistungsparameter sind Übertragungszeit, Aktualisierungszeit, Antwortzeit, tatsächlicher Sendezeitabstand, Anzahl der gesendeten Nachrichten, Anzahl der empfangenen Nachrichten, Anzahl der korrekt empfangenen Nachrichten und verschiedene Arten von Übertragungsfehlern. Der FTT-Manager dient auch zur Spezifikation von Testfällen, zur Steuerung des Testablaufs, zur Erfassung der Messdaten, zur Anzeige und zur Analyse der Messergebnisse.

In das FTT-System können zusätzliche Geräte integriert werden, wie z. B. ein Vektorsignalgenerator zur Erzeugung absichtlicher Störungen, um den Einfluss auf die 5G-Kommunikation zu untersuchen.

Mithilfe des Testsystems TSMA6 von Rhode & Schwarz können zusätzlich HF-bezogene Kenngrößen ermittelt werden. Dazu gehören die 5G-Signalabdeckung, das Signal-Rausch-

und-Interferenzverhältnis sowie die Synchronisation von 5G-Signalen von mehreren Zugangspunkten.

5. Beispielanwendung und weitere Schritte

Der Material Handling Demonstrator in der DEMAG Research Factory kann für Untersuchungen auf verschiedenen Ebenen vernetzter Anwendungen in den Bereichen Intralogistik und Material Handling genutzt werden. Zunächst kam der Material Handling Demonstrator gemäß Bild 2 zum Einsatz, um das Zeitverhalten des Datenaustauschs zwischen den verteilten Steuerungsanwendungen zu untersuchen. Zunächst wurden die Messungen für eine Geräteumgebung definiert, die bewusst unabhängig vom Material Handling Demonstrator gewählt wurde [6].

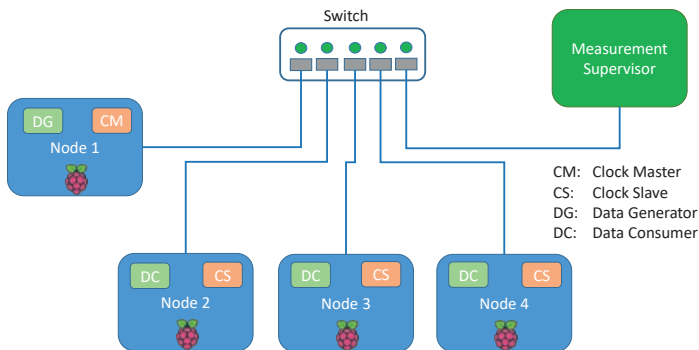


Bild 5: Prinzipieller Testaufbau zur Bewertung der Kommunikationseigenschaften

Die Geräteumgebung des Testsetups entsprechend Bild 5 besteht aus vier Embedded-Geräten, die über ein Kommunikationssystem miteinander verbunden sind. Das Embedded Device besteht aus einem Raspberry PI V1. Das verwendete Betriebssystem ist RASPBIAN STRETCH, erweitert um eine Komponente zur Synchronisation der lokalen Uhren im Netzwerk. Dabei hat einer der Knoten die Rolle des Clock-Masters (CM), die anderen sind Clock-Slaves (CS). Die Taktsynchronisation basiert auf IEEE 1588. Die verteilten Steuerungsanwendungen implementieren Datengeneratoren und Datenkonsumenten. Zusätzlich gibt es noch eine lokale Software zur Erzeugung einer CPU-Last und ein Werkzeug zur Erzeugung von zusätzlichem Netzwerkverkehr.

Für die Untersuchungen wurden verschiedene Szenarien definiert. Diese variieren die Zahl der Datenkonsumenten im System und deren Implementierung in einem oder in mehreren Geräten wie folgt:

1. Datengenerator und ein Datenkonsumenten befinden sich auf demselben Gerät
2. Datengenerator und drei Datenkonsumenten befinden sich auf demselben Gerät
3. Datengenerator und ein Datenkonsument befinden sich auf verschiedenen Geräten
4. Datengenerator und drei Datenkonsumenten befinden sich auf einem anderen Gerät

Weiterhin variieren die Testszenarien wie folgt in Anzahl und Größe der zu übertragenen Nutzdaten:

1. Eine Instanz 4 Oktette Nutzdaten
2. 10 Instanzen als Array 40 Oktette Nutzdaten
3. 100 Instanzen als Array 400 Oktette Nutzdaten
4. 1000 Instanzen als Array 4000 Oktette Nutzdaten

Als Referenz erfolgten diese Untersuchungen bei Vernetzung mit einem drahtgebundenen Ethernet-System. Untersuchungen dieser Szenarien bei Vernetzung über 5G werden Auskunft über die Leistungsfähigkeit der Kommunikationstechnologie geben.

Eine aus Anwendersicht anschaulichere Untersuchung am Material Handling Demonstrator zeigt die Auswirkung innovativer Assistenzfunktionen auf die Fertigungszeit. Die mithilfe der verteilten Steuerungsanwendungen realisierten Funktionen "Come to me", "Go To" und "Follow Machine" verkürzen die Taktzeit eines exemplarischen Montageprozess um ca. 18%. Die DEMAG Research Factory bietet eine optimale Möglichkeit der Vernetzung nicht nur in Bezug auf industrielle Kommunikation, sondern auch auf wissenschaftlicher und wirtschaftlicher Ebene.

- [1] G. Persico, H. Raddatz, M. Tekkalmaz, K. Meisberger, M. Riedl: Sichere Kommunikation für kollaborative Steuerungssysteme, KomMA 2020 -11. Jahreskolloquium "Kommunikation in der Automation", 29.10.2020 - inIT/Ifak - 2020
- [2] Sarah Willmann et. all, "Aspekte der Zuverlässigkeitsbewertung in ZDKI.", Bericht im BMBF-Förderprogramm "IKT 2020 – Zuverlässige drahtlose Kommunikation in der Industrie" (BZKI), V1.1, URL: file:///C:/Users/LRA~1/I-H/AppData/Local/Temp/ZDKI-FG1_Zu-verl%C3%A4ssigkeit_Druck.pdf, 2016
- [3] 5G-ACIA, "Integration of 5G with Time-Sensitive Networking for Industrial Communications.", White Paper, URL: https://5g-acia.org/wp-content/uploads/2021/04/5G-ACIA_IntegrationOf5GWithTime-SensitiveNetworkingForIndustrialCommunications.pdf, 2021
- [4] 5G-ACIA, "Using Digital Twins to Integrate 5G into Production Networks.", White Paper, URL: https://5g-acia.org/wp-content/uploads/2021/05/5G-ACIA_Using_Digital_Twins_to_Integrate_5G_into_Production_Networks-.pdf, 2021
- [5] A. Gnad, L. Rauchhaupt, S. Willmann, T. Oefler: "Verteiltes Testsystem zur Bewertung der Zuverlässigkeit von Kommunikationssystemen für Industrie 4.0 Anwendungen", KomMA 2017, ISBN 978-3-944722-63-4, Magdeburg, Nov. 2017
- [6] M. Riedl, G. Persico, H. Raddatz, D. Tran, M. Tekkalmaz : Kommunikationslösungen für die Integration verteilter Steuerungen in Logistiksysteme, AUTOMATION 2020, 30.06.-01.07.2020 - VDI Verlag GmbH - 2020

Ethernet-APL für hochverfügbare Sicherheitsanwendungen Vorteile für den Lebenszyklus von Prozessanlagen

M. Risser, Dr.-Ing. A. Ziegler, P. Salzmann, Dr.-Ing. M. Roser,
BASF SE, Ludwigshafen am Rhein

Kurzfassung

In diesem Beitrag werden die Vorteile der Ethernet-APL Technologie für hochverfügbare Sicherheitsanwendungen im Lebenszyklus von Prozessanlagen vorgestellt.

Das konservative Agieren in der Prozessautomatisierung, hauptsächlich aufgrund langer Anlagenlaufzeiten, ist allgemein bekannt. In aktuellen Prozessanlagen werden meist zwei Technologien zur Feldanbindung von Automatisierungssystem eingesetzt. Feldbuslösungen zur Regel- und Steuerung sowie klassische 4-20 mA Technik für PLT-Sicherheitsfunktionen. Um den Fortschritt der Digitalisierung nicht zu behindern ist hier jedoch ein Umdenken erforderlich. Es wird zunächst die Ethernet-APL Technologie vorgestellt, um im Folgenden dann die Vorteile in Bezug auf PLT-Sicherheitseinrichtungen näher zu erläutern. Hier wird insbesondere auf Beispiele in verschiedenen Lebenszyklusphasen eingegangen: Design und Installation (Messgenauigkeit, hohe Stückzahlen und geringe Lagerhaltungskosten durch gleiche Feldgeräte), Betrieb und Instandhaltung (ein Anlagenkonzept für safety und non-safety Anwendungen, vorausschauende Wartungsmöglichkeiten durch Zusatzdiagnose) und Anlagenänderungen (schnelle Anlagenanpassungen durch einheitliche Anlageninfrastruktur).

Es wird somit gezeigt, dass durch die Verwendung einer durchgängigen digitalen Kommunikation mit einem Feldgerät für safety und non-safety der Weg der Digitalisierung geebnet wird und gleichzeitig Aufwände reduziert werden können. Dabei werden die noch zu überwindenden Herausforderungen nicht verschwiegen.

1. Motivation

In der Prozessautomatisierung werden, insbesondere aufgrund langer Lebenszyklen und hoher Investitionskosten für neue Anlagen, neue Technologien eher konservativ und langsam adaptiert. Dieser Effekt wird durch die allgemein hohen Anforderungen an PLT-Sicherheitseinrichtungen noch verstärkt.

Dadurch sind in den vergangenen Jahren zwei große Lücken entstanden:

1. Die Prozessautomatisierung ist bei der Digitalisierung und dem Einsatz neuer Technologien im Vergleich zur Fabrik- und Gebäudeautomatisierung ins Hintertreffen geraten.
2. Innerhalb der Prozessautomatisierung entstand ein Technologiebruch zwischen Betriebs- (Feldbus) und Sicherheitseinrichtungen (4-20 mA).

Durch die Einführung von Ethernet-APL bietet sich nun die einzigartige Möglichkeit, beide Lücken gleichzeitig zu schließen.

Ethernet-APL verfügt über alle Eigenschaften die notwendig sind, um eine vollständige digitale Kommunikation zwischen der Feld- (einschließlich Ex-Zonen) und der Automatisierungsebene aufzubauen. Hinzu kommt, dass bei einer Investition in die neue Technologie, direkt von Anfang an auch der Bereich der PLT Sicherheitseinrichtungen mitberücksichtigt werden kann. Wenn heute neue Prozessanlagen gebaut werden, wird für die Integration von Feldgeräten für Prozessleitfunktionen (non-safety) auf die Feldbustechnologie gesetzt. Das bietet nicht nur Kostenvorteile durch geringeren Verkabelungsaufwand und schnellere Inbetriebnahmezeiten, sondern auch eine bessere Signalqualität der Loops. Weiterhin können so die wichtigen Zusatzinformationen aus den Feldgeräten weiterverarbeitet und ausgewertet werden (z.B. über den NOA-Kanal).

Anfängliche Probleme bei der Einführung von Feldbussystemen und eine hohe wahrgenommene Komplexität führten letztendlich zu einer geringen Akzeptanz und somit zum Scheitern in der PLT-Prozesssicherheit. Hier ist nach wie vor die 4-20 mA Technologie dominierend.

Dies hat zur Folge, dass in neuen Prozessanlagen (Greenfield) zwei Technologien geplant und gepflegt werden müssen. Dies bringt viele Nachteile mit sich, die sich negativ auf den Return on Investment (ROI) und die Wartbarkeit moderner Anlagen auswirken.

Weiterhin bieten moderne Feldgeräte, zusätzlich zum eigentlichen Prozessmesswert, eine hohe Anzahl an Diagnoseinformationen an. Diese müssen zur Realisierung vieler vielversprechender Industrie 4.0 Anwendungen ausgelesen- und in übergeordnete Systeme übertragen werden. Für die Anlagenverfügbarkeit sind hier gerade die safety Messstellen die entscheidenden.

Die derzeit noch vorrangig eingesetzte 4-20 mA Technologie limitiert den möglichen Fortschritt. [1]

2. Die Enabler Technologie: Ethernet-APL

Ethernet in Verbindung mit dem Advanced Physical Layer (Ethernet-APL) [2] ermöglicht eine leistungsstarke und durchgängige digitale Kommunikation vom Feld bis ins

Automatisierungssystem. Lange Kabellängen und Explosionsschutz (Zone 2/1/0) durch Eigen-sicherheit, mit Kommunikation und Stromversorgung über 2-Leiter-Technik, sind hier gegeben. Der Einsatz von sicherheitsgerichteten Black-Channel-Protokollen, wie z.B. Profisafe (SIL 3 zertifiziert), ermöglicht die Einbindung von PLT-Feldgeräten, die im Prozesssicherheitsumfeld zum Einsatz kommen. Durch dieses Übertragungsprinzip werden die Infrastrukturkomponenten selbst keine Unterstützung für PLT-Sicherheitsanwendungen benötigen. Diese Eigenschaften ermöglichen eine einheitliche Anlageninfrastruktur für Betriebs- und Sicherheitsfunktionen.

Die nachfolgende Abbildung (Bild 1) beschreibt, wie diese Technologie mit dem Trunk und Spur Prinzip funktioniert:

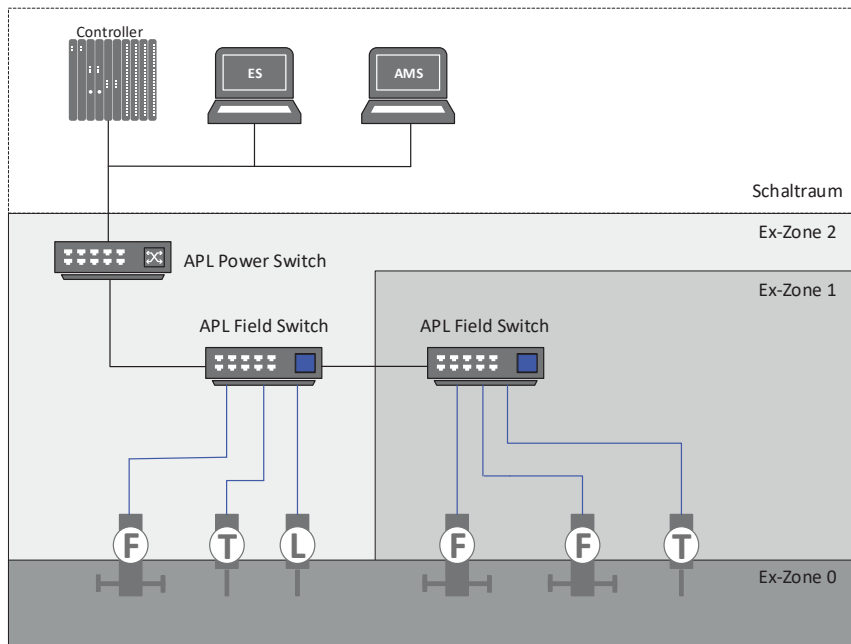


Bild 1: Ethernet-APL mit Trunk-und-Spur Technologie - in Anlehnung an [3]

Ein APL Power Switch ist das Bindeglied zwischen dem Automatisierungsnetzwerk und einem Ethernet-APL-Segment mit 2-Leitern. Mit bis zu 10 Mbits/s bietet dieser eine hohe Bandbreite (insbesondere im Vergleich zu aktuellen Lösungen für die Digitalisierung in der Prozesssicherheit, wie z.B. HART). Weiterhin übernimmt dieser Switch die Anpassung der Baudrate und der

Übertragungsphysik und stellt gleichzeitig die Energie für die APL-Field-Switche zu Verfügung. Die an die APL Field Switche angeschlossenen Feldgeräte werden so mit Energie versorgt. Separat gespeiste APL Field Switche ergänzen die Trunk-und-Spur-Topologie, sodass die APL Power Switche durch klassische Industrial Network Switche ersetzt werden. Die APL Field Switche dienen somit als Bindeglied zwischen Automatisierungssystem und Ethernet-APL, der zudem den Anschluss von Feldgeräten ermöglicht (siehe Bild 3 und 4 in Kapitel „Vorteile im Lebenszyklus von PLT-Sicherheitsfunktionen“). [3]

3. Vorteile im Lebenszyklus von PLT-Sicherheitsfunktionen

Mit dem Einsatz der Ethernet-APL Technologie zwischen Feld- und Automatisierungsebene, kann die Prozessindustrie die vollen Vorteile einer durchgängigen digitalen Kommunikation in der Automatisierungspyramide ausschöpfen.

3.1 Design und Installation

In klassischen sicherheitsgerichteten Sensor-Loops, welche die analoge 4-20 mA Übertragungstechnik nutzen, ist mit erhöhter Messunsicherheit zu rechnen. Das eigentlich schon digitale Nutzsignal im Sensor, muss zur analogen 4-20 mA Datenübertragung mittels D/A-Wandler umgewandelt werden, sodass es überhaupt erst übertragen werden kann. Sobald das Signal am Automatisierungssystem angekommen ist, wird es durch einen A/D-Wandler wieder zu einem digitalen Signal gewandelt. Durch diese Umwandlungen und dem fehlerbehafteten analogen Übertragungsweg, entstehen Fehler für das eigentliche Nutzsignal (vgl. Tabelle 1). Eine weitere zusätzliche Fehlerquelle ist das HART-Protokoll (Aufmodulation auf das 4-20 mA Signal). Wird dies zur Übertragung von zusätzlichen Parametern aus dem Feldgerät heraus verwendet, entsteht i.d.R. ein zusätzlicher Messfehler i.H.v. ca. 1% der Messspanne.

Mit einem ausschließlichen digitalen Übertragungsweg vom Feld- zur Automatisierungsebene können diese Messunsicherheiten vermieden werden. Weiterhin wird die Parametrierung einfacher, da z.B. keine Informationen über Messspannen mehr benötigt werden. Zusätzliche Diagnoseinformationen des Feldgeräts können nun auch einfach weitergeleitet (NOA-Kanal) und ausgewertet werden, ohne die Qualität des Nutzsignals negativ zu beeinflussen.

Insgesamt wird dadurch das mögliche Betriebsfenster vergrößert.

Tabelle 1: Einflüsse auf die Messgenauigkeit verschiedener Übertragungswege

Messkette	Prozess-an-schluss	Transmit-ter-eingang	Signal-ver-arbei-tung	Transmit-ter-ausgang	Speise-trenner	SSPS/PLS IO Modul	SSPS/PLS Logik
Bauteil	PT100	AD-Wandler	Prozessor	DA-Wandler	EX-Trenner	AD-Wand-ler	Prozessor
Signalart	analog	analog	digital	analog	analog	analog	digital
Fehler 4-20mA	mittel	mittel	vernachl.	signifikant	mittel	mittel	vernachl.
Fehler Di-gital	mittel	mittel	vernachl.	entfällt	entfällt	entfällt	vernachl.

Für die Planung bietet die ethernet-basierte Technologie eine hohe Flexibilität bei einer skalierbaren Verfügbarkeit. So können weiterhin klassische Sterntopologien aufgebaut werden, aber auch verfügbarkeitsorientierte Architekturen, wie z.B. eine Ringtopologie, sind ohne großen Aufwand planbar und umsetzbar. Die nachfolgende Abbildung zeigt eine Ringtopologie, welche eine bevorzugte Variante in der Prozessindustrie darstellt:

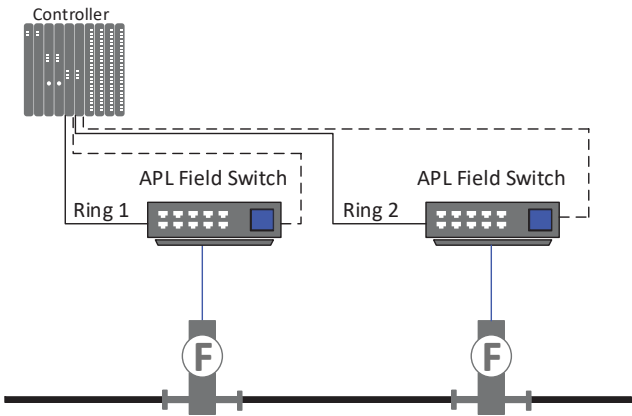


Bild 2: Ringtopologie für eine erhöhte Verfügbarkeit

Ein Aufbau mehrere Ringe, wie in Bild 2 dargestellt, bietet außerdem die Grundlage für redundante Feldstrukturen. So können bspw. 1oo2 und/oder 2oo3 Strukturen auf mehrere Ringe verteilt werden, um auch die Feldverfügbarkeit zu erhöhen.

Bild 3 zeigt einen schematischen Aufbau einer safety-Infrastruktur unter Verwendung von PROFIsafe.

3.2 Betrieb und Instandhaltung

Für den uneingeschränkten Betrieb einer Prozessanlage ist es zwingend erforderlich, dass die (PLT-) Betriebsmannschaft sich mit der Anlagenarchitektur bestmöglich auskennt. Durch die Vereinheitlichung der Architektur der betrieblichen- und sicherheitsrelevanten Messstellen ist ein geringer Schulungsaufwand des Betriebspersonals notwendig. Diese können sich damit vollständig auf eine Technologie konzentrieren.

Mit dem Zugriff auf immer mehr verfügbare Daten der Feldgeräte, können Diagnose- und prädiktive Wartungslösungen verbessert werden. Ein Beispiel ist die Implementierung von differenzierten Fehlerreaktionen basierend auf der Fehlermeldung des Gerätes.

Weiterhin können viele Wartungsaufgaben automatisiert werden. Daraus resultiert eine effizientere Wartbarkeit mit genaueren Wartungsanweisen wie sie bspw. in der NE 107 [4] beschrieben sind.

In Tabelle 2 ist der Vergleich der Diagnoseinformationen der klassischen 4-20 mA Übertragungstechnik mit der modernen digitalen gegenübergestellt.

Tabelle 2: Optimierte Diagnose

Technik	analog (4-20 mA)	digital
Prozesswert	ja	ja
Zusätzliche Prozesswerte	keine, maximal über das HART Protokoll möglich	ja, z.B.: dP: zusätzlich 2 x Absolutdruck CMM: zusätzlich Temperatur und Dichte
Meldung	NE43 Ausfallinformationen	detaillierter Gesundheitszustand
Reaktion	Auslösen des betroffenen Kanals	Differenziertes Auswerten und Reagieren von Alar-men mit Erhalt des Messwertes

Die klassische 4-20 mA Technik bringt neben dem eigentlich Nutzsignal nur wenige Diagnosemöglichkeiten mit. Außer der Ausfallinformationen gemäß NE 43 [5], ist es nur über das HART Protokoll möglich, zusätzliche Informationen aus den Feldgeräten auszulesen. Dieses Vorgehen ist recht aufwändig und es stehen nur bedingt weitere Diagnosemöglichkeiten zur Verfügung [6].

Ein digitales Protokoll kann neben dem Nutzsignal viele weitere Diagnoseparameter übertragen. So stehen z.B. einer Differenzdruckmessung (dP) zusätzlich die beiden Absolutdrücke und bei einem Coriolismassemesser (CMM) stehen zusätzlich noch die Temperatur und die Dichte zur Verfügung. Bei beiden Messprinzipien können diese Zusatzinformationen einfach und direkt über das digitale Protokoll ausgelesen werden.

Bei immer größer werdenden Mengengerüsten (gerade auch in PLT-Sicherheitseinrichtungen), ist dies von besonderer Bedeutung. Dadurch können unerwartete Anlagenstillstandszeiten minimiert, bzw. vermieden- und Abstellungszeiten verkürzt werden.

Abgerundet wird das Konzept mit einer vereinfachten Lagerhaltung, da identische Feldgeräte für Betriebs- und Sicherheitsfunktionen einsetzbar sind.

Zukünftige Anlagenänderungen, insbesondere das Umwidmen von Messstellen von einer betrieblichen- zu einer sicherheitstechnischen Funktion, lassen sich einfacher planen und realisieren. Hierzu sind keine aufwändigen Anlagenumbauten mehr notwendig (vgl. auch Abschnitt 3.3).

3.3 Anlagenänderungen

Weder in der Planungsphase noch während dem weiteren Lebenszyklus einer Prozessanlage ist es ungewöhnlich, dass sich Anforderungen durch neue Informationsquellen ändern. Dies kann dazu führen, dass betriebliche Funktionen zu sicherheitsrelevanten Funktionen werden und umgekehrt. Das macht in den aktuellen Anwendungen (4–20 mA) einen kompletten Austausch der Feldgeräte mit großem Planungs- und Dokumentationsaufwand notwendig. Das ist nicht nur teuer, sondern auch zeitaufwendig. Mit einer einheitlichen (sogar eine gemeinsame Architektur wäre denkbar) Kommunikationsinfrastruktur ist ein Tausch ohne großen Aufwand möglich.

Bild 4 zeigt, wie Feldgeräte aus dem betrieblichen Teil der Anlagen zu sicherheitsgerichteten Betriebselementen werden.

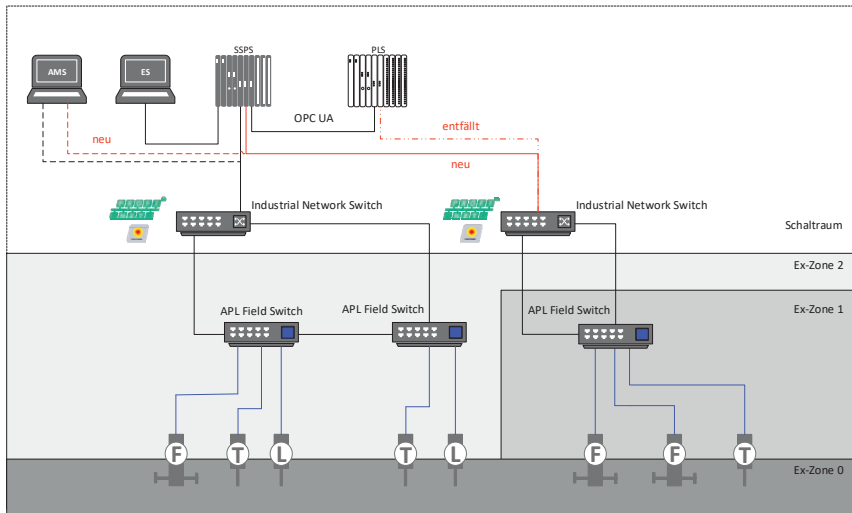


Bild 4: Ethernet-APL: Anlagenänderung - in Anlehnung an [3]

Es können dieselben Komponenten weiterverwendet werden. Es wird lediglich die Verbindung zwischen dem rechten Industrial Network Switch und dem PLS getrennt (Bild 4, rote Strich-Punkt-Punkt-Linie). Diese wird zur weiteren Verwendung im sicherheitsgerichteten Bereich an die SSPS angeschlossen (Bild 4, rote Linie). Die Netzwerkkomponenten können in diesem Fall weiterverwendet werden. Da nun ein Black-Channel-Protokoll zum Einsatz kommt, benötigen diese keine sicherheitstechnische Bewertung. Die Feldgeräte können auch, wie die bereits bestehenden SSPS Feldgeräte, an das AMS angeschlossen werden (Bild 4, rote Strich-Linie).

4. Herausforderungen beim Einsatz von Ethernet-APL

Für eine wirtschaftliche Investition in die Technologie, sollten mindestens 80% der Feldgeräte einer Anlage mit Ethernet-APL verfügbar sein. Um dies zu erreichen, ist eine hohe Akzeptanz am Markt unabdingbar. Durch die Befürwortung der NAMUR ist ein erster wichtiger Meilenstein auf dem Weg zum weltweiten Industriestandard bereits erreicht.

Gängige Messprinzipien müssen verfügbar sein, um eine hohe Abdeckung in Prozessanlagen zu erreichen. Die Feldgeräte müssen nach der IEC 61511 entwickelt sein, um diese in Sicherheitseinrichtungen einsetzen zu können. Damit alle Vorteile, die in den vorhergehenden Kapiteln beschrieben sind, auch umsetzbar sein werden, ist es zwingend erforderlich, dass das selbe Feldgerät für safety und non-safety einsetzbar ist.

Weitere Herausforderungen wird es im Bereich des Geräteschutzes geben. Die Geräte müssen vor Manipulation geschützt werden. Hier bietet sich ein Hard- und/oder Software Schreibschutz an. Die sicherheitsrelevanten Parameter eines Feldgerätes sollten außerdem mit einer Prüfsumme versehen werden, um schnell und zielgerichtet überprüfen zu können, ob diese weiterhin korrekt eingestellt sind. Dies bietet einen Schutz vor versehentliches Verstellen, als auch vor Manipulation der Feldgeräteeinstellungen.

Neben dem eigentlichen Messsignal sind für zukünftige Industrie 4.0 Anwendungen weitere Parameter des Feldgerätes von großer Bedeutung. Es werden azyklische Diagnoseabfragen durchgeführt, die die PLT-Sicherheitsfunktion beeinflussen können. Reine Diagnoseabfragen können bspw. mit aktiviertem Schreibschutz und ohne Beeinflussung der PLT-Sicherheitsfunktion durchgeführt werden. Wohingegen spezielle Testroutinen nur mit deaktiviertem Schreibschutz durchgeführt werden können. Dadurch kann die PLT-Sicherheitsfunktion beeinflusst werden, wenn z.B. ein kurzzeitiges Einfrieren des Messwerts für die Testroutine notwendig ist. Generelle Konzepte zu Diagnoseabfragen müssen erstellt werden.

Ist es notwendig ein Feldgerät zu tauschen, muss dies möglich sein, ohne dass Konfigurationsänderungen erforderlich sind (Beispiel: aktuell verbautes Feldgerät mit der Software Version 1.2, neues Feldgerät mit Software Version 2.4). Gerade bei Sicherheitsanwendungen wäre der Aufwand für die notwendige Dokumentation groß.

Generell wird für die Prozessindustrie Ethernet im Feld neu und gewöhnungsbedürftig sein. Ein weiterer wichtiger Baustein zum Erfolg wird sein, dass die Engineeringeinheiten und die Betriebsmannschaften zielgruppenorientierte Schulungen erhalten. Ohne eine Akzeptanz dieser, kann sich keine neue Technologie etablieren. Weitere begleitende Maßnahmen wie praxisnahe Handbücher zur Installation und der Wartung sind unabdingbar. Eine herstellerübergreifende Inbetriebnahmeprozedur ist ebenso wünschenswert.

5. Zusammenfassung und Ausblick

Ethernet-APL hat das Potential, sich als Lösung für die durchgängige und vor allem einheitliche Technologie der Feldgerätekommunikation durchzusetzen. Damit sind immense wirtschaftliche Vorteile verbunden.

Mit Unterstützung der NAMUR und der Hersteller, mit denen wir in diesem Bestreben zusammenarbeiten, ist BASF sicher, dass diese Technologie der nächste Industriestandard für die Integration von Feldgeräten sein kann und damit die genannten Herausforderungen überwindet.

BASF hat ein hohes Interesse an dieser Technologie. Der Einsatz in kommenden Projekten wird bereits jetzt geprüft und favorisiert. Weiterhin unterstützt BASF die Entwicklungsbemühungen der Hersteller mit Know-how und Erfahrung in der Endanwendung.

Quellen

- [1] Salzmann, Ziegler, Risser, Roser: Closing the GAP – A users view on Ethernet-APL in process safety, IACHEMA 2021
- [2] Ethernet-APL: <https://de.profibus.com/downloads/apl-white-paper>, aufgerufen am 19.04.2021
- [3] Ethernet-APL in the Field for high-availability Safety Applications: <https://www.slideshare.net/basf/ethernetapl-in-the-field-for-highavailability-safety-applications>, aufgerufen am 23.04.2021
- [4] NE 107: Selbstüberwachung und Diagnose von Feldgeräten
- [5] NE 43: Vereinheitlichung des Signalpegels für die Ausfallinformation von digitalen Messumformern mit analogem Ausgangssignal
- [6] Ziegler, Risser, Fischer, Roser: Nutzung erweiterter HART-Feldgerätefunktionalitäten im Lebenszyklus von PLT-Sicherheitseinrichtungen, VDI AUTOMATION 2020

I4.0-Verwaltungsschalen als Wegbereiter für optimierte Engineering- und Betriebsprozesse in der prozessnahen Gebäudeautomation

M. Krüttgen (M.Eng.), Prof. Dr. **J. Müller**, Technische Hochschule Köln

Kurzfassung

Das Paper beschreibt die aktuellen Arbeiten des Arbeitskreis 1.7 „Prozessnahe Gebäudeautomation“ des NAMUR e. V. und der Technische Hochschule Köln (TH Köln). Ziel der Forschungskooperation ist die Standardisierung und informationstechnische Beschreibung von industriell verbreiteten Anlagentypen der prozessnahen Technischen Gebäudeausrüstung (TGA). Grundlage ist das Konzept der Industrie 4.0-Verwaltungsschale, welche alle notwendigen Modellierungselemente zur Abbildung der angestrebten Spezifikationsergebnisse bereitstellt [1]. Zur Reduzierung des Aufwands im Engineering der Prozessführung wird das herstellerübergreifende MTP-Konzept (Module Type Package) angewendet [2]. Anlagenbilder (HMI) sowie Templates für Automationsfunktionen werden mit [3] in der Verwaltungsschale referenziert. Für das Plant Asset Management (PAM) erfolgt eine semantische Spezifikation von Asset-bezogenen Merkmalen für Anwendungen im Dokumentations-, Energie- und Instandhaltungs-Management. Die konfigurationsfreie Bereitstellung von aussagekräftigen und bewertbaren Key Performance Indicators (KPI) vervollständigt das PAM. Ergebnisse werden in einer NAMUR Empfehlung (NE) veröffentlicht. Entwickelte Software-Bibliotheken (HMI, Simulation) werden Open Source zur Verfügung gestellt. Durch die Anwendung von Informationsmodellen der I4.0-Verwaltungsschale und des MTP-Konzeptes werden die fehlenden Standards geschaffen. Eine konfigurationsfreie Bereitstellung von Asset-bezogenen Merkmalen und Fähigkeiten eines Heizkreises ist das Ergebnis dieses Papers, um das Engineering und die Betriebsführung zu optimieren.

1. Einleitung

In vielen technischen Prozessen der Medical Life Sciences, Halbleiter- und Fertigungsindustrie ist die Einhaltung von klimatischen Randbedingungen (Reinraumanforderungen, Temperatur, Feuchte etc.) eine essenzielle Voraussetzung für eine erfolgreiche Produktion. Raumlufttechnische Anlagen (RLT-Anlagen), Komponenten zur Erzeugung und Verteilung von Wärme und Kälte, Raumautomationssysteme etc., sind wesentliche Mittel zur Erreichung der gewünschten Konditionierung von Produktionsräumen. Die Automatisierung der TGA gewinnt daher in Pro-

duktionsprozessen zunehmend an Bedeutung. Trotz des Zeitalters der digitalen Transformation und Industrie 4.0 (I4.0), müssen sich Anwender mit einem aufwändigen händischen Engineering von Asset-Informationen in industriellen Anwendungen arrangieren (Bild 1).

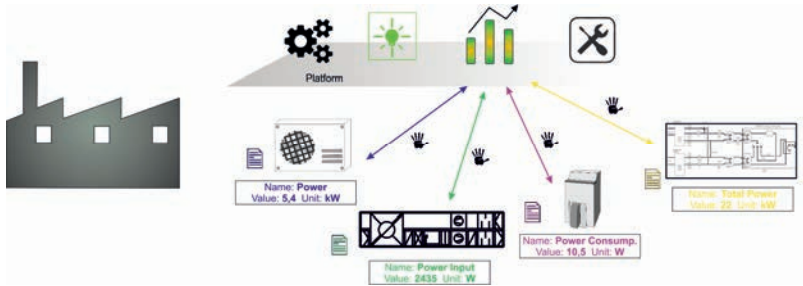


Bild 1: Fehlende oder uneinheitliche Standards für TGA-Anlagen am Beispiel des Energie-Monitorings

Für die informationstechnische Abbildung der prozessnahen TGA und Gebäudeautomation (GA) fehlen grundlegende Festlegungen und Standards. Merkmale oder Funktionen zur einheitlichen Handhabung der Anlagenkomponenten zur Prozessführung, Simulation und des PAM (z. B. Identifikation, Energiewerte, Instandhaltungsrelevante Informationen) sind nicht oder nur in geringem Maße verfügbar [4]. Untersuchungen in der Praxis [5] zeigen, dass aufgrund fehlender Standards, uneinheitlicher Informationsmodelle und einer Vielfalt von Subsystemen (in der Kommunikation und Anwendung), Engineering-Aufwände zur Datenintegration in z. B. Energie- oder Asset Monitoring Anwendungen oftmals höher sind als deren Nutzen. Finden Standardisierungsaktivitäten statt, so entstehen oftmals „Inseln“ von Asset-Typen, welche typübergreifend identische Eigenschaften semantisch und funktional unterschiedlich ausprägen.

Virtuelle Abbilder der TGA-Assets (Digitaler Zwilling, I4.0-Verwaltungsschale) sollen diese im I4.0-Umfeld unter verschiedenen Anwendungsszenarien repräsentieren. Sind diese virtuellen Abbilder vorhanden, so haben diese nur eine herstellerspezifische Ausprägung. Es werden z. B. in der Prozessführung TGA-Assets je nach Hersteller der Automationsstation uneinheitlich modelliert und automatisierungstechnisch unterschiedlich gehandhabt (fehlende Standards für HMI und Automationsfunktionen). Gleiches gilt für digitale Zwillinge von TGA-Anlagen in Simulationsumgebungen sowie standardisierte Merkmale für PAM- Anwendungen.

2. Verwaltungsschalen für die prozessnahe Technische Gebäudeausrüstung

TGA-Komponenten sind bedeutende Assets für die Optimierung von betrieblichen Prozessen, z. B. Energie- und Instandhaltungs-Management. Informationen der TGA-Komponenten müssen hierfür nahtlos in verschiedene PAM- Anwendungen integriert werden (Monitoring von Energie- und Asset-Zustandsinformationen, Dokumentationsmanagement etc.). Das Konzept der I4.0-Verwaltungsschale (engl. Asset Administration Shell, AAS) stellt dafür alle notwendigen Modellierungselemente zur Verfügung. I4.0-Verwaltungsschalen stellen die Verbindung zwischen dem physischen Asset und dem Internet der Dinge und Dienstleistungen dar (I4.0-Komponente bestehend aus einem Asset und seiner Verwaltungsschale [1]). Als virtuelle Abbilder von Anlagenkomponenten repräsentieren I4.0-Verwaltungsschalen die funktionalen Aspekte im Bereich Prozessführung, Simulation und PAM. Eine TGA-Anlage oder -Komponente wird durch standardisierte funktionale Aspekte z. B. Eigenschaften, Konfigurationsparameter, Zustände, Fähigkeiten und angebotene Services repräsentiert. Die maschinenlesbare Bereitstellung dieser funktionalen Aspekte erfolgt durch [3]. Ein zentraler Bestandteil der I4.0-Verwaltungsschale sind Teilmodelle, welche unterschiedlichste Funktionen des Assets (Simulation, HMI, Betrieb, Instandhaltung etc.) mit Hilfe einer strukturierten Menge von standardisierten Merkmalen und Fähigkeiten beschreiben [1].

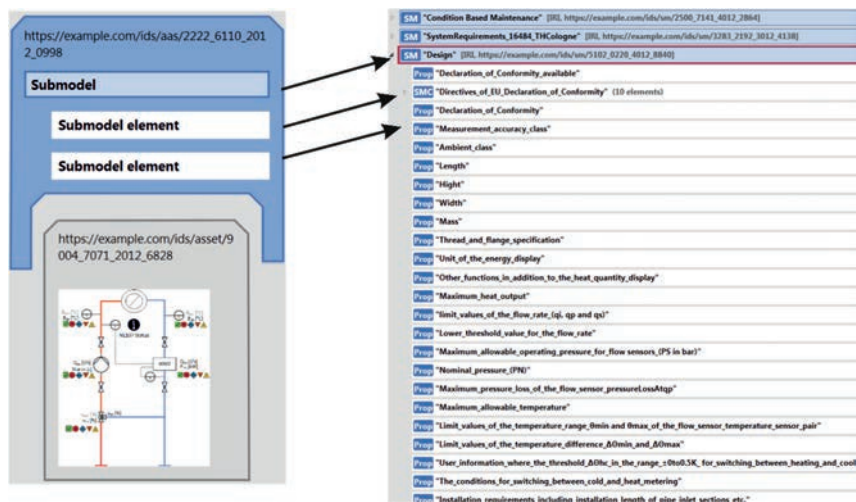


Bild 2: Teilmodell Design am Beispiel eines Heizkreises

Die von der TH Köln entwickelten Teilmodelle für die TGA (u. a. Pumpen) sind grundlegend für die Standardisierung der prozessnahen TGA durch I4.0-Verwaltungsschalen. Das Bild 2

Ein Anlagentyp soll für den Nutzer und Bediener durch die I4.0-Verwaltungsschale einheitlich unter den drei Aspekten der Prozessführung, Simulation und des PAM dargestellt werden (Bild 3). Durch die erwähnten Teilmodelle stehen die Asset-bezogenen Informationen für den optimierten Betrieb zur Verfügung. Standardisierte Merkmale und Funktionen lassen sich den Teilmodellen zuordnen. Der Aspekt der Prozessführung wird durch das MTP-Konzept nach [7] sichergestellt. In den derzeit veröffentlichten Blättern 1-4 [7] werden das Konzept, die Schnittstellen, eine Bibliothek für Datentypen und die Modellierung von Bedienbildern und Moduldiensten beschrieben. Im PAM erfolgt die semantische Beschreibung der TGA-Anlagen für Anwendungen im Dokumentations-, Energie- und Instandhaltungs-Management. Das PAM beinhaltet alle für das Asset notwendige Informationen, um die optimierte Betriebsphase vollständig abzubilden und liefert die Grundlagen für eine wertorientierte Assetführung nach [9]. Diese besagt, dass u. a. die Schaffung einer Wertsteigerung durch Verringerung der Erhaltungskosten und Reduzierung des Ersatzbedarfes zu einer erhöhten Wirtschaftlichkeit eines Assets führen kann. Anstehende und abgeschlossene Wartungen sowie die lückenlose Dokumentation des Assets sind maschinenlesbar wiederzufinden und können dem Bediener durch eine Plattform zur Verfügung gestellt werden [10]. Das Energy Management zeigt die retrospektiven und prognostizierten Energiebedarfe. Letztere werden benötigt, um eine prädiktive und netzdienliche Energieversorgung ermöglichen zu können.

Der Aspekt der Simulation wird durch digitale Zwillinge der jeweiligen Assets abgebildet. Ein digitaler Zwilling lässt sich als Repräsentant eines Assets in Simulationsumgebungen definieren, welcher dessen dynamisches Verhalten simuliert. Für die Erstellung von digitalen Zwillingen gibt es im Bereich der TGA unterschiedliche Softwaretools [11, 12]. Anhand von bereits existierenden Bibliotheken können verknüpfte und komplexe Anlagen- sowie Gebäudesimulationen mit unterschiedlichen Zielvorgaben erstellt werden. Die CARNOT-Bibliothek [13] ist eine Erweiterung der MATLAB-Simulink-Umgebung und ist auf die Anlagen- und Gebäudesimulationen zugeschnitten. Neben MATLAB-Simulink ist im wissenschaftlichen Kontext die Software-Umgebung Modelica bekannt [11]. Für den Einstieg ist OpenModelica als Open Source-Version erhältlich. Für die effiziente Simulationsbearbeitung mit Modelica stehen unterschiedliche Open Source Bibliotheken zur Auswahl, bspw. [14, 15]. Unabhängig von der verwendeten Software, werden die erarbeiteten digitalen Zwillinge in der Verwaltungsschale als Datei verlinkt und für übergeordnete Simulationaufgaben maschinenlesbar zur Verfügung gestellt.

3. Umsetzung in der prozessnahen TGA

Mit dem Arbeitskreis 1.7 des NAMUR e.V. werden Anlagentypen standardisiert, die in der täglichen Praxis immer wieder anzutreffen sind. Die Analyse der prozessnahen TGA und der wesentlichen Anwendungsszenarien mit den darauf aufbauenden Use Cases lassen sich aus Sicht der Anwender der prozessnahen Gebäudeautomation folgendermaßen zusammenfassen. Die Anlagentechnik kann durch standardisierte Merkmale, Funktionen und graphischen Darstellungen bedient werden. Dazu werden Best-Practice-Konfigurationen für optimierte Regelungs- und Automationsfunktionen verwendet. Die Automationsfunktionen werden im Vorfeld der Inbetriebnahme zur Qualitätssicherung durch eine virtuelle Inbetriebnahme nach [16] gegen simulierte Anlagen getestet. Falsche oder fehlende Parameter sowie Ablauffehler im Steuerungscode werden durch die virtuelle Inbetriebnahme sichtbar und das dynamische Verhalten der Anlage wird getestet. Standardisierte Informationen technischer Anlagen werden in PAM-Anwendungen integriert und die Performance von Anlagen wird durch einheitliche Zustandsinformationen und Performance-Indikatoren bewertet. Der erforderliche Engineering-Aufwand für die Konfiguration von Automationsfunktionen und Applikationen wird durch die geschaffenen Spezifikationen reduziert.

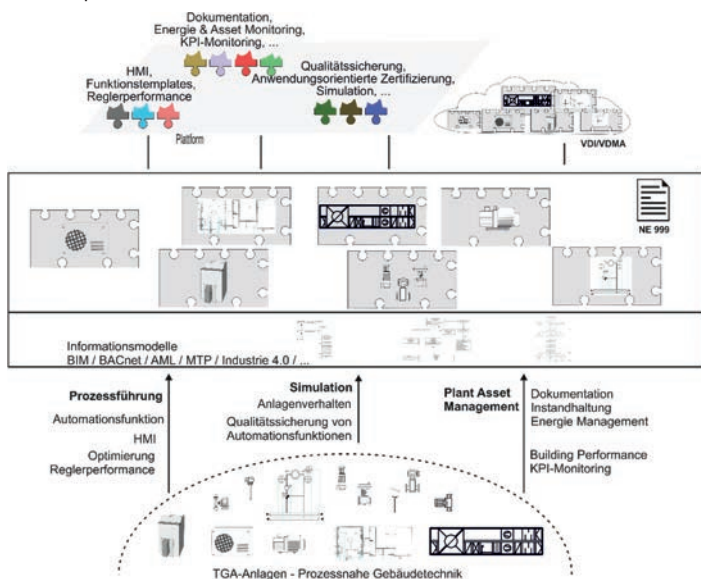


Bild 4: Überblick der Standardisierungsarbeit: Spezifikation von Verwaltungsschalen für die prozessnahe Technische Gebäudeausrüstung

In Bild 4 ist der Überblick zur Umsetzung der Standardisierungsarbeiten gezeigt, um auf die Informationen der prozessnahen TGA zugreifen zu können. Durch die Informationsmodelle werden die Spezifikationen bzgl. Prozessführung, Simulation und PAM maschinenlesbar und konfigurationsfrei zur Verfügung gestellt. Jeder Anlagentyp (RLT-Anlage, Heizkreis, Kältemaschine etc.) sieht für den Anwender gleich aus, unabhängig des gewählten Herstellers. Durch die Implementierung der geschaffenen Standards, kann ohne Engineeringaufwand auf die benötigten Informationen eines Assets zugegriffen werden. Eine erste prototypische Umsetzung der Verwaltungsschale für die prozessnahe TGA befindet sich aktuell im Aufbau und beschreibt zwei standardisierte Heizkreise. Das Bild 5 zeigt die beiden standardisierten Heizkreistypen A und B. Der Heizkreistyp A stellt die klassische Mengenregelung als Drosselschaltung dar, die bei Warmwasser-Systemen, Speicherbe- und entladungen, Fernheizungsanschlüssen, Luftkühlern mit Entfeuchtung, Lufterwärmern ohne Einfriergefahr, Niedertemperaturerzeugern und Wärmepumpen zum Einsatz kommen [17]. Der Heizkreistyp B stellt die Beimisch-Schaltung dar, die bei Heizkörper- und Flächenheizungen, Lufterwärmern mit Einfriergefahr, Niedertemperatur-Wärmeerzeugern und Wärmepumpen zum Einsatz kommen [17].

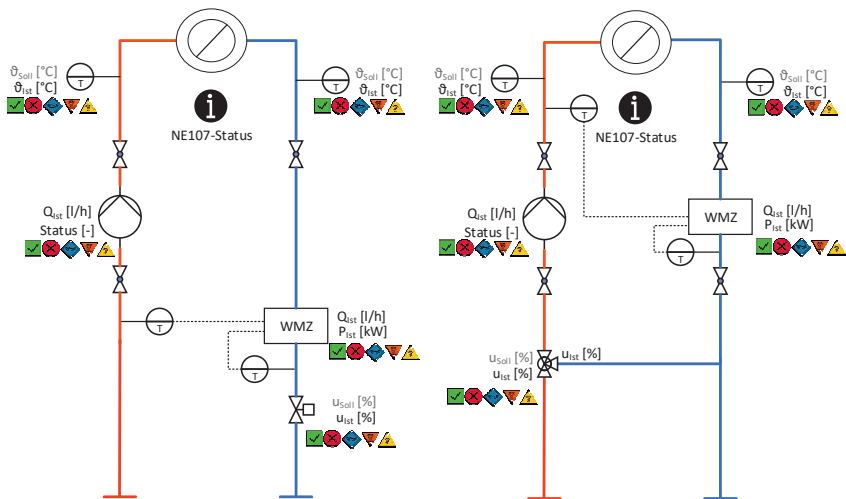


Bild 5: Schematische Darstellung der standardisierten Heizkreistypen A (li.) und B (re.)

Die farblichen Symbole an den Feldgeräten stehen für die Selbstüberwachung und Diagnose nach [18]. Die Selbstüberwachung und Diagnose unterstützt nach [9] die wertorientierte Assetführung. Neben der lagerichtigen und bestimmungsgemäßen Montage des Wärmemengen Zählers, ist die Lage der Fühler und Volumenstrommesseinrichtung für die Genauigkeit

des Messwertes entscheidend. Der Einbauort der Wärmemengenzähler erfolgt nach [19] und entspricht den Empfehlungen der Hersteller, bspw. [20]. Wie in Bild 5 dargestellt, wird die Volumenstrommesseinrichtung des Wärmemengenzählers, mit dem zu jedem Betriebs-Zeitpunkt maximal möglichen Volumenstrom durchfließen. Der auftretende Messfehler, bezogen auf den tatsächlichen Messwert, ist an den hier gezeigten Einbaustellen gering. Diese Details sind für die Standardisierung wichtig, um die Betriebsführung einfach, aussagefähig und messtechnisch genau zu gestalten. Ein ungenauer Messwert mit großer Abweichung ist für die Prozess- und anschließende Performancebewertung unzureichend.

Als Ergebnis aus den technischen Zeichnungen (P&ID) oder Skizzen wie in Bild 5, wurden mittels Engineering-Tools der Hersteller bspw. [21, 22] das HMI mit Funktionstemplates und Prozeduren erstellt. Das Bild 6 zeigt die Darstellung des Heizkreistypen A als HMI inkl. Templates in einer MTP-konformen Software-Umgebung [22]. Der Heizkreis ist dabei als PEA (Process Equipment Assembly, Modulare Prozesseinheit) zu bezeichnen und stellt Dienste im Sinne der Prozessführung zur Verfügung. Die Dienste werden in einem POL (Process Orchestration Layer, Prozessführungsebene) aufgerufen und erhalten in dieser Ebene die Möglichkeit des Eingriffs durch den Bediener oder die PLT (Prozessleittechnik).

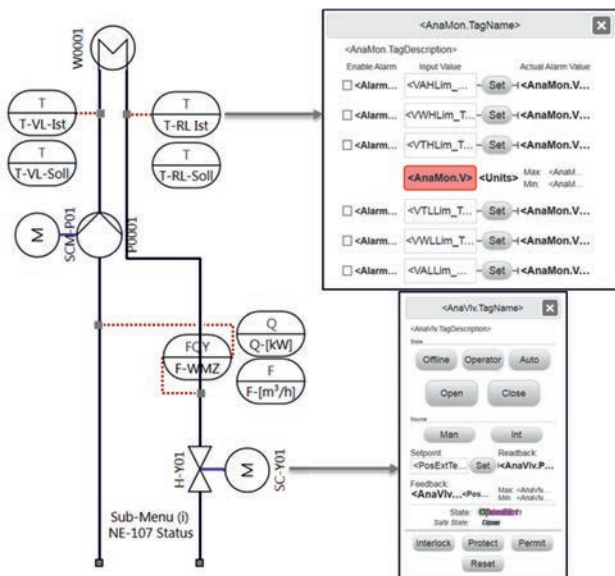


Bild 6: HMI und Funktionstemplates als Beispiel für den Heizkreistyp A

Eine erste prototypische Umsetzung des in Bild 5 und 6 gezeigten Heizkreises erfolgt derzeit an der TH Köln. Neben der Standardisierung der TGA-Anlage und dem hydraulischen Aufbau, ist die Umsetzung des MTP-Konzeptes mit Integration in die Verwaltungsschale der essenzielle Teil der Standardisierungsarbeit. Für die Automatisierung der Heizkreise nach den Vorgaben des MTP-Konzeptes wird das Blatt 4 [7] verwendet. Mittels Zustandsautomaten wird eine dienstbasierte Prozessführung ermöglicht, wodurch der Aufruf und die Ausführung von Funktionen vereinfacht wird. Ein Zustandsautomat führt auf Basis bestimmter Eingangsstatus festdefinierte Funktionen aus. Die dienstbasierte Prozessführung nutzt den Vorteil, dass im Zustandsautomaten nicht jede Feldgeräteansteuerung einzeln verarbeitet wird, sondern ein einzelner Dienst (bspw. Heizen) aufgerufen wird. Dieser Dienstauftrag führt zu einer festdefinierten Abfolge von Funktionen und Abläufen in der Ausführung, bspw. kontinuierliche Sollwertregelung. Der Zustandsautomat kann durch festgelegte Statuswörter (Start, Abbruch, Beenden, Stopp, Fertigstellung etc.) den Dienst in der Ausführung verändern, wie im Bild 7 dargestellt. Erst mit einem definierten Startbefehl, werden Dienste ausgeführt.

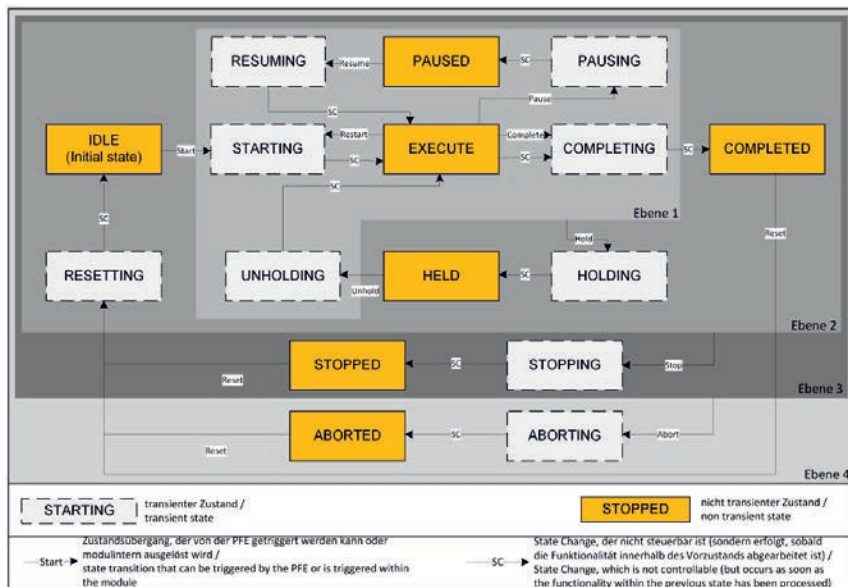


Bild 7: Dienste-Zustandsautomat [7]

Die hier behandelten Module sind stets als integrierbare Module zu verstehen, welche funktional vollständig spezifiziert in eine POL integriert werden. Neben den standardisierten HMI

und Funktionstemplates, wird die erläuterte dienstbasierte Prozessführung mittels Zustandsgraphen nach [7] erarbeitet. Durch den Parametersatz der Prozesswerte werden die übergreifenden Funktionen wie Antiblockierschutz und Frostschutz im Dienst Heizen umgesetzt. Die weiterführende Beschreibung der prototypischen Umsetzung erfolgt nach Fertigstellung des Demonstrators zur Namur Hauptsitzung.

Neben den bisher beschriebenen Prozessführungsaspekten ist die Prozessbewertung mittels KPI ein wichtiger Bestandteil für die Optimierung der Betriebsphase. KPI sind Kennzahlen, die durch einen aussagekräftigen Zahlenwert den Erfüllungsgrad des Prozesses bewerten. Als Beispiel soll der Performance-Indikator Ventilperformance gezeigt werden. Dieser beantwortet die Fragestellung, wie oft und lange das Regelventil größer 85% und kleiner 25%, in Abhängigkeit der Außentemperaturen, angesteuert ist. Der Performancewert gibt Aufschluss über die korrekte Betriebsweise und Auslegung des Ventils in einem Heizkreis. Darüber hinaus entspricht dieser der dauerhaften Performance-Überwachung zur richtigen Dimensionierung einer gebäudetechnischen Anlage nach [23]. Wenn das Ventil größer 85% bei einer Außentemperatur von -5°C und größer geöffnet ist oder kleiner 25% bei einer Außentemperatur von 5°C, dann wird die Laufzeit (T_{Valve}) bei diesem Sollwert gezählt.

$$KPI_{VP} = \frac{(T_{\text{Require}} - T_{\text{Valve}})}{T_{\text{Require}}} \cdot 100\%$$

Ist das Ventil 4.500 Stunden im Jahr in Betrieb (T_{Require}) und arbeitet davon 450 Stunden außerhalb der o. g. Grenzen (T_{Valve}), so ergibt das einen KPI_{VP} von 90%. Bei Ergebnissen unterhalb von 90% sollte das Ventil näher überwacht werden. Bei Ergebnissen unterhalb von 80% ist eine Überprüfung des Ventils, der Auslegung oder des Heizkreises umgehend empfehlenswert.

4. Ausblick

Die beschriebene Vorgehensweise zur Umsetzung eines Heizkreises mittels Informationsmodellierung der I4.0-Verwaltungsschale und des MTP-Konzeptes stellen den Anfang einer Reihe von Standardisierungen im Bereich der prozessnahen TGA dar. Mit der Fertigstellung der prototypischen Umsetzung eines Heizkreises, werden die Ergebnisse auf Kältekreise übertragen. Daran anschließend werden die restlichen Gewerke der Kostengruppe 400 [24] bearbeitet, u. a. Raumluftheiztechnische Anlagen, Sanitärtechnik und Raumautomation.



Bild 8: Zeitstrahl der zu bearbeitenden Anlagentypen

Der Zeitstrahl zeigt im Ergebnis der Standardisierungsarbeit eine Namur Empfehlung. Alle erzielten Ergebnisse der Standardisierungsarbeit werden als Open Source-Dateien zur Verfügung gestellt und veröffentlicht. Die Namur Empfehlung soll die gewonnen Arbeitsergebnisse und Entwicklungen dokumentieren. Neben der Dokumentation der Standardisierungsarbeit, dient die Namur Empfehlung als Anleitung zur Erstellung von Verwaltungsschalen in der prozessnahen TGA.

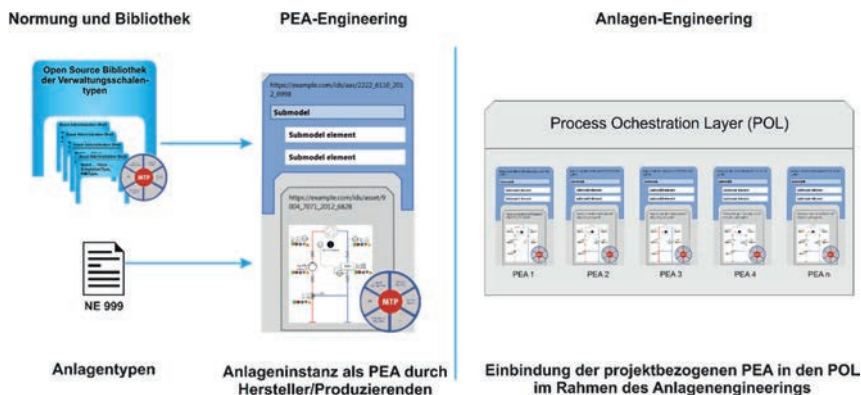


Bild 9: Integration der Forschungsergebnisse in das PEA- und Anlagen-Engineering

Die Namur Empfehlung, ergänzt um eine Open Source-Bibliothek von Anlagentypen als I4.0-Verwaltungsschale, unterstützt bei gemeinsamer Anwendung die Erstellung einer PEA. Diese PEA ist durch eine Verwaltungsschale vollständig beschrieben und beinhaltet ein vollumfängliches MTP-Konzept für den Bereich der Prozessführung und die darin enthaltenen Automations- und Regelungsfunktionen. Mit der fertigen PEA, unter Verwendung der Vorlagen, ist eine Integration in den POL konfigurationsfrei möglich. Bild 9 zeigt den Weg von der Standardisierung bis zum projektspezifischen Anlagenbau mithilfe von standardisierten Anlagentypen.

5. Literaturverzeichnis

- [1] Plattform Industrie 4.0: Details of the Asset Administration Shell, Part 1 V3 - The exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0RC01), November 2020
- [2] NE148: Anforderungen an die Automatisierungstechnik durch die Modularisierung verfahrenstechnischer Anlagen, NAMUR, 2013
- [3] IEC 62714: AutomationML, IEC, 2014
- [4] VDI 3814: Gebäudeautomation, in Überarbeitung, 2020
- [5] Schuhmann et.al.: Aufwand zur Integration von Maschinen in Asset Monitoring Anwendungen, Interne Studie BASF, 2019
- [6] Software: AAS Package Explorer; Copyright (c) 2018-2020 Festo AG & Co. KG et. al.; Version: v2021-02-25; Autoren: Michael Hoffmeister, Andreas Orzelski et. al.
- [7] VDI/VDE/NAMUR 2658 Blatt 1-4: Automatisierungstechnisches Engineering modularer Anlagen in der Prozessindustrie; März 2019-August 2020
- [8] Dr. Matthias Freund et.al.: „Das Module Type Package und die Verwaltungsschale-Ein Ansatz zur Integration“; Automation 2020 (Video-Präsentation), 30.06.2020
- [9] NE129 - NAMUR Empfehlung 129: Plant Asset Management; Version 12.10.2009
- [10] BIM im Betrieb durch lebenszyklusübergreifende Verfügbarkeit von Anlagendaten auf Basis von I4.0-Verwaltungsschalen; M. Ostermeier, R. Ulmer, N. Maisch, M. Both, J. Müller, AUTOMATION 2021, 22. Leitkongress der Mess- und Automatisierungstechnik, Publisher: VDI Verlag GmbH
- [11] Modelica Homepage; <https://www.modelica.org/documents>; Aufruf am 03.03.2021
- [12] IDA Indoor Climate and Energy (IDA ICE); EQUA Solutions AG, Untermüli 3, 6300 Zug (Hauptsitz) Switzerland; <https://equa.se/de/ida-ice>
- [13] SIJ Solar-Institut Jülich der FH Aachen, Solar Campus Jülich, Heinrich-Mußmann-Str. 5, 52428 Jülich, CARNOT Toolbox; <https://www.fh-aachen.de/forschung/solar-institut-juelich/carnot/>
- [14] Modelica Buildings Library: Lawrence Berkeley National Laboratory, <https://simulation-research.lbl.gov/modelica/>; Aufruf am 03.03.2021
- [15] AixLib: RWTH Aachen, <https://www.ebc.eonerc.rwth-aachen.de/cms/E-ON-ERC-EBC/Forschung/OPEN-SOURCE/~modh/AixLib58/>; Aufruf am 03.03.2021
- [16] VDI 3693: Virtuelle Inbetriebnahme, Teil 1: Modellarten und Glossar; August 2016
- [17] Siemens Building Technologies, "Hydraulik in der Gebäudetechnik"; Stand: 01/2019, Artikel: E10003-A38-H147

- [18] NE 107 – NAMUR-Empfehlung 107: Selbstüberwachung und Diagnose von Feldgeräten; Ausgabe: 2017-04-10
- [19] DIN EN 1434-6: Thermische Energiemessgeräte, Teil 6: Einbau, Inbetriebnahme, Überwachung und Wartung; August 2019
- [20] Landis+Gyr GmbH, Humboldtstraße 64, 90459 Nürnberg, Technische Beschreibung für Statische Wärme- und Kältezähler T550 (UH50...); 32 16 000 001 h, 23.08.2018
- [21] Software: *e!COCKPIT*; Wago Kontakttechnik GmbH & Co.KG, Hansastraße 27, 32423 Minden (Verwendete Version: 1.6.1.5)
- [21] Software: *MTP-Designer*; Phoenix Contact GmbH & Co.KG, Dringenauer Str. 30, 31812 Bad Pyrmont (Verwendete Version: 1.7.7412.22059)
- [23] EPBD 2018-European Directive on the Energy Performance of Buildings: Richtlinie (EU) 2018/844 vom 30. Mai 2018;
- [24] DIN 276 - Kosten im Bauwesen; Dezember 2018

Adaptive Massenstromregelung einer Pfannen-Gießmaschine mittels virtueller Sensorik

Patrick Ochudlo, M.Sc., Dr.-Ing. **Sebastian Stemmler**,
Univ.-Prof. Dr.-Ing. **Dirk Abel**,
Institut für Regelungstechnik (IRT) der RWTH Aachen University

Kurzfassung

In diesem Beitrag wird ein Automatisierungssystem für den gravimetrischen Ausguss von Metallschmelze am Beispiel des pfannenbasierten Formgusses vorgestellt. Hierfür wird der aus einer Gießpfanne ausfließenden Massenstrom mittels ein Extended Kalman-Filter (EKF) als Störgröße geschätzt. Zur Regelung des Massenstroms wird ein PI-Regler mit Gain-Scheduling entworfen. Durch das frequenzbasierte Tuning des Reglers mittels des Verfahrens des Open-Loop Shapings folgt der Massenstrom zuverlässig seinem Sollwert, ohne die Schmelze zum Schwappen anzuregen. Zusätzlich zur Massenstromregelung wird fortlaufend die Restkippmenge geschätzt, welche die Schmelzemasse beschreibt, die trägheitsbedingt nach Beginn des Rückkippens aus der Pfanne austritt. Sobald unter Berücksichtigung der prädizierten Restkippmenge das gewünschte Abgussgewicht erreicht ist, wird der Rückkippvorgang eingeleitet. Zur Validierung des Automatisierungssystems werden Versuche mit Eisenschmelze durchgeführt. Hierbei erreichen 91,4 % aller Abgüsse die geforderte Abgussqualität, nur rund 3,1 % der Abgüsse mindern durch signifikantes Unterschreiten des Zielgewichts als Ausschuss die Produktivität. Die Algorithmen zur Umsetzung des Automatisierungssystems werden im Zuge des Rapid Control Prototypings in der Softwareumgebung MATLAB/Simulink entwickelt, validiert und mittels Codegenerierung an einer Versuchsanalage implementiert.

1. Motivation

Maschinenhersteller in Hochlohnländern stehen vor der Herausforderung, ihre Maschinen kundenspezifisch zu optimieren und gleichzeitig Synergien zwischen ihren Entwicklungen zu schaffen. Hierbei müssen a-priori definierte Ziele trotz eines fest kalkulierten Kostenrahmens erreicht werden. Dieses Spannungsfeld wird als das Polylemma der Produktion bezeichnet [4]. Zur Auflösung des Polylemmas bedarf es einer konsequenten Produktivitätssteigerung durch Digitalisierung, simulationsgestützte Systementwicklung und Automatisierung der Maschinen. Diese Entwicklung ist auch im Gießereiwesen zu beobachten. Maschinen sollen individuell und möglichst schnell an die Produktionsbedingungen der Kunden adaptiert werden. Dabei besteht der Wunsch nach immer detaillierteren, teilweise nicht messtechnisch erfassbaren

Prozessinformationen zur Nutzung innovative Datenanalysen. Eine attraktive Möglichkeit zur Befriedigung der Anforderungen stellt die Etablierung von Automatisierungssystemen dar, welche frühzeitig mittels simulationsgestützter Tuningverfahren adaptierbar sind. Exemplarisch wird in diesem Beitrag der gravimetrische Formguss betrachtet, da dieser durch die direkte Formgebung des Gussteils hohe Anforderungen an die Automatisierung stellt.

Ziel des Formgusses ist ein gleichmäßiges Befüllen von Gussformen mit Schmelze. Das Einhalten eines Zielgewichtes ist hierbei von großer Relevanz. Während eine zu geringe Schmelzemasse zu Volumenfehlern im Gussteil und damit produktionsminderndem Ausschuss führt, stellt eine zu große Schmelzemasse Materialverlust und schlimmstenfalls eine Gefährdung des Personals dar. Der Abguss lässt sich für jeden Formguss in charakteristische Phasen aufteilen, die in Bild 1 dargestellt sind. Ausgangspunkt eines Gusses ist eine ruhende Pfanne. Die Pfanne wird angekippt, bis Schmelze aus der Gießschnauze in den Gießtrichter der Gussform fließt. Der Gießtrichter soll zu Beginn des Abgießens möglichst schnell befüllt werden, um ein Füllniveau auszuprägen, das im weiteren Verlauf des Abgusses konstant gehalten wird.

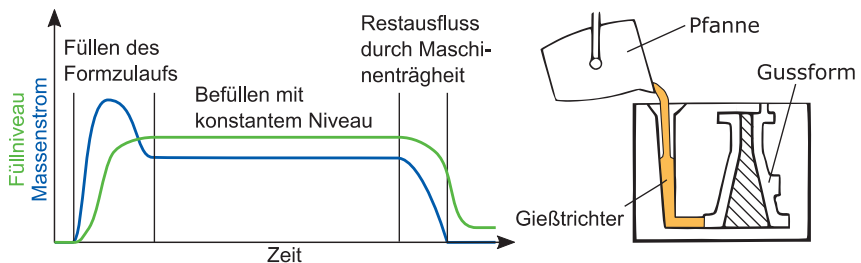


Bild 1: Qualitative Darstellung eines gewünschten Massenstrom- und Niveauverlaufes während eines Abgusses nach [9] (links) und Prinzipskizze eines Abgusses nach [5] (rechts)

Die Form wird durch das konstant hohe Füllniveau schnell, unter gleichmäßigen Bedingungen und ohne einen Einschluss von Luft, was die Entstehung von festigkeitsmindernden Lunkern erhöht, gefüllt [5]. Zum Ende des Abgusses wird die Pfanne zurückgekippt, sodass der Gießstrahl abhängig der aktuellen Maschinenbewegungen und des aktuellen Gießzustand verzögert abreißt und eine Restkippmenge aus der Pfanne ausfließt. Diese Restkippmenge muss bei der Bestimmung des optimalen Rückkippszeitpunktes berücksichtigt werden, um das Zielgewicht möglichst präzise zu erreichen [9]. Das Zielgewicht ist dabei mit einem kleinen Massenspuffer versehen, welcher als Reserve im Gießtrichter verbleibt.

2. Stand der Technik

Im industriellen Umfeld werden viele Gießprozesse noch immer gesteuert betrieben, wobei die Maschineneinstellungen so lange variiert werden, bis sich der gewünschte Massenstromverlauf aus Bild 1 einstellt. Dementsprechend hängt die Einrichtdauer maßgeblich von der Erfahrung des Maschinenführers ab. Für die Produktion großer Chargen mit konstanten Prozessbedingungen können so sehr gute Ergebnisse erzielt werden. Zur Produktion kleiner Losgrößen ist eine Steuerung jedoch problematisch, da hierdurch eine Vielzahl an Gießverlaufskurven für verschiedene Gussformen ermittelt werden müssen. Zusätzlich ändert sich das Ausflussverhalten durch variierende Prozessbedingungen wie einer Viskositätszunahme der abkühlenden Schmelze sowie Ablagerungen erstarrter Schmelze in der Pfanne und an der Gießschнауze. Somit besteht die Gefahr einer kontinuierlich abnehmenden Qualität der Abgüsse bei gesteuerten Ansätzen. Zur Verdeutlichung des dominanten Einflusses von Störgrößen im Gießvorgang wird in [9] eine Steuerung anhand der Invertierung eines Gießmodells mit einer Zwei-Freiheitsgrad-Regelung verglichen. Hierbei zeigt die Steuerung bei Versuchen mit Wasser bei störungsfreier Prozessführung zufriedenstellende Ergebnisse, die Abgussqualität bei gestörten Prozessbedingungen sinkt jedoch schon kurz nach der Kalibrierung rapide. Beim Gießen mit Schmelze ist der Prozess deutlich erhöhten Störeinflüssen ausgesetzt. Daher ist eine Regelung für Abgüsse mit gleichmäßig hoher Gießqualität notwendig. Regelgröße des in [9] entwickelten Regelungssystems stellt das Füllniveau im Gießtrichter dar, als Messgröße dient der durch ein Lasersystem gemessene Füllstand im Gießtrichter. Der Regelung der Füllstandshöhe ist eine Restkippmengenprädiktion überlagert. Die Prädiktion der Restkippmenge erfolgt hierbei anhand der Ausflussmodelle, welche für die Steuerung des Prozesses verwendet werden. Überschreitet die Summe aus abgegossenem Gewicht und der Restkippmengenprädiktion das Zielgewicht, wird das Rückkippen eingeleitet. Im Gießprozess ist die Messung der Füllstandshöhe im Gießtrichter nur schwer umzusetzen, da Störungen das Messsignal stark beeinflussen. Daneben erfordert dieses Vorgehen adäquate Modelle, welche das Übertragungsverhalten zwischen der Kippbewegung der Pfanne und der Füllhöhe des Gießtrichters beschreiben. Hierfür sind zeitintensive Identifikationsversuche erforderlich. Zur Reduktion des Identifizierungsaufwandes wird von [6] ein nichtlineares Ausflussmodell unter Betrachtung hydrodynamischer Effekte entwickelt. Die Autoren in [3] erweitern das Ausflussmodell um den Auftreffpunkt des Gießstrahls im Gießtrichter, indem der Ausfluss des Gießstrahls durch einen freien Fall der Schmelze modelliert wird. Mittels eines Kamerasystems wird der Auftreffpunkt des Gießstrahls gemessen und so eine Regelung des Auftreffpunkts umgesetzt. Das akkurate Treffen des Gießtrichters bildet eine notwendige Voraussetzung für einen akkuraten Abguss. Ein hinreichendes Qualitätskriterium im Gießprozess stellt der Auftreffpunkt jedoch nicht dar,

da das Abgussgewicht nicht direkt betrachtet wird. Für ein Automatisierungskonzept im Pfan-nengießprozess wird eine Messgröße benötigt, welche unabhängig von örtlichen Einschrän-kungen Informationen über das Gewicht der Schmelze liefert. Ein solches Messsystem stellt die Gewichtsmessung der Pfanne mittels eines Wägezellensystems dar. Zu diesem Zweck wird in [1] ein Extended Kalman-Filter (EKF) zur Massenstromschätzung basierend auf dieser Gewichtsmessung entwickelt. Dieses EKF wird in [10] zur Umsetzung einer Massenstromre-gelung genutzt. Überlagert schätzt eine Restkippmengenschätzung basierend auf den Zustän-den des EKFs die Restkippmenge und leitet das Rückkippen der Pfanne ein.

Als nachteilhaft an den bisherigen Entwicklungen wird die fehlende Adaptivität der Ausfluss-modelle an sich fortlaufend ändernde Prozessbedingungen betrachtet. Durch Ablagerungen in der Pfanne und Gießschnauze verändert sich das Ausflussverhalten der Schmelze fortlau-fend. Auch führt die Schmelze in der Pfanne eine komplexe Eigenbewegung aus, welche kom-plex zu modellierende Ausschläge auf der Wägezelle hervorruft. Daher wird in dieser Veröf-fentlichung ein Automatisierungssystem des Gießprozesses ohne hydrodynamische Modellie-rung des Massenstroms vorgestellt. In Kapitel 3 wird das zugrundeliegende Prozessmodell beschrieben und in Kapitel 4 wird ein Extended Kalman-Filter (EKF) basierend auf dem Pro-zessmodell entwickelt. Mithilfe der Messgrößen y_{Mess} aus dem Gießprozess schätzt das EKF hierbei Zustandsgrößen wie den Kippwinkel φ der Pfanne und den Massenstrom $\dot{m}_S$. In Kapi-tel 5 wird ein PI-Regler mit Gain-Scheduling entworfen, welcher den Massenstrom auf Basis eines Sollmassenstroms $\dot{m}_{S,\text{Soll}}$ regelt, ohne Eigenbewegungen der Schmelze anzuregen. Stellgröße stellt die Sollwinkelgeschwindigkeit ω_{Soll} dar. Zur Regelung des Abgussgewichtes wird in Kapitel 6 die Schätzung der Restkippmenge mittels eines Regressionsmodells vorge-stellt. Auf Basis der Restkippmengenschätzung wird das Rückkippen mit der Rückkippgeschwindigkeit $\omega_{\text{Rück}}$ eingeleitet, sobald die Summe aus abgegossener Schmelzemasse und prädi-zierter Restkippmenge das Zielgewicht $m_{S,\text{Soll}}$ überschreitet. Bild 2 zeigt das damit resul-tierende Gesamtsystem, welches in Kapitel 7 experimentell validiert wird. Abschließend wird in Kapitel 8 der Entwicklungsprozess des Automatisierungssystems erläutert, mithilfe welchen das System in einer modernen Simulationsumgebung entwickelt und bei der Inbetriebnahme unkompliziert in eine speicherprogrammierbare Steuerung (SPS) integriert wird.

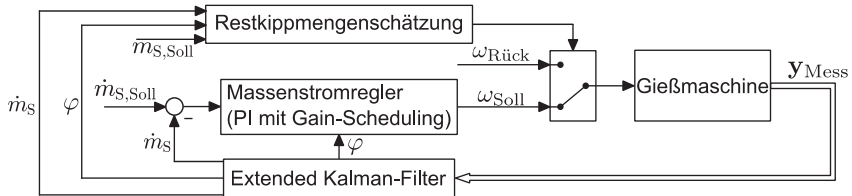


Bild 2: Wirkungsplan des vorgeschlagenen Ansatzes zur Regelung des Pfannengießens

3. Modellierung

Durch die fehlende Möglichkeit einer Messung des Massenstroms wird ein Prozessmodell zur Beschreibung des Massenstroms benötigt. Der hier vorgestellte Ansatz verzichtet auf eine hydrodynamische Modellierung. Grund hierfür ist die komplexe Bewegungsabfolge beim Gießen, welche Eigenbewegungen der Schmelze hervorruft, und über Gießzyklen hinweg entstehende Schmelzeablagerungen in der Gießpfanne und an der Gießschnauze. Durch die Schmelzeablagerungen eingebrachte Querschnittsänderungen, Turbulenzen und Ausflussumlenkungen machen die Modellierung des Schmelzeausflusses komplex und mit Unsicherheit behaftet.

Die zur Massenstromregelung relevante Dynamik des Gießprozesses setzt sich aus einer rotatorischen und einer translatorischen Bewegung zusammen, die antriebsseitig geregelt werden. Beide Bewegungen üben Trägheitskräfte auf die Pfanne und damit einen direkten Einfluss auf das Wägezellsignal aus. Die rotatorische Bewegung der Antriebe wird mittels eines Verzögerungsgliedes zweiter Ordnung mit der Sollwinkelgeschwindigkeit ω_{Soll} als Eingang und der gemessenen Winkelgeschwindigkeit ω als Ausgang durch

$$\ddot{\omega} = -p_{\omega,1} \dot{\omega} - p_{\omega,0} \omega + p_{\omega,0} \omega_{\text{Soll}} \quad (1)$$

beschrieben. Die Parameter $p_{\omega,1}$ und $p_{\omega,0}$ werden mithilfe von Identifikationsversuchen einmalig für die Maschine bestimmt. Die translatorische Beschleunigung $\ddot{z}$ der Gießpfanne ist direkt über eine Sollbeschleunigung $\ddot{z}_{\text{Soll}}$ vorgegeben.

Wird die Gussform mit Schmelze befüllt, soll der Schmelzevorrat im Gießtrichter konstant auf einem hohen Füllniveau gehalten werden. Dies geht mit der Bedingung einher, dass während des gesamten Gießprozesses ein kontinuierlicher Gießstrahl aus der Pfanne tritt und der Gießstrahl den Gießtrichter ausreichend zentral trifft, um ein Verspritzen der Schmelze zu vermeiden. Ein reproduzierbarer Auftreffpunkt der Schmelze im Prozess wird erreicht, indem die Pfanne um einen virtuellen Drehpunkt vor der Gießschnauze gedreht wird. Hierdurch bleibt die Position der Gießschnauze während des Gießvorgangs relativ zur Position des Gießtrichters konstant. Eine Rotation der Pfanne führt somit zu einer vertikalen Bewegung der Pfanne. Die

Bewegung der z-Achse wird mithilfe des Abstandes r_y zwischen Pfannenaufnahme und Gießschneuze in y-Richtung bzw. mithilfe des Abstandes r_z in z-Richtung, dem Pfannenwinkel φ und der Winkelgeschwindigkeit ω durch

$$\dot{z} = \begin{cases} \sqrt{r_y^2 + r_z^2} \left[\dot{\omega} \sin\left(\varphi + \arctan\left(\frac{r_y}{r_z}\right)\right) + \omega^2 \cos\left(\varphi + \arctan\left(\frac{r_y}{r_z}\right)\right) \right] & \text{während des Gießens} \\ 0 & \text{zwischen Abgüssen} \end{cases} \quad (2)$$

beschrieben. Die Informationen über die Dynamik werden zur Trägheitskompensation des Wägezellensignals verwendet. Hierzu wird die auf die Wägezelle wirkende Trägheitskraft und Gewichtskraft in z-Richtung modelliert. Trägheitskräfte durch die Rotation der Pfanne werden aufgrund der geringen Winkelbeschleunigungen im Prozess vernachlässigt. Hieraus ergibt sich die resultierende Kraft

$$F = [(m_S + m_P + m_{Tot}) \left(g + \frac{d}{dt} \dot{z}\right) - \dot{m}_S \dot{z}] \quad (3)$$

auf die Wägezelle. Sie ist abhängig von der Gravitationskonstanten g , der Bewegung in z-Richtung, der Schmelzmasse m_S , der Pfannenmasse m_P und der Totlast m_{Tot} , welche prozessbedingte Massen beschreibt, die sich zwischen dem Wägesystem und der Pfanne befinden. Die Dynamik der Wägezelle wird als Verzögerungsglied zweiter Ordnung

$$\ddot{F}_{Mess} = -p_{F,1} \dot{F}_{Mess} - p_{F,0} F_{Mess} + p_{F,0} F, \quad (4)$$

mit der wirkenden Kraft F und der gemessenen Kraft F_{Mess} , modelliert. Die Parameter $p_{F,1}$ und $p_{F,0}$ werden mithilfe von Identifikationsversuchen ermittelt. Das Gesamtprozessmodell ergibt sich unter Nutzung von Gl. (1) bis Gl. (4) zu:

$$\frac{d}{dt} \begin{pmatrix} \varphi \\ \omega \\ \dot{\omega} \\ z \\ \dot{z} \\ m_S \\ F_{Mess} \\ \dot{F}_{Mess} \\ \dot{m}_S \end{pmatrix} = \begin{pmatrix} \omega \\ \dot{\omega} \\ -p_{\omega,1} \dot{\omega} - p_{\omega,0} \omega + p_{\omega,0} \omega_{Soll} \\ \dot{z} \\ \ddot{z} + \ddot{z}_{Soll} \\ \dot{m}_S \\ F_{Mess} \\ -p_{F,1} \dot{F}_{Mess} - p_{F,0} F_{Mess} + p_{F,0} [(m_S + m_P + m_{Tot}) \left(g + \frac{d}{dt} \dot{z}\right) - \dot{m}_S \dot{z}] \\ 0 \end{pmatrix}. \quad (5)$$

Die Modellierung des Massenstroms wird ohne Informationen über seine Dynamik durchgeführt. Daher wird im Modell die Annahme getroffen, der Massenstrom sei konstant. Die Modellierung des Massenstroms mithilfe dieser Annahme verhindert eine Berechnung rein auf Basis des Prozessmodells. Mittels der Gewichtsmessung durch die Wägezellen ermöglicht jedoch ein Zustandsschätzer die Schätzung des Massenstroms in Abhängigkeit des Gewichtssignals.

4. Extended Kalman-Filter

Mithilfe des vorgestellten Prozessmodells wird ein Extended Kalman-Filter (EKF) entwickelt, um den austretenden Massenstrom zu schätzen. Der Schätzung einer Zustandsgröße durch das EKF wird in Bild 3 veranschaulicht. Die Ausgangssituation stellt die Schätzung des Zustands bestehend aus ihrem Mittelwert und ihrer Standardabweichung im letzten Zeitschritt dar. Mittels des Runge-Kutta Verfahrens 4.Ordnung wird eine nichtlineare Prädiktion der Verteilung anhand eines Prozessmodells vorgenommen. Anhand des aktuellen Messwerts und der a-priori festgesetzten Messunsicherheit wird die Verteilung der Messung festgelegt. Anhand der Informationen durch Prädiktion und Messung wird so der Zustand korrigiert [7]. Der Messgrößenvektor $\mathbf{y}_{\text{Mess}}$ setzt sich aus Messgrößen der Gießmaschine und dem Pfannengewicht m_{Mess} inklusive Schmelze zusammen:

$$\mathbf{y}_{\text{Mess}} = (\varphi \quad \omega \quad z \quad \dot{z} \quad m_{\text{Mess}})^T. \quad (6)$$

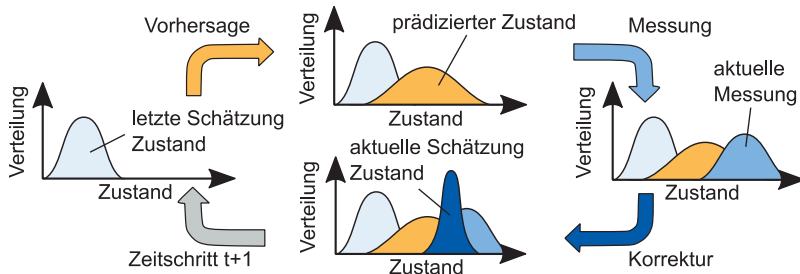


Bild 3: Schematische Darstellung der Funktionsweise eines Kalman-Filters [7]

Hierbei findet die Messung über den gesamten Gießvorgang statt. Die ersten vier Messgrößen sind im Prozessmodell beschrieben, das Pfannengewicht hingegen ist nicht in Gl (5) enthalten. Das Pfannengewicht wird das von der Wägezelle gemessene Kraftsignal um den Gewichtseinfluss der Totlast korrigiert und in ein Gewichtssignal transformiert. Es lässt sich mithilfe von

$$m_{\text{Mess,EKF}} = \frac{F_{\text{Mess}}}{g} - m_{\text{Tot}} \quad (7)$$

berechnen. Hierdurch ist ein Update der gewichtsbezogenen Zustände anhand des Wägezellensignals möglich. Wie in Bild 1 dargestellt unterscheidet sich der benötigte Massenstrom abhängig von der Phase im Gießprozesses, wodurch in den verschiedenen Phasen unterschiedliche Bewegungen ausgeführt werden. Diese Bewegungen führen neben Trägheitskräften zu veränderten Störgrößen auf dem Wägezellensignal. Dieses varierte Messrauschen wird durch eine Adaption der Messkovarianzmatrix im EKF kompensiert. Hierfür wird die Messkovarianzmatrix manuell für verschiedene Phasen des Gießprozesses parametrisiert. Durch die

Information über die aktuelle Gießphase wird zwischen den jeweiligen Werten umgeschaltet. Die Schätzung des Massenstroms wird beispielhaft an einer Abfolge von zwei Abgüssen in Bild 4 verdeutlicht. Das Wägezellsignal (hellblau) zeigt zwischen den Abgüssen ein deutlich verstärktes Rauschverhalten. Der Ausschlag der Wägezelle geht auf die Bewegung der Formbahn zurück, welche die nächste Gussform vor die Gießmaschine fährt. Diese Bewegung ruft Erschütterungen auf der Wägezelle hervor. In Bereichen hoher Dynamik, welche vorwiegend zu Beginn des Abgusses und während des Rückklippens stattfinden, zeigt der Massenstrom kurzzeitig hohe Ausschläge. Die Ausschläge sind in der Berücksichtigung der Messung begründet und können vom EKF aufgrund der Modellierung des Massenstroms als Störgröße nicht gänzlich entfernt werden. Da die Regelung des Massenstroms in diesen Bereich nicht aktiv ist, verringern die Ausschläge die Regelgüte jedoch nicht.

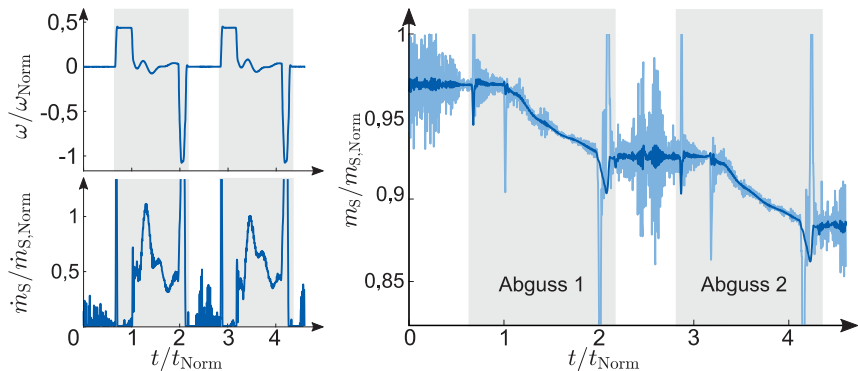


Bild 4: Normierter Verlauf der Winkelgeschwindigkeit (links oben), des geschätzten Massenstroms (links unten) und des Massensignals (rechts) über zwei Abgüsse

5. Massenstromregelung

In der ersten Phase des Gießvorgangs wird die Maschine gesteuert gekippt, bis Schmelze aus der Gießpfanne austritt. Der initiale Ausfluss von Schmelze aus der Gießschnauze wird von einem Kamerasystem erkannt und ein Aktivierungssignal an den Massenstromregler gesendet. Anschließend wird der Massenstrom mithilfe eines PI-Reglers mit Gain-Scheduling geregelt. Als Stellgröße wird, wie in Bild 2 dargestellt, die Sollwinkelgeschwindigkeit der Gießmaschine genutzt, die wiederum von der unterlagerten Antriebsregelung umgesetzt wird. Das baulich bedingte nichtlineare Ausflussverhalten der Pfanne in Abhängigkeit des Kippwinkels macht die Definition der Regelparameter für mehrere Stützstellen notwendig. Daher wird die

Auslegung des Reglers in 5° Schritten des Kippwinkels durchgeführt. Zwischen diesen diskreten Werten werden die Regelparameter interpoliert. Zur Parametrierung des Reglers wird das Verfahren des Open-Loop Shapings verwendet [2]. Ziel der Parametrierung ist es, dass der Regler schnell auf eine niederfrequente Regelabweichung reagiert, hochfrequente Massenstromänderungen durch Eigenbewegungen der Schmelze und durch Prozessrauschen jedoch vernachlässigt. Durch die Trägheit der Maschine ist der Regler nicht in der Lage, solche hochfrequenten Störgrößen zu kompensieren. Vielmehr erzeugt eine Reaktion auf diese Störeinflüsse hohe Winkelbeschleunigungen, welche Schwappbewegungen der Schmelze in Ausflussrichtung anregen. Hierdurch werden Wellen zur Gießschnauze transportiert. Ein Wellenberg an der Gießschnauze hebt den Massenstrom schlagartig an, ein Wellental senkt den Massenstrom ab. Die zeitverzögerte Reaktion des Reglers auf diese lokalen Ereignisse beunruhigt den Gießprozess stark und soll vermieden werden. Zur systematischen Regelparametrierung ist ein Zusammenhang zwischen der gemessenen Winkelgeschwindigkeit und dem durch das EKF geschätzten Massenstrom zu modellieren. Auf Basis experimenteller Daten wird ein Streckenmodell der Gießmaschine identifiziert. Da der Massenstrom vom EKF geschätzt wird, wird die Dynamik des EKFs implizit in der Identifikation berücksichtigt. Der offene Regelkreis, bestehend aus dem identifizierten Streckenverhalten und der Reglerstruktur, wird aufgestellt und ein Wunschverhalten für diesen vorgegeben. So werden die Regelparameter systematisch anhand des Frequenzgangs ausgelegt. Der Erfolg der Regelung ist in Bild 5 beispielhaft an einem Abguss dargestellt. Der Sollmassenstrom wird zum Füllen des Gießtrichters zunächst höher gewählt als für ein konstantes Füllniveau benötigt und nach kurzer Zeit abgesenkt. Der Massenstrom zeigt den in Bild 1 abgebildeten gewünschten Verlauf. Nach einem initialen Füllen des Gießtrichters wird der Massenstrom abgesenkt und so in Näherung das Schmelzeniveau im Gießtrichter konstant gehalten. Die niederfrequente Auslegung des Reglers verhindert das akkurate Folgen dynamischer Massenstromänderungen. Dies wird akzeptiert, da die Regelung ruhig agiert, kein Schmelzeschwappen angeregt und die Regelgüte zufriedenstellend ist. Im Bereich des Rückklippens kommt es durch ein Fehlen von Prozessinformationen zu einem kurzzeitig starken Anstieg in der Schätzung des Massenstroms. Da das Rückklippen gesteuert stattfindet, wird die Regelgüte nicht beeinflusst.

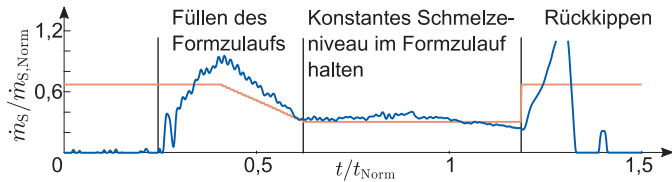


Bild 5: Beispielhafter normierter Verlauf des Massenstroms (blau) geregelt anhand des Sollmassenstroms(orange)

6. Regelung des Abstellgewichts

Wie im Stand der Technik beschrieben ist das Einleiten des Rückkippens der Gießpfanne vor Erreichen des Zielgewichts notwendig. Hierzu wird anhand eines funktionalen Zusammenhangs fortlaufend die Restkippmenge m_{Rest} berechnet. Überschreitet die Summe aus abgegossenem Gewicht geschätzt durch das EKF und die prädizierte Restkippmenge das Zielgewicht, wird das Rückkippen eingeleitet. Hierbei wird die Restkippmengenschätzung anhand eines Regressionsmodells mit rekursiver Parameterschätzung umgesetzt. Zur Festlegung der Modellstruktur ist in Bild 6 auf der linken Seite die Auswertung im Gießprozess ermittelte Restkippmengen in Abhängigkeit des Massenstroms $\dot{m}_S$ dargestellt. Der Massenstrom zeigt einen dominanten Einfluss auf die Restkippmenge. Die Messdaten weisen auf einen linearen Zusammenhang zwischen den Größen über einen weiten Arbeitsbereich hin. Bei vergleichsweise großen oder kleinen Massenströmen verschlechtert sich die Genauigkeit der linearen Annahme. Dies könnte auf ein nichtlineares Pfannenverhalten hindeuten, jedoch auch einer geringen Menge an Messdaten geschuldet sein. Auf der rechten Seite ist der Einfluss des Kippwinkels φ auf die Restkippmenge dargestellt. Zur Analyse ist bereits der identifizierte Einfluss des Massenstroms auf die Restkippmenge entfernt worden.

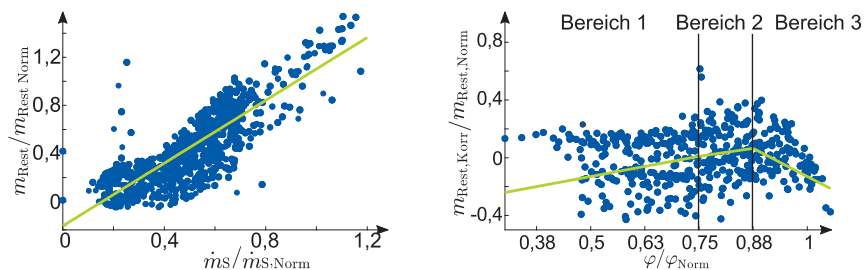


Bild 6: Restkippmenge in Abhängigkeit des Massenstroms (links) und des Winkels (rechts)

Die Messdaten werden in drei Bereiche aufgeteilt, in welchen eine stückweise lineare Abhängigkeit der Restkippmenge vom Kippwinkel erkennbar ist. Im Bereich kleiner und großer Kippwinkel wird die Restkippmenge anhand des Einflusses des Massenstroms überschätzt. Dazwischen ist kaum eine Korrektur der Restkippmenge abhängig vom Kippwinkel notwendig. Mithilfe dieser Erkenntnisse wird der Zusammenhang zwischen der Restkippmenge m_{Rest} und den Restkippmengenparameter $p_{j,i}$, des Massenstroms $\dot{m}_S$ und des Kippwinkels φ durch

$$m_{\text{Rest}} = p_{m,1} \dot{m}_S + p_{\varphi,1} \varphi + p_{m,0} + p_{\varphi,0} \quad (8)$$

beschrieben. Der Einfluss des Massenstroms auf die Restkippmenge wird über einen längeren Zeithorizont optimiert. Hierdurch können Veränderungen des Ausflussverhaltens durch ablaagerungsbedingte Geometrieänderungen berücksichtigt werden. Die Parameter $p_{m,1}$ und $p_{m,0}$ werden mittels eines Recursive Least Square Algorithmus geschätzt. Der Einfluss des Kippwinkels hingegen wird über einen kurzen Zeithorizont gelernt. Mittels der Differenz zwischen der geschätzten Restkippmenge und der tatsächlichen Restkippmenge werden die Parameter $p_{\varphi,1}$ und $p_{\varphi,0}$ zur Kompensation des Schätzfehlers angepasst.

7. Validierung des Automatisierungssystems

Zur Validierung des vorgestellten Automatisierungssystems werden Gießversuche mit Eisen-schmelze durchgeführt. Hierzu wird das Wägezellsignal in der Ruhephase zwischen den Abgüssen ausgewertet und so das Abgussgewicht messtechnisch erfasst. Bei der Vorstellung der Ergebnisse werden alle Massenangaben mit einem Normierungsgewicht verrechnet. Das Zielgewicht wird zwischen 56 % und 280 % des Normierungsgewichts variiert und stellt somit eine große Spanne an möglichen Abgusszenarien dar. Ein Gießversuch gilt als erfolgreich, wenn die abgegossene Masse um weniger als 4 % des Normierungsgewichts vom Zielgewicht abweicht. Werden die 4 % Abweichung unterschritten, besteht die Gefahr von Volumenfehlern. Diese Bauteile werden als Ausschuss bewertet. Ein Übergießen des Zielgewichts erzeugt unnötigen Materialverlust, welcher im Nachgang eingeschmolzen werden muss. Ein Übergießen des Zielgewichts hat jedoch geringere Auswirkungen auf die Produktivität des Gießvorgangs als ein Untergießen. Daher werden zusätzlich Abgüsse als erfolgreich angesehen, welche bis zu 10 % übergossen wurden. Die Ergebnisse der Versuche sind in Bild 7 dargestellt.

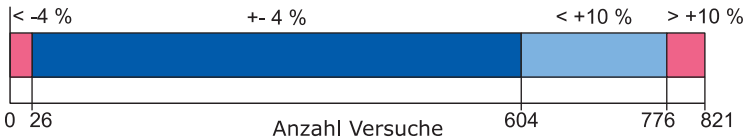


Bild 7: Versuchsauswertung der Abweichung zwischen Abguss- und Zielgewicht (normiert)

Mithilfe des entwickelten Automatisierungssystems ist eine präzise Regelung des Gießprozesses möglich. Von insgesamt 821 Versuchen waren 26 Bauteile potenziell Ausschuss und 45 um mehr als 10 % übergossen. Die Parametrierung des Automatisierungssystems wird mit der Zielsetzung einer Minimierung des Ausschusses durchgeführt. Dies geht mit einer leichten Übergießen des Zielgewichts einher. Im Mittel wird das Zielgewicht um rund 2,4 % übergossen. Diese Tatsache ist jedoch nicht problematisch, da ein Übergießen zu keinem erhöhten Ausschuss führt. Eine Messung des Bauteilgewichts war im Zuge der Versuchsdurchführung nicht möglich. Daher kann der tatsächliche Ausschuss durch Einflüsse wie neben den Gießtrichter gegossene Schmelze nicht ausgewertet werden.

8. Hardwarenahe Umsetzung

Die Automatisierung der Gießmaschine erfolgt mittels einer SoftSPS der SIMATIC S7-1500 Software Controller Reihe der Siemens AG, München. Das in diesem Beitrag vorgestellte Regelungskonzept wird separat von der restlichen Automatisierungslösung in der Entwicklungsumgebung MATLAB/Simulink entwickelt. Mittels des SIMANTIC Target 1500S werden aus den Modellen automatisiert für ODK-fähige CPUs anwendbare Modulbausteine kompiliert. Die Trennung von Regelung und Hardware ermöglicht eine Entwicklung im Rahmen des Rapid Control Prototyping und damit einer stringenten Realisierung des Automatisierungssystems zunächst simulativ und abschließend auf der Maschine [8]. Hierzu werden einzelne Komponenten sowie die Gesamtregelung in einer simulativen Umgebung entwickelt und parametrieren. Im Anschluss findet eine automatisierte Codegenerierung statt. Dabei wird der in MATLAB/Simulink entwickelte Quellcode in mehreren Teilschritten in einen für die CPU der SoftSPS kompilierbaren Quellcode umgewandelt. Die Automatisierungslösung wird erst im fortgeschrittenen Entwicklungsstadium an der realen Maschine getestet. Hierdurch wird der Entwicklungsaufwand an der Maschine stark reduziert.

9. Fazit und Ausblick

Mit dem hier beschriebenen Ansatz ist es möglich, einen Gießprozess trotz einer Vielzahl an Störgrößen mit überschaubarem Einstellaufwand zu regeln. Der vergleichsweise geringe Einstellaufwand im Prozess wird durch adaptive Methoden und den Verzicht auf eine hydrodynamische Modellierung des Prozesses realisiert. Durch die Zustandsschätzung mittels eines EKFs werden Prozessinformationen gesammelt, welche bisher nicht nutzbar waren. Dies hilft Gießmaschinenherstellern, im Zuge der Digitalisierung erweiterte Prozessüberwachungs- und Diagnosewerkzeuge anzubieten. Durch die konsequente Nutzung von Simulationssoftware im Entwicklungsprozess sind Validierungsversuche bereits vor der Inbetriebnahme der Maschine möglich. Softwareseitige Probleme sind frühzeitig identifizierbar, was die Inbetriebnahmezeit verkürzt und die Versuche mit Schmelze vor Ort sicherer gestaltet. Dadurch ist es möglich, im Polylemma der Produktion zu bestehen. Limitiert ist das Automatisierungssystem durch die vorherrschenden Störgrößen. Die Schätzung des Massenstroms als Störgröße erfordert Vertrauen in die Gewichtsmessung mittels des Wägezellensignals. Zur optimalen Schätzung darf das Wägezellensignal ausschließlich von weißem Rauschen überlagert sein. Konstante oder sich langsam ändernde Einflüsse auf die Wägezelle sind zu modellieren, da die Trennung verschiedener Störgrößen auf das Wägezellensignal ohne Modellwissen nicht möglich ist und hierdurch die Schätzung des Massenstroms verschlechtert wird.

10. Literaturverzeichnis

- [1] Noda, Y.; Matsuo, Y.; Terashima, K.; Zheng, Y. (2008): A Novel Flow Rate Estimation Method using Extended Kalman Filter and Sensor Dynamics Compensation with Automatic Casting Pouring Process.- IFAC Proceedings Volumes 2. 710–715.
- [2] Doyle, J. (2013): Feedback Control Theory.- Mineola, Dover Publications.
- [3] Noda, Y.; Kofu, Y.; Fukushima, R.; Aichi, T.; Terashima, K.; Suzuki, M.; Ohta, K. (2012): High-precision Falling Position Control to Outflow Liquid from Ladle in Automatic Pouring Machine.- Proceedings of the 70th World Foundry Congress. 478–483.
- [4] Brecher, C. (2011): Integrative Produktionstechnik für Hochlohnländer.- Berlin, Springer.
- [5] Bergner, O.; Fehrmann, M.; Hillebrand, T.; Ignatowitz, E.; Kluge, M.; Kinz, U.; Lämmlin, G.; Steinmüller, A. (2017): Metalltechnik.- 11. Aufl. Haan-Gruiten, Verlag Europa-Lehrmittel.
- [6] Noda, Y.; Terashima, K. (2006): Nonlinear Modeling with Hydrodynamics and Flow Control Using Inverse Pouring Dynamics of Tilting-Ladle-Type Automatic Pouring Process.-Proceedings of the 67th World Foundry Congress. 207/1-10.
- [7] Jurić, D. (2015): Object Tracking: Kalman Filter with Ease.-CodeProject.
- [8] Abel, D.; Bollig, A. (2006): Rapid Control Prototyping.- 1. Aufl. Berlin, Springer.
- [9] Yano, K.; Kaneko, M.; Noda, Y.; Terashima, K. (2007): Supervisory control of automatic pouring process considering residual pouring quantity.- 2007 European Control Conference (ECC) 2045–2050.
- [10] EP2425914B1.

Maximizing Efficiency and Transparency in Batch Processing with Simulation-driven Predictive Decision Support

Maximale Effizienz und Transparenz in der Batchproduktion durch simulationsbasierte Entscheidungsunterstützung

Dr.-Ing. **Christian Sonntag**, **Dominik Wolff**, **Dominik Bleidorn**,
Erik Balling, INOSIM Software GmbH, Dortmund;
Ferdinand Marx, CSL Behring GmbH, Marburg

Abstract

The complexity of modern industrial batch plants makes it nearly impossible for plant personnel and production managers to predict their future behavior accurately over longer periods of time. This contribution presents the *INOSIM Foresight* system for predictive decision support that employs highly accurate dynamic material flow simulation to create a transparency and situational awareness that allows the plant operators and management to operate and plan more efficiently, productively, and safely, which can lead to savings in the millions. Intuitive predictive graphical dashboards support plant personnel in a variety of ways. For example, they provide concise guidance to improve production (e.g. by counteracting negative events before they occur and by predicting key KPIs days or even weeks ahead), and they enable the staff to create production, personnel, and in-process maintenance plans that are efficient and non-conservative. The system is currently being deployed at a new, innovative pharma batch plant of CSL Behring GmbH in Marburg, Germany.

Kurzfassung

Die Komplexität in der modernen Batchfertigung macht es den Betreibern nahezu unmöglich, über längere Zeiträume zuverlässige Vorhersagen über das zukünftige Anlagenverhalten zu treffen. Dieser Beitrag stellt das neue System *INOSIM Foresight* für die prädiktive Entscheidungsunterstützung in komplexen Batchanlagen vor, das durch die Nutzung hochgenauer dynamischer Materialflusssimulation zuverlässige Vorhersagen liefert. Somit können der Anlagenbetrieb und die Ressourcennutzung genauer und effizienter als bisher geplant werden, was zu Einsparungen in Millionenhöhe führen kann. Das System, das zurzeit an einer industriellen Pharma-Batchanlage von CSL Behring GmbH in Marburg implementiert

wird, nutzt intuitive grafische Dashboards, um dem Betriebspersonal prädiktive Informationen zu liefern, mit denen sie verschiedene Aspekte des Anlagenverhaltens optimieren können, wie z.B. die Ressourcen-, Mitarbeiter- und Wartungsplanung. Zudem ermöglicht es eine prädiktive Anlagenüberwachung sowie eine prädiktive Störungserkennung und -vermeidung.

1. Introduction

Modern industrial batch plants are complex cyber-physical systems of systems [1] in which dozens of different products may be produced, often in parallel, each using complex dynamic production recipes. While modern batch production is in many cases highly automated and numerous approaches for batch plant control and optimization have been developed, see e.g. [2], many tasks that arise in batch production require manual actions and interventions, including manual steps in single production operations, e.g. harvesting a filter press, and medium- and long-term production decisions by plant management. The large complexity of modern batch plants makes it nearly impossible for operations personnel and production managers to predict their future behavior accurately over longer periods of time, particularly in very dynamic plants with many parallel activities that are interwoven and interdependent based on highly complex constraints on resources, unit availability, and production sequencing. Consequently, decisions are often based on experience and intuition instead of facts, which may lead to large inefficiencies, reduced flexibility, unexpected (expensive) downtimes, and conservatism in production, resource, and operations planning.

This contribution presents the *INOSIM Foresight* system for predictive decision support that employs highly accurate dynamic material flow simulation to predict the behavior of complex batch plants more accurately than is possible with existing tools, both on shorter time horizons (e.g. for the exact scheduling of routine operations) and over longer periods of time (e.g. for strategic planning). A key feature of the system is that the underlying digital twin is able to represent even the most detailed and complex constraints and dynamics within the plant faithfully - in fact, the usefulness of predictive decision support grows with the complexity of the plant. The accuracy of the resulting predictions create a transparency and situational awareness that allows plant operators and management to operate more efficiently and safely:

- Resource and personnel planning can be made more accurate and less conservative than before.
- Detailed scheduling of manual routine operator interactions can be accurately predicted and visualized (similar to a flight departure board at an airport).
- Integrated preventive in-process maintenance can be planned accurately using predictions of maintenance slots and of the effects of maintenance-induced delays.

- Predictive process supervision allows to predict the effects of process variations, such as delays and other disturbances, to detect undesired plant states, process violations, or errors before they occur, and predictive what-if analysis can be used to compute the best possible corrective actions in advance.
- Important KPIs can be predicted hours, days, or even weeks ahead.

The *INOSIM Foresight* system is based on the *INOSIM* simulation software suite [3]–[5]. Via generic data connectors, the system retrieves data about the current state of the real plant for initialization of the digital twin from the plant IT/OT systems (e.g. MES and DCS systems, real-time data management platforms, cloud-based data lakes, etc.), ensuring that all variations that have occurred during production are automatically considered within the system, with no need for manual interactions by the operator. The data architecture is open and independent on specific IT/OT providers, thus ensuring wide applicability within the processing industry. Predictive decision support is provided using intuitive graphical dashboards that can be distributed throughout the plant and displayed on both, web screens and mobile devices. The *INOSIM Foresight* system is developed and implemented by the company group INOSIM in cooperation with CSL Behring GmbH, using an existing simulation model that was developed in the design and engineering stages, and is being deployed at one of their newest innovative pharma batch plants in Marburg, Germany, in the scope of a pilot project.

The remainder of this contribution is structured as follows: Section 2 outlines challenges and requirements for generating accurate predictions in industrial batch plants and motivates the use of highly accurate material flow simulation for this purpose. Section 3 gives an overview of the *INOSIM Foresight* system for predictive decision support, and section 4 illustrates the main use cases of the system and their benefits. The contribution closes with conclusions and an outlook in section 5.

2. Challenges for Predictive Decision Support for Industrial Batch Plants

Complex industrial batch plants exhibit highly complex discrete behaviors in which many parallel activities occur at the same time, and these activities are usually interwoven and interdependent based on complex constraints arising e.g. from unit interconnectivity, from resource limitations, from unit availability, and from production sequencing. A typical example is a plant in which different products are produced simultaneously, and each product is based on a different production recipe with distinct processing times and recipes. The products run through a number of production lines in parallel, with several batches being processed in the same line at once. The production processes rely on a variety of utility systems that e.g. prepare base chemicals, additives, and intermediate products, that provide cleaning-in-place

(CIP) services for the production units, or that provide cooling, heating, and waste removal services. All of these units and systems are subject to a variety of external constraints that arise e.g. from the limited availability of personnel with potentially different and specialized skill levels, complex restrictions on the interconnectivity and availability of production and utility systems, complex constraints on material flows, planned and unplanned downtimes, and more. In complex systems with discrete dynamics, even minor changes can have large ramifications on the system behavior. Examples are e.g. that detailed constraints on the preproduction of different base chemicals in utility systems or on the availability of cleaning skids for specific production units can lead to delays or even to reordering of the production plans and sequences within the plant, or that complex constraints on material flows within the plant or complex operator schedules may lead to dynamically varying production delays. The effects of these complex constraints may even cascade through the plant, e.g. when constraints on utility systems lead to changes in the production plan, which in turn leads to unforeseen requirements on the preproduction of base materials. These effects are not directly apparent but may affect the production potentially even days later. Thus, plant design and plant operations must be done conservatively to take these uncertainties into account.

3. The INOSIM Foresight System for Predictive Decision Support

Figure 1 provides a conceptual overview of the *INOSIM Foresight* system for predictive decision support. The system is designed such that it is usable by plant personnel (e.g. plant operators, maintenance personnel, production planners, production managers, etc.). Key information for decision support is provided using intuitive graphical dashboards. Multiple dashboards with varying granularity can be provided for different target groups.

The workflow controller component provides a customizable execution environment for simulation tool chains for predictive decision support (comprising data import from plant IT/OT systems, data preprocessing, model initialization, model simulation, information extraction from simulation results, and visualization for plant personnel).

The data integrator component was developed as a generic data hub to import real-plant data into the system, with connectors available for the widely established industrial data platform *Osisoft PI*, for SQL databases (which enables access to a wide variety of IT/OT such as most MES systems), and for file-based data formats. In addition, INOSIM has developed the *INOSIM BICON* (*Business Intelligence CONnector*) component that allows the user to automatically export simulation results in a standardized data format that can be visualized in dashboards that are implemented in business intelligence suites such as *Tableau*, see e.g. [6]. The predicted data can be combined with historic data sets from the real plant before visualization.

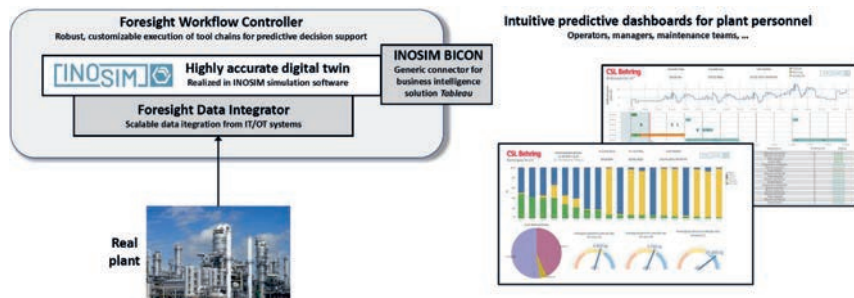


Figure 1: Conceptual overview of the INOSIM Foresight software architecture.

Figure 2 illustrates a typical simulation-driven workflow with *INOSIM Foresight*. Simulations that are initialized to the plant state, thus taking all conditions of the plant and its environment into account automatically, and that are executed regularly (*reality-predictive simulations*) are at the core of the predictive decision support system as they run frequently to compute up-to-date predicted data sets. The *Foresight* architecture supports an arbitrary number of reality-predictive simulations of models that may have different objectives. For example, a model for frequent short-term predictions is used to predict manual actions within the plant for the predictive operator task board (similar to a flight departure board) while a longer-term simulation computes predicted KPIs for strategy decisions, and several other long-term simulations predict average expected performance values for production variants for the continuous optimization of production parameters. The simulations can run in parallel or interleaved, and at different frequencies. The dashboards are automatically updated after reality-predictive simulations to show updated information e.g. for the next months or until the end of a production campaign. Since the system does not influence the plant directly, it is suited for environments with strict regulatory requirements, such as pharma production.

In addition, different stakeholders (e.g. logistics planners or the production manager) can run what-if scenario simulations that are initialized with the current state of the plant and that can be used e.g. to determine optimized production and maintenance plans or to find responses to predicted undesired behaviors before they occur. These can be performed on any computer within the network or on (local) cloud servers that host INOSIM simulation software. The focus of these what-if scenario exploration simulations is on ease of use so that any employee can easily perform simulations without being an expert in the modeling and simulation domain.

Model parametrization for what-if analysis runs is based on tailored software interfaces or on MS Excel sheets that hide the complexity of simulation-based analysis and allow the user to adapt those model parameters of interest, such as the number, unit allocation, and starting

times of maintenance orders, or the downtimes of utilities. Simulations run at the click of a button, and the user can inspect the results, e.g. via dashboards tailored for what-if analysis. At a later stage, we plan to provide web-based interfaces for model parametrization and execution. Expert users will be able to use INOSIM software directly to ensure that they have maximum freedom in their analysis.

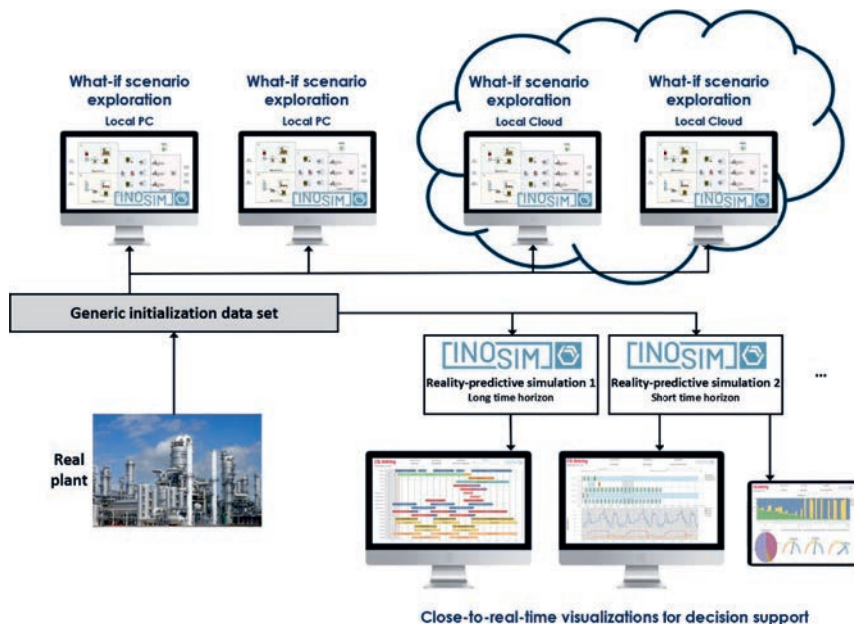


Figure 2: Overview of a simulation-driven workflow with INOSIM Foresight.

A key feature of the *Foresight* architecture that makes this parallel use of reality-predictive and what-if simulations by a variety of different stakeholders possible is the approach to model initialization. The *Foresight* data integrator can generate a generic initialization data set from the real plant at any time. The format of this data set has intentionally been kept simple – it is just a text file. This file is provided to the simulation engine for model initialization, but it can also be distributed to any INOSIM installation for what-if analyses.

Plant operators will only accept a new solution for predictive decision support if they are always aware of the validity and relevance of the information that is displayed. Each *Foresight* dashboard provides visual elements that let them understand this information at a single glance. As an example, consider the predictive operator task board shown in Figure 3.

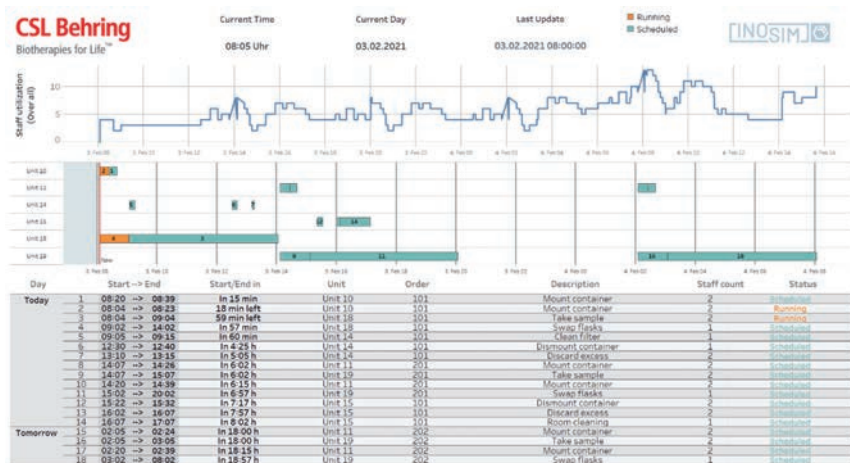


Figure 3: Exemplary predictive operator task board (top: predicted staff utilization over time, center: Gantt representation of manual operations, bottom: list of the manual operations).

It provides the following information:

- **Last Update:** This element tells the user when the last simulation run was executed, i.e. when the displayed information was last updated (Feb. 3, 2021, at 08:00 in the example). If the last update was too long ago, e.g. because of a system failure, the field is colored in red to indicate to personnel that the displayed information is outdated.
- **Current Time, Current Day:** These elements inform the user about the current day and time. This time point is in addition visualized by a vertical red line in Gantt views.

The *Foresight* dashboards are customizable. The visualizations can easily be adapted so that they show only information that is relevant e.g. in certain parts or zones of the plant, or that is only of interest to specific members of plant personnel.

4. Use Cases

This section describes the concrete use cases and benefits that the partners have developed and are currently implementing at an industrial batch plant of CSL Behring in Marburg, Germany. The plant in Marburg is a progressive green-field plasma processing facility with over 400 million USD invest volume. It contains 4 production lines that produce 6 products.

4.1. Predictive Operator Task Board

While automation plays an important role in modern batch plants, many of the processing steps still require human interaction and attention. The complexity of the behaviors and of the interactions within batch plants makes it challenging to predict and plan operator movements

in advance. This can lead to inefficiencies, large variations in the durations of process steps, negative impacts on production quality, and frustrations of the plant personnel due to a lack of situational awareness, e.g. by being “at the right place at the wrong time”.

The example dashboard in Figure 3 visualizes upcoming manual operations within the plant and provides an intuitive and highly customizable way to quickly grasp where personnel has to be at what time, and other related information. It visualizes the upcoming operations both graphically within a Gantt representation as well as via a list-based view. Each task is visualized as a row in the list, including the exact date and start/end times, the related order and equipment unit, a description and status of the manual operation, and the required staff count for the task. The column “Start/End in” indicates when the next manual interaction will be necessary for a task, or how long the allowed residual time is for time-critical tasks. The predictive operator task board allows the plant personnel to plan their movements and activities within the plant with high confidence and enables optimal scheduling, thus saving time and money. It increases the work motivation of the operations personnel, plant robustness, and work capacity while reducing deviations and the amount of staff required for e.g. night shifts.

4.2. Predictive In-Process Maintenance Support

Classical maintenance planning can severely affect the performance of industrial production since it leads to long shut downs of parts of or even of the complete plant, and to underutilization of the available maintenance resources. Integrated in-process maintenance has become an important goal in the processing industry since it holds large potential for efficiency and productivity improvements, in particular because it can eliminate the need to shut down the complete plant for maintenance if maintenance slots can be identified a priori.

Predictive in-process maintenance with *Foresight* enables maintenance planners to design plans that inhibit the production flow as little as possible. Planners can identify and test potential maintenance slots and identify their impact on production immediately, and maintenance crews are always aware about the timing of upcoming maintenance slots so that their utilization can be maximized.

Figure 4 shows an example of a predictive maintenance planning board. The Gantt chart is color-coded so that maintenance orders and cleaning orders stand out from all other orders. Alternatively, a list-based view of upcoming maintenance orders can be inserted as well, similar to the one in Figure 3. The *Foresight* system supports the complete maintenance planning and execution process, for example as described in the following scenario:



Figure 4: Exemplary predictive maintenance planning dashboard.

a) Maintenance planning with scenario simulation: A maintenance planner uses the predictive maintenance planning board to identify suitable time slots for new maintenance orders. The planner then schedules the new maintenance slot (which may be longer than the chosen free time slot) within the simulation and uses scenario simulation to determine and assess the effect of the new order on the production schedule, with the goal to find a maintenance scheduling strategy that will execute the maintenance order with as little negative effect as possible. The planner can use what-if analyses with schedule adaptations (e.g. by moving orders to different units or splitting the maintenance task) to try to minimize any negative effects. Once the maintenance planner has determined a suitable strategy, it is released to the shop floor for execution.

b) Maintenance visualization and execution: Maintenance personnel obtain accurate information about upcoming maintenance orders from maintenance dashboards that visualize the predicted production and maintenance schedule throughout the plant in real time.

4.3. Predictive Process Supervision

Process supervision is an area that profits immensely from predictive information since predictions let you identify and counteract problems and adverse effects of decisions and of complex effects of production long before they occur. The *Foresight* solution for predictive decision support provides three tools for predictive process supervision:

Detailed Visualization of Production Schedules

Foresight provides accurate predictions and visualizations of the concrete production schedule down to the level of single production operations, ranging from hours up to weeks. This enables plant personnel to gain situational awareness at a glance, and to quickly identify production bottlenecks or opportunities to schedule another batch.

Predictive Conflict Detection and Mitigation

Prediction with *Foresight* shines if unexpected situations (whose repercussions are difficult to predict) occur during production. These situations include e.g.

- Minor or major variations of nominal production, such as delays, that are not recognizable as perilous but may amplify and lead to critical states down the line (such as violations of safety thresholds, vessels that run full or empty, or even shut-downs)
- Equipment failures or breakdowns
- The arrival of high-priority rush orders
- Resource limitations (e.g. operators, utilities, raw materials)

With *Foresight* decision support, the effects of such unexpected events can be accurately predicted, and any negative consequences can be mitigated before they occur. *Foresight* provides semi-automated tools that can be used to identify critical events and to communicate this information to the relevant stakeholders to support plant robustness and resilience:

- Errors, conflicts, or other detrimental situations that the reality-predictive simulation identifies over the prediction horizon (e.g. that vessels run full or empty or that a maximum process hold time will be exceeded) are visualized within a dashboard.
- Warning or error messages that are generated during the simulation can in addition be visualized within a suitable Gantt view, by highlighting the affected orders and operations with a warning color.
- A suitable dashboard is planned that intuitively visualizes the effects of delays, breakdowns, or failures in a Gantt view that displays and compares the schedules of nominal and disturbed production and highlights the differences, e.g. with suitable warning colors.

Predictive KPI Cockpits

Key Performance Indicators (KPIs) are simple, expressive measures that represent quality aspects relating to plant performance, robustness, productivity, yield, or efficiency. While KPIs are widely established tools for process supervision, they usually only reflect the current state of the plant, e.g. in KPI dashboards ("KPI cockpits") that are provided by MES systems. Prediction with *Foresight* enables the user to compute the important KPIs of the plant hours, days, or even weeks ahead. This means that if any of the predicted KPIs go into the red,

indicating that plant performance will be deteriorating, there is still time to correct the problem before it occurs, enabling the personnel to increase plant robustness and to produce more consistently close to the golden batch. Figure 5 shows an example of a predictive KPI cockpit that provides key information on important aspects of the plant, such as graphical representations of equipment utilization, of waiting times, or of average production rates.



Figure 5: Exemplary predictive KPI cockpit dashboard.

5. Conclusions and Outlook

This contribution gives an overview of the *INOSIM Foresight* predictive decision support system that is developed jointly by *INOSIM* and CSL Behring GmbH and is deployed and demonstrated within a pilot project at a new batch production plant that CSL Behring GmbH operates in Marburg, Germany. Based on a highly accurate material flow simulation that was already successfully employed in the design and engineering phases of the plant and that takes the most detailed relevant plant dynamics into account, the system provides transparency and situational awareness going beyond what is possible with available planning tools. Thus, the system enables the user to significantly increase the efficiency, robustness, and resilience while improving the quality of production, resource, and operations planning, saving time and money already in the short run.

The partners are currently working on the development and deployment of the full-fledged *Foresight* system at the industrial batch plant of CSL Behring GmbH in Marburg. After deployment, the focus will be on realizing the full potential of the system as a basis for the deployment of automated methods and tools for plant diagnosis, monitoring, and optimization,

such as AI- based approaches for plant self-optimization or the simulation-driven system for optimal production scheduling that is developed in the *OptiProd.NRW* project [7].

About INOSIM

INOSIM is a group of German SMEs that provide innovative solutions and consulting services around their high-performance discrete-event material flow simulation software, with the goal to maximize productivity, resource efficiency, and safety in the processing industry.

About CSL Behring

CSL Behring is a global biotherapeutics leader that develops and delivers innovative biotherapies that are used to treat serious and rare conditions such as coagulation disorders, primary immune deficiencies, hereditary angioedema, and respiratory disease.

6. References

- [1] S. Engell and C. Sonntag, Eds., *Proposal of a European Research and Innovation Agenda on Cyber-physical Systems of Systems, 2016-2025*. Strategic Roadmap Document of the EU project CPSoS, Process Dynamics and Operations Group (DYN), TU Dortmund, Germany, 2016.
- [2] D. Bonvin and G. Francois, "Control and Optimization of Batch Chemical Processes," in *Coulson and Richardson's Chemical Engineering (Fourth Edition)*, 2017, pp. 441–503.
- [3] INOSIM Software GmbH, INOSIM Simulation Software for Process Engineering and Bioengineering. Dortmund, Germany, 2021.
- [4] C. Sonntag, D. Wolff, and M. Zajac, "A Simulation-based Support System for the Planning and Design of Modular Logistics in Chemical Production Processes," presented at the Proc. 12th European Congress of Chemical Engineering (ECCE 12), Florence, Italy, 2019.
- [5] K. Sulzbacher, P. Balling, and G. Schembecker, "INOSIM Bio - New Approaches for Bioprocess Simulation and Optimization," in *Computer-aided Chemical Engineering*, 2013, vol. 32, pp. 865–870.
- [6] X. Qin, Y. Luo, N. Tang, and G. Li, "Making Data Visualization More Efficient and Effective: A Survey," doi: <https://doi.org/10.1007/s00778-019-00588-3>.
- [7] "The OptiProd.NRW Project: Increased Productivity and Sustainability with Digital Production Optimization from NRW, 2020 - 2022, funded by the European Regional Development Fund (ERDF)." www.optiprod.nrw/en.

Vergleich und Zusammenführung des Open Process Automation Standard (OPAS) und der VDI/VDE/NAMUR 2658

Comparison and Alignment of the Open Process Automation Standard (OPAS) and the VDI/VDE/NAMUR 2658

M. Blumenstein, Prof. Dr.-Ing. **A. Fay**,
Helmut-Schmidt-Universität, Hamburg;
A. Stutz, Dr.-Ing. **M. Maurmaier**, Siemens AG, Karlsruhe;
Prof. Dr.-Ing. **M. Barth**, Hochschule Pforzheim, Pforzheim

Kurzfassung

Der Open Process Automation Standard (OPAS) und die Richtlinie VDI/VDE/NAMUR 2658 zielen auf die Standardisierung offener Automatisierungssysteme ab. Ziel des vorliegenden Beitrags ist der Vergleich und die Zusammenführung dieser Standards. Dazu wird ein Automatisierungskonzept basierend auf zwei Arten von Komponenten vorgestellt und validiert. Dieses ermöglicht durch den Einsatz standardisierter Schnittstellen eine einfache, herstellerunabhängige Austauschbarkeit von Komponenten. Weiterhin wird ein Engineering-Workflow präsentiert, der auf dem vorgestellten komponentenbasierten Automatisierungskonzept aufbaut.

Abstract

The Open Process Automation Standard (OPAS) and the VDI/VDE/NAMUR 2658 guideline aim at standardizing open automation systems. The objective of this paper is to compare and align these standards. Therefore, an automation concept based on two types of components is presented and validated. Using standardized interfaces, this approach enables the simple, vendor-independent exchangeability of components. Furthermore, a corresponding engineering workflow is presented building on the presented component-based automation approach.

1 Motivation und Einleitung

Mit den aktuellen Entwicklungen im Kontext von Industrie 4.0 verändert sich auch die Prozessindustrie. Es besteht zunehmend der Wunsch, Produktionsanlagen nicht mehr wie bisher geschlossen und proprietär zu gestalten, sondern Offenheit in Bezug auf Interoperabilität sowie die Austauschbarkeit von Daten und Geräten zu gewährleisten [1].

In diesem Zusammenhang entstanden der Open Process Automation Standard (OPAS) [2] sowie die Richtlinie VDI/VDE/NAMUR 2658 [3], die das Module Type Package (MTP)-Konzept beschreibt. Beide Standards zielen darauf ab, die in künftigen Automatisierungssystemen geforderte Offenheit bereitzustellen. Allerdings haben die beiden Standards ihren Ursprung in verschiedenen Domänen und verfolgen im Detail verschiedene Zielstellungen. Um Unsicherheiten zu vermeiden, wann welcher Standard eingesetzt werden soll, und um die Interoperabilität zwischen OPAS- und MTP-basierten Anlagen sicherzustellen, besteht der Wunsch, die Konzepte beider Standards zusammenzuführen [4].

In [1] wird ein Demonstrator vorgestellt, der Ansätze aus OPAS und der VDI/VDE/NAMUR 2658 nutzt. Diese existieren allerdings nebeneinander und werden nicht tatsächlich zusammengeführt. Damit können mögliche Synergien, die sich aus einer solchen Zusammenführung ergeben, nicht vollständig ausgenutzt werden. Aufgrund dessen ist es Ziel des vorliegenden Beitrags, einen detaillierten Vergleich der OPAS- und MTP-Konzepte durchzuführen und einen Ansatz zu deren Zusammenführung aufzuzeigen.

2 Analyse und Vergleich der Standards

Als Basis für einen Vergleich zwischen OPAS und der VDI/VDE/NAMUR 2658 wurden charakteristische Merkmale beider Standards herausgearbeitet und geclustert. Im Ergebnis zeigt sich, dass ein Vergleich und eine Zusammenführung vor allem hinsichtlich der Aspekte Modularität, Interoperabilität, Anlagenkoordination und Konfiguration sinnvoll erscheinen. Der Aspekt der Modularität befasst sich dabei mit der Definition der kleinsten *Elemente*, aus denen sich eine Automatisierungssoftware zusammensetzt. Im Rahmen der Interoperabilitäts-Betrachtung werden Schnittstellen und Interaktionsmechanismen zwischen den definierten *Elementen* untersucht. Der Aspekt der Anlagenkoordination behandelt eine Laufzeitarchitektur für Anlagen, die aus *Elementen* bestehen. Schließlich befasst sich der Aspekt der Konfiguration mit dem Engineering von Anlagen, die auf den drei zuvor untersuchten Aspekten beruhen. Die wesentlichen Charakteristika der Standards in Bezug auf die genannten Aspekte werden in den Subkapiteln 2.1 bis 2.4 verglichen.

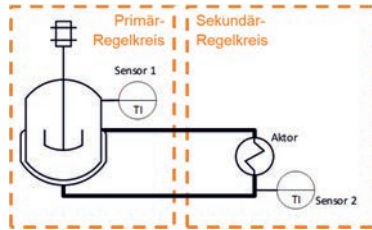


Bild 1: Anwendungsbeispiel einer kaskadierten Temperaturregelung

Zur Veranschaulichung der Inhalte dieses Beitrags wird das in Bild 1 dargestellte Anwendungsbeispiel einer kaskadierten Temperaturregelung genutzt. Ziel ist es, in einem Primärregelkreis die Temperatur innerhalb eines Reaktors zu regeln. Der Sekundärregelkreis wird verwendet, um die Temperatur des genutzten Heizmediums exakt einzustellen.

2.1 Modularität

Die Konzepte beider Standards basieren auf der Modularisierung von Anlagenfunktionalität. Dabei wird die Automatisierungstechnik von Prozessanlagen in mehrere, voneinander weitestgehend unabhängige Bestandteile zerlegt. Im Rahmen dieses Beitrags sollen die kleinsten Bestandteile, aus denen sich die Automatisierungssoftware zusammensetzen kann, als *Elemente* bezeichnet werden. Ein *Element* wird dabei als eine vollständig implementierte, unteilbare Softwareeinheit verstanden, die einen bestimmten Teil der Funktionalität einer Prozessanlage kapselt. Zudem zeichnet sich ein *Element* dadurch aus, dass es stets auf einem einzigen Controller ausgeführt wird. Mehrere *Elemente* können hingegen beliebig auf verschiedene Controller verteilt werden. Ein solcher Controller wird in OPAS Distributed Control Node (DCN) genannt [5]. Im Kontext der VDI/VDE/NAMUR 2658 werden Controller, die darauf implementierte Automatisierungssoftware und alle notwendige Hardware zur Ausführung einer Prozessfunktion als eine Einheit betrachtet. Diese Einheiten werden gemäß VDI 2776 [6] als Process Equipment Assemblies (PEAs) bzw. Functional Equipment Assemblies (FEAs) bezeichnet.

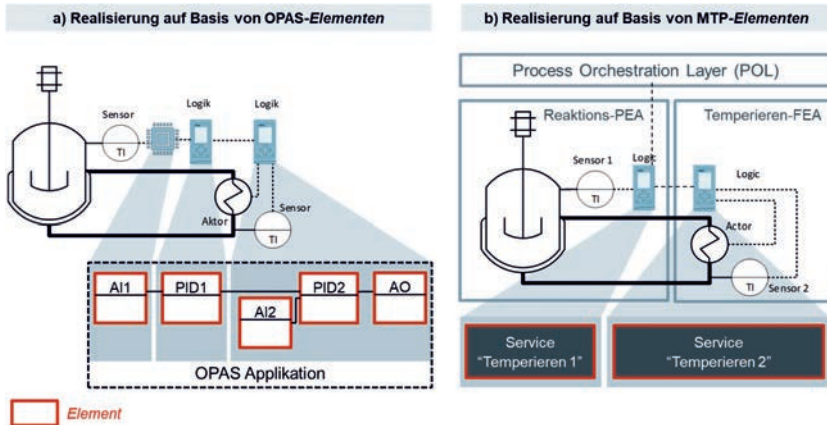


Bild 2: Realisierung der Beispielanlage a) mit OPAS-Elementen b) mit MTP-Elementen

Anhand der o.g. Definition können die *Elemente* aus Sicht von OPAS und der VDI/VDE/NAMUR 2658 identifiziert werden. Im Ergebnis zeigt sich, dass im Falle von OPAS Funktionsbausteine als *Elemente* betrachtet werden können. Im Falle der VDI/VDE/NAMUR 2658 sind Services als *Elemente* anzusehen. Dieser Unterschied in der *Element*-Definition beider Standards wird in Bild 2 anhand der betrachteten Beispielanlage veranschaulicht.

Es wird deutlich, dass OPAS eine feingranulärere Untergliederung der Anlagenfunktionalität vornimmt als die VDI/VDE/NAMUR 2658. Im Falle der OPAS-Realisierung setzt sich die Anlagenfunktionalität aus fünf Funktionsbausteinen zusammen, die als *Elemente* zu betrachten sind und folglich beliebig auf verschiedene DCNs verteilt sein können. Im Falle der MTP-Realisierung werden lediglich die Primärregelung und die Sekundärregelung jeweils als ein *Element* angesehen.

2.2 Interoperabilität

In Bezug auf den Aspekt der Interoperabilität gilt es, die Schnittstellen und Interaktionsmechanismen zwischen den in Kapitel 2.1 identifizierten *Elementen* von OPAS und der VDI/VDE/NAMUR 2658 zu untersuchen und zu vergleichen. Bei diesem Vergleich fällt auf, dass beide Standards die Kommunikation zwischen *Elementen* auf Signalebene unterstützen. Im Falle von OPAS werden hierzu sogenannten OPAS Connectivity Framework Function Blocks (OCFBs) eingesetzt (siehe Bild 3a). Diese ermöglichen es, einen Prozesswert inkl. einer Reihe von Metadaten von einem Funktionsbaustein über Controller-Grenzen hinweg zu einem anderen Funktionsbaustein zu übertragen.

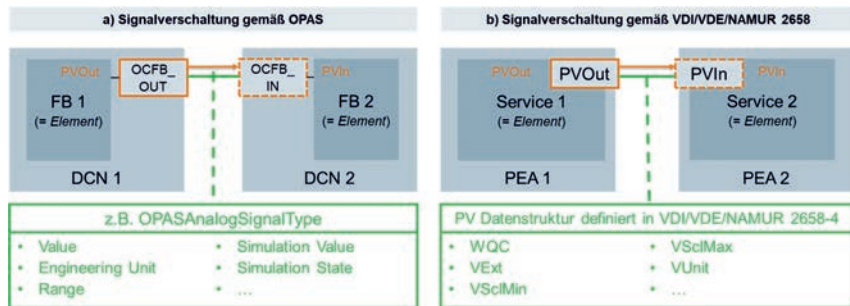


Bild 3: Signalverschaltung a) nach dem OPAS-Konzept b) nach dem MTP-Konzept

Ein ähnliches Konzept wird im Falle der VDI/VDE/NAMUR 2658 mittels Prozesswert-Eingängen und Prozesswert-Ausgängen umgesetzt (siehe Bild 3b). Diese können im Interface eines Service enthalten sein und ermöglichen service-übergreifende Signalverschaltungen. Neben diesen Verschaltungsmöglichkeiten auf Signalebene bietet die VDI/VDE/NAMUR 2658 zusätzlich die Möglichkeit, ganze Servicefunktionalitäten abstrahiert über ein einziges Interface abzurufen. Dies kommt einer Verschaltungsmöglichkeit auf *Element*-Ebene gleich. Dieser Mechanismus wird im MTP-Kontext genutzt, um Servicefunktionalität in einen übergeordneten Process Orchestration Layer (POL) zu integrieren [7] [8]. Gemäß [9] kann dieser Mechanismus aber auch genutzt werden, um Services in Form einer dezentralen Orchestrierung in andere Services zu integrieren. Letztere Möglichkeit ist in Bild 4 dargestellt.

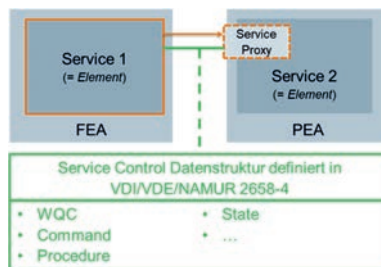


Bild 4: Dezentrale Orchestrierung eines MTP-Service

2.3 Anlagenkoordination

OPAS konzentriert sich nach heutigem Stand vorrangig auf kontinuierlich arbeitende Anlagen. Diese sind dazu konzipiert, einen bestimmten Arbeitspunkt für eine lange Zeit zu halten. Sequentielle Abfolgen (Procedural Control im Sinne der ISA 88 [10]) sind dabei nur zum

Anfahren, Wechseln des Arbeitspunktes und Herunterfahren notwendig. Aufgrund dessen werden in OPAS vorerst nur Konzepte vorgesehen, die zur Verschaltung kontinuierlich verarbeiteter Prozesswerte (z.B. für Regelkreise oder Interlocks) dienen (Basic Control im Sinne der ISA 88 [10]). Die Realisierung prozeduraler Anlagensteuerungen ist mit den OPAS-Ansätzen zwar möglich, aber es sind keine entsprechenden Konzepte und Bausteine im Standard enthalten.

Die VDI/VDE/NAMUR 2658 konzentriert sich nach heutigem Stand auf die Automatisierung von Batch-Anlagen. Dabei ist die Ausführung von Prozessfunktionen in einer bestimmten Reihenfolge von zentraler Bedeutung. Das MTP-Konzept sieht hierzu die Orchestrierung von Services durch einen Process Orchestration Layer (POL) vor [8]. Neben dieser im Standard beschriebenen zentralen Orchestrierung werden in weiterführenden Veröffentlichungen zudem die Möglichkeiten der dezentralen Orchestrierung [9] sowie der Choreografie [11] [12] beschrieben. Kontinuierlich verarbeitete Prozessfunktionalität im Sinne der Basic Control (gemäß [10]) sind aus Sicht des MTP-Konzepts zum Großteil in den PEAs zu kapseln und sind in dieser Form nicht Teil des Standards. Sollen diese Funktionalitäten modul-übergreifend realisiert werden, kann dies mittels Prozesswertverschaltungen (siehe auch Kapitel 2.2) umgesetzt werden.

2.4 Konfiguration

Im Rahmen des Konfigurationsaspekts soll ein Engineering-Workflow für Anlagen bestehend aus *Elementen* untersucht werden. Im Zuge des Engineerings wird ein Satz an Konfigurationssinformationen in Form einer eindeutigen, maschinenlesbaren, plattformunabhängigen Beschreibung eines Automatisierungssystems erzeugt. Diese beinhaltet die Auswahl der notwendigen *Elemente*, deren Parametrierung und Verschaltung. Zudem erfolgt eine Zuordnung der *Elemente* zu den verwendeten Controllern. Im Wesentlichen finden sich diese Informationen auch in den OPAS- und MTP-Konzepten wieder. Im Detail sind jedoch die folgenden Unterschiede zu beachten:

- Aufgrund der verschiedenen *Element*-Definitionen (siehe auch Kapitel 2.1) wird die Konfiguration im Falle von OPAS auf Ebene der Funktionsbausteine angesetzt, während im Falle des MTP-Konzepts Services konfiguriert werden.
- Im Falle der VDI/VDE/NAMUR 2658 gilt es, neben der einfachen Verschaltung und Parametrierung von Variablen auch sequentielle Logik zu konfigurieren. Dies ist in OPAS nicht intrinsisch vorgesehen (siehe auch Kapitel 2.3).
- Im Falle der VDI/VDE/NAMUR 2658 ist die Zuordnung von Funktionalität zu einem Controller zum Zeitpunkt der Konfiguration bereits dadurch festgelegt, dass die verfügbaren Services auf einem Controller vorimplementiert sind. OPAS sieht vor, dass

Funktionsbausteine zum Zeitpunkt der Konfiguration von Funktionsbausteintypen instanziiert und einem ausführenden DCN zugeordnet werden.

3 Zusammenführung der Standards

Ziel dieses Kapitels ist es, Ansätze zur Zusammenführung von OPAS und der VDI/VDE/NAMUR 2658 aufzuzeigen. Dazu werden in den Subkapiteln 3.1 bis 3.4 die identifizierten Unterschiede der Standards aufgegriffen und Konzepte zu deren Auflösung beschrieben.

3.1 Modularität

In Kapitel 2.1 wurde aufgezeigt, dass die beiden Standards verschiedene Granularität bei der Modularisierung von Anlagenfunktionalität vorsehen. Im Rahmen dieses Beitrags wird ein Mittelweg zwischen diesen Granularitätsstufen vorgeschlagen. Hierfür werden zwei verschiedene *Komponenten* (Components) als *Elemente* vorgeschlagen:

- *Control Components (CCs)*: CCs kapseln Logik zum Zugriff auf Sensoren und Aktoren eines Produktionsprozesses. Sie sind folglich dadurch charakterisiert, dass sie direkt mit dem Prozess interagieren können. Die Granularität der gekapselten Funktionalität ist dabei nicht exakt festgelegt. Eine CC kann die Logik zum Zugriff auf einen einzelnen Sensor oder Aktor beinhalten, aber auch die Logik eines ganzen Regelkreises mit mehreren Sensoren und Aktoren. CCs orientieren sich dabei stark an der Definition für Einzelsteuereinheiten (Control Modules) im Sinne der ISA 88 [10].
- *Service Components (SCs)*: SCs kapseln die Servicelogik im Sinne der VDI/VDE/NAMUR 2658, welche nicht bereits durch CCs abgedeckt ist. Dies umfasst insbesondere die Service-Zustandsmaschine, das Serviceinterface sowie den Teil der regulatorischen und verriegelnden Funktionen, der nicht in CCs enthalten ist. SCs sind charakterisiert dadurch, dass sie keinen direkten Zugriff auf den Prozess besitzen. Zur Interaktion mit dem Prozess werden CCs in das Programm einer SC integriert.

Bild 5 stellt eine exemplarische Realisierung des Anlagenbeispiels unter Nutzung der vorgestellten *Komponenten* dar. Der Sekundärregelkreis ist dabei komplett durch die CC „PIDCtrl“ realisiert. Die Primärregelung umfasst die SC „Temperieren“ und eine CC „Sensor 1“, die die Logik zum Zugriff auf den Temperatursensor des Reaktors enthält. Die SC „Temperieren“ nutzt dabei die CCs „PIDCtrl“ und „Sensor 1“ für den Zugriff auf den Produktionsprozess.

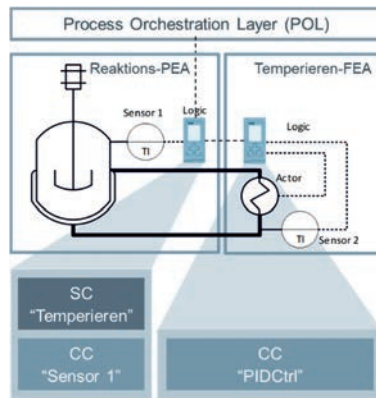


Bild 5: Softwarearchitektur auf Basis von Control Components und Service Components

3.2 Interoperabilität

Im Rahmen der Interoperabilitäts-Betrachtung gilt es, Schnittstellen und Interaktionsmechanismen für die in Kapitel 3.1 definierten *Komponenten* zu konzipieren. Dabei liegt der Fokus auf der Integration von CCs in SCs. Die Interaktion zwischen mehreren SCs im Anlagenkontext wird in Kapitel 3.3 behandelt. Eine direkte Interaktion zwischen mehreren CCs wird im Rahmen dieses Beitrags nicht betrachtet.

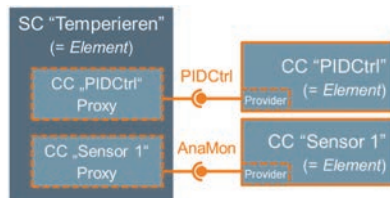


Bild 6: Integration von Control Components in eine Service Component

Die Integration einer CC in eine SC basiert auf dem gleichen Prinzip wie die Integration eines Service in einen anderen, die in Kapitel 2.2 als dezentrale Orchestrierung eingeführt wurde (siehe Bild 4). Bild 6 stellt die Übertragung dieses Prinzips auf die *Komponenten* der Beispielanlage dar. Die Funktionalität der CCs wird dabei abstrahiert und auf eine standardisierte Schnittstelle abgebildet. Im Rahmen dieses Beitrags werden hierfür die Schnittstellen der VDI/VDE/NAMUR 2658-3 [13] vorgeschlagen. Diese sind geeignet, um die Funktionalitäten verschiedener CCs zu abstrahieren. Im Falle der Beispielanlage kann die Funktionalität der CC „Sensor 1“ auf die AnaMon-Schnittstelle abgebildet werden, für die Abstraktion der CC „PIDCtrl“ ist die PIDCtrl-Schnittstelle geeignet. Diese Abstraktion der CC-Funktionalität kann

mittels eines Funktionsbausteins in der CC umgesetzt werden. Dieser wird im Folgenden als Provider-Baustein bezeichnet.

Um CC-Funktionalität in eine SC zu integrieren, werden sogenannte Proxy-Bausteine genutzt. Diese sind überall dort in das Programm der SC zu integrieren, wo CC-Funktionalität benötigt wird. Jeder Proxy-Baustein wird dabei einem Provider-Baustein zugeordnet und spiegelt dessen Schnittstelle in die SC hinein. Das bedeutet, ein Proxy-Baustein muss die gleiche Schnittstelle besitzen wie der Provider-Baustein, den er repräsentiert. Er stellt zudem durch Interaktion mit dem entsprechenden Provider-Baustein sicher, dass exakt die gleichen Werte an beiden Schnittstellen anliegen. Somit kann ein Proxy-Baustein im Programm der SC in gleicher Weise verwendet werden, als wäre die zugehörige CC selbst in der SC implementiert.

Dieser Mechanismus wird im Folgenden als Provider-Proxy-Pattern bezeichnet. Dieses Pattern ist mit verschiedenen Kommunikationstechnologien umsetzbar. Die präferierte Technologie in den OPAS- und MTP-Konzepten ist nach heutigem Stand OPC UA. Das vorgestellte Pattern wurde daher auf Basis einer OPC UA Client/Server-Architektur validiert (siehe Kapitel 4). Dabei stellt jeder Provider- bzw. Proxy-Baustein eine OPC UA Client-Funktionalität bereit. Mittels dieser können die Bausteine die für sie notwendigen Werte aus dem OPC UA-Server des jeweiligen Kommunikationspartners auslesen.

3.3 Anlagenkoordination

Um die in Kapitel 3.1 vorgestellten *Komponenten* zu einer Gesamtanlage zusammenzufügen, wird ein zweistufiger Ansatz vorgeschlagen. Dessen Prinzip ist in Bild 7 anhand der betrachteten Beispielanlage dargestellt.

Im ersten Schritt wird die sogenannte *Komponenten*-Orchestrierung durchgeführt. Ziel dabei ist es, aus *Komponenten* Services im Sinne der VDI/VDE/NAMUR 2658-4 zusammenzufügen. Ein solcher Service setzt sich stets aus genau einer SC und einer beliebigen Anzahl an CCs zusammen. Im Zuge der *Komponenten*-Orchestrierung werden allen für die Ausführung des Service notwendigen Proxies innerhalb der SC entsprechende CCs gemäß des in Kapitel 3.2 beschriebenen Mechanismus zugeordnet.

Schritt 2, die Anlagen-Orchestrierung, zielt darauf ab, die in Schritt 1 erstellen Services zu einer Gesamtanlage zusammenzusetzen. Dabei ist zu beachten, dass sich aus *Komponenten* aufgebaute Services nach außen hin genau so verhalten wie alle anderen Services, die der VDI/VDE/NAMUR 2658 folgen. Sie stellen zudem die gleichen Schnittstellen bereit. Somit kann für die Anlagen-Orchestrierung auf die Mechanismen zurückgegriffen werden, die im MTP-Umfeld bereits existieren.

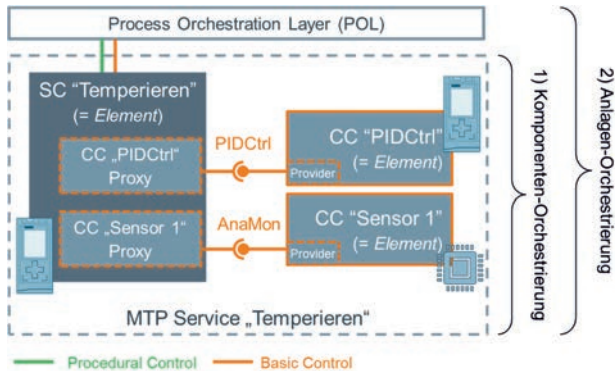


Bild 7: Konzept der Anlagenkoordination

3.4 Konfiguration

Ziel der Konfigurations-Betrachtung ist es, einen Konfigurations-Workflow für die in Kapitel 3.3 vorgestellte Anlagenarchitektur zu entwickeln. Der Ansatz für einen solchen Workflow ist in Bild 8 dargestellt.

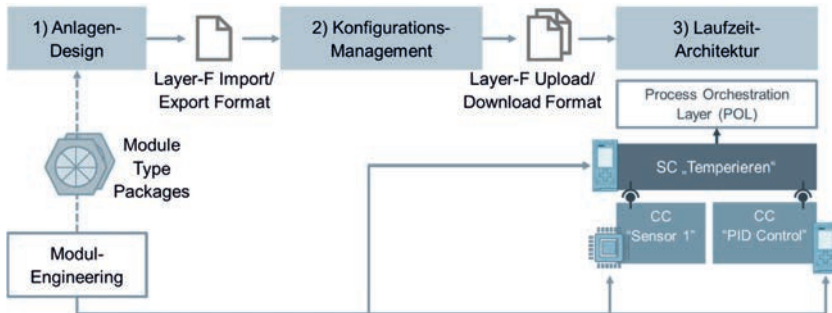


Bild 8: Konfigurationsmechanismus

Es werden im Wesentlichen drei Schritte durchlaufen. Im ersten Schritt, dem Anlagen-Design, wird eine genaue Beschreibung der Anlagenarchitektur erstellt. Bestehende Engineering-Workflows wie z.B. in [14] können dabei als Grundlage dienen. Gemäß Kapitel 2.4 muss im Zuge dessen die Auswahl der notwendigen SCs und CCs, deren Parametrierung, Verschaltung und Zuordnung zu Controllern erfolgen. Abhängig von der Domäne und der Charakteristik der gewünschten Anlage kann hierbei die Anlagenfunktionalität von Grund auf entwickelt werden (OPAS-Philosophie) oder auf bereits vorimplementierte Controller zurückgegriffen werden (MTP-Philosophie). Auch eine Mischung beider Varianten ist möglich. In Bezug auf das

vorliegende Beispiel wird angenommen, dass die drei benötigten *Komponenten* bereits im Zuge eines Modul-Engineerings, welches nicht direkt zum Konfigurations-Workflow gehört, auf drei Controllern vorimplementiert wurden. In diesem Fall werden die MTPs dieser Controller im Zuge des Anlagen-Designs einbezogen und zu einer Gesamtanlagenfunktionalität verknüpft. Angelehnt an den OPAS-Konfigurations-Workflow [5] wird im Anschluss an das Anlagen-Design ein sogenanntes Layer-F-Import/Export-Format erzeugt, welches die Anlagenkonfiguration auf herstellerunabhängige Weise in einer AutomationML-Datei abbildet.

Der zweite Schritt, das Konfigurations-Management, stellt anschließend sicher, dass jeder Controller der Anlagenarchitektur genau die Konfigurations-Informationen erhält, die er zur Ausführung der ihm zugeordneten Funktion benötigt. Die controllerspezifischen Konfigurations-Informationen werden gemäß OPAS [5] als Layer-F-Upload/Download-Formate auf die einzelnen Controller heruntergeladen. Diese Konfigurations-Informationen können beispielsweise Orchestrierungsinformationen für den Process Orchestration Layer oder Informationen über die Verschaltung von *Komponenten* sein. Wird die OPAS-Konfigurations-Philosophie verfolgt, können auch Informationen darüber enthalten sein, welche *Komponenten* auf welche Controller heruntergeladen werden müssen.

Nach diesem Konfigurations-Prozess ist die Laufzeit-Architektur einer Anlage zusammengestellt und kann in Betrieb genommen werden.

4 Validierung

Zur Validierung der in Kapitel 3 vorgestellten Konzepte wurde ein Demonstrator entwickelt, dessen Aufbau in Bild 9 veranschaulicht ist. Dabei wurde die in Bild 1 eingeführte Beispielanlage anhand der in diesem Beitrag vorgestellten Konzepte umgesetzt. Hierzu wurden die SC „Temperieren“ sowie die CCs für die Sekundärregelung („PID Ctrl“) und den Temperatursensor des Reaktors („Sensor 1“) auf drei SIMATIC S7 1515-2 PN (PLCSim-Instanzen) realisiert. Zusätzlich wurde eine CC „Sensor 1**“ als Node.js®-Applikation implementiert, die ersatzweise für die CC „Sensor 1“ in das System integriert werden kann.

Die Verschaltungen zwischen den *Komponenten* wurden durch flexible OPC UA-Verbindungen umgesetzt, die auf dem in [15] vorgestellten Ansatz basieren. Der zu regelnde Temperaturprozess wurde als Simulation auf einer weiteren PLCSim-Instanz umgesetzt, die den *Komponenten* die jeweils benötigten Werte bereitstellt. Auf diese Weise konnten die in den Kapiteln 3.1 und 3.2 vorgestellten Konzepte sowie die *Komponenten*-Orchestrierung aus Kapitel 3.3 validiert werden. Zudem wurde der in Kapitel 3.4 vorgestellte Konfigurations-Workflow vereinfacht umgesetzt. Die Modellierung der Anlagenarchitektur mittels eines Layer-F-Import/Export-Formats wurde händisch gezeigt. Im Schritt des Konfigurations-Managements wurden die

gerätespezifischen Konfigurationen mittels UAExpert von Unified Automation unter Nutzung von OPC UA-Methoden auf die verschiedenen Controller heruntergeladen.

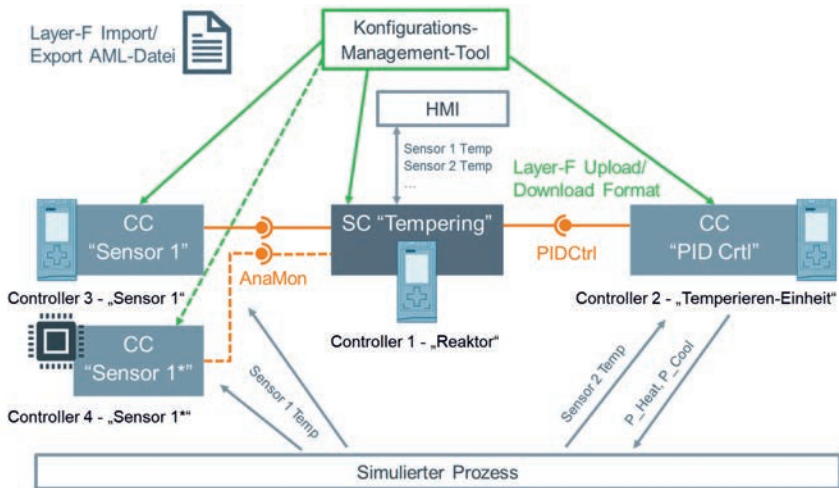


Bild 9: Aufbau des erstellten Demonstrators

Am Beispiel des Austauschs der CC „Sensor 1“ gegen die CC „Sensor 1*“ konnte anhand dieses Demonstrators gezeigt werden, dass das vorgestellte Konzept eine schnelle und einfache Austauschbarkeit von *Komponenten* ermöglicht. Zudem wurde deutlich, dass eine plattform-unabhängige Implementierung von *Komponenten* möglich ist. Die Node.js®-basierte CC „Sensor 1*“ konnte ohne großen Aufwand in das aus klassischen PLCs bestehende System eingebunden werden.

5 Zusammenfassung und Ausblick

Ziel dieses Beitrags ist der Vergleich und die Zusammenführung des Open Process Automation Standard (OPAS) und der Richtlinie VDI/VDE/NAMUR 2658. Wesentliche Unterschiede beider Standards stellten sich dabei in Bezug auf die Modularität, Interoperabilität, Anlagenkoordination und Konfiguration von Prozessanlagen heraus. Um die identifizierten Unterschiede zu überwinden, wurden *Control Components* (CCs) und *Service Components* (SCs) als kleinste Bestandteile eingeführt, aus denen sich die Automatisierungssoftware einer Prozessanlage zusammensetzen kann. Eine SC und beliebig viele CCs können dabei unter Nutzung der *Komponenten*-Orchestrierung zu MTP-Services zusammengefügt werden. Um dabei die

Integration von CC-Funktionalität in SCs zu realisieren, wurde das sogenannte Provider-Proxy-Pattern eingeführt. Die aus *Komponenten* aufgebauten MTP-Services können mittels der Anlagen-Orchestrierung nach klassischen MTP-Orchestrierungsmethoden zu einer Gesamtanlage zusammengefügt werden. Zudem wurde ein Konfigurations-Workflow aufgezeigt, der das Engineering einer aus *Komponenten* aufgebauten Anlage realisiert. Die erstellten Konzepte wurden anhand eines Demonstrators validiert, der den Anwendungsfall einer kaskadierten Temperaturregelung umsetzt. An diesem Demonstrator konnte der einfache, plattformunabhängige Austausch von *Komponenten* am Beispiel eines Sensortauschs gezeigt werden. Zusammenfassend zeigt der vorliegende Beitrag, dass die Zusammenführung von OPAS und der VDI/VDE/NAMUR 2658 unter Nutzung der vorgestellten Konzepte möglich und sinnvoll ist. Allerdings wurden im Zuge der Erarbeitung der Inhalte dieses Beitrags auch weiterführende Forschungs- und Entwicklungsbedarfe identifiziert. So ist es beispielsweise sinnvoll, in Bezug auf das Modularitäts-Konzept ein Regelwerk zum Entwurf von SCs und CCs zu entwickeln. Im Rahmen der Interoperabilität ist zu validieren, ob die vorgeschlagenen CC-Schnittstellen für den praktischen Einsatz ausreichend sind oder inwiefern diese erweitert werden müssen. In einer Weiterführung der Betrachtung zur Anlagenkoordination kann neben einer Integration von CCs in SCs auch eine Integration von CCs in CCs untersucht werden. Im Zuge der Konfigurations-Thematik gilt es, u.a. eine Toolchain zu entwickeln, welche Anlagenbauer bei der Umsetzung des vorgestellten Engineering-Workflows unterstützt.

Literatur

- [1] Krauß, M., Klinger, A., Kotsch, C., Caigny, J. de, Grote, W., Norris, J., Haller, A., Schubert, J. u. Pamar, A.: Open Automation Demonstrator for the Chemical Industry. Synergies of O-PAS, MTP and NOA. *atp magazin* (2020) 06-07, S. 60–67
- [2] 2020-02. *O-PAS™ STANDARD, VERSION 2.0*. <https://publications.opengroup.org/i201m>, abgerufen am: 17.11.2020
- [3] *VDI/VDE/NAMUR 2658: Automatisierungstechnisches Engineering modularer Anlagen in der Prozessindustrie*. https://www.vdi.de/suche?tx_solr%5Bq%5D=VDI%2FVDE%2FNAMUR+2658, abgerufen am: 17.11.2020
- [4] Leeuw, V. de: Industry Standards Groups to Converge Their Work, 2017. <https://www.arcweb.com/blog/industry-standards-groups-converge-their-work>, abgerufen am: 11.01.2020
- [5] P201-1:2020-02. *O-PAS™ Standard, Version 2.0: Part 1 – Technical Architecture Overview (Informative)*. <https://publications.opengroup.org/i201m>

- [6] VDI 2776-1: *Verfahrenstechnische Anlagen - Modulare Anlagen - Grundlagen und Planung modularer Anlagen*. <https://www.vdi.de/richtlinien/details/vdi-2776-blatt-1-verfahrenstechnische-anlagen-modulare-anlagen-grundlagen-und-planung-modularer-anlagen>, abgerufen am: 20.01.2021
- [7] 2019-10. VDI/VDE/NAMUR 2658-1: *Automation engineering of modular systems in the process industry - General concept and interfaces*. <https://www.vdi.de/richtlinien/details/vdivdenamur-2658-blatt-1-automatisierungstechnisches-engineering-modularer-anlagen-in-der-prozessindustrie-allgemeines-konzept-und-schnittstellen>, abgerufen am: 08.06.2020
- [8] VDI/VDE/NAMUR 2658-4: *Automatisierungstechnisches Engineering modularer Anlagen in der Prozessindustrie: Services für Process Equipment Assemblies*
- [9] Stutz, A., Maurmaier, M., Spaethe, F. u. Hill, P.: Einführung, Entwicklung und Anwendung von sub-elementaren Services für modulare Anlagen. In: 20. Leitkongress der Mess- und Automatisierungstechnik, S. 809–822
- [10] 1995-10. ANSI/ISA–88.01–1995: *Batch Control: Part 1: Models and Terminology*. gmpua.com/GAMP/ISA-88.pdf, abgerufen am: 06.06.2020
- [11] Stutz, A., Fay, A., Barth, M. u. Maurmaier, M.: Orchestration vs. Choreography - Functional Association for Future Automation Systems. In: 21st IFAC World Congress 2020
- [12] Stutz, A., Fay, A., Barth, M. u. Maurmaier, M.: Choreographies in Microservice-Based Automation Architectures. Next Level of Flexibility for Industrial Cyber-Physical Systems. In: ICPS 2020 : IEEE International Conference on Industrial Cyber-Physical Systems
- [13] 2019-03. VDI/VDE/NAMUR 2658-3: *Automation engineering of modular systems in the process industry - Library for data objects*. <https://www.vdi.de/richtlinien/details/vdivdenamur-2658-blatt-3-automatisierungstechnisches-engineering-modularer-anlagen-in-der-prozessindustrie-bibliothek-fuer-datenobjekte>, abgerufen am: 08.06.2020
- [14] Blumenstein, M., Stutz, A. u. Maurmaier, M.: Rollenbasierte Entwurfsmethodik für modulare Anlagen auf Basis von sogenannten Super-Services. In: AUTOMATION 2019, S. 197–214
- [15] Stutz, A. u. Maurmaier, M.: Dezentralisierte Kommunikationsarchitektur für direkte Modul-zu-Modul-Interaktion auf Basis von standardisierten Konstrukten der VDI/VDE/Namur 2658. In: AUTOMATION 2019, S. 343–360

Rezeptbasierte Orchestrierung einer modularen kontinuierlichen Prozessanlage – Ergebnisse und Erkenntnisse aus dem Forschungsprojekt ENPRO2-ORCA

Recipe-based orchestration of a modular continuous process plant – Results and insights from the research project ENPRO2-ORCA

Andreas Stutz, Dr. Mathias Maurmaier,
Siemens AG, Karlsruhe;
Manfred Eckert, Heiko Mannheim, Hidir Ercan Akar,
Merck KGaA, Darmstadt

Kurzfassung

Wandelbare Prozesseinheiten sind Grundlage, um den neuen Marktanforderungen nach kürzerer Time-to-Market mit wirtschaftlichen Maßnahmen entgegenzuwirken. In diesem Beitrag wird ein Service-Konzept für die Abstraktion und Umsetzung wandelbarer Control Module Services sowie deren Nutzung im Rahmen einer modularen kontinuierlichen Prozessanlage und einem Client-Server-basierten POL-System evaluiert. Beide Bestandteile sind Ergebnisse einer Kooperation innerhalb des Forschungsprojekt ENPRO 2.0 ORCA zwischen den Firmen Merck KGaA und der Siemens AG. Die Ergebnisse und Erkenntnisse werden in diesem Beitrag dargestellt und bewertet.

Abstract

Adaptable process units are the basis for facing the new market challenges for shorter time to market with economic solutions. This paper presents a service concept for the abstraction and implementation of adaptable control module services and evaluates their use in the context of a modular continuous process plant and a client-server-based POL-system. Both components are results of a cooperation within the research project ENPRO 2.0 ORCA between the companies Merck KGaA and Siemens AG. The results and findings are presented and discussed in this paper.

1. Einführung und Motivation

Flexibilität und Wandelbarkeit in intelligenten modularen Prozesseinheiten ist eine Voraussetzung, um den steigenden Anforderungen in verfahrenstechnischen Mehrzweck-Systemen auf wirtschaftlichem Wege zu begegnen. Durch die Wiederverwendung einer Vielzahl von Komponenten (Component, COMP), modularer Funktions- (Functional Equipment Assembly, FEA) und Prozesseinheiten (Process Equipment Assembly, PEA) wird der Engineering-Aufwand und das finanzielle Risiko in der Produkt- und Prozessentwicklung reduziert. Aus diesen modularen Einheiten, die jeweils in einem Modul Type Package (MTP) beschrieben sind, wird die modulare Prozessanlage (Modular Plant, MP) aufgebaut. [1]

In Bezug auf die Automatisierung ist hierfür ein übergeordnetes koordinierendes System notwendig, der sogenannte Process Orchestration Layer (POL). Die Hauptaufgabe des POL liegt in der Orchestrierung der modularen Einheiten.

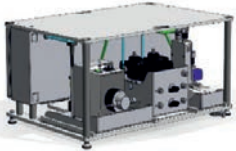
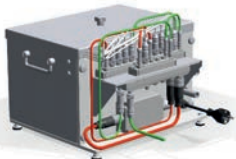
In den vergangenen drei Jahren wurde im Rahmen des Forschungsprojekts ENPRO 2.0 ORCA [2] an Konzepten der effizienten Orchestrierung modularer Prozessanlagen gearbeitet. Hierbei lagen die Schwerpunkte einerseits auf einem Konzept zur Service-orientierten Automatisierung modularer Prozesseinheiten und andererseits auf deren effiziente Orchestrierung als modulare Prozessanlage. Die beiden Schwerpunktbereiche wurden im Rahmen von anwendungsnahen Demonstratoren parallel zum Projektfortschritt untersucht und die Erfüllung vorab definierter Anforderungen bewertet. Die Anforderungen an die Orchestrierung modularer Anlagen wurde bereits veröffentlicht [3].

In diesem Beitrag präsentieren die Merck KGaA und die Siemens AG Ergebnisse und Erkenntnisse ihrer Kooperation im Rahmen des Forschungsprojekts ENPRO 2.0 ORCA. In Anlehnung an die Projektstruktur unterteilt sich dieser Beitrag in drei Hauptkapitel. Kapitel 2 präsentiert den im Fokus stehenden Demonstrator der Fa. Merck – die modulare kontinuierliche Prozessanlage (MKP). Dieses Realbeispiel dient im Rahmen der Konzeptentwicklungen als wiederkehrendes Anwendungsbeispiel zur Verdeutlichung der Sachverhalte. Kapitel 3 präsentiert ein entwickeltes Service-Konzept, welches die Wandelbarkeit modularer Prozesseinheiten für das Labor verbessert. Das Konzept findet mehrfach Anwendung in der MKP. Des Weiteren beschreibt Kapitel 4 die Anwendung einer Client-Server-Architektur zur Orchestrierung der modularen kontinuierlichen Prozessanlage. Hierbei wird ein prototypisches POL-System auf Basis des Prozessleitsystems SIMATIC PCS 7 gezeigt, welches große Teile der Anforderungen, veröffentlicht in [3], prototypisch umsetzt. Sowohl in Kapitel 3.3 als auch 4.4 wird jeweils ein kurzes Zwischenfazit gezogen und Erkenntnisse dargelegt, bevor in Kapitel 5 eine Gesamtzusammenfassung sowie ein Ausblick auf die kommende Zusammenarbeit gegeben wird.

2. Die modulare kontinuierliche Prozessanlage

Ziel bei der Konzeption und dem Bau der modularen kontinuierlichen Prozessanlage (MKP) war es, ein wandelbares Mehrzweck-System zu realisieren. Mit diesem Mehrzweck-System soll es möglich sein, schnell und einfach ein Versuchsaufbau aufzubauen und anzupassen, ohne dass die Mitarbeiterinnen und Mitarbeiter im Labor und Technikum besondere Kenntnisse in der Automatisierungstechnik benötigen. Die Herausforderung liegt somit im Optimum hinsichtlich Wandelbarkeit und Handhabbarkeit des Systems. Die MKP wurde im Rahmen von ENPRO 2.0 ORCA als eine von mehreren Demonstrationsanlagen verwendet [5]. Die MKP besteht aus vier PEA-Typen - einer Dosier-PEA, einer Rührwerk-PEA, einer Reaktion-PEA und einer Ausschleus-PEA, von welchen jeweils mehrere Instanzen zur Verfügung stehen. Diese PEA-Typen sind sowohl konstruktiv als auch software-technisch auf größtmögliche Wandelbarkeit ausgelegt. Jede PEA besteht aus mehreren FEAs, die einfach durch Lösen einer Arretierung und dem Trennen des Netzwerksteckers getauscht werden können. Auf diese Weise können bspw. Pumpen mit unterschiedlichen Mechanismen einfach ein- und ausgebaut werden. In Tabelle 1 sind die entsprechenden PEA-Typen in tabellarischer Form dargestellt.

Tabelle 1: Darstellung der vier PEA-Typen der MKP und die Anzahl ihrer Services

	
Dosier-PEA mit 2 Services	Rührwerk-PEA mit 3 Services
	
Reaktion-PEA mit 12 Services	Ausschleus-PEA mit 2 Services

Die dargestellten PEA-Typen können in zahlreichen Kombinationen zu einer modularen Gesamtanlagen zusammengesetzt werden. Eine Prozessfunktion wird in der Automatisierungssoftware durch einen Automation Service ansteuer- und überwachbar, kurz orchestrierbar gemacht. Bisher kapselten Automation Services abgeschlossene Prozessfunktionen [5]. Vergleicht man dies mit der IEC 61512 [6] können durch Automation Services die Elemente Phase, Operation oder Unit Procedure des Procedural Models realisiert werden.

Um eine noch höhere Wandelbarkeit zu erreichen ist es sinnvoll, nicht nur abgeschlossene Prozessfunktionen austauschbar zu machen, sondern dies auch auf die Einzelsteuerebene auszuweiten. Aufgrund dessen wird im folgenden Kapitel das MTP-Servicekonzept auf die Einzelsteuerebene ausgeweitet, die in der IEC61512 [6] der Control Module Ebene entspricht. Die Control Module Ebene ist Teil des physikalischen Modells der IEC61512 [6].

3. Konzeptvorstellung wandelbarer Control Module Services

Das Ziel der Control Module Services liegt darin, automatisierungstechnische Funktionen wandelbar zu gestalten, so dass Sensoren und Aktoren, mit denen die Services arbeiten, ausgetauscht werden können. Durch die Konfigurationsmöglichkeiten können einfach zu nutzende Automation Services umgesetzt werden, die Labormitarbeiterinnen und -mitarbeiter mit minimalen Automatisierungskenntnissen an die Prozessanforderungen und -bedingungen eigenständig anpassen können, ohne die Automatisierung umprogrammieren zu müssen. Diese Control Module Services steigern die Wandelbarkeit der PEA und können mit Hilfe der in der VDI/VDE/NAMUR 2658-4 [5] standardisierten Schnittstellen umgesetzt werden. Im nächsten Unterkapitel wird die Grundidee und das Umsetzungskonzept vorgestellt.

3.1 Umsetzungskonzept von Control Module Services

Der Ansatz zur Umsetzung von Control Module Services liegt in der Kombination des Service-Konzepts der VDI/VDE/NAMUR 2658-4 [5] und zur Laufzeit konfigurierbarer Signalbaugruppen [7]. Entwurfsgrundlage ist das Modell der Eingabe (E), Verarbeitung (V) und Ausgabe (A), wodurch eine Automatisierungsfunktion in drei Bestandteile aufgeteilt werden kann. Fokus eines Control Module Service liegt auf der Umsetzung einer solchen Teilfunktion. Durch eine Kombination der drei Bestandteile des EVA-Modells sind bis zu sieben sinnvolle Kombinationen denkbar (E, A, V, EV, VA, EA, EVA). Unter der Verwendung spezieller Hardware-Baugruppen können so höchst wandelbare Service-Systeme umgesetzt werden.

Damit die wandelbaren Services in den POL integriert werden können, sind die in der VDI/VDE/NAMUR 2658-4 [5] spezifizierten Schnittstellen umzusetzen. Die Richtlinie spezifiziert hierfür Schnittstellen in den Klassen Service Control, Prozeduren,

Konfigurationsparameter, Prozedurparameter, Prozesswerte und Reportwerte, die sowohl zustandsbasierte als auch kontinuierliche Automatisierungsfunktionen abrufbar machen. In Tabelle 2 werden die sechs Schnittstellen-Klassen und deren Interpretation und Anwendbarkeit für Control Module Services dargestellt.

Tabelle 2: Auflistung der Schnittstellen-Klassen für Automation Services nach [4] und deren Interpretation im Kontext von Control Module Services

Schnittstellen-Klasse	Interpretation und Anwendungsmöglichkeit
Service Control	Diese Schnittstelle kapselt den standardisierten Zustandsautomat eines Service. Die Automatisierungsfunktion auf Einzelsteuerebene hängt von den gekapselten Bestandteilen ab.
Prozeduren	Ein Service besitzt eine oder mehrere Prozeduren. Bei Control Module Services können mit Hilfe von Prozeduren unterschiedliche Variationen oder Parametersätze für die Funktion über eine Prozedur-ID einfach auswählbar gemacht werden. Eine Prozedur-ID kann somit stellvertretend als eine Variation bzw. eine Konfiguration interpretiert werden.
Prozedurparameter	Jeder Prozedur können bei Bedarf dedizierte Parameter angehängt werden, die Prozedurparameter genannt werden. Mittels dieser Parameter können optionale Diagnose-Funktionen, Parameter für Skalierung und die Einheit an den Control Module Service übergeben werden.
Konfigurationsparameter	Konfigurationsparameter bieten eine dritte Möglichkeit zur Einstellung bzw. Auswahl der Konfiguration. Im Vergleich zu Prozedurparameter gelten Konfigurationsparameter in gleichem Maße für alle Prozeduren. Diese sind im Zuge der Inbetriebnahme zu setzen und können für weitere optionale Features eines Control Module Service eingesetzt werden.
Prozesswerte	Prozesswerte sind Schnittstellen, die kontinuierlichen Datenaustausch ermöglichen. Sie werden in eingehende und ausgehende Prozesswerte unterschieden. Da Control Module Services meist Sensor- oder Aktor-Services sind, haben diese meist offene Prozesswert-Eingänge und -Ausgänge. Die Prozesswertanschlüsse können im Rahmen der Orchestrierung über Broker-basierte oder Peer-to-Peer-basierte Mechanismen verbunden werden.
Reportwerte	Reportwerte dienen Aufzeichnungssystemen im Laborumfeld. Messwerte oder Stellsignale der Control Module Services können für Auswertungen der Versuchsreihen als Reportwerte bereitgestellt werden. Mittels eines standardisierten Freeze-Mechanismus ist sichergestellt, dass von einem Service immer eine konsistente Menge an Werten abgerufen werden kann.

Die hier vorgestellten Interpretations- und Anwendungsmöglichkeiten der standardisierten Schnittstellen-Klassen auf das Konzept der Control Module Services werden im nächsten Kapitel durch ein praktisches Anwendungsbeispiel verdeutlicht. Das Anwendungsbeispiel zeigt einen Mess-Service, der mit Hilfe einer konfigurierbaren analogen Eingangssignalbaugruppe umgesetzt wird.

3.2 Anwendungsbeispiel wandelbarer Mess-Services

Grundlage für den wandelbaren Messen-Service ist eine zur Laufzeit konfigurierbare analoge Eingangssignal-Baugruppe (z.B. SIMATIC ET200SP AI 4xU/I 2-wire ST [7]). Im Falle des betrachteten Mess-Services bedeutet Wandelbarkeit, die Möglichkeit zu haben unterschiedliche Sensoren mit unterschiedlichem Messprinzip und Messbereich anbinden zu können. Die genannte Signalbaugruppe besitzt dazu vier in Messprinzip, Messbereich und Glättung konfigurierbare Kanäle. Neben den kanalgranularen Funktionen können auch baugruppenweite Diagnose- sowie Vorverarbeitungsfunktionen eingestellt werden.

Ein solcher Mess-Service kann nun in unterschiedlicher Form umgesetzt werden – ein Service mit vier Kanälen oder vier Services mit je einem Kanal. Beide Fälle haben gewisse Einschränkungen, die sich auf die Flexibilität und Wandelbarkeit auswirken. Im Falle von einem Service und vier Kanälen können die Kanäle nur in Summe ein- bzw. ausgeschaltet werden. Ein Umkonfigurieren ist somit nicht kanalgranular möglich. Im Falle von vier Services mit einem Kanal ist die Konfigurierbarkeit der Baugruppeneinstellungen eingeschränkt. Diese Einstellungen sind lediglich an einem der vier Services vorhanden.

Zum Entwurf des Mess-Service gilt es zunächst zu entscheiden, ob es einen Mess-Service mit vier Kanälen oder vier Services mit einem Kanal geben soll. In diesem Beitrag wird der Service für Kanal 0 dargestellt in der Variante vier Services mit einem Kanal. Zudem wurden folgende Entwurfsentscheidungen getroffen:

- Das Messprinzip und der Messbereich werden in Form von sechs Prozeduren abgebildet (siehe Tabelle 3).
- Das Signal von Kanal 0 wird als ausgehender Prozesswert am Service bereitgestellt.
- Die Skalierung, die Einheit des ausgehenden Prozesswerts sowie die Glättung des Signals werden als Prozedurparameter abgebildet.
- Die Diagnosefunktionen und Störfrequenzunterdrückung der Baugruppe werden als Konfigurationsparameter abgebildet und gelten für alle Services der verwendeten Baugruppe.

Somit ergibt sich die in Tabelle 3 dargestellte Service-Signatur für den Mess-Service von Kanal 0.

Tabelle 3: Service-Spezifikation eines Control Module Service in Form eines wandelbaren
Mess-Service

MEAS-CH0-Service* ¹		Prozedur 1 (U, +/-5V)	Prozedur 2 (U, +/- 10V)	Prozedur 3 (U, 1..5V)	Prozedur 4 (U, 0..10V)	Prozedur 5 (I, 0..20mA)	Prozedur 6 (I, 4..20mA)
Prozedurparameter							
ScaleMin	REAL	Alle Prozeduren besitzen diese vier Prozedurparameter					
ScaleMax	REAL						
Unit* ³	DINT (Enum)						
Smoothing* ⁴	DINT (Enum)						
Konfigurationsparameter* ²							
Noise Suppression* ⁵	DINT (Enum)	Für alle Prozeduren von Kanal 0 bis Kanal 3 Service gültig.					
PowerFailure (L+)	BOOL						
ShortCircuit (M)	BOOL						
WireBreak	BOOL						
Underflow	BOOL						
Overflow	BOOL						
Ausgehender Prozesswert							
ProcessValue	REAL	Der Prozesswert ist für alle Prozeduren relevant					
Reportwerte							
ValueIndication	REAL	Die Reportwerte sind für alle Prozeduren relevant					
MinTowIndica- tion	REAL						
MaxTowIndica- tion	REAL						
Kommentar							
(* ¹) am Beispiel des Analog-Eingabemoduls - AI 4xU/I 2-wire ST (6ES7134-6HD01-0BA1)							
(* ²) Nur am CH0-Service, CH1-3 hat die gleichen Einstellungen (Baugruppen-bedingt)							
(* ³) siehe VDI/VDE/NAMUR 2658-3							
(* ⁴) Enumeration aus (keiner, schwacher, mittlerer, starker) Glättung							
(* ⁵) Enumeration aus (60Hz, 50Hz, 16,6Hz) Störfrequenzunterdrückung							

Beim Start einer Prozedur wird im Zustand „Starting“ die entsprechende Konfiguration an die konfigurierbare Baugruppe übertragen und diese aktiviert. Im Zustand „Execute“ wird der erfasste Messwert über eine Normierung und Skalierung vorverarbeitet und am Prozesswertausgang bereitgestellt. Es ist kein direktes Umschalten zwischen Prozeduren möglich, da eine

Umschaltung auch ein Umbau des Sensors zur Folge hätte. Im Zustand „Completing“ wird der Kanal deaktiviert und die Konfiguration zurückgesetzt. Im Zustand „Completed“ kann der Sensor hardware-technisch entfernt oder getauscht werden. Der Zustand „Resetting“ setzt die beiden Schleppanzeigen (Min- und MaxTowIndication) auf Default zurück. Eine Übernahme der Prozedur- und Konfigurationsparameter ist ausschließlich im Zustand „Idle“ erlaubt, da der Kanal in „Starting“ aktiviert und in „Completed“, „Stopped“ und „Aborted“ deaktiviert wird. Während der Kanal aktiviert ist, ist ein Umkonfigurieren nicht möglich.

Diese Art von Services ist in der modularen kontinuierlichen Prozessanlage z.B. in der Reaktion-PEA und der Rührwerk-PEA verbaut und in Anwendung. Somit kann die Anlage ohne automatisierungsspezifische Kenntnisse umkonfiguriert und mit unterschiedlichen Sensoren bestückt werden. Die dargestellte exemplarische Umsetzung ist eine von vielen Möglichkeiten, einen Control Module Service umzusetzen. Aktor-Services mit entsprechend konfigurierbaren Ausgabebaugruppen sowie die Einbindung eines konfigurierbaren PID-Controllers sind ebenfalls denkbar.

3.3 Erkenntnisse aus der Anwendung von Control Module Services

Die Vorteile und Möglichkeiten von Control Module Services als Sensor- sowie als Aktor-Service haben sich bestätigt. Die Wandelbarkeit hinsichtlich der nutzbaren Sensoren und Aktoren ohne die Voraussetzung tiefergehender Automatisierungskennnisse eröffnen Mitarbeitern im Labor- und Technikum neue Möglichkeiten. Konfigurierbare Signalbaugruppen wie die oben erwähnte konfigurierbare analoge Eingangssignal-Baugruppe (z.B. SIMATIC ET200SP AI 4xU/I 2-wire ST) gibt es in verschiedensten Baugruppenklassen sowie für verschiedene Steuerungs- und Remote-IO-Familien. Die Control Module Services in Kombination mit austauschbaren Komponenten resultieren in modulare Einheiten mit größtmöglicher Wandelbarkeit.

4. Konzeptvorstellung eines client-server-basierten Process Orchestration Layer

Wie sich in einer ersten Studie zur Umsetzung des Process Orchestration Layer [8] gezeigt hat, hängt die Ausstattung und Zusammensetzung stark von dem entsprechenden Anwendungsfall und den Bedürfnissen ab. Aus diesem Grund wurden im Forschungsprojekt ENPRO 2.0 ORCA [2] zusammen mit Merck drei Komplexitätsstufen eines POL untersucht – (1) reines Bedien- und Beobachtungssystem, (2) rezeptbasierte Orchestrierung in einer Einzelplatzkonfiguration und (3) eine Client-Server-Architektur in einer Mehrplatzkonfiguration. In diesem Beitrag wird ausschließlich die umfangreichste Stufe (Stufe 3) betrachtet, da diese auf den beiden Stufen 1 und 2 aufbaut und diese erweitert. Grundlage des POL-Systems ist das

Prozessleitsystem SIMATIC PCS 7 mit prototypischen Erweiterungen, die die Umsetzung des Module Type Package auf die PCS 7 Datenhaltung durchführen.

4.1 Systemarchitektur einer Mehrplatzkonfiguration

Im Zuge der Stufe 3 wurde ein virtualisiertes Client-Server System basierend auf drei virtualisierten Rechnern – Engineering Station (ES), Operator Station Server/Client (OS) und Batch Server/Client (Batch) aufgesetzt. Mittels Remote-Desktop Verbindung kann das eingerichtete Client-Server-System von beliebigen Rechner-Stationen im Merck-Labor-Netzwerk erreicht werden. Die Ausführung des Batch-Systems erfordert eine SIMATIC S7-400 Station, die als POL-Controller dient. Aufgabe dieser Steuerung ist die PLC-seitige Ausführung von projektierten Chargen-Rezepten der Batch-Runtime. Alle Systemkomponenten des Client-Server-Architektur-basierten POL-Systems sind mit ihren Zusammenhängen in der Systemdarstellung in Bild 1 dargestellt. Alle PEAs, die für die Ausführung der Anlage notwendig sind, sind mit dem Netzwerk zu verbinden und entsprechend des nachfolgenden Engineering-Workflows in das POL-Projekt einzubinden, zu laden und auszuführen.

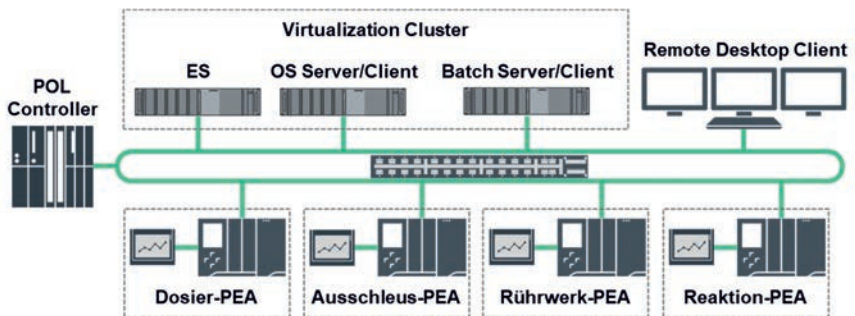


Bild 1: Stufe-3 Systemarchitektur des prototypischen POL Systems der evaluierten modularen kontinuierlichen Prozessanlage

4.2 Engineering-Workflow bei einer Mehrplatzkonfiguration

Ansatz für die Ausgestaltung des POL Engineering Workflow ist eine Kombination aus dem bereits bekannten SIMATIC PCS 7 Engineering Workflow und MTP-spezifischen Ergänzungen. Diese Entscheidung basiert auf der Annahme, dass auch in der Zukunft ein hybrider Ansatz aus teils konventionellen und teils modularen Anlagenteilen bestehen wird. Aus diesem

Grund wird der SIMATIC PCS 7 Engineering Workflow um die Anwendung und Ausnutzung von MTP-unterstützenden Automatismen ergänzt. Um die Informationen des MTP in einem PCS 7 Projekt zu nutzen, wird dies an Stelle der klassischen Projektierung in der technologischen Hierarchie eingefügt. Mit Hilfe des PCS 7 Integration and Assignment Tools werden die Informationen des Module Type Package interpretiert und in das PCS 7 Projekt übertragen. Es ergibt sich der in Bild 2 dargestellte Engineering Workflow.

Die Schritte 1-5 stellen die konventionellen Schritte des PCS 7 Engineering-Workflow in einer sehr komprimierten Form dar. Diese Projektierung kann in ein Projektarchiv abgelegt werden, womit die Schritte 1-5 wiederverwendbar gestaltet werden. Falls ein Template vorhanden ist, sollte dies genutzt werden und es kann direkt bei Punkt 6 des Workflows begonnen werden. Schritte 20-22 sind für ein neues Rezept zu wiederholen, wenn die Anlage in ihrer Struktur unverändert bleibt. Schritte 6-22 sind zu wiederholen, wenn ein Modul hinzugefügt oder entfernt werden soll. Ein Modultauch mit einer PEA identischen Typs ist einfach durch eine Konfigurationsänderung im SIMATIC WinCC Tag Management umzusetzen. Hier ist am Kommunikationskanal der Verbindung lediglich eine andere OPC UA Server Endpoint Adresse einzutragen.

Durch die Kombination des PCS 7 Workflows und des MTP Workflows wurde ein sich ergänzender Workflow aus Brownfield und Greenfield Ansätzen entwickelt.

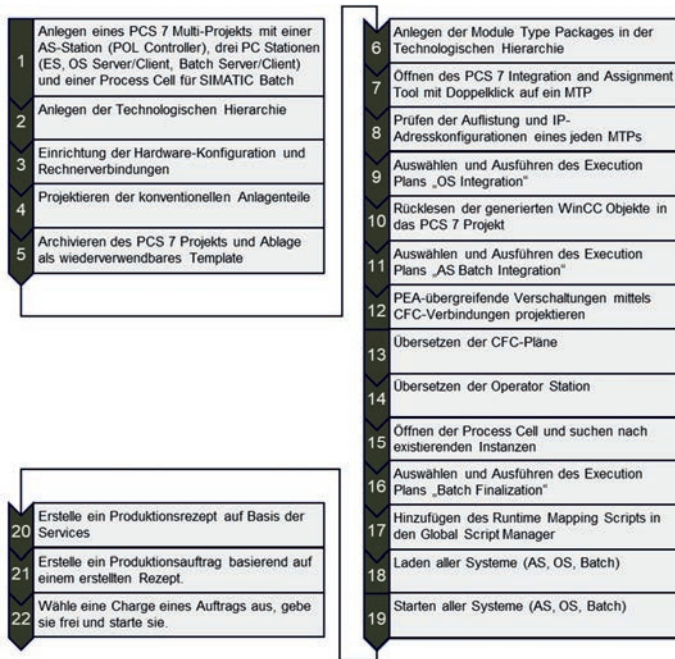


Bild 2: Workflow-Darstellung des kombinierten PCS 7- und POL-Engineering-Workflow des prototypischen POL-Systems

4.3. Szenarienbasierte Evaluation des Process Orchestration Systems

Im Rahmen des Forschungsprojekts wurden 13 Testszenarien definiert, die wiederum in 6 Haupt-Szenarien zusammengefasst sind. In Summe umfassen die 13 Testszenarien 172 Testschritte. Eine Auflistung der Testszenarien sowie die Ergebnisse der Testschritte sind in Tabelle 4 dargestellt. Eine Auflistung der detaillierten Testschritte ist aus Platzgründen nicht möglich. Die Ergebnisse der Testschritte sind nach „ja“, „nein“ und „nicht getestet“ aufgeschlüsselt mit folgender Bedeutung:

„ja“ bedeutet, dass der Testschritt durchgeführt wurde und erfolgreich war

„nein“ bedeutet, dass der Testschritt durchgeführt wurde und nicht erfolgreich war

„nicht getestet“ bedeutet, dass die für den Testschritt notwendige PEA-Funktionen nicht zur Verfügung stand.

Die Zahl in der entsprechenden Ergebnisspalte in Tabelle 4 spiegelt die Anzahl der mit „ja“, „nein“ und „nicht getestet“ bewerteten Testschritte wider. Es wird deutlich, dass 79% der Testschritte erfolgreich abgeschlossen wurden. 6% der Testschritte wurden nicht erfolgreich abgeschlossen. Diese lassen sich teils auf die Notwendigkeit eines erneuten „Übersetzen“ und „Laden“ zurückführen oder auf nicht implementierte Features im prototypischen POL-System. Eine Übersetzung ist in den meisten Systemen heute notwendig, da die Engineering-Daten nicht automatisch in die Laufzeit-Systeme übertragen werden. Die restlichen 15% („nicht getestet“) lassen sich zum einen darauf zurückführen, dass zu testende Funktionen in den von Merck bereitgestellten Demonstrationsmodulen nicht eingebaut sind und zum anderen darauf, dass die Tests nur mit Wasser gefahren wurden. In dem von Merck durchgeführten Test wurden zusätzlich 15 Verbesserungsvorschläge sowohl für den Workflow als auch die Benutzeroberfläche des POL-Systems aufgelistet. Darüber hinaus wurden 8 Funktionen positiv bewertet, die nicht in dem eines POL-Systems geforderten Minimalumfang enthalten sind.

Tabelle 4: Auflistung der Testszenarien und die Ergebnisse der einzelnen Testschritte

Testszenario		Testergebnisse (Anzahl der Testschritte)		
#	Titel	Ja	Nein	Nicht getestet
Hauptszenario „Neuanlage“				
N1	Neuanlage	18	5	0
Hauptszenario „Kapazitätsänderung“				
K1	Kapazitätsänderung mit nicht baugleicher, aber identische Funktion	9	0	2
K2	Kapazitätsänderung mit nicht baugleicher und unterschiedlicher Funktion	8	1	2
K3	Kapazitätsänderung mit baugleicher und identischer Funktion (NumberingUp)	8	1	2
Hauptszenario „Funktionsänderung“				
F1	Funktionsänderung mit baugleicher und unterschiedlicher Funktion	8	1	2
F2	Funktionsänderung mit nicht baugleicher und unterschiedlicher Funktion	8	1	2
Hauptszenario „Modulrevision“				
M1	Modulrevision mit nicht baugleicher, aber identischer Funktion	9	0	2
M2	Modulrevision mit nicht baugleicher und unterschiedlicher Funktion	8	1	2
M3	Modulrevision mit baugleicher und identischer Funktion	6	0	2
M4	Modulrevision mit baugleicher und unterschiedlicher Funktion	8	1	2
Hauptszenario „Produktänderung“				
P1	Produktänderung	8	0	1
Hauptszenario „Operator-Bedienung“				
CM	Operator-Bedienung der Einzelsteuerebene (Control Module)	22	0	4
SM	Operator-Bedienung der Serviceebene (Service Module)	16	0	2
Zusammenfassung				
Spaltensumme (absolut)		136	11	25
Spaltensumme (relativ)		79%	6%	15%

4.4. Erkenntnisse aus der Anwendung des entwickelten Process Orchestration Layer

Trotz des positiven Ergebnisses, dass eine Orchestrierung mittels des prototypischen POL-System möglich ist, zeigen sich in der Effizienz bei der Benutzung dennoch einige Schwachstellen. Das prototypische POL-System baut auf SIMATIC PCS 7 und den beiden Softwarepaketen SIMATIC OpenOS und SIMATIC Batch auf. SIMATIC OpenOS ist eine Systemerweiterung zur Integration von Dritt-Steuerung in Form eines Projektierungswerkzeugs und einer Laufzeiterweiterung. Es ist Teil des automatisierten Integrationsprozesses, der durch ein eigens für das Projekt entwickeltes Tool vollständig im Hintergrund genutzt wurde. SIMATIC Batch ist die Basis für die rezeptbasierte Orchestrierung der Anlage.

Sowohl die Nutzung der Systeme als auch die Einrichtung und Nutzung der Client-Server-Struktur in einem sich häufig ändernden System erfordern tiefergehende Automatisierungskennntnisse und Kenntnisse im Umgang mit den Systemen. Laborpersonal besitzt normalerweise diese Kenntnisse nicht.

Alle nicht erfolgreichen Tests (6%) sind auf die Notwendigkeit eines Übersetzungs- und Ladevorgangs zurückzuführen. Strukturelle Veränderungen durch Kapazitätsänderung, Funktionsänderung oder Modulrevision mündet in der Mehrzahl der Szenarien darin, dass das POL-System nach der Offline-Änderung neu übersetzt und geladen werden muss. Diese Vorgänge sind auf die Art heutiger Prozessleitsysteme zurückzuführen.

Zusammenfassend lässt sich sagen, dass hoch wandelbare Prozessorchestrierungssysteme grundlegende Änderungen in ihrer Umsetzung und Ausgestaltung erfahren müssen. Die Trennung zwischen Engineering und Laufzeit verschwindet zunehmend. Informationsmodelle im Engineering und der Laufzeit müssen möglichst identisch sein. Je nach Anwendungsfall werden unterschiedliche Teilfunktionen im POL-System erforderlich. Heutige Prozessleitsysteme, die eine solide Basis für zukünftige POL-Systeme bieten können, sind für den Einsatz in Produktionsumgebungen sicherlich gut geeignet. Nicht zuletzt auch daher, da dort eigene Automatisierungsspezialisten eingesetzt werden. Im Umfeld von schnell aufzubauenden Tests zur Prozessentwicklung sind meist die Labor- und Technikums-Angestellten diejenigen, die auch die Anlage zusammenstellen, automatisieren und betreiben. Aus diesem Grund sind die Anforderungen hinsichtlich Benutzerfreundlichkeit nochmal stärker in den Fokus zu rücken. Darüber hinaus wurde eine Koexistenz neuartiger POL-Systeme mit Bestandssystemen z.B. Labor-Infrastruktursysteme, die teils auch mit Prozessleittechnik umgesetzt sind, diskutiert. Hier haben sich auch Lücken aufgetan, die im Rahmen von ORCA jedoch nicht näher untersucht werden konnten.

5. Zusammenfassung und Ausblick

Dieser Beitrag präsentiert eine Auswahl der Ergebnisse und Erkenntnisse aus der Zusammenarbeit von Merck und Siemens. Innerhalb des Forschungsprojekts ORCA aber auch bilateral arbeiten Merck und Siemens am Module Type Package Konzept. Die gemeinsame Entwicklung von Service-Konzepten und der Austausch über Umsetzungsmöglichkeiten bildeten die Basis für die automatisierungstechnische Realisierung der modularen kontinuierlichen Prozessanlage.

Die im Rahmen der Service-Konzepten entwickelten Control Module Services bieten eine Vielzahl von Möglichkeiten, hoch wandelbare automatisierte Einheiten zu realisieren. Die Services sind derart konzipiert, dass sie ohne tiefergehende Automatisierungskenntnisse bedient und verwendet werden können. Mit Hilfe dieses Konzepts wurde die Grundlage für eine beschleunigte Prozessentwicklung im Labor- und Technikumsumfeld gelegt.

Die Evaluierung des auf SIMATIC PCS 7 basierenden prototypischen POL-Systems hat gezeigt, dass eine Orchestrierung möglich ist. Eine differenzierte Betrachtung zwischen Produktion und Prozessentwicklung zeigt, dass etablierte Workflows und Mechanismen in einer modularen Welt nur bedingt effizient sind, wenn sich die Systeme häufig ändern. Demzufolge eignet sich das prototypische POL-System vorzugsweise für ein Produktionsumfeld, jedoch nicht für Labor und Technikum.

Daraus abgeleitet ergeben sich neue Forschungsfragen in Richtung funktional modularisierter POL-Systeme, die die Anforderungen wie Benutzerfreundlichkeit, Einfachheit und Anpassbarkeit an die Bedürfnisse des Kunden und dessen Problemstellung besser unterstützen. Einige der hier genannten Punkte und die noch unbeantworteten Fragen sollen in einer möglichen Fortsetzung des Forschungsprojekts weiter untersucht werden.

Die Autoren danken dem Bundesministerium für Wirtschaft und Energie sowie dem Projektträger Jülich für die finanzielle Unterstützung im Rahmen der ENPRO 2.0 Initiative (FKZ 03ET1517).

Literatur

- [1] VDI 2776-1, "Process engineering plants - Modular plants - Fundamentals and planning modular plants", 2020, www.vdi.de/2776
- [2] ORCA Project Website, http://enpro-initiative.de/ENPRO+2_0/ORCA-p-275.html, 13.10.2019
- [3] A. Klose et al., "Orchestration Requirements for Modular Process Plants in Chemical and Pharmaceutical Industries", Chemical Engineering & Technology, Vol. 42, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2282-2291, 2019
- [4] Bittorf L. et al.; „Demonstration of service-controlled process equipment assemblies and the efficient orchestration of modular plants“, AUTOMATION 2020, Baden-Baden (Online)
- [5] VDI/VDE/NAMUR 2658-4, "Automation engineering of modular systems in the process industry - Modelling of module services", 2020, www.vdi.de/2658
- [6] IEC 61512-1; "Batch control - Part 1: Models and terminology"; Edition 1.0; 1997, doi: 10.31030/8529860
- [7] Gerätehandbuch Analogeingabemodul AI 4x U/I 2-wire ST (6ES7134-6HD01-0BA1); A5E03572831-AF
- [8] Stutz A., Maurmaier M., Korner C., Berenz M.; „The Process Orchestration Layer - An Overview of typical realization possibilities for the process control systems for modular production facilities“, AUTOMATION 2020, Baden-Baden (Online)

Integriertes Vorgehensmodell für die Planung und Realisierung von modularen Plug & Produce Systemen in der Pharmaproduktion

Adel Jebari (M. Eng), **Holger Mettler** (M.A.),
Exyte Central Europe GmbH, Stuttgart

Kurzfassung

In diesem Beitrag wird ein integriertes Vorgehensmodell zur Planung und Umsetzung von modularen und computerisierten Produktionssystemen und -architekturen in der regulierten Pharmaproduktion vorgestellt. Ausgangspunkt ist das standardisierte Design von Produktionsumgebungen (z.B. Reinräume) mittels vorgefertigter, architektonischer Elemente und vorgefertigten Anlagen für Produktionseinrichtungen. Dieses flexible architektonische Konzept (ExyCell®) bietet auch die Möglichkeit der Integration von flexiblen, modularisierten Automations- und IT-Komponenten zur Steuerung und Überwachung der Herstellprozesse (Plug & Produce). Durch die Berücksichtigung von Standards wie NAMUR Module Type Package (MTP) und der VDI Standards (VDI 2776, 2658) wird die Entwurfsplanung und Realisierung von Anlagen in Modulbauweisen erleichtert. Somit zeigt dieser Beitrag auch die Digitale Transformation im Engineering und dem Planen von modularen und computerisierten Anlagen für die Pharmaindustrie. Ein wichtiger Aspekt ist hierbei die Berücksichtigung der gesetzlich vorgeschriebenen GMP-Anforderungen in der Planung und im Engineering dieser computerisierten und digital vernetzten Anlagen. Eine der möglichen Anwendungen von solchen modularen Konzepten ist insbesondere der Aufbau von flexiblen Produktionsplattformen für neuartige Wirkstoffe (z. B. auf Basis der mRNA-Technik) bei der ein kürzeres Time-to-Market eine wichtige Rolle spielt.

Abstract

In this paper an integrated process model for the design and realization of modular and computerized systems and architectures in the pharmaceutical production is introduced. Pre-configured and pre-fabricated architecture modules (e.g. for the construction of cleanrooms) allows also the application and integration of flexible and modular automation and IT components. The paper shows a model which consider the MTP Namur standard and GMP requirements.

Motivation

Die pharmazeutische Industrie zeigt einen zunehmenden Trend zur Produktindividualisierung und damit zu einer immer breiteren Produktpalette, was zu kürzeren Produktlebenszyklen sowie zu kürzeren Produktionszeiten führt [1]. Dieser Trend zeigt sich vor allem in der weltweit immer weiter steigenden Nachfrage an Impfstoffen gegen Viruserkrankungen [2] und aktuell in der Corona-Pandemie-Situation. Hieraus ergibt sich die Notwendigkeit in schneller Zeit neue flexible und hochgradig automatisierte Wirkstoffanlagen zu planen und zu realisieren.

In Anlagen der Pharmaindustrie kommen häufig Anlagen (Skids) unterschiedlicher Hersteller zum Einsatz. Bei der automationstechnischen Integration der verfahrenstechnischen Systeme in pharmazeutischen oder biotechnologischen Produktion (z.B. Abfüllsysteme, Reinstmedien, Bioreaktoren, Chromatographiesysteme, Verpackungsanlagen etc.) ergibt sich oft das Problem der fehlenden Kompatibilität der Schnittstellen. Das führt in vielen Fällen dazu, dass die mechanische Einbindung von Systemen deutlich schneller erfolgt als die automatisierungstechnische Einbindung in ein Prozessleitsystem.

Die Notwendigkeit der modularen Automation

Die Integration der Automatisierung verschiedener Systeme von verschiedenen Anbieter wird überwiegend durch Remote-IO-Systeme oder teilweise verteilte Steuerungssysteme konventionell realisiert. Diese stellen nur einfache Funktionen über proprietäre Schnittstellen zur Verfügung, während die meisten Systemfunktionen und Aufzeichnungen von Produktionsparameter in zentralen Prozessleitsystemen (PLS) realisiert werden. Jedes System verfügt unter anderem über eine herstellerspezifische Logik, Programmierung und Bedienung. Neben Übertragungsprotokollen müssen auch die von den Anlagen erzeugten Alarmer und Meldungen sachgemäß ins Leitsystem integriert werden. Dasselbe gilt für die Visualisierung sämtlicher Steuerparameter. Solche konventionellen Konzepte im Anlagenbau und in der Planung der Automatisierung erweisen sich als unzureichend für eine rasche, flexible und anpassbare automatisierte Herstellung zunehmend individualisierter Produkte, wie beispielsweise neuartige Diagnose- und Therapieformen (Gentherapien, personalisierte Medizin) [3].

Um auf den globalen Märkten wettbewerbsfähig zu bleiben, müssen im Pharma-Anlagenbau die Hauptansätze von Industrie-Zukunftsinitiativen (wie Industrie 4.0, Pharma 4.0, Plug & Produce), i. d. R. Flexibilität und Modularisierung [4], bereits bei der Planung berücksichtigt werden. Bisher monolithisch betrachtete Prozessanlagen müssen in diesem Zusammenhang als Module gedacht werden, die wiederum mittels modularer Automation zusammengesetzt werden können. Voraussetzung für diesen modulatorientierten Ansatz sind somit standardisierte, herstellerübergreifende Schnittstellen. Sie gewährleisten die effiziente, flexible Konfiguration,

Kommunikation und Integration von modularen Anlagen für die Wirkstoffproduktion. Dabei unterscheidet sich grundsätzlich die Planung pharmazeutischer Anlagen in vieler Hinsicht vom Engineering anderer verfahrenstechnischer Anlagen (wie z. B. einer Chemieanlage). Zu berücksichtigende Aspekte sind die von Anfang an durchdachten Konzepte für den Material- und Personalfluss, qualifizierungsfähige Apparate/ Komponenten und Reinigungsprozesse, Computer System Validierung, Cyber Security und generell die gesetzlichen Anforderungen an die GxP-regulierte Industrie, bei der die Qualitätssicherung aufgrund der Produkt- und Patientensicherheit die Rahmenbedingungen für das Engineering setzen.

Anforderungen an modulare Plug & Produce-Systeme im regulierten Umfeld

Bei der Planung und Realisierung von modularen Plug & Produce-Systemen im regulierten GxP-Umfeld ergeben sich folgende regulatorische Betrachtungen:

- Ingenieurstechnische Normen
- Neue Normen für modularisierte und automatisierte Plug & Produce-Systeme
- Normen im GxP-regulierten Umfeld (Good Manufacturing Practice, GMP)

Generelle, ingenieurtechnische Anforderungen, die im Anlagenbau im Allgemeinen zu berücksichtigen sind z.B.: Maschinen-Richtlinie, Druckgeräte-Richtlinie, ATEX-Produktrichtlinie, Niederspannungsrichtlinie, Elektromagnetische Verträglichkeit-Richtlinie, ATEX-Betriebsrichtlinie oder die REACH-Verordnung.

Im Bereich der modularisierten und automatisierten Plug & Produce-Systeme beschreibt die Richtlinie VDI 2776 [5] die Anwendung eines modularen Anlagenkonzepts in der Prozessentwicklung sowie in der Engineering- und Betriebsphase entlang des Anlagenlebenszyklus. Sie definiert dabei aber keine Standardausrüstung oder Instrumente, die im modularen Anlagenkonzept verwendet werden sollen.

Die VDI-Richtlinie 2658 [6] spezifiziert das automatisierungstechnische Engineering modularer Anlagen in der Prozessindustrie. Die Zielgruppen dieser Richtlinie besteht dabei aus dem Modulhersteller, dem Werkzeughersteller und dem Modulintegrator.

Die Richtlinien VDI 2776 und VDI 2658 betrachten jedoch weder GMP-Aspekte noch den Einsatz von architektonischen, vorgefertigten Elementen.

Spezielle GMP-Anforderungen, die den Anlagenbau für pharmazeutische Produktion im europäischen Raum betreffen, werden vor allem (aber nicht ausschließlich) in den Kapiteln des EU-GMP-Leitfadens Teil I [7] formuliert. Diese Kapitel legen Anforderungen in der pharmazeu-

tischen Anlagenplanung zu den Themen Materialien, Hygienische Konstruktion, Anlagensteuerung, Technische Dokumentation, Kalibrierung, Instandhaltung und Reinigung von Anlagen fest:

- Kapitel 1: Qualitätsmanagement
- Kapitel 3: Räumlichkeiten und Ausrüstung
- Kapitel 4: Dokumentation
- Kapitel 5: Produktion
- Anhang 11: Computergestützte Systeme
- Anhang 15: Qualifizierung und Validierung

Ähnliche GMP-Anforderungen sind zudem in den amerikanischen FDA 21 CFR Part 210 [8], 211 [9] bzw. Part 11 [10] Richtlinien formuliert auf die hier nicht näher eingegangen wird.

Vorgefertigte, architektonische Elemente

Für den Aufbau einer raschen, flexiblen und modularen GMP-Herstellungsanlage können standardisierte Designs von Produktionsumgebungen im Form von Modulen verwendet werden. Realisiert werden diese mittels vorgefertigten architektonischen Elementen und vorgefertigten Systemen für Reinraumproduktion und Produktionseinrichtungen wie in der nächsten Bild 1 dargestellt in einem Deckenkonzept [11, 12].



Bild 1: Vorgefertigtes architektonisches Element [ExyCell®]

Bei dem entwickelten Konzept handelt es sich um ein spezielles Deckenelement (ExyCell®), welches als Backbone eingesetzt werden kann. Es enthält die gängigen Versorgungsleitungen, die in einer pharmazeutischen Produktionsanlage benötigt werden, wie z.B.:

- (HVAC) Zu- und Abluft, Filter Fan Units
- Schwarzmedien Verteilung
- Clean Utility Verteilung
- Beleuchtungselemente
- Rauchabzüge
- Raumüberwachungselemente
- Verteilerkästen/ Remote I/Os für die Automatisierung

Module Engineering mittels MTP

Damit sich die oben beschriebenen ExyCell®-Module einfach und schnell kombinieren lassen, braucht es aber auch standardisierte Schnittstellen auch für die Automatisierung der Prozessanlagen und Anbindung an die Prozessleittechnik innerhalb dieser architektonischen Module. Das „Module Type Package“ übernimmt diese Aufgabe für die Automationsanbindung – ähnlich dem Druckertreiberprinzip – für die Computerisierung und Digitalisierung. Dabei handelt es sich um eine Organisationsdatei, die Funktionalitäten sowie Visualisierungseigenschaften der Benutzerinterfaces beschreibt. Diese standardisierte Schnittstelle ermöglicht in der Automatisierungstechnik eine anlagenweit einheitliche Systemstruktur, so dass im modularen Kontext eine firmenunabhängige Implementierung in der Pharma- und Biotechindustrie erfolgt. Für die Beschreibung einer modularen Anlage, wird der Begriff Modular Plant (MP) verwendet. Eine MP besteht aus einer oder mehrere zusammengesetzte Process Equipment Assembly (PEA), die als modulare Prozesseinheit festgelegt wird. Folglich besteht eine PEA aus einer oder mehrere Functional Equipment Assembly (FEA), das eine modulare Funktionseinheit beschreibt, die eine verfahrenstechnische Funktion ausführt. Wiederum besteht eine FEA aus einer Zusammensetzung von Components (COMP), die aus verschiedenen Komponenten bestehen kann. Das Engineering mit MTP erfolgt in zwei Schritten. Diese bestehen aus dem Modul- und Anlagenengineering. [6]

In dem Modulengineering werden die Funktionalitäten des Moduls bestimmt. Es wird die Steuerungslogik entworfen und die Bedienbilder des Prozesses generiert. Es existiert eine Bibliothek, die alle Funktionen des Moduls und Beschreibungen der Bedienbilder enthält. Somit kann der Modulbauer entscheiden, welche Dienste dem Nutzer zur Verfügung stehen. Die Beschreibungsdatei MTP wird nach der Konzeption des Moduls automatisch generiert.

Das Module Type Package (MTP) bietet eine Beschreibung aller Dienste, eine abstrakte Layout-Spezifikation der PEA für das Human-Machine Interface (HMI) sowie die mit der PEA ausgetauschten Anzeige- und Steuerungsvariablen und den Kommunikationskanal, den die PEA verwendet. Generiert wird das MTP vom PEA-Hersteller als eine Containerdatei, die intern Automation ML zur Beschreibung nutzt. AutomationML (Automation Markup Language) ist ein offener Standard der ein neutrales, XML-basiertes Datenformat für die Speicherung und zum Austausch von Anlagenplanungsdaten bietet. AutomationML hat das Ziel den Austausch von Engineering-Daten in einer heterogenen Tool-Landschaft von modernen Engineering-Tools für verschiedene Planungsdisziplinen wie zum Beispiel Mechanisches Design, Elektrisches Design, SPS-Programmierung, HMI-Entwicklung, oder Robotersteuerung zu ermöglichen. [6] Die modulinterne Automatisierung, z.B. Steuerungen und interne Sicherheitslogiken, ist hinter den Diensten vollständig gekapselt und somit außerhalb der PEA nicht bekannt. Dieser Ansatz ermöglicht es einer anlagenweiten übergeordneten Prozessorchestrierungsschicht (mittels MTP), die von der PEA veröffentlichten Dienste zu integrieren und aufzurufen, um ein Produktrezept zu erfüllen.

Beim Engineering muss der PEA-Hersteller die Prozessfähigkeiten der PEA spezifizieren, analysieren und bewerten und dann die Automatisierung einschließlich der angebotenen Dienstleistungen planen. Der PEA-Lieferant hat nur sehr wenige oder gar keine Informationen über den zukünftigen verfahrenstechnischen Prozess oder die verwendeten Substanzen, mit denen das Modul verwendet wird, dennoch müssen die PEA und ihre Automatisierung so ausgelegt sein, dass sie im spezifizierten Betriebsbereich sicher arbeiten und kritische Prozessbedingungen verhindern [13].

GMP-Aspekte bei der Planung

In der Planung der Plug & Produce-Einheiten im Pharmaumfeld müssen neben den beschriebenen Engineering Aspekten auch die gesetzlichen Vorgaben für die Herstellung von Wirkstoffen berücksichtigt werden. In der pharmazeutischen Produktion steht vor allem die GMP-Konformität im Vordergrund. Diese sichert die Qualität des Produkts, den beabsichtigten Zweck und somit die einhergehende Produkt- und Patientensicherheit und Datenintegrität. In Bezug auf GMP sind folgende Aspekte im Module Engineering von besonderer Wichtigkeit (vgl. Bild 2):

- Materialien (Beanspruchbarkeit, Fertigungsanforderungen, Wirtschaftlichkeit)
- Hygienische Konstruktion (reinigungsgerechte Gestaltung von Bauteilen, Komponenten)

- Modulsteuerung (Messung qualitätsrelevanter Prozessparameter)
- Qualifizierung/Validierung (Tests zur Eignung des vorgesehenen Zwecks)
- Technische Dokumentation (Nachvollziehbarkeit, Nachweisbarkeit, Unterstützung des Qualitätsmanagements, Unterstützung des Dokumentenmanagements, Unterstützung des Wissensmanagements, Wartung und Service)
- Kalibrierung (Sicherheit, Vergleichbarkeit, Behördliche Anforderung)
- Instandhaltung (Erhalt des funktionsfähigen Zustands)
- Reinigung (Sicherung Verfügbarkeit, Sicherung Produktqualität)

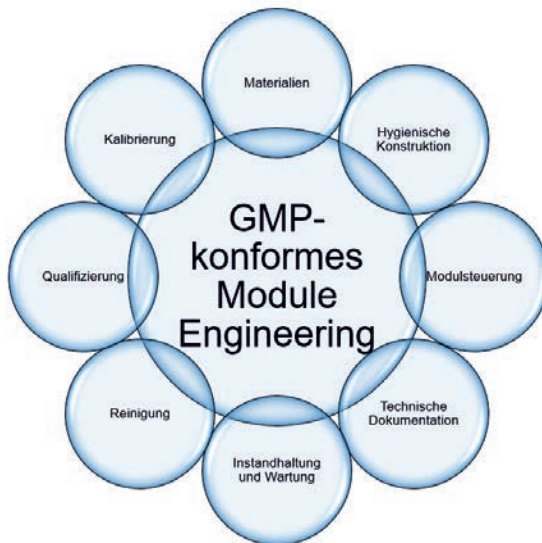


Bild 2: GMP-konformes Module Engineering [Exyte Darstellung]

Vorgehensmodell für Plug & Produce-Systeme im GxP-regulierten Umfeld

Bild 3 zeigt ein Modell auf, das prozessorientiert die Integration von Anforderungen aus dem modularen Anlagenbau in den Planungsprozess von Anlagen und automatisierten Systemen in der Pharmaproduktion einbezieht. Dabei spielt die Modulautomatisierung eine entscheidende Rolle innerhalb des Engineering- und Implementierungsprozesses.

Das Modell setzt als Grundlage die Erfüllung von Anforderungen an den pharmazeutischen Anlagenbau voraus. Auch die Dokumentation (Validierungsmasterplan, Spezifikationen, Factory Acceptance Tests – FATs, Site Acceptance Tests – SAT, Qualifizierungsdokumentation,

etc.) sollte bereits im großen Maße standardisiert vorliegen. In Form von Durchführbarkeitsstudien (Feasibility Study) werden die grundlegenden Eigenschaften der modularen Pharmaanlage in Form von Anforderungsspezifikationen, Lastenhefte bzw. User Requirements (URS) erarbeitet und risikobasiert bewertet.

Des Weiteren sind im Modell standardisierte Prozessmodule (Modulvarianten) bereits vorkonstruiert, vor-automatisiert und vorgefertigt und können mit standardisierten Schnittstellen (Automatisierung durch MTP) und dezentraler Intelligenz durch den Modullieferanten kurzfristig geliefert bzw. angepasst werden.

Neu integrierte Elemente im klassischen Engineering-Workflow sind dabei:

- Prozessentwicklung (Process Design)
- Konfiguration (Configure)
- Zusammenstellung (Plug)
- Orchestrierung (Produce)
- Umstrukturierung (Rearrange)

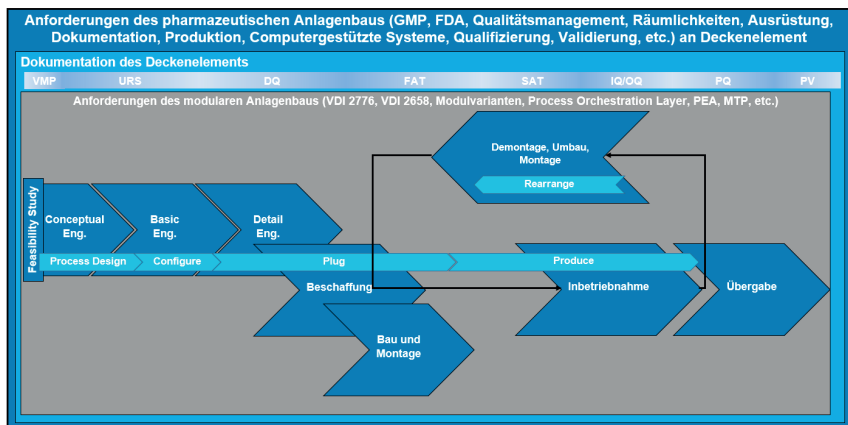


Bild 3: Vorgehensmodell für Plug & Produce-Systeme im GxP---regulierten Umfeld
[Exyte Darstellung]

Die Phase der Prozessentwicklung (Process Design) wird in die Konzept- und der Entwurfsphase (Conceptual Engineering, Basic Engineering) integriert. Sie enthält die Entwicklung von einzelnen Betriebsschritten (z.B. Reaktion, Lösen, Rühren, Reinigen), die vorgefertigte stan-

dardisierte Prozessmodule verwenden (Module Engineering). Die Ausarbeitung eines Backbone-Konzepts entfällt durch die standardisierten Anschlüsse, die durch die vorgefertigten, architektonischen Elemente (Deckenmodul) gegeben sind.

Die Prozessmodule müssen im nächsten Schritt durch den Austausch von entsprechenden Teilmodulen auf die jeweiligen Prozessbedingungen konfiguriert werden (Configure). Dieser Schritt lässt sich in die Entwurfsphase (Basic Engineering) einordnen. So lassen sich einzelne Arbeitsschritte durch standardisierte Schnittstellen schnell und einfach zu einem Gesamtprozess zusammenstellen (Plug). Die Phase des Zusammenstellens erstreckt sich dabei von der Ausführungsphase (Detail Engineering) über die Beschaffungsphase bis hin zum Bau und der Montage.

Aufgrund der dezentralen Intelligenz der Module müssen die Module im Sinne der Automatisierung lediglich von einer prozessführenden Einheit orchestriert werden (Produce). Diese Orchestrierung wird in die Phase der Inbetriebnahme integriert.

Die oben beschriebene Vorgehensweise und die Schnittstellen ermöglichen zudem eine flexible, schnelle und effiziente Anpassung des Gesamtprozesses (Rearrange) in einer späteren neu definierten Engineering-Phase, die als „Demontage, Umbau und Montage“ bezeichnet wird. Neben den üblichen Mitarbeiterschulungen müssen für diese Phase auch Schulungen zum Wechsel von Modulen ausgearbeitet werden, die bei der Konfiguration der Anlage in Kraft treten. Ein Konzept für den Modultauch im Sinne von Plug & Produce wird in [14] geliefert. Es besteht im ersten Schritt aus einem Sicherungsprozess, einer Außerbetriebnahme und anschließendem Umbau. Im nächsten Schritt wird das Modul integriert, von der Prozessführungsebene erkannt und mit der Geräteverwaltung verbunden, sodass die Eignung des neu anzuschließenden Moduls überprüft werden kann. Ist dieses geeignet, so werden die neuen Konfigurationssätze zusammengestellt und von einem Operator bestätigt. Die neuen Parameter werden in das Modul übertragen und verifiziert. Im letzten Schritt folgt eine Wiederinbetriebnahme des neuen Moduls.

Fazit

Das integrierte Vorgehensmodell zur Planung und Umsetzung von modularen und computerisierten Produktionssystemen- und Architekturen im regulierten Life Science Umfeld dient als erster Anhaltspunkt, um das Design und Engineering im modularen Anlagenbau mit den neuen Konzepten der modularen Automatisierung zu verknüpfen. Weitere Themen wie die Anbindung der modularisierten Systeme oder der Betrieb als Cloud-Anwendung und die Planung der Sicherheit der Datenverarbeitung vernetzten modularen Komponenten (Cyber Security by

Design) sollen dann das Modell sukzessive erweitern und in übergeordneten Planungsprozess einlaufen.

Literaturverzeichnis

- [1] C. Bramsiepe und G. Schembecker, „Die 50 %-Idee: Modularisierung im Planungsprozess,“ *Chemie Ingenieur Technik*, Bd. 84, Nr. 5, pp. 581-587, 2012.
- [2] S. Hofmann, Milliardenmarkt Impfstoffe, 19 08 2020. [Online]. Available: <https://www.handelsblatt.com/unternehmen/industrie/milliardenmarkt-impfstoffe-enorme-nachfrage-allein-die-erste-impfwelle-verspricht-20-milliarden-dollar-umsatz/26108338.html?ticket=ST-6675633-ajvtSgo3NouTjJnpnJzL-ap5>. [Zugriff am 17.01 2021].
- [3] G. Moelgaard und A. Geipel-Kern, Personalisierte Medizin - Konsequenzen für die Produktion, 22 04 2014. [Online]. Available: <https://www.process.vogel.de/personalisierte-medizin-konsequenzen-fuer-die-produktion-a-443484/>. [Zugriff am 30.03.2021].
- [4] Modulbasierte Produktion in der Prozessindustrie - Auswirkungen auf die Automation im Umfeld von Industrie 4.0, ZVEI, 2015. [Online]. Available: https://www.zvei.org/fileadmin/user_upload/Presse_und_Medien/Publikationen/2015/maerz/White_Paper__Module-Based_Production_in_the_Process_Industry_-_Effects_on_Automation_in_the__Industrie_4.0__Environment/ZVEI-White-Paper-Modulare-Automation.pdf. [Zugriff am 24.03.2021].
- [5] VDI/ VDE 2776: *Verfahrenstechnische Anlagen; Modulare Anlagen*, 2020.
- [6] VDI/VDE/Namur-Richtlinie 2658: *Automatisierungstechnisches Enigeering modularer Anlagen in der Prozessindustrie*, 2020.
- [7] *Anlage 2 zur Bekanntmachung des Bundesministeriums für Gesundheit zu § 2 Nr. 3 der Arzneimittel- und Wirkstoffherstellungsverordnung vom 27. Oktober 2006 - Leitfaden der Guten Herstellungspraxis Teil I.*
- [8] *21 CFR Part 210 - CURRENT GOOD MANUFACTURING PRACTICE IN MANUFACTURING*,, 1996.

- [9] *21 CFR Part 211 CURRENT GOOD MANUFACTURING PRACTICE FOR FINISHED PHARMACEUTICALS*, 1996.
- [10] *21 CFR Part 11 Electronic Records, Electronic Signatures*, 2003.
- [11] CHEManager - Siemens und Exyte wollen Aufbau intelligenter Biotech-Anlagen beschleunigen, 09.06.2020. [Online]. Available: <https://www.chemanager-online.com/news-opinions/nachrichten/siemens-und-exyte-wollen-aufbau-intelligenter-biotech-anlagen-beschleunige>. [Zugriff am 02.12.2020].
- [12] Siemens Pharma - Siemens und Exyte schließen Partnerschaft für integrierte Lösungen bei raschem Aufbau intelligenter Biotech-Anlagen, 09.06.2020. [Online]. Available: <https://press.siemens.com/global/de/pressemitteilung/siemens-und-exyte-schliessen-partnerschaft-fuer-integrierte-loesungen-bei-raschem>. [Zugriff am 02.12.2020].
- [13] C. Fleischer-Trebes, J. Wittmann, N. Kockmann, T. Bieringer und C. Bramsiepe, „Sicherheitstechnische Aspekte bei Planung und Bau modularer Produktionsanlagen,“ *Chemie Ingenieur Technik*, Bd. 87, Nr. 9, pp. 1258-1269, 2015.
- [14] L. Nothdurft und U. Epple, „Plug&Produce für den Feldgerätetausch,“ in *AUTOMATION*, Baden-Baden, 2019.

Modellgetriebene Entwicklung für service-orientierte Automatisierungssysteme in der Prozessindustrie

B.Sc. **M. Waßmer**,

Siemens AG, Karlsruhe, Karlsruher Institut für Technologie, Karlsruhe;

Dr.-Ing. **M. Maurmaier**, M.Sc. **A. Stutz**,

Siemens AG, Karlsruhe;

M.Sc. **V. Pazmino Betancourt**, M.Sc. **B. Liu**,

Forschungszentrum Informatik, Karlsruhe;

Prof. Dr.-Ing. Dr. h. c. **J. Becker**,

Karlsruher Institut für Technologie, Karlsruhe,

Forschungszentrum Informatik, Karlsruhe

Kurzfassung

Modulare Prozesseinheiten sollen flexibel für unterschiedliche Anlagen und für verschiedene Anwendungsszenarien ausgelegt werden. Dazu ist es nötig, eine universell einsetzbare Softwarelösung zu entwickeln, was einen erheblichen Anstieg der Komplexität und des Umfangs bedeutet. Modellgetriebene Entwicklung kann dazu beitragen, den neuen, hohen Anforderungen gerecht zu werden. Dabei wird die Struktur und das Verhalten einer modularen Prozesseinheit durch passende Modellierungssprachen auf einer einfachen, abstrakten Ebene beschrieben und aus dieser modellbasierten Beschreibung SCL-Code generiert. Dies erlaubt eine einfachere und schnellere Entwicklung von Automatisierungssoftware für modulare Prozesseinheiten.

Abstract

Modular Process Units shall be designed to be flexible for different plants and various use cases. For that it is necessary to develop software solutions that can be used universally, which leads to enormous complexity and scope. Model-driven development can help to meet those new, high demands. Thereby, the structure and the behavior of a Modular Process Unit is described by a suitable modelling language on an easy, abstract layer and SCL-code is generated from this model-based description. This allows an easier and quicker development of automation software for Modular Process Units.

1. Motivation

Die Anforderung auch in der Prozessindustrie die Time-to-Market stark zu verkürzen gibt es bereits einige Jahre [1], Durch modulare Anlagen kann diese Anforderung erfüllt werden. Modulare Anlagen werden aus vordefinierten Anlagenteilen, Process Equipment Assemblies (PEAs), zusammengesetzt, was einen schnelleren Engineeringprozess, eine Skalierung der Anlage sowie flexible Änderungen ermöglicht. Um die Investitionskosten für modulare Anlagen zu reduzieren, sollen PEAs wiederverwendbar sein und in unterschiedlichen Anlagenkonfigurationen zum Einsatz kommen. [2],[3] Dadurch kann die benötigte Automatisierungssoftware, wie bei monolithischen Anlagen üblich, nicht mehr auf nur einen Anwendungsfall ausgerichtet werden. Vielmehr sind universell einsetzbare Softwarelösungen gefragt. Dies führt zu einer Steigerung der Komplexität und des Umfangs der Automatisierungssoftware.

In diesem Beitrag wird dargelegt, wie mithilfe von modellgetriebener Entwicklung eine effiziente und einfache Entwicklung der Automatisierungssoftware erreicht werden kann. Dabei wird der Entwicklungsprozess von der Analyse benötigter Services auf Basis des Rohrleitungs- und Instrumentenfließschema (R&I) bis hin zum fertigen Quellecode in SCL gemäß IEC 61131 betrachtet.

2. Heutiger Entwicklungsprozess für modulare Prozesseinheiten

Für den Automatisierungstechniker ist das R&I Ausgangspunkt für die Entwicklung von Steuerungssoftware für PEAs. Im R&I ist enthalten, welche Aktoren und Sensoren in einer PEA verbaut sind und welche Prozessleittechnikaufgabe sie erfüllen sollen.

Zunächst analysiert der Entwickler anhand des R&Is und ggf. weiterer Anforderungen, welche Services die PEA bereitstellen soll. Zur Identifizierung von Services und Prozeduren können Methoden wie in [4] beschrieben verwendet werden. Üblicherweise wird die Spezifikation von Services und ihren Prozeduren in nicht formalisierten Dokumenten oder Tabellen dokumentiert. Wichtige Bestandteile dieser Spezifikation sind die Bedingungen für Zustandsübergänge und das Verhalten der Prozeduren der Services in den einzelnen Zuständen. Die Bedingungen und das Verhalten werden dabei in Prosa festgehalten. Auch Regelungen werden meist mit Prosa beschrieben. Verriegelungen für Aktoren werden oft in Prosa oder in einer Ursache-Wirkungs-Matrix (Cause&Effect-Matrix) festgehalten.

Nach der Spezifikation der PEA erfolgt die eigentliche Softwareentwicklung unter Beachtung der VDI/VDE/NAMUR 2658. Zur Unterstützung der Entwicklung der Automatisierungssoftware können Baustein-Bibliotheken und Entwurfsmuster [5] verwendet werden. Zusammengefasst besteht der Entwicklungsprozess aus mehreren kreativen, manuellen Schritten, in welchen die

Ergebnisse des vorausgegangenen Entwicklungsschritts als Dokument in den nächsten Schritt eingehen.

3. Grundlagen: Modellgetriebene Entwicklung von Automatisierungssoftware

In der modellgetriebenen Entwicklung geht es darum, mithilfe von Modellen ein System zu beschreiben und einen daraus ausführbaren Code zu generieren. Ein Modell beschreibt immer nur einen begrenzten Aspekt eines Systems, weshalb häufig domänenspezifische Sprachen (eng. domain specific languages, kurz DSLs) eingesetzt werden. DSLs werden spezifisch für ihre Domäne formuliert. Dabei kann eine DSL auf Basis einer bestehenden Modellierungssprache entwickelt werden, indem diese vereinfacht oder erweitert wird.[6] Das Modell, das zur Beschreibung der DSL bzw. der aufgegriffenen Modellierungssprache dient, wird Metamodell genannt. Das Modell, das mithilfe der DSL bzw. der Modellierungssprachen erstellt wird, wird Benutzermodell genannt. Dieses Modell bildet einen bestimmten Aspekt der Realität ab, in diesem Fall der Automatisierungssoftware einer PEA.

Modellgetriebene Entwicklung ermöglicht durch eine einfache und abstrakte Beschreibung eine Reduzierung der Komplexität. Aus diesen Beschreibungen wird der Code der Software für die Zieldomäne automatisiert generiert. Durch die Codegenerierung können Redundanzen und unkreative Entwicklungsanteile vermieden werden, was zu einer kürzeren Entwicklungszeit führt. Zudem kann die Entwicklung von Software mit hoher Qualität durch das Aufgreifen von Best Practices und Know-how von erfahrenen Entwicklern erreicht werden. [6][7]

Für die erfolgreiche Verwendung von modellgetriebener Entwicklung in der Automatisierungstechnik gibt es zahlreiche Beispiele [8 bis 10]. Auch der Gedanke, aus dem Rohrleitungs- und Instrumentierungsfließschema (R&I) Informationen zu entnehmen und daraus Code zu generieren, ist nicht neu. So beschreibt zum Beispiel [11] ein Vorgehen für die Entwicklung von Steuerungscode nach IEC 62424 und IEC 61499 mit modellgetriebener Entwicklung mit der Verwertung von Informationen aus dem R&I. Allerdings gibt es nach heutigem Stand noch kein Konzept, mithilfe dessen aus Modellen Steuerungscode für PEAs gemäß der Richtlinie VDI/VDE/NAMUR 2658 generiert werden kann.

4. Anwendungsbeispiel Rührreaktor

Das Konzept wird anhand eines Rührreaktors veranschaulicht. Im Bild 1 dargestellten R&I Schaubild der Rührreaktors sind die Sensoren und Aktoren sowie deren grundsätzliche Wirkzusammenhänge zu sehen. Ein Rührreaktor mit dem dargestellten R&I umfasst die Services Befüllen, Entleeren, Zirkulieren, Mischen, Rühren und pH-Wert einstellen. Ziel ist es, die

Automatisierungssoftware für diese Services formalisiert mittels Modelle zu beschreiben und daraus den SCL Code zu generieren.

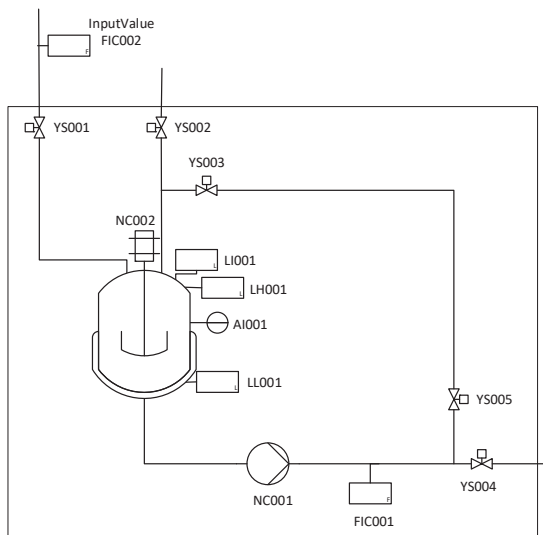


Bild 1: R&I des Beispiel-Use Cases

4. Zieldomänenanalyse

Ziel des in diesem Beitrag vorgestellten Entwicklungsprozesses ist es, die Automatisierungssoftware einer PEA aus abstrakten Modellen zu generieren. In diesem Abschnitt werden die Bestandteile der zu generierenden Software vorgestellt. In der Automatisierungssoftware einer PEA sind die Akteure, Sensoren und Services als Control Modules bzw. Service Modules abgebildet. Treiberbausteine, Control Module Bausteine und Service Module Bausteine werden durch Bibliotheken bereitgestellt. Die benötigten Control Module Bausteine müssen instanziiert und mit ihren zugehörigen Treiberbausteinen verschaltet werden. Benötigte Verriegelungen für Akteure werden durch eine entsprechende Verschaltung der Eingänge der Bausteine implementiert.

Die Service Modules Bibliothek bietet Bausteine an, die bereits die Struktur von Service und Parametern abbilden und nur noch beschalten werden müssen. In einem Bibliotheksbaustein ist der Zustandsautomat implementiert, wobei Zustandsübergänge mit logischen Ausdrücken zu beschalten sind und in den Zuständen das Verhalten des Service zu implementieren ist. Für Services werden in manchen Fällen weitere Bausteine wie Integratoren oder Timer benötigt. Lange und komplizierte Berechnungen werden häufig in Funktionen ausgelagert.

5. Der Modellgetriebene Entwicklungsprozess

Der in diesem Beitrag vorgestellte modellgetriebene Entwicklungsprozess nutzt das R&I als Eingangsartefakt. Zielartefakt ist der Quellcode der Automatisierungssoftware der PEA in SCL. Dabei werden sowohl die Struktur als auch das Verhalten modelliert und der entsprechende Quellcode generiert. Der Entwicklungsprozess lässt sich in drei grobe Schritte unterscheiden. Im ersten Schritt wird die PEA als Ganzes beschrieben. Dies umfasst die Aktoren und Sensoren, deren Verriegelung sowie die äußere Betrachtung der Services. Im zweiten Schritt werden die Services jeweils von innen betrachtet und deren Parameter, Prozeduren und die enthaltenen Regelkreise beschrieben. Zuletzt wird das Verhalten in den Prozeduren der Services beschrieben. Die wesentlichen Schritte des Engineering-Prozess sind in Bild 2 dargestellt.

Für Sensoren und Aktoren bedarf es einer Beschreibung, die in anderen Modellen aufgegriffen werden kann. Dazu werden die sogenannten *Elementmodelle* eingeführt. Instanzen dieser Elementmodelle werden im Benutzermodell verwendet, um Ansteuerungen für die Aktoren anzugeben oder Aktor- und Sensorwerte zu lesen zu schreiben.

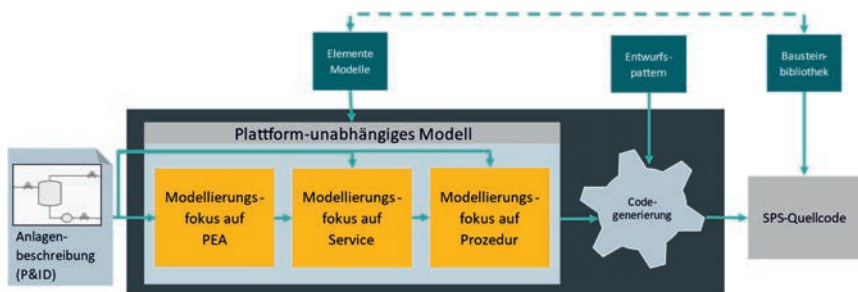


Bild 2: Modellgetriebene Entwicklung der Automatisierungssoftware für PEAs

5.1. Elemente Modelle

Die VDI/VDE/NAMUR 2658 definiert in Blatt 3 verschiedene Aktor- und Sensortypen. Diese Aktor- und Sensortypen werden durch jeweils ein Elementmodell abgebildet. Für jeden Aktor oder Sensor der PEA wird im Benutzermodell eine Instanz des entsprechenden Elementmodells instanziiert. Um klar identifizieren zu können, um welchen Aktor oder Sensor es sich handelt, erfolgt die Bezeichnung nach der Prozessleittechnikaufgabe aus dem R&I.

Bild 3 zeigt die resultierenden Modellierungsebenen. Zur Modellierung des Elementmodells wird eine DSL basierend auf den Konzepten des Klassendiagramms aus der Unified Modelling Language (UML) vorgeschlagen. Die Elementmodelle können als Klassen angesehen werden, die beschreiben, wie eine Instanz aufgebaut sein muss.

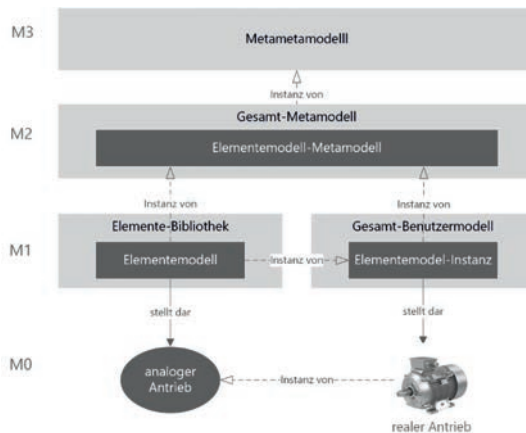


Bild 3: Modellierungsebene

Wie eine Klasse in UML hat auch ein Element des Elementmodells einen Namen, Attribute und Methoden. Der Name des Elementmodells entspricht dem der Aktor- bzw. Sensortypen nach VDI/VDE/NAMUR 2658. Im Elementmodell wird nun allerdings zwischen zwei Attributstypen und zwei Methodentypen unterschieden. Der erste Attributstyp wird "Attribut" genannt. Er beschreibt einen Zustand eines Sensors oder eines Aktors, der ausgelesen werden kann. Jeder Sensor besitzt zudem im Modell einen Messwert. Parameter, die eingestellt werden können, wie zum Beispiel die Drehzahl eines analogen Antriebs, werden durch den zweiten Typ an Attributen, "Signale" genannt, abgebildet.

Die zwei Methodentypen unterscheiden sich in ihrem Gebrauch. Methoden, die nur "Aktionen" genannt werden, dienen zur Beschreibung einer Ansteuerung im normalen Betrieb. Die sogenannten "Verriegelungsaktionen" bilden die Verriegelungsfunktionen Permit, Interlock und Protect nach VDI/VDE/NAMUR 2658 ab.

Da sich die Elemente Modelle auf die VDI/VDE/NAMUR 2658 beziehen, können die Elementmodelle als Bibliothek ausgeliefert werden. Der Endnutzer muss sich nicht um die Erstellung der Elementmodelle kümmern. Diese verwendet er lediglich zu Instanziierung der Sensor- und Aktormodelle im Benutzermodell.

Für die Programmierung von Services werden oftmals Timer, Integratoren oder Regler benötigt. Um diese im Benutzermodell nutzen zu können, werden auch für sie Elemente Modelle benötigt. Sie werden durch das gleiche Metamodell wie zur Modellierung der Sensoren oder

Aktoren beschrieben. Der einzige Unterschied ist, dass sie keine Verriegelungsaktionen besitzen.

Auch für Services gibt es ein Elementemodell. Anhand von Attributen eines Service kann abgefragt werden, in welchem Zustand er sich befindet und welche Prozedur gerade aktiv ist. Jeder steuerbare Zustandsübergang aus dem Zustandsautomat der VDI/VDE/NAMUR 2658 wird durch eine Aktion abgebildet. Für jeden Service, der im Benutzermodell angelegt wird, muss eine Service-Element-Instanz erstellt werden.

5.2. Darstellung des Entwicklungsprozesses

Im ersten Schritt des modellgetriebenen Entwicklungsprozesses werden Modelle erstellt, die die PEA als Ganzes betrachten. Zuerst muss für jeden Aktor und jeden Sensor, der im R&I ausgeführt ist und nicht rein manuell betrieben wird, eine Elementinstanz des jeweiligen Sensor- oder Aktortyps erzeugt werden.

Danach wird modelliert, welche Services die PEA besitzen soll, und welche Aktoren und Sensoren für den jeweiligen Service benötigt werden. Die PEA nach Bild 1 verfügt beispielsweise über den Service Entleeren, der für die Entnahme von Flüssigkeit aus dem Tank zu verwenden ist. Zur Durchführung des Entleeren-Service wird die Pumpe NC001 betätigt und das Ventil YS005 muss geöffnet werden. Der Messwert Durchflusssensor FIC001 wird zur Regelung des Durchflusses und zur Berechnung der entnommenen Menge verwendet. Die Werte des Füllstandsensors LI001 und des unteren Schutzschalters des Tanks LH001 werden benötigt, um den Service anzuhalten oder zu beenden, sobald der Tank leer ist.

Die Zuordnung der Sensoren und Aktoren zu den einzelnen Services erfolgt in einer formalisierten Tabelle. In den Zeilen werden alle Services aufgelistet und in den Spalten die Aktoren und Sensoren. Durch Setzen eines Kreuzes in der entsprechenden Zelle wird eine Relation zwischen Service und Aktor bzw. Sensor im Modell erstellt.

Im nächsten Schritt folgt die Modellierung der Verriegelungen für Aktoren. Beispielsweise soll verhindert werden, dass die Pumpe NC001 gegen ein geschlossenes Ventil (YS004 und YS005) pumpt. Solche Verriegelungen sind unabhängig vom Service und dienen der Sicherheit der Anlage. Sie werden als Ursache-Wirkungs-Matrix dargestellt. Zur Definition der Ursachen kann die logische Verknüpfung mehrerer Zustände oder Wertvergleiche nötig sein. Diese können in ergänzenden Tabellen formuliert werden. Hierdurch wird die eigentliche Ursache-Wirkungs-Matrix nicht zu komplex. Die Effekte sind die Verriegelungsaktionen der Aktoren. Die Relation zwischen einer Ursache und ihrer Wirkung kann durch Setzen von Kreuzen in der Matrix angegeben werden.

Das nächste Modell muss für jeden Service der PEA einmal erstellt werden. Es beschreibt, welche Prozeduren er besitzt und welche Aktoren, Sensoren, Hilfselemente und Parameter dafür benötigt werden. Der Service Entleeren hat zum Beispiel die Prozedur *Entleeren nach Volumen*. Dazu werden, wie für den Service, die Pumpe NC001, das Ventil YS005, der Füllstands sensor LI001, der untere Schutzschalter LL001 und der Durchflusssensor FIC001 benötigt. Als Hilfselement wird ein Integrator verwendet, der die entnommene Stoffmenge aus dem Durchfluss berechnet. Durch Prozedurparameter kann festgelegt werden, welche Menge dem Tank entnommen werden soll.

Auch diese Darstellung der Struktur eines Service erfolgt durch eine Tabelle. Für jede Prozedur wird eine Zeile in der Tabelle hinzugefügt und in den Spalten vermerkt, was zur Durchführung der Prozedur benötigt wird. In weiterführenden Tabellen können die Parameter näher spezifiziert werden.

Zudem kommen für Services Regelkreise zum Tragen. Für die Modellierung wird eine Darstellung angelehnt an einen Wirkungsplan (IEC 60050-351), wie es zur Darstellung von Regelkreisen üblich ist, verwendet. Der Beispiel-Service Entleeren enthält einen Regelkreis, der den Durchfluss regelt. Der Sollwert wird durch einen Prozedurparameter vorgeben, der Istwert durch den Durchflusssensor FIC001 erfasst und der Regelausgang auf den Eingang der Pumpe NC001 zur Einstellung der Drehzahl verschaltet.

Im Schritt drei liegt der Modellierungsfokus auf den Prozeduren und des dahinter liegenden Zustandsmodells der VDI/VDE/NAMUR 2658. Für jede Prozedur wird jeder steuerbare Zustandsübergang des Zustandsautomaten und ihr Verhalten in den Zuständen beschrieben. Es wird für die Zustandsübergänge zwischen Zustandsübergangsbedingungen und Triggern unterschieden. Bei Erfüllung einer Zustandsübergangsbedingungen darf in den nächsten Zustand gewechselt werden. Eine Zustandsübergangsbedingung ist zum Beispiel, dass der Service Entleeren nur gestartet bzw. fortgesetzt werden darf, wenn der Tank einen Mindestfüllstand hat.

Beim Eintreten einer Triggerbedingung wird in den entsprechenden Zustand gesprungen. Trigger werden vor allem dazu verwendet, bei dem Auftreten von Fehlern oder kritischen Zuständen in einen Fehlerzustand zu springen. Zum Beispiel soll beim Entleeren-Service beim Auftreten eines Fehlers an der Pumpe NC001 oder bei Auslösen des unteren Schutzschalters LL001 der Service angehalten werden.

Zur Formulierung von Trigger- oder Zustandsübergangsbedingungen werden logische Ausdrücke oder mathematische Vergleiche verwendet. Die Darstellung von logischen Ausdrücken erfolgt grafisch mithilfe von Logikgattern. Um die Konsistenz zu wahren, werden auch Vergleiche grafisch mit ähnlichen Blöcken dargestellt.

Das Verhalten in den Zuständen wird für jede Prozedur und jeden Zustand einzeln modelliert. Zur Modellierung des Verhaltens wird in diesem Beitrag eine DSL basierend auf dem Strukturgramm nach Nassi-Shneiderman (DIN 66261) vorgeschlagen. Wie bei Nassi-Shneiderman werden parallele Abläufe durch parallele Rechtecke und Sequenzen durch aufeinanderfolgende Rechtecke modelliert. Schleifen und Verzweigungen werden wie im Nassi-Shneiderman dargestellt.

Ergänzend zu Nassi-Shneiderman werden für die unterschiedlichen Aktionen verschiedene Rechtecke verwendet. Es wird zwischen Aktionen von Aktoren, Hilfselementen und Services unterschieden sowie zwischen Aktionen mit und ohne Feedback. Eine Aktion mit Feedback bedeutet, dass die entsprechende Aktion nicht nur angestoßen wird, sondern auch abgewartet werden soll, bis ein Feedbacksignal, dass die Aktion beendet wurde, gesetzt wird. Beispielsweise wird bei Ausführen der Feedbackaktion „Ventil öffnen“ das Ansteuerungssignal gesetzt und erst fortgefahren, wenn am Ausgang das Signal, das ein offenes Ventil anzeigt, gesetzt wurde. Bei Aktionen ohne Feedback wird direkt nach Setzen des Ansteuerungssignals mit der nächsten Aktion fortgefahren. Ein Beispiel für eine Aktion ohne Feedback ist das Setzen eines Ventils in den Automatikmodus.

Da es immer erforderlich sein kann, dass das Einstellen eines gewissen Zustands Voraussetzung für das Ausführen der nächsten Aktion ist, wird ein neuer Block, das WAIT-FOR-Statement, eingeführt. Die in diesem Statement festgelegte Bedingung muss erfüllt sein, damit die nächste Aktion ausgeführt werden kann. Das Wait-For-Statement ist vergleichbar mit einer Transitionsbedingung einer Schrittkette. Beispiele dafür sind, dass ein Rührer so lange aktiv bleiben soll, bis ein Timer abgelaufen ist, oder solange Säure einer Flüssigkeit zugeführt wird, bis ein gewisser pH-Wert erreicht ist.

Um benutzerdefinierte Erweiterungen für die Logikimplementierung zu erlauben, gibt es die sogenannten Funktionen. Dazu muss im Modell das Interface der Funktion modelliert werden. Eingangsparameter sind erlaubt, die Funktionen besitzt aber keinen Rückgabewert. Die Funktion kann wie eine Aktion in der DSL verwendet werden, aber zusätzlich Übergabeparameter entgegennehmen. Bei der Codegenerierung wird ein leerer Baustein mit dem definierten Interface generiert. Die benötigte Logik kann der Benutzer später im SPS-Programmierungswerkzeug in den leeren Baustein in SCL implementieren.

Nach Durchlaufen aller Modellierungsschritte ist die PEA vollständig modelliert. Mithilfe eines Codegenerators wird der SCL-Code der Automatisierungssoftware generiert.

6. Prototypische Implementierung

Das Konzept wurde prototypisch basierend auf Eclipse implementiert und anhand des vorgestellten Beispiel-Use Case validiert. Das Eclipse Modelling Framework (EMF) wurde zur Implementierung des Metamodells, Sirius zur Implementierung eines Tools zum Erstellen des Benutzermodells und Aceleo zur Implementierung des Codegenerators verwendet. Der generierte Code besteht aus Bausteinen, die in SCL erstellt wurden und nach TIA Portal V16 importiert werden können. Mithilfe eines TIA-Addins wurde ein komfortables Importieren der generierten Bausteine und das Einordnen in eine entsprechende Ordnerstruktur umgesetzt.

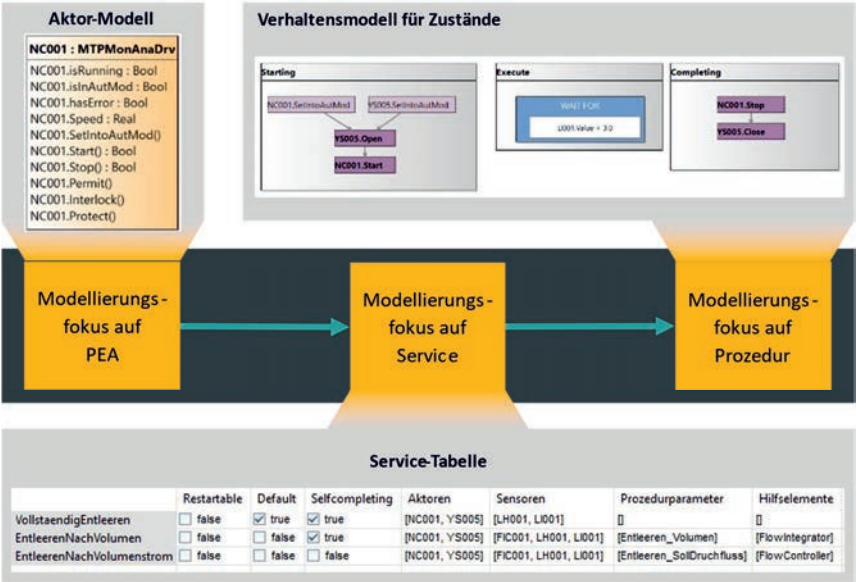


Bild 4: Auswahl an Screenshots aus der prototypischen Implementierung

Um das Konzept zu veranschaulichen, ist in Bild 4 für jeden Modellierungsfokus ein Screenshot aus der prototypischen Implementierung zu sehen. Oben links ist ein Beispiel für eine Elemente Modell-Instanz, einen analogen Antrieb, zu sehen. Rechts daneben ist die Darstellung von Verhalten in drei ausgewählten Zuständen zu sehen. In der untersten Zeile wird eine Service-Tabelle mit drei Prozeduren gezeigt.

Die Validierung wurde mit mehreren einfachen Use Cases wie in Bild 1 dargestellt und einer realen PEA aus der Siemens SmartAutomation Plant Karlsruhe durchgeführt. Der aus dem

jeweiligen Benutzermodell generierte SCL-Code kann in TIA Portal importiert und dort übersetzt werden. Die Vollständigkeit und logische Richtigkeit des Codes wurde durch Reviews und teilweise durch Prozesssimulationen validiert.

Mithilfe dieser prototypischen Umsetzung konnte bewiesen werden, dass es mit dem definierten modellgetriebenen Prozess möglich ist, eine PEA in diesem Umfang so zu beschreiben, dass daraus ein vollständiger, richtiger Code zur Ansteuerung einer PEA generiert werden kann.

7. Fazit und Ausblick

In diesem Beitrag wurde dargelegt, wie mit Hilfe eines modellgetriebenen Prozesses die Struktur und das Verhalten der Automatisierungssoftware einer PEA gemäß VDI/VDE/NAMUR 2658 beschrieben und der SCL-Code vollständig generiert werden kann. Die vorgestellten Modelle greifen, wo möglich und sinnvoll, bereits bestehende und in der Praxis übliche Darstellungsformen auf oder verwenden diese als Basis für eine DSL. Durch das Formalisieren und Erweitern der bestehenden Methoden zur Spezifikation von PEAs dienen die Modelle nicht nur der Dokumentation einer PEA, sondern vielmehr direkt der Softwareentwicklung.

Die Modelle erlauben eine abstraktere und einfachere Beschreibung des PEA, als es mit Quellcode möglich ist, und erleichtern damit die Softwareentwicklung. Informationen werden von einem Modell in das Nächste übergeben. Unkreative und sich wiederholende Code-abschnitte können dem Benutzer gänzlich durch den Codegenerator abgenommen werden.

Zudem sind heutige Methoden und Best-Practices für die Entwicklung von Software in der Automatisierung von PEAs in den modellgetriebenen Entwicklungsprozess verwoben. Das bietet den Vorteil, dass ein Programmierer diese nicht erst erlernen muss, sondern durch Befolgen des vorgestellten Entwicklungsprozesses automatisch umsetzt.

Die Validierung anhand einer realen PEA zeigte, dass mithilfe des vorgestellten modellgetriebenen Entwicklungsprozesses Automatisierungssoftware für PEAs schneller, einfacher und in hoher Qualität entwickelt werden kann.

Für die prototypische Implementierung wurden open-source Tools verwendet. Aus konzeptioneller Sicht ist eine Integration in kommerzielle Engineering- und Entwicklungswerkzeuge wie beispielsweise Comos, einer Siemens Plant Engineering Softwarelösung, realisierbar.

Die Modelle zu Beginn des Entwicklungsprozesses orientieren sich bereits am R&I. Weiterführend könnte deshalb eine computergestützte Analyse des R&Is durchgeführt und die dabei extrahierten Informationen in die ersten Modelle überführt werden. Konkret kann das bedeuten, dass alle relevanten Aktoren und Sensoren automatisch im R&I identifiziert und für die entsprechenden Elementmodelle automatisch instanziiert werden. Darüber hinaus sind

Konzepte denkbar, mithilfe derer Services und Prozeduren automatisch erkannt werden und in die entsprechenden Modelle des Entwicklungsprozesses einfließen.

- [1] Bramsiepe, C. u. Schembecker, G. (2012): Die 50%-Idee: Modularisierung im Planungsprozess.- Chemie Ingenieur Technik 5 (84), 581-587, Weinheim
- [2] NE148: Anforderungen an die Automatisierungstechnik durch die Modularisierung verfahrenstechnischer Anlagen, Namur, Leverkusen 2013
- [3] Bernshausen, J., Haller, A., Holm, T., Hoernicke, M., Obst, M., u Ladiges, J. (2016): Namur Modul Type Package–Definition: atp magazin, 58(01-02), 72-81, Essen
- [4] Stutz, A., Maurmaier, M. u. Unser, M. (2019): Service Design Principles zur Unterstützung des Modulbauers bei der Ermittlung möglicher Modul-Services gemäß VDI/VDE/Namur 2658, 883-894, Automation, Baden-Baden
- [5] Stutz, A. u. Maurmaier, M. (2018): How object-oriented design principles enhance the development of complex automation programs-A Best Practice Paper on how to develop service-interlaces for process modules as defined in VDI/VDE/NAMUR 2658.- 2018 IEEE 14th International Conference on Automation Science and Engineering (CASE) ,156-159, München
- [6] Stahl, T., Völter, M. (2005): Modellgetriebene Softwareentwicklung: Techniken, Engineering, Management, Heidelberg, dpunkt.verlag.
- [7] Selic, B. (2003): The pragmatics of model-driven development.- IEEE software, 20(5), 19-25, Los Alamitos
- [8] Maurmaier, M.: Modellgetriebene Entwicklung von Automatisierungssystemen (Dissertation). Stuttgart 2011
- [9] Schmidt, J. P., Zeller, A., u. Weyrich, M. (2017): Modellgetriebene Entwicklung service-orientierter Anlagensteuerungen.- at-Automatisierungstechnik, 65(1), 26-36. München
- [10] Liu, B., Glock, T., Betancourt, V. P., Kern, M., Sax, E., u. Becker, J. (2020): Model Driven Development Process for a Service-oriented Industry 4.0 System.- 2020 9th International Conference on Industrial Technology and Management (ICITM), 78-83, Oxford
- [11] Pang, C., Vyatkin, V., u. Dai, W. (2014): IEC 61499 based model-driven process control engineering.- Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA) , 1-8, Barcelona

Durchgängig konsistente Darstellung von Verriegelungen im Engineering prozesstechnischer Anlagen

F. El Sakka, Univ.-Prof. Dr.-Ing. A. Fay,
Helmut-Schmidt-Universität, Hamburg

Kurzfassung

Während des Engineerings von prozesstechnischen Produktionsanlagen werden Anforderungen an Funktionalitäten (exemplarisch werden hier die Verriegelungen betrachtet) in Lösungsvorschläge überführt und implementiert. Die Dokumentation von Engineering-Ergebnissen in geeigneten Beschreibungsmitteln dient als Kommunikationsgrundlage zwischen den beteiligten Gewerken im Engineering, wobei der Übergang von einem zum anderen Beschreibungsmittel fehleranfällig ist und Mehraufwand mit sich bringt. Im späteren Verlauf des Lebenszyklus einer Anlage wird der Wechsel von detaillierten Darstellungen auf höhere Abstraktionsstufen immer relevanter. Von den Autoren wird in diesem Beitrag ein Konzept für eine durchgängig konsistente Darstellung von Verriegelungen vorgestellt. Mithilfe eines entwickelten Beschreibungsmittels wird auf drei unterschiedlichen Detaillierungsstufen die Darstellung und das Editieren von Inhalten zu Verriegelungen ermöglicht. Als Basis für die unterschiedlichen Detaillierungsstufen wurde ein Informationsmodell definiert, welches die Inhalte der Verriegelungen in neutraler Form beschreibt. Mithilfe des Aufstellens von Regeln wird eine regelbasierte und automatisierte Aggregation und Dekomposition der Inhalte zwischen den unterschiedlichen Ebenen ermöglicht und die kausale Konsistenz der Inhalte über die unterschiedlichen Ebenen sichergestellt.

1. Motivation und Hintergrund

Die Planung und Inbetriebnahme prozesstechnischer Anlagen, auch Engineering genannt, erfordert als interdisziplinärer Prozess ein erhebliches Maß an Fachwissen aus unterschiedlichen Ingenieurdisziplinen, die auch als Gewerke bezeichnet werden [1]. Die Qualität der hierbei erzeugten Informationen hat nicht nur wesentlichen Einfluss auf die Geschwindigkeit der Anlagenerrichtung, sondern auch auf die Qualität und die Kosten der späteren Betriebsbetreuung über den Lebenszyklus der Anlage [2].

Das Zusammenwirken von unterschiedlichen Gewerken, in Form des Austausches von Engineering-Ergebnissen über Kommunikationsschnittstellen zwischen den Gewerken, stellt eine Herausforderung dar. Aufgrund hoher Abhängigkeiten zwischen den einzelnen Gewerken ist

ein regelmäßiger Austausch von Planungsständen wichtig. Insbesondere der Austausch von Engineering-Ergebnissen aus der Verfahrenstechnik (VT), welche für die Auslegung der Prozessabläufe und die Definition funktionaler und sicherheitstechnischer Anforderungen verantwortlich ist, und der Prozessleittechnik (PLT), die für die automatisierungstechnische Realisierung der Anforderungen zuständig ist, hat einen wesentlichen Einfluss auf den Erfolg eines Engineering-Projektes und sorgt zugleich für nicht unerhebliche Aufwände [1]. Um die Aufwände zu minimieren und die Qualität der Engineering-Ergebnisse zu erhöhen, werden hohe Anforderungen an die Kommunikationswege und die verwendeten Beschreibungsmittel (BSM), Methoden und Werkzeuge im Engineering gestellt.

Das Engineering einer Anlage in der Prozessindustrie ist nach der NA35 in sieben Planungsphasen eingeteilt [3]. Im Zuge der fortschreitenden Detaillierung von Engineering-Ergebnissen über die Planungsphasen kommt es zu einem Wechsel der Beschreibungsmittel. Textuelle Beschreibungen der Verfahrenstechnik über kausale Zusammenhänge zwischen Elementen einer Produktionsanlage werden im Laufe des Conceptual Engineerings und des Basic Engineerings in Rohrleitungs- und Instrumentierungs-Fließbilder (R&I-Fließbilder) überführt [3]. Aus den R&I-Fließbildern und zusätzlichen Anforderungen der Verfahrenstechnik, die oftmals verbal während Funktionsdurchsprachen an die Prozessleittechnik übermittelt werden, erstellt die Prozessleittechnik wiederum Cause-and-Effect-Matrizen (C&E), welche in Abschnitt 2 näher beschrieben werden. Die Cause-and-Effect-Matrizen werden zu Beginn des Detail Engineerings in Funktionspläne (FUP) überführt, die als Vorlage für die Software-technische Realisierung im Prozessleitsystem dienen.

Der beschriebene Wechsel von Beschreibungsmitteln ist aufgrund der oftmals händischen Übertragung zwischen den Beschreibungsmitteln fehleranfällig. Zudem kann eine Sicherstellung der Konsistenz der übermittelten Inhalte nur durch sehr hohe Aufwände erreicht werden. Sowohl bei der Inbetriebnahme als auch im weiteren Verlauf des Lebenszyklus' einer prozesstechnischen Anlage wird umgekehrt die Überführung der Darstellung von detaillierten Engineering-Ergebnissen hin zu einem höheren Abstraktionsniveau bei Umbauten oder Erweiterungen einer Anlage immer relevanter. Eine veraltete und unübersichtliche Dokumentation führt in diesen Situationen dazu, dass sich bestimmte Funktionsweisen einer bestehenden Anlage nur durch hohe Aufwände nachvollziehen lassen [4].

Dies betrifft insbesondere die Spezifikation und Realisierung von Verriegelungen einer prozesstechnischen Anlage und das Nachvollziehen dieser im Laufe des Lebenszyklus'. Verriegelungen, die in Beschreibungsmitteln mit einem hohen Detaillierungsgrad dargestellt werden, bspw. in Funktionsplänen, sind oftmals für die Verfahrenstechnik in dieser Form nicht mehr

verständlich, und die bisherigen Darstellungsformen daher für Projekt- und Funktionsdurchsprachen nicht ausreichend [5].

Die Herausforderungen in der Kommunikation, die aufgrund der unterschiedlich verwendeten Terminologien innerhalb der Gewerke auftreten, können durch den Einsatz von geeigneten Beschreibungsmitteln reduziert werden, da sie als unabhängige Notationsformen agieren [6, 7].

Um die beschriebenen Probleme zu adressieren, wurde von den Autoren ein Konzept zur Darstellung von Verriegelungen auf verschiedenen Abstraktionsebenen, mit Sicherstellung einer durchgängigen Konsistenz von Engineering-Ergebnissen, entwickelt.

In Abschnitt 2 werden zunächst die bisher in der Praxis verwendeten Beschreibungsmittel und deren erweiterte Formen vorgestellt. Das entwickelte Konzept zur durchgängig konsistenten Darstellung von Verriegelungen wird in Abschnitt 3 vorgestellt. In Abschnitt 4 folgt die Vorstellung einer prototypischen Implementierung des Konzeptes und eine Evaluation. Abschnitt 5 gibt abschließend eine Zusammenfassung und einen Ausblick.

2. Beschreibungsmittel zur Darstellung von Verriegelungen – state-of-the-art

Wie in Abschnitt 1 aufgeführt, finden während des Engineerings prozesstechnischer Produktionsanlagen unterschiedliche Beschreibungsmittel für Verriegelungen Anwendung. Unter Verriegelungen werden Teile eines Automatisierungssystem verstanden, die kritische Eingaben des Nutzers oder Ausgaben der Steuerung unterbinden und so die Ausführung einer Aktion verhindern, wenn zuvor definierte Bedingungen vorherrschen oder nicht vorherrschen. Verriegelungen bilden die letzte Instanz zur Vermeidung kritischer Situation in einer Anlage und erfordern ein hohes Maß an Fachwissen sowie ein besonders gründliches und systematisches Planen [8].

Zur Darstellung von Ursache-Wirkungs-Zusammenhängen gibt es verschiedene Beschreibungsmittel, die unterschiedliche Informationen mit variierenden Detaillierungsgraden darstellen. Als Ursachen werden hierbei Ereignisse gesehen, die in einer prozesstechnischen Anlage eine Reaktion eines technischen Systems auslösen. Die Wirkung ist als Reaktion auf eine Ursache definiert und stellt somit die Konsequenz der logischen Aktion der Ursache dar [5]. Das Überschreiten eines gewissen Grenzwertes eines Drucksensors in einem Kessel könnte eine typische Ursache für eine Verriegelung sein, wobei das Öffnen eines Ventils die Wirkung in Form einer Reaktion auf das Eintreten des zu hohen Drucks im Kessel ist.

Im Rahmen dieser Forschungsarbeit wurden verschiedene in der Praxis verwendete, Beschreibungsmittel für Ursache-Wirkungs-Zusammenhänge analysiert. Als Hauptkriterien wurden hierbei der Informationsumfang der Darstellung, die Verständlichkeit der Darstellung, der

Detaillierungsgrad für einzelne Ursache-Wirkungs-Zusammenhänge und die Erweiterbarkeit des jeweiligen Beschreibungsmittels mit einer Skala von „niedrig“ bis „sehr hoch“ bewertet. Zudem wurden als Nebenkriterien untersucht, ob das Beschreibungsmittel häufig in der Prozessindustrie Anwendung findet und demnach ein bekanntes Beschreibungsmittel für die Beteiligten ist, ob es Funktionsbausteine, oftmals auch in Form von einzelnen Funktionsblöcken als Typicals bezeichnet, darstellen kann und ob es eine Aggregation und Dekomposition der abgebildeten Inhalte ermöglicht. Untersucht wurde neben der Cause-and-Effekt-Matrix [5], welche das gängigste Beschreibungsmittel für die Darstellung von Verriegelungen darstellt und verschiedene Varianten umfasst, auch die ebenfalls matrixbasierte Dependency-Structure-Matrix (DSM) und Multi-Domain-Matrix (MDM) [9]. Zudem wurden verschiedene Fließbilder wie das R&I-Fließbild nach [10], das System Control Diagram (SCD) nach [11] und die Darstellung einer Process Control Engineering Aufgabe (PCE-Aufgabe) nach [12] analysiert. Des Weiteren wurden Beschreibungsmittel aus der Fehleranalyse wie die Fault Tree Analysis (FTA) und die Failure Mode and Effect Analysis (FMEA) [13], sowie die Darstellung in Funktionsplänen nach [14] untersucht.

Wie in Bild 1 zu erkennen ist, erfüllt keines der untersuchten Beschreibungsmittel alle Hauptkriterien mit einer hohen Bewertung. Stattdessen weisen die verschiedenen Beschreibungsmittel, insbesondere die Cause-and-Effekt-Matrix, die Design-Structure-Matrix und die Funktionspläne, Vorteile in einzelnen Eigenschaften auf. Aufgrund der Kürze dieses Beitrages sollen im Folgenden die wesentlichen Vorteile der zuvor genannten Beschreibungsmittel vorgestellt werden.

Die Cause-and-Effekt-Matrix ermöglicht die Darstellung eines sehr hohen Informationsumfanges, mithilfe dessen ein guter Gesamtüberblick über Ursache-Wirkungs-Zusammenhänge innerhalb einer Anlage oder eines Teilanlagenabschnittes gegeben werden kann. Zudem weist die Cause-and-Effekt-Matrix die höchste Bewertung in der Erweiterbarkeit auf, da die Darstellung von Informationen weder durch die in [5] beschriebene Struktur noch durch gestalterische Grundsätze eingeschränkt ist. Dem Detaillierungsgrad der Cause-and-Effekt-Matrix wurde aufgrund der eingeschränkten Darstellungsoptionen an den Schnittpunkten der Matrix eine mittlere Bewertung zugeordnet. Die höchste Bewertung in diesem Kriterium haben die Funktionspläne, die in einer hohen Anzahl für eine prozesstechnische Produktionsanlage erstellt werden und einzelne Signalverarbeitungen mithilfe von Funktionsbausteinen beschreiben.

Analyse der Beschreibungsmittel		BSM in der Prozessindustrie	Darstellung					Erweiterbarkeit des Beschreibungsmittels
			Informationsumfang der Darstellung	Verständlichkeit der Darstellung	Darstellungsoptionen			
					Detaillierungsgrad der Darstellung	Typicals	Aggregation und Komposition	
Matrix basiert	C&E	X	●	◐	◐	(X)	-	●
	DSM / MDM	-	◐	◑	◐	-	X	◐
Fließbilder	R&I	X	◑	◐	◑	-	-	◐
	SCD	X	◐	◑	◑	X	-	◑
	PCE-Aufgabe	X	◐	◑	◑	(X)	-	◑
Fehleranalyse	FTA	-	◐	◑	◐	(X)	X	◐
	FMEA	-	◑	◐	◐	-	X	◐
Sonstige	FUP	X	◑	◑	●	X	X	◐
Legende: ◐ : Niedrig ◑ : Hoch X : vorhanden - : nicht vorhanden ◐ : Mittel ● : sehr Hoch (X) : teilweise vorhanden								

Bild 1: Ergebnisse der Analyse unterschiedlicher Beschreibungsmittel für Ursache-Wirkungs-Zusammenhänge

Aufgrund der komplexen Darstellungsweise ist die Verständlichkeit der Funktionspläne niedrig zu bewerten, insbesondere für Gewerke außerhalb der Prozessleittechnik. Die Design-Structure-Matrix zeichnet sich mit ihrer symmetrischen Struktur durch eine hohe Verständlichkeit der Darstellung aus, und somit die höchste Bewertung in diesem Hauptkriterium.

Das Potential der matrixbasierten Beschreibungsmittel wurde in [15] aufgegriffen und ein erweiterter Ansatz einer Cause-and-Effekt-Matrix vorgestellt, bei dem die einzelnen Ursache-Wirkungs-Zusammenhänge nach der jeweiligen Funktionsweise aggregiert werden. Dieser Ansatz unterstützt die Darstellung von Verriegelungen auf unterschiedlichen Detaillierungsgraden durch eine Hierarchisierung der Informationen.

Um die genannten Herausforderungen bei der Kommunikation zwischen den Gewerken zu adressieren, sollte ein Beschreibungsmittel unabhängig vom vorherrschenden Planungsstand einsetzbar sein und eine Konkretisierung von Informationen im Laufe des Engineerings ermöglichen. Mithilfe der Aggregation und Dekomposition von Informationen kann zudem der Informationsumfang und der Detaillierungsgrad je nach anwendendem Gewerk angepasst werden.

Aus der Analyse geht hervor, dass in der Praxis des Engineerings prozesstechnischer Anlagen sowie in der Forschung kein Beschreibungsmittel existiert, das planungsstand- und gewerk-unabhängig verwendet werden kann und sowohl einen hohen Informationsumfang als auch eine hohe Verständlichkeit und einen entsprechenden Detaillierungsgrad aufweist. Vor diesem Hintergrund wurde die Notwendigkeit der Entwicklung eines Konzeptes erkannt, welches die geforderten Eigenschaften besitzt.

3. Konzept zur durchgängig konsistenten Darstellung von Verriegelungen

Um die Herausforderungen und Probleme im Engineering zu adressieren, wurde von den Autoren ein Konzept zur Darstellung von Verriegelungen auf verschiedenen Abstraktionsebenen, mit Sicherstellung einer durchgängigen Konsistenz von Engineering-Ergebnissen, entwickelt, welches über den gesamten Lebenszyklus einer Anlage Anwendung finden kann. Hierzu wurde ein Beschreibungsmittel entwickelt, welches auf drei Ebenen mit unterschiedlichen Detaillierungsstufen die Darstellung und das Editieren von Inhalten zu Verriegelungen ermöglicht. Als Basis für die unterschiedlichen Darstellungen des Beschreibungsmittels wurde ein Informationsmodell definiert, welches die Inhalte der Verriegelungen im Engineering von prozesstechnischen Anlagen in neutraler Form beschreibt. Mithilfe der Aufstellung von Regeln wird eine regelbasierte und automatisierte Aggregation und Dekomposition der Inhalte zwischen den unterschiedlichen Ebenen realisiert und somit die kausale Konsistenz der Inhalte über die unterschiedlichen Ebenen sichergestellt. Im Folgenden werden das Beschreibungsmittel, das Informationsmodell und die Regeln zur Aggregation und Dekomposition vorgestellt.

Beschreibungsmittel

Die Ebenen des Beschreibungsmittels gliedern sich nach den unterschiedlichen Informationsbereichen in eine Übersichtsebene (Ebene 1), welche eine abstrahierte Sicht auf die Verriegelungen einer Anlage oder Teilanlage ermöglicht, eine Geräteebene (Ebene 2) und eine

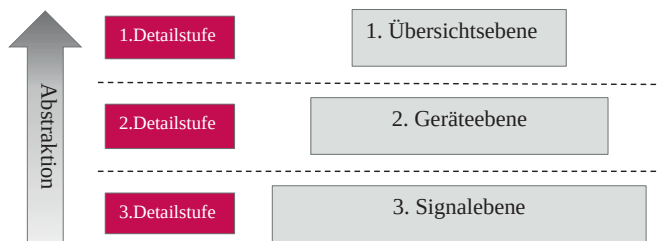


Bild 2: Abstraktion innerhalb des entwickelten Beschreibungsmittels

Signalebene (Ebene 3). Wie in Bild 2 dargestellt, wächst der Abstraktionsgrad hierbei von der Signalebene, welche die höchste Detaillierungsstufe darstellt, über die Geräteebene bis zur Übersichtsebene an.

In der ersten Ebene, der Übersichtsebene, werden räumliche Informationen einer Anlage untergliedert in Teilanlagenabschnitte und dazugehörige R&I-Fließbildnummern. Diese werden in einer Design-Structure-Matrix dargestellt und durch Operatoren an den Schnittpunkten verbunden. Die Operatoren setzen sich aus der absoluten Häufigkeit von kausalen Zusammenhängen aus den unteren Darstellungsebenen zusammen. Bild 3 zeigt hierzu eine konzeptionelle Darstellung der Ebene 1. Um insbesondere bei Brown-Field-Projekten eine Auskunft über die grundlegende Funktionsweise einer Anlage zu geben, wurden im Rahmen des Konzeptes, in Anlehnung an die in [16] definierten state types, verschiedene Szenarien einer Anlage definiert. Die Anlagenszenarien geben Auskunft darüber, für welches Szenario ein Element, wie bspw. ein verfahrenstechnischer Apparat oder ein prozessleittechnisches Gerät, in einem Anlagenabschnitt vorgesehen war oder genutzt wird. Definiert wurden die Anlagenszenarien *Anfahren* (A), *Normalbetrieb* (N), *Störung/Stop* (S), *Herunterfahren* (H) und ein vom Anwender frei definierbares Anlagenszenario (Z), welches bspw. Sonderoperationen, wie den Wechsel eines Katalysators, beschreibt. Die Anlagenszenarien *Normalbetrieb* (N) und *Störung/Stop* (S) lassen sich hierbei in weitere Unterkategorien einteilen, die bestimmte Grundoperationen im Prozessablauf oder Arten von Störungen beschreiben. Mithilfe der Einteilung in die beschriebenen Anlagenszenarien kann dem Anwender in der Übersichtsebene die absolute Häufigkeit der kausalen Zusammenhänge für ein oder mehrere Anlagenszenarien dargestellt werden und so Auskunft über die grundlegenden Zusammenhänge über verschiedene Teilanlagenabschnitte und in R&I-Fließbildern untergliederte Aggregate der Anlage gegeben werden. Die Operatoren werden durch den Anfangsbuchstaben des jeweiligen Anlagenszenarios und eine Ziffer der absoluten Häufigkeit symbolisiert, die durch einen Doppelpunkt getrennt werden, bspw. steht „A:4“ für vier Zusammenhänge des Anlagenszenarios *Anfahren*. Durch die Verwendung einer Design-Structure-Matrix werden zwingend alle R&I-Fließbilder und Teilanlagen miteinander in Relation gesetzt, wodurch auch das Nichtvorhandensein von Zusammenhängen visualisiert wird. Hierdurch ermöglicht die Übersichtsebene eine schnelle und gute Übersicht über eine Anlage und stellt die höchste Abstraktion des Beschreibungsmittels dar.

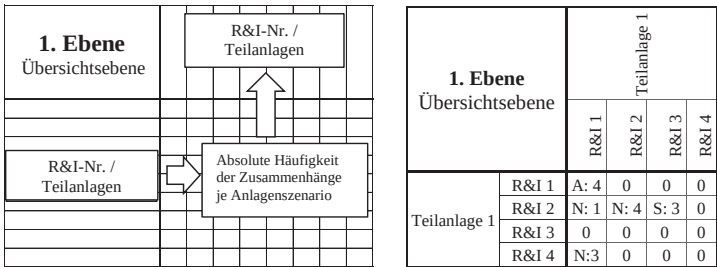


Bild 3: Konzept der Ebene 1 – Übersichtsebene (links: Prinzip; rechts: Beispiel)

In der Geräteebene, der Ebene 2, werden die einzelnen automatisierungstechnischen Einrichtungen, die typischerweise in einem R&I-Fließbild dargestellt sind, aufgeführt und nach dem Prinzip einer Cause-and-Effekt-Matrix in Ursachen (Sensoren und Aktoren) und Wirkungen (Aktoren) unterteilt. Bild 4 zeigt eine konzeptionelle Darstellung der Ebene 2. Den Geräten werden beschreibende Informationen zugeordnet, wie beispielsweise eine Bezeichnung, die R&I-Fließbildnummer, sowie Grenzwerte und Sicherheitsstellungen des Gerätes und das jeweils betrachtete Anlagenszenario. Die kausalen Zusammenhänge zwischen den Geräten werden durch unterschiedliche Operatoren in den jeweiligen Schnittpunkten zwischen den Ursachen und Wirkungen dargestellt. In der Geräteebene werden hierzu zwei unterschiedliche Operatoren verwendet. Mithilfe eines Operators der undefinierten Verbindung, welcher durch ein „X“ symbolisiert ist, können generelle Ursache-Wirkungs-Zusammenhänge bereits in den frühen Planungsphasen dokumentiert werden. Der Operator der undefinierten Verbindung stellt somit einen Hinweis für einen weiter zu spezifizierenden Zusammenhang an einen anderen Anwender, beispielsweise aus einem anderem Gewerk, dar. Im weiteren Planungsverlauf gilt es diese undefinierte Verbindung weiter zu konkretisieren. Durch den Operator der definierten Verbindung werden bereits konkretisierte Zusammenhänge zwischen Geräten mithilfe des Eingangssignals des Ursache-Gerätes und der Funktion des Wirkungs-Gerätes, getrennt durch einen Schrägstrich, beschrieben. Um die Zugehörigkeit von komplexen Zusammenhängen zu beschreiben, die sich aus mehreren Verbindungen zwischen Geräten zusammensetzen, werden innerhalb der Operatoren kleingestellte Identifikationsnummern (ID-Nummern) aufgeführt. Negierte Signale werden durch eine Darstellung des Signals innerhalb einer Klammer beschrieben. Der Operator „true/close₁“ wird bspw. verwendet, wenn das Ursache-Gerät das Signal „true“ sendet und dies in Kombination mit weiteren Zusammenhängen der ID-Nummer „1“ die Funktion „close“ beim Wirkungs-Gerät auslöst.

Um in Anlehnung an [15, 17] eine informelle Beschreibung von Zusammenhängen zu ermöglichen, können auf der Geräteebene zusätzliche Kommentare eingeführt werden, die bspw. Anforderungen der Verfahrenstechnik in einer frühen Engineering-Phase an eine bisher undefinierte Verbindung dokumentieren. Hierdurch sollen Informationen über grundlegende Wirkweisen und Ihre Begründung in der Anlage über den Lebenszyklus erhalten bleiben.

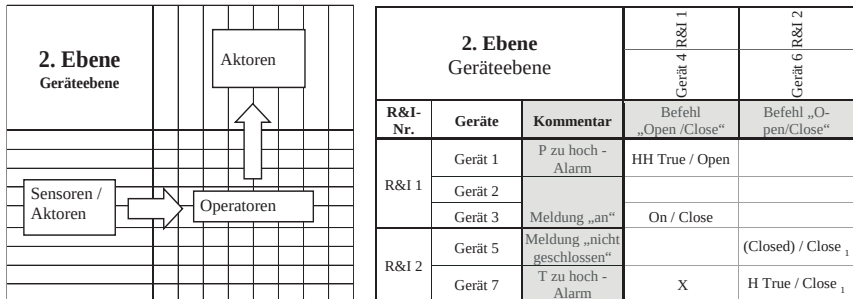


Bild 4: Konzept der Ebene 2 – Geräteebene (links: Prinzip; rechts: Beispiel)

In der Signalebene, der Ebene 3, werden die zugehörigen Ein- und Ausgangssignale und Funktionen der jeweiligen Elemente der Geräteebene in kausale Beziehungen gesetzt. Die Signalebene beinhaltet den höchsten Detaillierungsgrad des Beschreibungsmittels und umfasst Informationen, die typischerweise in einem Funktionsplan (FUP) aufgeführt werden. Wie bereits bei der Geräteebene wird auch bei der Signalebene das Beschreibungskonzept einer Cause-and-Effekt-Matrix übernommen. Bild 5 zeigt eine konzeptionelle Darstellung der Ebene 3. Die Operatoren an den Schnittpunkten der Kombinationen aus Signalen und Funktionen sind aufgrund der unterschiedlichen Komplexität analog zur Geräteebene in zwei verschiedene Operatoren unterteilt. Ist eine einfache Verbindung zwischen einem Signal und einer Funktion vorhanden, so wird ein „X“ an den Schnittpunkten der Matrix-Darstellung hinterlegt. Das „X“ ist gleichzusetzen mit einem einfachen ODER (OR). Bei der Verbindung mehrerer Signale zu einer logischen Verknüpfung, die eine Funktion auslöst, wird der Operator einer komplexen Verbindung verwendet. Dieser beinhaltet neben einer Zusammenhangs-Identifikationsnummer (Zusammenhangs-ID), welche die Zuordnung der Signale zu einer komplexen Verbindung in einem Funktionsbaustein repräsentiert, auch die Art des Funktionsbausteins bzw. des verwendeten Typicals und ggf. einen alternativen Ausgang, welcher durch die Zusammenhangs-ID des nachfolgenden Funktionsbausteins gekennzeichnet ist. Die Zusammenhangs-ID, die Art des Funktionsbausteins und der alternative Ausgang werden in drei Zeilen übereinander aufgeführt. Letzteres findet Anwendung, wenn eine Verkettung von

Funktionsbausteinen eines Zusammenhanges vorliegt und der Signalausgang eines Funktionsbausteins als Eingang für einen weiteren Funktionsbaustein fungiert. Hierdurch können auch komplexere Wirkweisen, wie eine Verriegelung über mehrere Eingangssignale und mehrere Funktionsbausteine, abgebildet werden. Die Zusammenhangs-IDs werden auf der Signalebene anders als auf der Geräteebene durch einen Hashtag dargestellt, bspw. „#002“. Ähnlich zur Geräteebene werden negierte Eingangssignale durch eine Klammer um die Zusammenhangs-ID im Operator der komplexen Verbindung oder durch eine Klammer um den

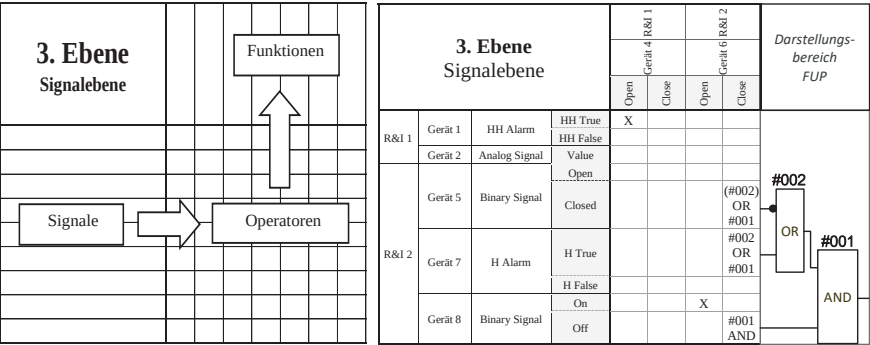


Bild 5: Konzept der Ebene 3 – Signalebene (links: Prinzip; rechts: Beispiel)

Operator der einfachen Verbindung dargestellt. Da sich Funktionspläne als detaillierte Darstellung von komplexeren Verbindungen etabliert haben, wird auf der Signalebene die Möglichkeit gegeben sich einzelne Verbindungen als Funktionspläne anzuzeigen. Die Signalebene stellt die detaillierteste Sicht im Beschreibungsmittel auf einzelne Signale, signalverarbeitende Operatoren und die resultierenden Funktionen dar.

Regeln zur Aggregation und Dekomposition

Das Konzept sieht eine Möglichkeit zur Aggregation und Dekomposition von Informationen zwischen den beschriebenen Ebenen des Beschreibungsmittels vor. Dies ermöglicht es einerseits verschiedene Abstraktionsformen der vorhandenen Informationen zu generieren und andererseits die Konsistenz bei Veränderungen in den Eingaben auf allen Ebenen sicherzustellen.

Von der Ebene 2 zur Ebene 3 wird eine Dekomposition von Informationen ermöglicht. Hierzu werden die auf Ebene 2 getätigten Eingaben über die Geräte und deren Eigenschaften, die Art des Operators, sowie bereits definierte Eingangssignale des Ursache-Gerätes und Funktionen des Wirkungs-Gerätes in die Ebene 3 transferiert. Wird bspw. in der Geräteebene eine

nicht komplexe und definierte Verbindung hinzugefügt, so wird die Eingabe des Eingangssignals, bspw. „HH-True“, zu der Ursachen-Seite der Signalebene und die ausgelöste Funktion, bspw. „Open“, zu der Wirkungs-Seite der Signalebene hinzugefügt. In der Schnittstelle zwischen dem hinzugefügten Eingangssignal und der Funktion wird ein Operator der einfachen Verbindung in der Signalebene, dargestellt durch ein „X“, hinzugefügt. Wird in der Geräteebene eine undefinierte Verbindung eingegeben, wird in der Signalebene das entsprechende Feld der Schnittstelle durch eine Umrandung hervorgehoben. Diese signalisiert dem Anwender, dass es hier einen Bedarf zur weiteren Spezifizierung der Verbindung gibt. Umgekehrt findet von der Ebene 3 zur Ebene 2 eine Aggregation von Informationen statt. Wird beispielsweise die zuvor genannte Umrandung der Schnittstelle einer nicht definierten Verbindung in der Signalebene weiter spezifiziert und sowohl Eingangssignal als auch die Funktion bestimmt, so wird diese Information in die Geräteebene übertragen und das zuvor gesetzte „X“ der undefinierten Verbindung wird durch einen Operator der definierten Verbindung ersetzt. Dieser stellt das in der Signalebene ausgewählte Eingangssignal und die Funktion durch einen Schrägstrich getrennt dar. Somit werden aus der Signalebene sowohl die Art des Operators als auch die Eingangssignale und Ausgangsfunktionen und die Zusammengehörigkeit von komplexen Zusammenhängen in die Geräteebene transferiert.

Von der Geräteebene zur Übersichtsebene findet eine Aggregation von Informationen statt. Hierbei werden die Zusammenhänge aus der Geräteebene für eine Teilanlage oder eine R&I-Fließbildnummer anhand der jeweiligen Anlagenszenarien aufsummiert. Die Übersichtsebene stellt damit eine generierte Sicht dar, die mithilfe der Inhalte der Geräteebene erstellt wird.

Als formale Notation der Regeln wurde die Semantic Web Rule Language (SWRL) [18] gewählt. Ein Beispiel hierfür ist nachfolgend aufgeführt. Es beschreibt die Zuordnung eines Anlagenszenarios zu einem Operator aufgrund der Verbindung zu einem Gerät, dem ein entsprechendes Anlagenszenario zugewiesen wurde.

$$\text{OperatorEbene2}(\text{?o1})^{\wedge}\text{hasInterface}(\text{?o1},\text{?p1})^{\wedge}\text{PlantHierarchyItem}(\text{?p1})^{\wedge}\text{hasPlantSzenario}(\text{?p1},\text{?s1}) \\ \rightarrow \text{hasPlantszenario}(\text{?o1},\text{?s1})$$

Informationsmodell

Um eine datentechnische Realisierung des Beschreibungsmittels zu ermöglichen und einen objektorientierten Ansatz zu unterstützen, wurde ein Informationsmodell für das beschriebene Beschreibungsmittel und die aufgeführten Inhalte definiert. Das Informationsmodell basiert auf dem Datenmodell Computer Aided Engineering eXchange (CAEX), welches in [12] definiert ist und einen generischen objektorientierten Ansatz für die Modellierung von Bibliotheken von Objektklassen und einer Instanz-Hierarchie ermöglicht [19]. Mithilfe des Bibliothekskonzept

von CAEX wird zudem die Auslegung von undefinierten und definierten Operatoren unterstützt. Das zugrundeliegende XML-Format bietet bei der Implementierung des Konzeptes Vorteile durch die Lesbarkeit für Maschinen und Menschen.

Das im Rahmen dieser Forschungsarbeit definierte Informationsmodell erweitert das CAEX-Datenmodell um die Objektklasse der unterschiedlichen Operatoren und erweiterten Interfaces dieser für die Geräteebe und die Signalebe. Bild 6 stellt das definierte Informationsmodell dar. Die RoleClass-Bibliothek wurde um die Operatoren der Geräteebe und der Signalebe erweitert. Für die in der Signalebe beschriebenen Funktionsbausteine sind in der SystemUnitClass-Bibliothek Standard Funktionsbausteine wie bspw. OR, AND, XOR und unternehmensspezifische Funktionsbausteine, bzw. Typicals definiert worden. Die Signaltypen, die als Verbindungen zwischen den einzelnen Operatoren agieren, wurden in der InterfaceClass-Bibliothek ergänzt. Da davon auszugehen ist, dass die innere Logik der verwendeten Funktionsbausteine durch unternehmensinterne Richtlinien und Guidelines bekannt ist, kann die innere Logik der Funktionsbausteine nach dem in [19] aufgeführten Vorgehen vernachlässigt werden. Die einzelnen Funktionsbausteine werden somit als Blackboxes gesehen, wodurch die Modellierung erleichtert und gleichzeitig weniger fehleranfällig gegenüber dem Hinzufügen weiterer unternehmensspezifischer Funktionsbausteine ist [19]. Den Funktionsbausteinen können die in der erweiterten InterfaceClass-Bibliothek vordefinierten Ein- und Ausgangssingale zugeordnet werden.

Im Rahmen einer Expertenbefragung von Planern der Prozessleittechnik wurde das Konzept evaluiert. Es wurden der Informationsumfang, die Verständlichkeit und der Detaillierungsgrad der Darstellung für jede Ebene des Beschreibungsmittels einzeln und aller Ebenen insgesamt bewertet. Zudem wurde auch die Möglichkeit zur Integration des Konzeptes in den Workflow der Prozessleittechnik während der Abwicklung von PLT-Planungsprojekten für jede Ebene des Beschreibungsmittels und insgesamt bewertet. In Bild 9 sind die Ergebnisse der Evaluation dargestellt. Die Ergebnisse der Evaluation stellen mit einer durchschnittlichen Bewertung von 7,6 Punkten bis 8,1 Punkten von möglichen 10 Punkten eine positive Bewertung des dargestellten Ansatzes dar. Als Kritikpunkt wurde die Notwendigkeit zum Erlernen der neuen Syntax, insbesondere bei den Operatoren der Ebene 3, genannt. Bei besonders komplexen Verbindungen kann hier die Verständlichkeit leiden. Das Anzeigen von Zusammenhängen in Funktionsplan-Darstellung wurde jedoch von den Befragten als positiv für die Verständlichkeit der Signalebene bewertet.

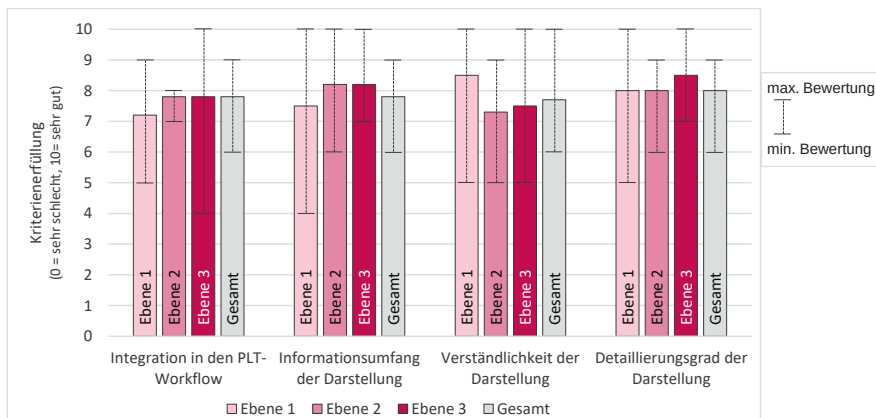


Bild 9 Evaluation des Ansatzes durch eine Expertenbefragung von PLT-Planern – n=6

5. Zusammenfassung und Ausblick

In diesem Beitrag wurde ein Konzept für eine durchgängig konsistente Darstellung von Verriegelungen über den Lebenszyklus einer prozesstechnischen Produktionsanlage vorgestellt. Hierzu wurden die Hintergründe und die Motivation der Forschungsarbeit sowie die Defizite in der aktuellen Darstellung von Verriegelungen in der Praxis dargelegt. Das vorgestellte Konzept ermöglicht mithilfe der Darstellung von Verriegelungen auf drei unterschiedlichen Detaillierungsstufen zugleich eine detaillierte Sicht auf die genaue Funktionsweise einzelner Verriegelungen und eine abstrahierte Übersicht über das Gesamt-Zusammenwirken von

Verriegelungen über verschiedene Teile einer Anlage. Letzteres ist für Sicherheitsbetrachtungen und Umbaumaßnahmen im Zuge von Brown-Field-Projekten im fortschreitenden Verlauf des Lebenszyklus einer Anlage von besonderer Relevanz und kann u.a. auch zur Einschätzung von Engineering-Aufwänden genutzt werden.

Durch die Definition von Regeln wird die kausale Konsistenz der Inhalte bei einer automatisierten Aggregation und Dekomposition über die unterschiedlichen Ebenen sichergestellt. Durch die Zusammenarbeit mit Engineering-Abteilungen eines großen deutschen Chemiekonzerns konnte die Praxisrelevanz während der Konzeptionierung sichergestellt werden. Anhand der Evaluationsergebnisse der Expertenbefragung in der Fachdisziplin der Prozessleittechnik wurde die Darstellungsform des Beschreibungsmittels bestätigt und das Konzept insgesamt als durchweg positiv für die Integration in den Planungs-Workflow der Prozessleittechnik bewertet.

Mithilfe eines Imports von Planungsdokumenten gilt es die Aufwände für manuelle Eingaben zu minimieren. Die zukünftige Umsetzung des Konzeptes in einem Software-Werkzeug soll sowohl von Planern der Verfahrenstechnik als auch von Planern der Prozessleittechnik und dem Betreiber einer Anlage verwendet werden und als interdisziplinäres Beschreibungsmittel über den Lebenszyklus einer Anlage Verwendung finden.

Die weiteren Ergebnisse der Forschungsarbeit sollen in die Normungsarbeiten der NAMUR und der GMA zur Erweiterung des NAMUR-Containers in Form einer neuen Version der NE150 und der VDI-Richtlinie 3697 zur herstellernerutralen Beschreibung und technischen Umsetzung von Verriegelungslogiken eingebracht werden.

Danksagung

Für die inhaltlichen Impulse und das substantielle Feedback während fachlicher Diskussionen im Rahmen der Forschungsarbeit bedanken sich die Autoren bei Herrn Dr. Marco Boll, Herrn Dr. Andreas Steinweg, Herrn Dr. Hisham Mubarak, Herrn Horst Seel, Herrn Dr. Kai Krüning und Herrn Dr. Matthias Müns.

6. Literaturangaben

- [1] Bindel, T. und Hofmann, D. *Projektierung von Automatisierungsanlagen: Eine effektive und anschauliche Einführung*, 3rd ed. Wiesbaden: Springer Vieweg, 2017.
- [2] Tauchnitz, T. *Mammutprojekt integriertes Engineering: Los geht's!*, Chemie Technik, S.67–70, Oct. 2013.
- [3] NA 35: *PLT-Planung und -Abwicklung in der Prozessindustrie*, NAMUR, 2019.
- [4] Weber, K. *Dokumentation verfahrenstechnischer Anlagen: Praxishandbuch mit Checklisten und Beispielen*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.
- [5] IEC 62881: *Ursache-Wirkungstabelle*, IEC, 2018.
- [6] Ahrens, W., Felleisen, M, Schnieder, E. und Chouikha, M. *Formale Prozessbeschreibungen - gestern, heute und morgen*. atp - Automatisierungstechnische Praxis. 42(9), S. 24-32, 2000.
- [7] Diedrich, C. und Mühlhause, M. *Modellansätze für die digitale Fabrik*, at-Automatisierungstechnik, vol. 59, no. 1, 2011.
- [8] Fischer, S. *Entwurf und Verifikation von Ablaufsteuerungen*, Dortmund, Techn. Univ., Diss., Aachen: Shaker, 2015.
- [9] Lindemann, U., Maurer, M. und Braun, T. *Structural Complexity Management: An Approach for the Field of Product Design*, Berlin Heidelberg: Springer, 2009.
- [10] DIN EN ISO 10628-1: *Schemata für die chemische und petrochemische Industrie: Teil 1: Spezifikation der Schemata*, Deutsches Institut für Normung e.V., 2014.
- [11] IEC PAS 63131: *System control diagram*, IEC, 2017.
- [12] DIN EN 62424: *Festlegung für die Darstellung von Aufgaben der Prozessleittechnik in Fließbildern und für den Datenaustausch zwischen EDV-Werkzeugen zur Fließbilderstellung und CAE-Systemen*, Deutsches Institut für Normung e.V., 2017.
- [13] Werdich, M. *FMEA - Einführung und Moderation: Durch systematische Entwicklung zur übersichtlichen Risikominimierung*, 2. Auflage, Wiesbaden, 2011.
- [14] IEC 61131-3: *Programmable controllers: Programming languages*, IEC, 2013.
- [15] Fischer, S., Sonntag, C. und Engell, S. *Systematic Formalization of Control Requirements using Hierarchical Cause-Effect Charts*, in *Computer Aided Chemical Engineering*, vol. 31, pp. 1642–1646, Oxford: Elsevier, 2012.
- [16] ANSI/ISA–88.00.01: *Batch Control Part 1: Models and Terminology*, 2010.
- [17] Fischer, S., Hüfner, M, Sonntag, C. und Engell, S. *Systematic Generation of Logic Controllers in a Model-Based Multi-Formalism Design Environment*, 18th IFAC World Congress, 2011.

- [18] Horrocks, I., Patel-Schneider, P., Boley, H., Tabet, S., Grosz, B. und Dean, M. *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*. [Online]. Available: <http://www.w3.org/Submission/SWRL/>
- [19] Drath, R. und Ingebrigtsen, I. *AutomationML in the Oil&Gas Industry: Automated code generation by means of the IEC PAS 63131*, 5th AutomationML User Conference, 2018.
- [20] NE 159: *Standardisierte NAMUR-Schnittstelle zwischen CAE-Systemen für den Austausch von verfahrenstechnischen Daten*, NAMUR, 2018.

Testen 4.0 in der Automatisierungstechnik: Agiles modellbasiertes Testen vernetzter Systeme und Komponenten

Testing 4.0 in Industrial Automation: Agile Model-Based Testing of Networked Systems and Components

A. Löcklin, M. Müller, Prof. Dr.-Ing. Dr. h. c. **M. Weyrich**,
Institut für Automatisierungstechnik und Softwaresysteme,
Universität Stuttgart;
C. Kotsch, Dr.-Ing. **K. Krüning**, BASF SE, Ludwigshafen am Rhein;
M. Rentschler, Balluff GmbH, Neuhausen auf den Fildern;
Prof. Dr.-Ing. **C. Ebert**, Vector Consulting Services GmbH, Stuttgart

Kurzfassung

Die Flexibilisierung von Produktionsanlagen ist notwendig, um in einem volatilen Marktumfeld erfolgreich sein zu können. Durch die Bereitstellung und Pflege geeigneter Schnittstellen und Betriebsmodi, sowie dem Umbau von Anlagen, kann auf geänderte Anforderungen an die Produktion reagiert werden. Für die Realisierung derartiger hochflexibler und rekonfigurierbarer Anlagen ist ein erhöhter Engineering- und Testaufwand notwendig. Jede zusätzliche Funktion samt aller damit verbundener Abhängigkeiten müssen vor Inbetriebnahme mehrfach von Herstellern, Integratoren und Betreibern überprüft werden. Damit der resultierende erhöhte Testaufwand auch zukünftig handhabbar bleibt, werden in dieser Veröffentlichung folgende Ansätze vorgestellt, um ein effizientes und wirkungsvolles Testen im Bereich der flexiblen Produktion zu ermöglichen: Zum einen lässt sich durch eine auf Richtlinien basierende Modularisierung von Anlagen und Komponenten der Testaufwand reduzieren. Auch sorgen Agile Teams für eine verbesserte Kommunikation zwischen Entwicklungs- und Testabteilungen. Überdies ermöglicht die Anwendung von Digitalen Zwillingen einen effektiveren Informationsaustausch zwischen den Stakeholdern und Modellbasiertes Testen ermöglicht Vorteile bei der Testautomatisierung. Alle Ansätze sind jeweils eigenständig wirksam, aber zusammengekommen ergeben sich zusätzliche Synergieeffekte.

1. Herausforderungen für das Testen vernetzter Systeme und Komponenten

Geänderte Marktanforderungen haben in den letzten Jahren für eine stärkere Fokussierung der Flexibilität von Produktionsanlagen gesorgt [1–4]. Allerdings sorgen die Steigerungen von Flexibilität und Effizienz der Produktion in der Prozessindustrie und diskreten Fertigung für eine Erhöhung des Testaufwands bei den eingesetzten Komponenten und Systemen [5, 6]. Diese müssen über viele unterschiedliche Schnittstellen verfügen, um flexibel in verschiedensten Nutzungsszenarien eingesetzt werden zu können. Die umfassende Vernetzung aller an der Produktion beteiligten Assets ermöglicht darüber hinaus Effizienzsteigerungen durch die Realisierung flexibler Systemreaktionen bei unerwarteten Ereignissen wie etwa Ausfällen in der Lieferkette. Die dadurch entstehenden zusätzlichen internen und externen Abhängigkeiten sorgen für mehr Aufwand bei der Komponentenentwicklung und können Schwierigkeiten bei der Inbetriebnahme verursachen. Bei allen Beteiligten, also Komponentenherstellern, Systemintegratoren und Systembetreibern, steigen die notwendigen Testaufwände für einen effizienten und in getesteten Grenzen flexiblen Produktionsbetrieb [5, 6]. Die Nutzung des vollen Potentials flexibler Produktionsanlagen und die Erreichung maximaler Effizienz ist nur möglich, wenn die notwendigen Testaufwände noch bewältigt werden können.

In diesem Beitrag werden zunächst in Kapitel 2 Lösungsansätze vorgestellt, welche dann an einer Beispielanlage aus der Prozessindustrie in Kapitel 3 und 4 erläutert werden. Am Ende folgt ein Ausblick in Kapitel 5.

2. Testaufwand durch effizienteres Engineering reduzieren

Nicht mehr zu bewältigende Testaufwände können die Inbetriebnahme verzögern oder zu Teilfreigaben führen, die nicht einen flexiblen, sondern nur einen in sehr starren Grenzen definierten Betrieb erlauben. Der Grundsatz von exponentiell steigenden Fehlerbeseitigungskosten, je später Fehler entdeckt werden, gilt auch hier. Je später im Lebenszyklus einer Anlage eine Funktion überprüft wird, desto aufwändiger ist die Beseitigung möglicher Fehler, beispielsweise da Experten und Messgeräte nicht mehr verfügbar oder Zugangsmöglichkeiten eingeschränkt sind [6]. Deshalb müssen frühzeitig Maßnahmen zur besseren Bewältigung des Testaufwands ergriffen werden. Hierzu bieten sich folgende Strategien an:

- Testaufwand reduzieren
- Testdurchführung beschleunigen und Testen effizienter gestalten
- Testpriorisierung auf Basis von Kritikalitätsbewertungen

Im Folgenden werden die ersten beiden Strategien weiter betrachtet, eine Testpriorisierung sollte zusätzlich immer erwogen werden. Die Überprüfung des vollen Funktionsumfangs von

Komponenten und Teilsystemen stellt eine Bestätigung der Anlagenflexibilität dar und erfordert die Durchführung aller Tests. Daher sollten im Rahmen einer Priorisierung keine Testfälle eingespart werden. Um Testaufwand zu reduzieren und die Testdurchführung zu beschleunigen werden vier Ansätze vorgestellt. Durch die in Kapitel 2.1 beschriebene Modularisierung und Standardisierung von Komponenten wird eine Reduktion von Testaufwand möglich. Eine Verbesserung der Effizienz von Testprozessen kann erreicht werden mittels der früheren Einbindung von Testexperten im Zuge agiler Entwicklungsprozesse (Kapitel 2.2), der einfacheren Bereitstellung von Information durch Digitale Zwillinge (Kapitel 2.3) sowie einer Erhöhung von Testautomatisierung (Kapitel 2.4).

2.1. Wiederverwendung und Modularisierung reduzieren Testaufwand

Der Testaufwand einer Komponente oder eines Systems ist abhängig vom Funktionsumfang, weil jede bereitgestellte Funktion getestet werden muss. Eine Reduktion des Funktionsumfangs ist meist nicht möglich, da sich dieser aus den Marktanforderungen ergibt. Die verbleibende Möglichkeit zur tatsächlichen Reduktion von Testaufwänden stellt die Wiederverwendung von Komponenten, Teilsystemen oder Anlagen dar. Bei Wiederverwendung von Software müssen bereits durchgeführte Tests nicht wiederholt werden. Bei der Wiederverwendung von Hardware müssen lediglich Mängel im Herstellungsprozess durch Tests abgesichert werden.

Wiederverwendung profitiert von einer durchdachten Modularisierung von Strukturen auf Software-, Hardware- und Funktionsebene. Ein häufig wiederverwendetes Modul rechtfertigt hohe Investitionen in Qualitäts- und Effizienzoptimierungen. Je häufiger ein bestimmtes Modul eingesetzt wurde, desto weniger Qualitätsmängel treten bei einer erneuten Wiederverwendung und der zur Freigabe notwendigen Integrations- und Systemtests auf. Und jeder durch die Wiederverwendung von Komponenten eingesparte kundennahe Abnahme- oder Factory Acceptance Test spart besonders zeitkritischen Testaufwand. Darüber hinaus ergeben sich für Standardkomponenten eine bessere Ersatzteil- und Knowhow- bzw. Expertenverfügbarkeit im Fehlerfall.

Aufgrund der in der Domäne der Produktions- und Automatisierungstechnik großen Heterogenität von Komponenten und Anwendungsszenarien erfordert Modularisierung auch eine von ingenieurwissenschaftlichen Prinzipien getriebene Standardisierung, um der Bildung von modularisierten aber trotzdem schlecht wiederverwendbaren Insellösungen entgegenzuwirken. Die in Kapitel 4.1 beschriebene Reduktion von Testaufwand durch Modularisierung am Beispiel einer Anlage aus der Prozessindustrie nutzt daher den Module-Type Package (MTP) Standard.

2.2. Agile Teams verbessern Kommunikation

Für eine bessere Bewältigung des Testaufwands kann die Effizienz von Testprozessen erhöht werden. Verzögerungen im Bereich des Testentwurfs werden häufig durch die mangelnde Testbarkeit von Anforderungen verursacht. Diese auch häufig als *testability* bezeichnete Eigenschaft von Anforderungen drückt aus, wie einfach Tests und notwendige Testdaten, wie zum Beispiel Grenzwerte, aus Anforderungen abgeleitet werden können. Je präziser eine Anforderung vorliegt, desto höher deren Testbarkeit. Die möglichst frühe Einbindung von Testexperten in einem agilen Engineering Prozess kann die Testbarkeit von Anforderungen verbessern, da diese mehr Fokus auf klar definierte Grenzwerte und Verhalten in Ausnahmefällen legen. Dadurch wird der Interpretationsspielraum bei der späteren Testfallerstellung reduziert und Missverständnisse vermieden. Darüber hinaus erhalten Testexperten bereits in frühen Projektphasen ein vollständigeres Bild des Testobjekts und können rechtzeitig organisatorische Abhängigkeiten erkennen und verteilte Testaufgaben sowie die Entwicklung von Testautomatisierungslösungen einplanen.

2.3. Digitale Zwillinge für effizienten Informationsaustausch

Ein häufiger trivialer Grund für Verzögerungen von Engineering- und Testprozessen stellt die mangelnde Verfügbarkeit von Informationen dar [6, 7]. Das Digitale Zwilling Konzept hilft verstreut vorliegende Informationen an einer zentralen Stelle zu bündeln [8, 9]. Mittels teilautomatisierter Informations-Synchronisierung zwischen realer Anlage und der digitalen Beschreibung wirkt ein Digitaler Zwilling der schleichenden Obsoleszenz von Komponenten- bzw. Systemdokumentation entgegen. Für einen automatisierbaren Informationsaustausch zwischen Abteilungen oder Unternehmen müssen semantische Lücken, also wenn etwa gleiche Dinge unterschiedlich benannt oder gemessen werden, durch Standards zur Informationsbeschreibung abgebaut werden. Die inzwischen durch die Industrial Digital Twin Association (IDTA) [10] vorangetriebene Entwicklung der Verwaltungsschale (engl. Asset Administration Shell) stellt hierbei einen Ansatz für die Verknüpfung verschiedener Standards und formalisierter Informationsbereitstellung durch einen Digitalen Zwilling dar.

2.4. Modellbasierte Testautomatisierung

Neben der Kostenoptimierung ist ein Hauptziel von Testautomatisierung die beschleunigte Bewältigung von Testaufwänden. Dabei werden zwischen der Automatisierung der Testausführung und der Testfallgenerierung unterschieden. Eine *automatisierte Ausführung* von Testfällen bedingt einen zusätzlichen Entwicklungsaufwand und lohnt sich nur bei wiederkehrenden Testaufwänden. Die in diesem Beitrag betrachtete *automatisierte Generierung* von Tests kann

hingegen auch bei einmaliger Testausführung eine Aufwandseinsparung erzielen. Ein im Bereich Softwareentwicklung sehr erfolgreich eingesetzter Ansatz zur Testfallgenerierung stellt das Modellbasierte Testen dar. Anstatt Testfälle manuell zu entwerfen, wird beim Modellbasierten Testen die Abstraktionsebene beim Testfallentwurf erhöht und Testexperten überführen Anforderungen und Spezifikationen in formalisierte Modelle, etwa Zustandsautomaten. Der anschließende Entwurf von konkreten Testfällen kann dann automatisiert mit Hilfe von Softwarewerkzeugen erfolgen, Testfälle werden nicht mehr entworfen, sondern generiert. Für die Analyse der Modelle nutzen die Softwarewerkzeuge Techniken wie etwa Pfadtraversierung oder Mutationsgenerierung und berechnen gleichzeitig Metriken der Testabdeckung, um automatisiert ausreichend Testfälle aus den Modellen zu generieren. Empirischen Studien zufolge können mit modellbasierter Testfallgenerierung Testaufwände schneller bewältigt werden [11, 12] und es werden mehr kritische Fehler gefunden [13, 14], allerdings entstehen initiale Mehraufwände bei der Einführung von Modellbasierten Testverfahren [12, 15].

2.5. Bessere Bewältigung von Testaufwand

Die in den Kapitel 2.1 bis 2.4 vorgestellten Ansätze zur besseren Bewältigung von Testaufwänden stellen einzelne Puzzlestücke für das Testen vernetzter Systeme und Komponenten dar. Eine reine Beschleunigung von Testprozessen mit den in Kapitel 2.2 bis 2.4 beschriebenen Maßnahmen kann scheitern, wenn Testaufwände aufgrund immer komplexerer Anlagen unendlich ansteigen. Durch die in Kapitel 2.1 beschriebene Wiederverwendung und standardisierte Modularisierung von Komponenten und Systemen kann Testaufwand wieder reduziert werden. Eine höhere Agilität durch die in Kapitel 2.2 beschriebene frühe Einbindung von Testexperten in das Requirements Engineering erleichtern die Testplanung und verbessern die Testbarkeit von Anforderungen. Bei abteilungs- und unternehmensübergreifenden Großprojekten ist hierfür ein reibungsloser Austausch stets aktueller Informationen notwendig, einen Ansatzpunkt hierfür stellt, wie in Kapitel 2.3 beschrieben, die Verwendung von Digitalen Zwillingen dar. Die durch Digitale Zwillinge verbesserte Modellverfügbarkeit sowie Fortschritte bei der Einführung von modell-getriebenem Engineering machen zudem die Einführung von Modellbasierter Testfallgenerierung greifbar, welche zu einer Aufwandseinsparung führt.

Alle vier Ansätze wirken auch für sich genommen und wenn alle beschriebenen Puzzlestücke zusammenwirken ergeben sich Synergieeffekte. Dies wird im Folgenden beispielhaft an einer Demonstrationsanlage aus der Prozessindustrie erläutert. Je modularer und formalisierter eine Anlage aufgebaut wurde, desto leichter können Agile Entwicklungsprozesse angewendet, Digitale Zwillinge erstellt und modellbasierte Testfallgenerierung etabliert werden. Die Abkürzung Testen 4.0 soll dabei nicht als Revolution aufgefasst werden, sondern damit wird verdeutlicht,

dass flexible, vernetzte Anlagen große Herausforderungen im Bereich des Testens mitbringen und Testprozesse entsprechend weiterentwickelt werden müssen.

3. Vorstellung des Open Automation Demonstrators aus dem Bereich Modularisierung

Für die weitere Erläuterung der in Kapitel 2 dargestellten Ansätze dient als Praxisbeispiel die in Bild 1 dargestellte Demonstrationsanlage aus dem Bereich der Prozessindustrie. Als Demonstrator für die modulare Automation in der chemischen Industrie wurde ein Testsystem gebaut, das im Fachzentrum für Automatisierungstechnik der BASF Ludwigshafen für die Evaluierung neuer Prozessleitsysteme sowie neuer Leitsystemversionen eingesetzt wird. Mit diesem Demonstrator können alle gängigen Funktionen eines Prozessleitsystems einer realen Anlage getestet werden. Das Testsystem wurde zunächst ohne den Fokus Modularisierung vor etwa zehn Jahren aufgebaut. Ziel war es, alle wesentlichen Funktionalitäten, die in den Anlagen und am Verbundstandort Ludwigshafen implementiert sind, in ein kompaktes Testsystem zu integrieren und auf dem Stand der Technik zu halten. Im Rahmen der Erneuerung des Demonstrators wurde die in Kapitel 4.1 beschriebene Modularisierung mittels Module-Type Package (MTP) realisiert [16].

Das Testsystem besteht aus vier Wasserbehältern, die mit Verrohrung verbunden sind. Jeder Tank hat ein Auslassventil und zwei Einlassventile. Alle Tanks besitzen eine Kühlfunktion und zwei der Tanks zusätzlich eine Heizfunktion. Um das Medium zwischen den Tanks zu transportieren, sind zwei Pumpen im Testsystem verbaut und fest verrohrt.

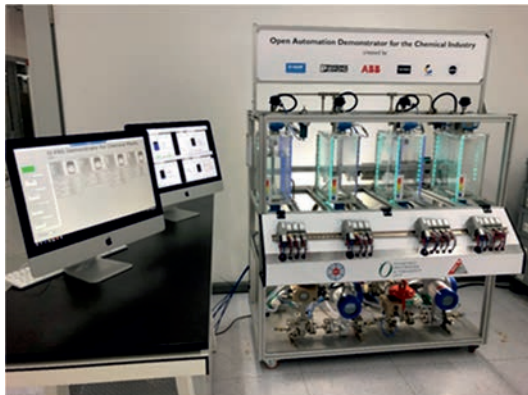


Bild 1: Open Automation Demonstrator for the Chemical Industry

Der Demonstrator deckt zwei Anlagentypen ab: Kontianlagen und Batchanlagen. Ein einfaches Rezept einer Batchanlage ist, aus den Vorratsbehältern eins und zwei, im Behälter drei eine Reaktion in einem bestimmten Verhältnis hervorzurufen. Die Reaktion wird durch Kühlen und/oder Heizen hervorgerufen. Der Behälter vier dient als Tanklager. Die Ansteuerung der Module erfolgt über Services. Dieses einfache Rezept zeigt schon sehr deutlich die Anforderung an die Automatisierung: Während in einem herkömmlichen Leitsystem die Funktionalität auf einem sehr niedrigen Abstraktionsniveau in Form von Funktionsbausteinen und Logikgattern implementiert wurde, ist bei der modularen Automation wie in Bild 2 dargestellt die Software-Kapselung auf einem hohen Abstraktionsniveau ein zentrales Ziel.

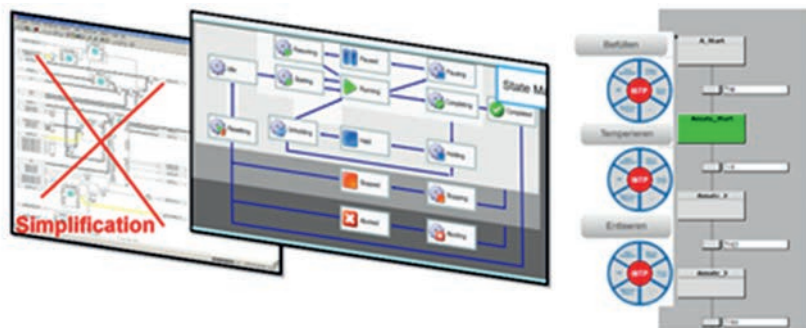


Bild 2: Vorkonfektionierte Services ersetzen individuelle Logik

4. Testen 4.0 am Beispiel des Open Automation Demonstrators

Die in Kapitel 2 vorgestellten Ansätze zum effizienten und wirkungsvollen Testen von vernetzten Komponenten und Systemen werden im Folgenden weiter vertieft.

4.1. Modularisierung mittels Module-Type Package

Ein verfahrenstechnischer Umbau von Bestandsanlagen ist sehr teuer. Oberstes Ziel ist es daher, keinerlei verfahrenstechnische Änderungen wie Zusatzarmaturen, Verrohrung oder zusätzliche Absperrventile einzubauen und so nah wie möglich an der existierenden Implementierung zu bleiben. Die nachträgliche softwaretechnische Modularisierung erleichtert die Skalierung einer Gesamtanlage, den Test der Anlagen im Vorfeld, wie auch das Ergänzen zusätzlicher Funktionen mit der Sicherheit, dass vorhandene Funktionen nicht verändert wurden. Der Migrationsaufwand für das Leitsystem reduziert sich, da elementare Steuerungsfunktionen von Modulen gekapselt sind.

Über die Kapselung der Software erreicht man eine funktionale Modellierung. Hierbei werden in den Modulen Services implementiert, z. B. für das Temperieren und Transportieren. Vergleicht man nun das herkömmliche Leitsystem mit dem Modularisierungsansatz, so entsprechen die Services der Module den Grundfunktionen eines Batchs (siehe ISA 88), jedoch mit einem höheren internen Automatisierungsgrad versehen und dadurch im Rahmen von Systemtests besser testbar.

Bei dem Umbau des Demonstrators in ein modular automatisiertes System ist die erste Aufgabe, Module zu definieren. Durch den Zuschnitt der Module des Demonstrators auf Softwareebene ergeben sich vier Module, mit jeweils einem Behälter, mit zugehörigem Equipment, wie Aktorik und Sensorik. Die Definition der Module auf Softwareebene zeigt, dass aufgrund der Brownfield-Automatisierung und der fehlenden Zusatzarmaturen, sowie einem nicht modularen verfahrenstechnischen Aufbau, gewisse Einstiegshürden zu nehmen waren. Die Zuordnung der Pumpen für den Transport des Mediums zwischen den Behältern stellte hierbei die größte Hürde dar, diese wurden als geteiltes Equipment realisiert, das mehreren Modulen zur Verfügung steht.

Durch die modulare Automatisierung kann das Skalierungsverhalten einer Bestandsanlage verbessert werden. Die neu zu implementierenden Funktionen können im Vorfeld nahezu vollständig getestet werden. Die Stillstandszeit der Anlage reduziert sich deutlich oder ist gar nicht mehr notwendig. Anlagenteile können schneller erweitert werden, neue Funktionalitäten können in laufende Anlagen per Plug & Produce implementiert werden, ohne auf die Verfügbarkeit der Anlage zu verzichten.

4.2. Agiles Testen mit Test-Driven Requirements Engineering

Requirements Engineering und Test werden häufig getrennt behandelt. Das hat zur Folge, dass in vielen Projekten Anforderungen erst sehr spät auch aus der Testperspektive analysiert werden, häufig erst, wenn erste Systembestandteile bereits teilweise realisiert sind. Durch diese gängige Praxis ergeben sich zwei Nachteile: Unzureichende Anforderungsqualität wird zu spät entdeckt und die Testfall-Erstellung ist aufwändiger und fehleranfälliger aufgrund eines Mangels an Kontextinformationen aus der Requirements-Engineering Phase. [17]

Test-Driven Requirements Engineering (TDRE) setzt hier an. Durch den Einbezug von Testingenieuren beim Requirements Engineering steigt die Qualität von Anforderungen sowie die Verzahnung mit der Testfall- und Testautomatisierungs-Entwicklung. Bei der Umsetzung von TDRE werden Kundenanforderungen direkt in Testfälle übersetzt. Eine Automatisierung dieses Vorgangs ist bei formalisiert vorliegenden Anforderungen möglich, also wenn Anforderungen über eine konsistente Struktur samt messbaren Vor- und Nachbedingungen verfügen [18].

Bei der im Kapitel 4.1 beschriebenen Modularisierung mittels Module-Type Package werden beispielsweise standardisierte Zustandsautomaten genutzt. Darauf aufbauend können, wie in Kapitel 4.4 beschrieben, Modelle entwickelt und für eine automatisierte Testfallgenerierung eingesetzt werden.

Ein gutes Traceability Konzept, also die Verknüpfung von Anforderungen, Tests und sonstigen Entwicklungs-Artefakten, ist eine Grundvoraussetzung für erfolgreiches TDRE. Die in Kapitel 4.3 beschriebene Verwaltungsschale kann hierfür als zentraler Zugriffspunkt für Entwicklungs-Artefakte dienen. Erfolgreich angewendet ermöglicht Test-Driven Requirements Engineering eine Reduktion von Doppelarbeit durch eine verbesserte Vollständigkeit, Genauigkeit und Klarheit von Anforderungen. Auch das unzulässige Ausblenden von Problembereichen, etwa durch die Annahme schwer zu rechtfertigender Randbedingungen, kann reduziert werden. Gemäß [18] können durch Test-Driven Requirements Engineering die Durchlaufzeiten und Kosten von Testprozessen um bis zu 30% gesenkt werden.

4.3. Verwaltungsschale als standardisierter Digitaler Zwilling

Die Grundfunktionalität Digitaler Zwillinge liegt in der virtuellen Repräsentation von Assets und der Bereitstellung von untereinander verknüpften domänenspezifischen Modellen [8]. Digitale Zwillinge können für unterschiedliche Zwecke eingesetzt werden [8], in [19] werden unterschiedliche Anwendungszwecke im Rahmen der Verifikation und Validierung analysiert. Wie in Abschnitt 2.3 beschrieben, stellt bereits die Beschleunigung des Informationsaustauschs einen großen Nutzen von Digitalen Zwillingen dar. Die Bereitstellung von Informationen ist eine Grundanforderung und -herausforderung bei Absicherungs- und Testprozessen [6]. Nur wenn alle Informationen rechtzeitig und umfassend vorliegen, kann Testaufwand effizient bewältigt werden.

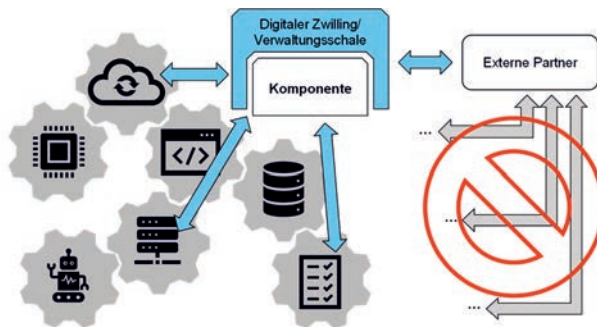


Bild 3: Digitale Zwillinge betten sich bestehende Unternehmens-IT ein und bündeln mehrere Schnittstellen, was die Zusammenarbeit mit externen Partnern erleichtert

Die Hauptaufgaben eines Digitalen Zwillings liegen in der domänenübergreifenden, Gegenstands-individuellen (instanzspezifischen) Informationsbereitstellung und Unterstützung beim Lifecycle Management. Hierzu betten sich Digitale Zwillinge, wie in Bild 3 dargestellt, in die bestehende Unternehmens-IT [9] ein und ermöglichen eine neue instanzspezifische Repräsentation. Insbesondere bei der Zusammenarbeit und Kommunikation mit externen Partnern kann dann ein Digitaler Zwilling als zentrale Schnittstelle dienen. Dadurch können Schnittstellen zu verschiedenen IT-Systemen gebündelt und die Kommunikation mit externen Partnern vereinfacht werden.

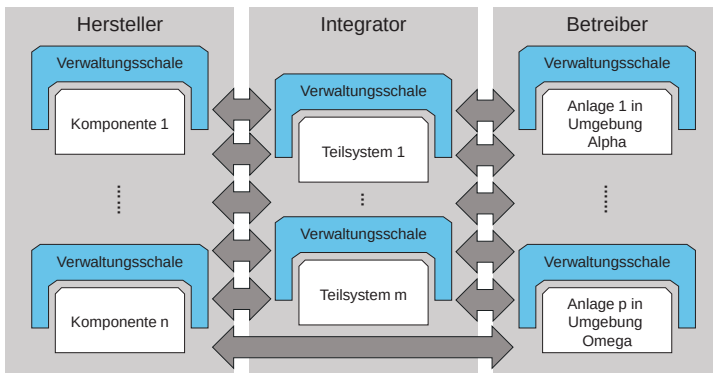


Bild 4: Hersteller, Integratoren und Betreiber nutzen eigene Digitale Zwillinge

Die Verwaltungsschale (englisch Asset Administration Shell) [10] wurde als grundlegendes Konzept für die Realisierung von Digitalen Zwillingen entwickelt. Wie in Bild 4 dargestellt, stimmen Hersteller, Integratoren und Betreiber ihre jeweiligen Digitalen Zwillinge auf ihre Anforderungen und Anwendungen ab. Während herstellerseitig vor allem die Verwaltung und übersichtliche Darstellung von Spezifikationselementen und Einsatzszenarien im Fokus steht, dienen Digitale Zwillinge auf Betreiber-Seite zur Vereinfachung der Kommunikation in Wartungsphasen und im Fehlerfall. Digitale Zwillinge, die sich am Ansatz der Verwaltungsschale orientieren, nutzen bestehende Standards und stellen standardisierte Submodelle zur Verfügung. Durch die Standardisierung der Inhalte und Schnittstellen ermöglicht die Verwaltungsschale den einfachen Informationsaustausch zwischen den Digitalen Zwillingen sowie zwischen Ihren Herstellern, Integratoren und Betreibern.

Für den in Kapitel 3 vorgestellten Open Automation Demonstrator wurde für einen Flüssigkeitstank eine Verwaltungsschale erstellt. Hierin wurden, wie in Bild 5 dargestellt, UML-Modelle gespeichert, welche die Funktionsweise des Tanks beschreiben, und Testfälle, welche

der Überprüfung des durch den Tank angebotenen Service „Temperieren“ dienen. Die Verwaltungsschale der Komponente Tank würde im Falle eines realen Produkts vom Hersteller erstellt und könnte sowohl Integratoren als auch Betreibern zur Verfügung gestellt werden. Mit den hinterlegten Informationen ist eine Überprüfung der Einhaltung von Standards sowie mittels Ausführens der Testfälle eine Überprüfung der korrekten Funktion möglich. Die Verwaltungsschale wurde mit dem Open Source Tool „AASX Package Explorer“ [20] erstellt.

4.4. Testautomatisierung mittels Modellbasiertem Testen erhöhen

Ein Synergieeffekt durch die in Kapitel 4.1 beschriebene Modularisierung nach MTP ist die erleichterte formalisierte Beschreibung der Funktionsweise von Komponenten und deren Services. Die durch ISA 88 und VDI/VDE/NAMUR 2658 ermöglichte standardisierte Beschreibung von Zustandsautomaten und Bezeichnern unterstützt bei der initialen Modellierung, da existierende Modelle anderer Komponenten zu großen Teilen übernommen und wiederverwendet werden können. Die in Bild 5 gezeigten UML-Modelle beschreiben den Service „Temperieren“ des in Kapitel 3 vorgestellten Open Automation Demonstrators. Durch die standardisierte Beschreibung der Services ist eine hohe Übertragbarkeit der erstellten UML-Modelle auf andere Services und Anlagen gewährleistet.

Eine verbesserte Modularisierung und Standardisierung sorgen für eine größere Verfügbarkeit von Modellen und damit Synergieeffekte im Bereich des modellgetriebenen Engineerings. Modelle erleichtern die Kommunikation zwischen Entwicklerteams sowie über Unternehmensgrenzen hinweg, da zwischen nicht formalisierten Anforderungen und fertiger Software bzw. Hardware durch Anforderungs-, Blockdefinitions- und Zustandsdiagramme weitere Beschreibungsformen auf verschiedenen Abstraktionsniveaus verfügbar sind.

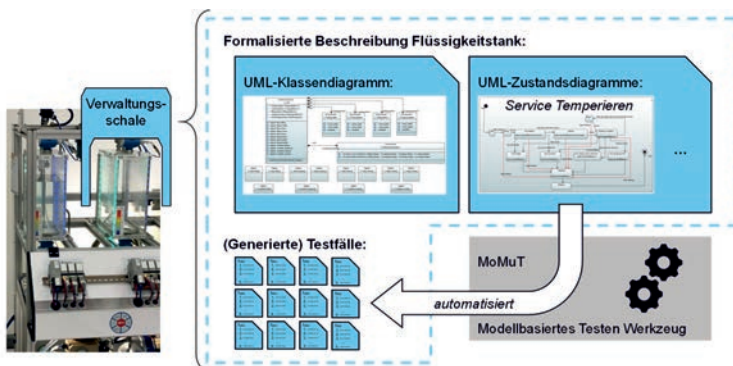


Bild 5: Verwaltungsschale des Flüssigkeitstanks umfasst Modelle und die daraus automatisiert generierten Testfälle

Modelle können wie in Kapitel 2.4 erläutert für Testautomatisierung eingesetzt werden. Ausgangspunkt bei jeder modellbasierten Testautomatisierung ist die Auswahl eines geeigneten Werkzeugs, welches automatisiert aus Modellen Tests generieren kann. Werkzeuge des modellbasierten Testens unterstützen eine oder mehrere Modellierungssprachen wie etwa UML, BPMN, SysML oder proprietäre, werkzeugspezifische Sprachen, einen Überblick bietet [21]. Als Werkzeug zur automatisierten Testgenerierung wurde zur Erprobung der in dieser Veröffentlichung beschriebenen Inhalte das Open Source Werkzeug MoMuT [22–24] eingesetzt. Die Abkürzung MoMuT steht für Model-based Mutation Testing for UML und bezeichnet inzwischen eine ganze Familie von Werkzeugen zur Testfallgenerierung. MoMuT kann unter anderem direkt aus UML Zustandsdiagrammen automatisch Testfälle generieren. Im Falle des Flüssigkeitstanks des Open Automation Demonstrators wurde für den entsprechend modellierten Service „Temperieren“ mit 17 generierten Testfällen eine vollständige Testabdeckung erreicht. Die generierten Testfälle umfassen eine Schritt-für-Schritt-Anleitung zur Testdurchführung mittels der über die grafische Benutzeroberfläche zugänglichen Bedienelemente.

Zur Untersuchung der Potenziale von modellbasiertem Testen mit dem UML-orientierten Werkzeug MoMuT wurden die Funktionen des in Kapitel 3 vorgestellten Open Automation Demonstrators formalisiert beschrieben. Für einen Wassertank wurde der Service „Temperieren“ als UML-Zustandsdiagramm modelliert. Für die Definition der im Zustandsdiagramm genutzten Variablen und Funktionsaufrufe wird zudem ein UML-Klassendiagramm benötigt. Beide Diagramme wurden mit dem Open Source UML-Editor Eclipse Papyrus [25] erstellt und folgen der durch ISA 88 bzw. IEC 61512 vorgegebenen Bezeichnungen. Neben Zuständen des Regelbetriebs wurden auch Fehlerzustände modelliert.

In Bild 5 ist das Vorgehen bei der modellbasierten Testautomatisierung nochmals dargestellt. Der in Kapitel 3 vorgestellte Open Automation Demonstrator verfügt über insgesamt vier Flüssigkeitstanks. Für einen davon wurde, wie in Kapitel 4.3 beschrieben, eine Verwaltungsschale realisiert. Die Verwaltungsschale speichert neben allgemeinen Informationen und Dokumenten auch die für die modellbasierte Testautomatisierung notwendigen UML-Diagramme sowie die daraus generierten Testfälle. Die Modellierung in UML wurde manuell durchgeführt. Da es sich bei dem betrachteten Tank um eine standardisierte, zu Module-Type Package konforme Komponente handelt, besteht der Großteil des Modellierungsaufwands in der Übersetzung von bestehenden Beschreibungsformen in UML. Module-Type Package sorgt für eine bessere Wiederverwendbarkeit von Software, dies trifft auch für die Modellierung mit UML zu, die entstandenen UML Modelle lassen sich effizient für ähnliche Komponenten wiederverwenden.

Das für den Service „Temperieren“ erstellte Zustandsdiagramm folgt den Vorgaben des Werkzeugs MoMuT und kann von diesem eingelesen werden. Vollautomatisiert generiert MoMuT anschließend Testfälle, welche dann ebenfalls in der Verwaltungsschale gespeichert werden. Es lassen sich an diesem Beispiel vier unterschiedliche Ebenen von Testautomatisierung aufzeigen. So kann das Ziel einer Testautomatisierung darin bestehen, vorliegende Dokumentation und Modelle auf Konformität zu Normen und Richtlinien zu prüfen, also die Modelle oder die Dokumentationsinfrastruktur zu überprüfen. Dies stellt die Eingangsstufe in das automatisierte Testen dar.

Dem übergeordnet ist die in diesem Kapitel beschriebene Testautomatisierung. Mit der hier vorgestellten Methodik werden automatisiert Tests für eine einzelne Komponente, einen einzelnen MTP Container, einen einzelnen Service, etc. generiert. Dadurch wird die Testentwurfphase beschleunigt.

Die nächste, dritte Stufe ist die automatisierte Testgenerierung für Anlagen, welche aus mehreren Komponenten bestehen bzw. mehrere MTP Container umfassen und mehrere, voneinander abhängige Services anbieten. Vor einer Testgenerierung müssen dann die Modelle einzelner Komponenten bzw. Services zu einem Anlagenmodell zusammengeführt werden. Methoden der Modellorchestrierung werden in [26, 27] diskutiert.

Stufe 4 richtet sich anschließend an die automatisierte Testausführung. Bei Stufe 2 und 3 werden Testfälle automatisch generiert und müssen manuell abgearbeitet werden, was bei sich wiederholenden Tests ineffizient ist. Für eine Vollautomatisierung des Testens müssen die einzelnen Testfallschritte über Skripte automatisiert ausführbar und auswertbar gemacht werden. Es existieren hierfür unterschiedliche Ansätze, weit verbreitet ist beispielsweise das Keyword-Driven Testing, ein Überblick wird in [5] gegeben. Je nach Testgegenstand muss auch das Testbett, also die Umgebung in der die Tests durchgeführt werden, automatisiert werden. Gängige Konzepte sind dabei „Hardware-in-the-Loop“ (HiL) und „Software-in-the-Loop“ (SiL). Die automatisierte Testausführung und Auswertung ist unabhängig davon, ob Testfälle manuell oder automatisiert erstellt werden. Modellbasierte Testfallgenerierung der Stufen 2 und 3 erzeugt Testfälle mit hoher Systematik hinsichtlich der Überdeckung und Benennung, dadurch ergeben sich Vorteile für die Automatisierung der Testausführung [15].

5. Zusammenfassung und Ausblick

Flexible, vernetzte Produktionsanlagen und ihre Komponenten weisen eine erhöhte Komplexität auf [3, 4]. Aufgrund zusätzlicher Schnittstellen und Betriebsmodi können verschiedene Einsatzzwecke unterstützt und die Flexibilität einer Anlage erhöht werden. Jede zusätzliche Funktion und Abhängigkeit erhöht aber den Engineering- und Testaufwand [5, 6].

Um notwendige Testaufwände auch zukünftig bewältigen zu können, wurden in dieser Veröffentlichung vier Ansätze vorgestellt:

- Die nach Standards orientierte Modularisierung von Komponenten und Anlagen mittels Module-Type Package (MTP) ermöglicht eine Reduktion von Testaufwand. Durch Wiederverwendung steigt die Qualität sowie die Knowhow- und Ersatzteilverfügbarkeit, gemeinsame Standards verhindern die Bildung von Insellösungen.
- Ein Wandel hin zu agileren Test- und Entwicklungsteams verbessert die Kommunikation. Durch die Umsetzung von Test-Driven Requirements Engineering (TDRE), also der Einbindung von Testexperten in frühen Projektphasen, kann die *testability* von Anforderungen gesteigert werden, da Testexperten viel Erfahrung und Fokus auf Grenzwerte und Ausnahmesituationen einbringen. Durch eine Steigerung der Qualität und Testbarkeit von Anforderungen können daraus effizienter Tests abgeleitet und entworfen werden.
- Um den Informationsfluss zu verbessern eignet sich der Einsatz von Digitalen Zwillingen. Die Verwaltungsschale eignet sich zur zentralen Informationsbereitstellung auch über Unternehmensgrenzen hinweg und wirkt dem Problem unterschiedlich aufgefasstem, unvollständiger oder obsoletter Dokumentation entgegen. Ein effizienter Informationsfluss reduziert Wartezeiten oder Unklarheiten beim Testfallentwurf.
- Für eine bessere Bewältigung von Testaufwand kann zudem Testautomatisierung erwogen werden. Es lassen sich die automatisierte Testausführung und Testgenerierung unterscheiden. Modellbasierte Testfallgenerierung hat auch bei selten ausgeführten Tests einen positiven Einfluss auf den Testaufwand.

Jeder Ansatz kann einzeln angewendet werden, zusammengekommen werden zusätzliche Synergieeffekte erzielt. Eine nach Standards durchgeführte Modularisierung erleichtert das Requirements Engineering, sowie die Erzeugung von Digitalen Zwillingen und formalisierter Modelle für Modellbasiertes Testen. Agilität und ein Abbau von Abteilungsgrenzen verbessern das Erkennen von wiederverwendbaren Lösungen sowie das gemeinsame Arbeiten an Dokumentation und Modellen. Digitale Zwillinge fördern die Standardisierung von Inhalten und erleichtern die Informationsbereitstellung in Form von Dokumenten und Modellen. Modelle des Modellbasierten Testens verbessern dank einer gesteigerten Anschaulichkeit und höherem Abstraktionsniveau die Kommunikation zwischen Experten und fördern die Standardisierung. Das Testen vernetzter Komponenten und Systeme ist aufwändiger, durch die Anwendung und Kombination der beschriebenen Ansätze wird die Weiterentwicklung von Testprozessen und ein effizientes und wirkungsvolles Testen im Bereich der flexiblen Produktion erreicht.

Danksagung

Die Autoren bedanken sich bei allen Mitgliedern des VDI/VDE GMA Fachausschusses 7.25 und Veranstaltungsteilnehmern für die Diskussion der Inhalte im Rahmen des VDI Expertenforums „Testen vernetzter Systeme und Komponenten“.

6. Literaturverzeichnis

- [1] J. Bernshausen, A. Haller, T. Holm, M. Hoernicke, M. Obst und J. Ladiges, „Namur Modul Type Package – Definition“, *atp*, Jg. 58, S. 72, 2016, doi: 10.17560/atp.v58i01-02.554.
- [2] H.-P. Wiendahl *et al.*, „Changeable Manufacturing - Classification, Design and Operation“, *CIRP Annals*, Jg. 56, Nr. 2, S. 783–809, 2007, doi: 10.1016/j.cirp.2007.10.003.
- [3] T. Müller, N. Jazdi und M. Weyrich, „Intelligentes Rekonfigurationsmanagement selbstorganisierter Produktionssysteme in der diskreten Fertigung“ (de), *VDI-Kongress AUTOMATION*, 2020, doi: 10.18419/opus-10928.
- [4] T. Müller, N. Jazdi, Jan-Philipp Schmidt und M. Weyrich, „Cyber-Physical Production Systems: enhancement with a self-organized reconfiguration management“ (en), *Procedia CIRP*, 2020, doi: 10.13140/RG.2.2.12562.48323/1.
- [5] M. Weyrich *et al.*, „Testen vernetzter Systeme für die Industrie 4.0“, *atp*, Jg. 60, Nr. 03, S. 26–33, 2018, doi: 10.17560/atp.v58i03.2343.
- [6] *Testen vernetzter I4.0-Systeme: Grobplanung verteilter Testprozesse*, 4004 Blatt 1, VDI/VDE, Apr. 2020.
- [7] D. Méndez Fernández und S. Wagner, „Naming the pain in requirements engineering: A design for a global family of surveys and first results from Germany“, *Information and Software Technology*, Jg. 57, S. 616–643, 2015, doi: 10.1016/j.infsof.2014.05.008.
- [8] B. Ashtari Talkhestani *et al.*, „An architecture of an Intelligent Digital Twin in a Cyber-Physical Production System“, *at - Automatisierungstechnik*, Jg. 67, Nr. 9, S. 762–782, 2019, doi: 10.1515/auto-2019-0039.
- [9] R. Heidel, M. Hoffmeister, M. Hankel und U. Döbrich, Hg., *Industrie4.0 Basiswissen RAMI4.0: Referenzarchitekturmodell mit Industrie4.0-Komponente*, 1. Aufl. Berlin, Wien, Zürich, Berlin, Offenbach: Beuth Verlag GmbH; VDE Verlag GmbH, 2017.
- [10] IDTA, *IDTA - Industrial Digital Twin Association e.V.* [Online]. Verfügbar unter: <https://id-twin.org/> (Zugriff am: 4. Mai 2021).
- [11] C. Brandes, „Testkosten senken, Qualität steigern“, *IT Management*, Juli-August, 2015.

- [12] M. Winter, T. Roßner, C. Brandes und H. Goetz, *Basiswissen modellbasierter Test: Aus- und Weiterbildung zum ISTQB® Foundation Level–Certified Model-Based Tester*, 2. Aufl. Heidelberg: dpunkt.verlag, 2016.
- [13] C. Schulze, D. Ganesan, M. Lindvall, R. Cleaveland und D. Goldman, „Assessing model-based testing: an empirical study conducted in industry“ in *Companion Proceedings of the 36th International Conference on Software Engineering*, Hyderabad India, 2014, S. 135–144, doi: 10.1145/2591062.2591180.
- [14] A. Marques, F. Ramalho und W. L. Andrade, „Comparing Model-Based Testing with Traditional Testing Strategies: An Empirical Study“ in *IEEE Seventh International Conference on Software Testing, Verification and Validation Workshops*, OH, USA, 2014, S. 264–273, doi: 10.1109/ICSTW.2014.29.
- [15] V. Gudmundsson, M. Lindvall, L. Aceto, J. Bergthorsson und D. Ganesan, „Model-based Testing of Mobile Systems – An Empirical Study on QuizUp Android App“, *Electron. Proc. Theor. Comput. Sci.*, Jg. 208, S. 16–30, 2016, doi: 10.4204/EPTCS.208.2.
- [16] M. Krauss et al., „Open Automation Demonstrator für die chemische Industrie“, *atp*, Jg. 62, 6-7, S. 60–67, 2020, doi: 10.17560/atp.v62i6-7.2476.
- [17] C. Ebert, *Systematisches Requirements Engineering: Anforderungen ermitteln, dokumentieren, analysieren und verwalten*, 6. Aufl. Heidelberg: dpunkt.verlag, 2019.
- [18] C. Ebert und R. Ray, „Test-Driven Requirements Engineering“, *IEEE Softw.*, Jg. 38, Nr. 1, S. 16–24, 2021, doi: 10.1109/MS.2020.3029811.
- [19] A. Löcklin, M. Müller, T. Jung, N. Jazdi, D. White und M. Weyrich, „Digital Twin for Verification and Validation of Industrial Automation Systems – a Survey“ in *25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2020, S. 851–858, doi: 10.1109/ETFA46521.2020.9212051.
- [20] M. Hoffmeister A. Orzelski et al., *admin-shell-io*. [Online]. Verfügbar unter: <https://github.com/admin-shell-io> (Zugriff am: 8. April 2021).
- [21] Z. Micskei, *Model-based testing overview, tools and projects*. [Online]. Verfügbar unter: http://mit.bme.hu/~micskeiz/pages/modelbased_testing.html (Zugriff am: 3. Mai 2021).
- [22] *MoMuT – Model-based Mutation Testing*. [Online]. Verfügbar unter: <https://momut.org/> (Zugriff am: 4. Mai 2021).
- [23] W. Krenn, R. Schlick, S. Tiran, B. Aichernig, E. Jobstl und H. Brandl, „MoMut:UML Model-Based Mutation Testing for UML“ in *IEEE 8th International Conference on Software Testing, Verification and Validation (ICST)*, Graz, Austria, 2015, S. 1–8, doi: 10.1109/ICST.2015.7102627.

- [24] R. Schlick, „MoMuT – Eine Transfer-Geschichte über modellbasiertes Testen“ (en), *Lecture Notes in Informatics (LNI)*, S. 195–200, 2020, doi: 10.18420/SE2020_56.
- [25] Papyrus. [Online]. Verfügbar unter: <https://www.eclipse.org/papyrus/> (Zugriff am: 8. April 2021).
- [26] A. Zeller und M. Weyrich, „Composition of Modular Models for Verification of Distributed Automation Systems“, *Procedia Manufacturing*, Jg. 17, S. 870–877, 2018, doi: 10.1016/j.promfg.2018.10.139.
- [27] M. Grochowski *et al.*, „Formale Methoden für rekonfigurierbare cyber-physische Systeme in der Produktion“, *at - Automatisierungstechnik*, Jg. 68, Nr. 1, S. 3–14, 2020, doi: 10.1515/auto-2019-0115.

NAMUR Open Architecture

Security for NOA

Dr.-Ing. **Markus Runde**, BASF, Ludwigshafen;
Dr. rer. nat. **Christian Barth**, **FESTO**, **Denkendorf**;
Dr. rer. nat. **Gerd Wartmann**, WSC, Hockenheim;
Dipl.-Ing. (FH) **Mirko Brcic**, Endress+Hauser, Reinach, Schweiz

Kurzfassung

Die NAMUR Open Architecture hat zum Ziel die Datenoffenheit von Prozessautomatisierungsanlagen zu fördern, ohne Nachteile hinsichtlich Security. Um diesem Anspruch gerecht zu werden sind Maßnahmen zu ergreifen, hinsichtlich Architektur des NOA-Konzepts. Dazu zählen aber ebenso technologische als auch prozedurale Lösungen, die in das NOA Security Konzept einfließen. Grundlage hierfür sind anerkannte Standards der Security die für die Prozessautomatisierung adaptiert werden.

Abstract

The aim of NAMUR Open Architecture is to promote the data openness of process automation systems, without any disadvantages in terms of security. To meet this demand, measures must be taken regarding the architecture of the NOA concept. However, this also includes technological as well as procedural solutions that are incorporated into the NOA security concept. This is based on recognized standards of security that are adapted for process automation.

1. NAMUR Open Architecture – Eine Übersicht

Moderne Methoden der Digitalisierung ermöglichen die Umsetzung verschiedenster Anwendungsfälle, um aus Feldgerätedaten Mehrwerte zu generieren. Der von der NAMUR zusammen mit dem ZVEI entwickelte NOA Ansatz ermöglicht dabei eine effiziente Auslese der Rohdaten über einen von der Prozesskontrolle getrennten Kanal. Der erste Kernaspekte dabei ist ein einheitliches Informationsmodell, durch welches die Daten verschiedener Hersteller über eine Semantik direkt vergleichbar und kombinierbar sind. Der zweite zentrale Punkt ist eine Systemarchitektur, die sicherstellt, dass die Nutzung der Daten keinerlei Rückwirkung auf den Industriellen Prozess hat. Dieses Sicherheitskonzept steht im Fokus des folgenden Beitrags.

1.1. Motivation & Zielsetzung

Einer der Grundgedanken des NOA Konzeptes ist es, kompatibel zu etablierten Automatisierungslösungen zu sein. Auf diese Weise kann auf langjährig erprobte Lösungen zurückgegriffen werden um höchste Ansprüche an Verfügbarkeit, Integrität und Sicherheit zu gewährleisten. Außerdem ermöglicht der Ansatz neben neuen Anlagen auch Bestandsanlagen, die als Langzeitinvestition ausgelegt sind, mit Digitalisierungslösungen anzureichern.

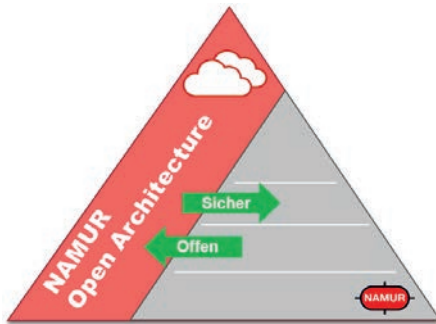


Bild 1: NOA-Basiskonzept mit M+O-Bereich (rosa) und klassischer Automatisierung mit CPC-Bereich [1]

Die klassische Automatisierungspyramide welche die Core-Process-Control Umgebung (CPC) umfasst, wird bei NOA um einen Bereich erweitert, der sich aus IT-nahen Komponenten aufbaut, und die Funktion des Monitoring & Optimization (M+O) übernimmt (siehe Bild 1). Die Schnittstellen zwischen den beiden Umgebungen müssen dabei besondere Anforderungen erfüllen, um zum einen den CPC Bereich vor unerwünschter Einflussnahme zu schützen, und zum anderen die erhobenen Daten standardisiert nutzbar zu machen [2].

1.2. Basisarchitektur

Die Trennung der IT-fokussierten (M+O) und OT-fokussierten Umgebung (CPC) wird im Kern dadurch gewährleistet, dass es eine Datenflusskontrolle gibt um Informationen rückwirkungs-frei aus dem CPC in die M+O-Umgebung zu transferieren. Die Anforderungen dieses NOA-Security Gateways sind in der Namur Empfehlung 177 zusammengefasst [3]. Wichtig ist, dass in Richtung M+O-Umgebung die Feldgerätedaten über das OPC UA Protokoll zur Verfügung gestellt werden. Um Erkenntnisse aus der M+O-Umgebung wieder zurück ins CPC fließen zu lassen, gibt es eine eigene Schnittstelle. Die NOA-Verification-of-Request bestimmt dabei die Anforderungen an diesen Rückkopplungsmechanismus und ist derzeit noch in Arbeit. Einen Überblick über den Informationsfluss innerhalb und zwischen den Umgebungen zeigt die Bild 2.

2.1. NOA Security – Zonen & Schutzziele

Das NOA Security Konzept definiert NOA-Security Zonen. Bild 3 zeigt die NOA Basis Architektur aus Bild 2 in vereinfachter Form und ordnet diese den NOA Security Zonen CPC (Core Process Control, *deutsch Kern-Prozessautomatisierung*) sowie M+O On-Premises und M+O Off-Premises (Monitoring & Optimization, *deutsch Überwachung und Optimierung*) zu.

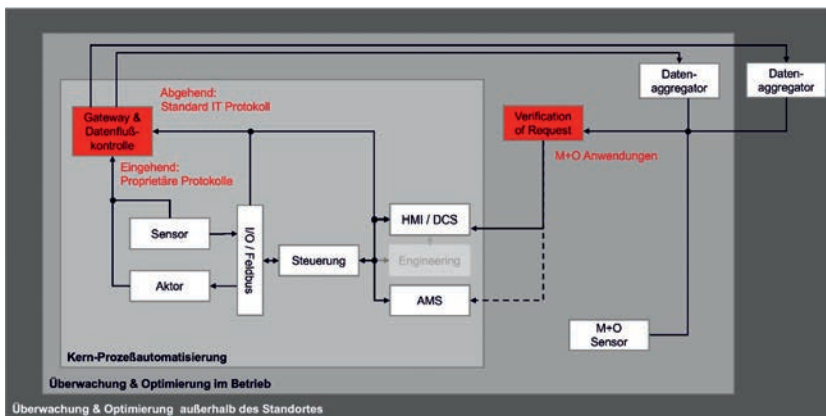


Bild 3: NOA-Security Zonen [4]

Die CPC Security Zone einer Automatisierungsanlage dient der Durchführung des eigentlichen industriellen Produktionsprozesses. Komponenten und Geräte in dieser Zone nutzen häufig proprietäre technologische Lösungen. An die Daten und den Informationsflüssen in dieser CPC Security Zone werden höchste Ansprüche an ihre Verfügbarkeit und Integrität gestellt.

In den M+O Security Zonen erfolgt eine Verarbeitung, Anzeige und Speicherung der Betriebsdaten, die in einer CPC Zone erzeugt werden. Sie werden zur Diagnose und Optimierung der eingesetzten CPC-Komponenten genutzt. Die Verarbeitung der Daten in diesen M+O Zonen erfolgt mit standardisierten IT-Werkzeugen.

In einer M+O Security Zone On-Premises wird in der Regel dem Schutzziel Datenintegrität der höchste Stellenwert eingeräumt. Das Schutzziel der Datenvertraulichkeit steht an zweiter Stelle. In einer zentralen M+O Zone Off-Premises, realisiert zum Beispiel in einer IT-Cloud, wird demgegenüber dem Schutzziele der Vertraulichkeit der Daten ein größerer Stellenwert eingeräumt als der der Datenintegrität. M+O Zonen greifen nicht in das unmittelbare Geschehen einer Automatisierungsanlage ein. Dem Schutzziel der Verfügbarkeit der Daten und Informationen rangiert deshalb in einer M+O Zone an dritter Stelle.

Entscheidend für das Securitykonzept von NOA ist es, dass aller Datenfluss aus der CPC Security Zone in eine M+O Security Zone keine Rückwirkung auf den Betrieb in einer CPC Zone hat. Dieser rückwirkungsfreie Datentransfer wird erreicht durch das NOA Security Gateway (siehe Kapitel 3). Gezielte Optimierungseingriffe aus einer M+O Zone in die CPC Zone müssen kontrolliert über ein getrenntes Gateway mit entsprechender Verifikationsaufgabe (Verification of Request) erfolgen. Dies ist nicht Gegenstand dieses Beitrags.

2.2. Definition von Security-Profilen

Cyber-Security Maßnahmen für industrielle Automatisierungsanlagen werden von dem Standard IEC 62443-3-3 [6] beschrieben. Dieser Standard legt umfangreiche Schutzmaßnahmen für die sieben zentralen IT-Schutzfelder einer Automatisierungsanlage fest, beispielsweise für die Identifizierung und Authentisierung von Zugriffen (siehe Tab. 1). Die jeweiligen Schutzmaßnahmen werden in 4 Stufen, Security Level SL1 bis SL4, eingeteilt, wobei SL4 die Stufe mit den umfassendsten Schutzmaßnahmen ist. Damit trägt man der unterschiedlichen Schwere und Intensität von IT-Angriffen Rechnung. Die Auswahl der Schutzstufen erfolgt auf der Basis einer Risikobetrachtung, bei der der jeweilige Schutzbedarf der betrachteten Sicherheitszone entscheidend ist.

Tabelle 1: Systemanforderungen Identifizierung und Autorisierung [3,6]

SR 1. Identifikation und Autorisierung menschlicher Benutzer		SL 1	SL 2	SL 3	SL 4
		NOA BASIC		NOA EXTENDED	
SR 1.1	Das Automatisierungssystem muss die Fähigkeit haben, alle menschlichen Nutzer zu identifizieren und zu authentifizieren. Diese Fähigkeit muss an allen Schnittstellen, die menschlichen Nutzern Zugriff auf das Automatisierungssystem gewähren, eine Identifizierung und Authentifikation dergestalt durchsetzen, dass die Trennung von Aufgaben und das Prinzip der minimal erforderlichen Rechte entsprechend den geltenden IT-Sicherheitsleitlinien und Vorgehensweisen unterstützt wird.	X	X	X	X
RE(1)	Das Automatisierungssystem muss dazu befähigt sein, alle menschlichen Nutzer eindeutig zu identifizieren und zu authentifizieren.		X	X	X
RE(2)	Das Automatisierungssystem muss die Fähigkeit haben, eine Multifaktor-Authentifikation für den Zugriff aller menschlichen Nutzer auf das System zu verwenden, die über ein nicht vertrauenswürdigen Netz erfolgen.			X	X
RE(3)	Das Automatisierungssystem muss die Fähigkeit haben, eine Multifaktor-Authentifikation für den Zugriff aller menschlichen Nutzer auf das System zu verwenden.				X

Das NOA Securitykonzept fasst die 4 Stufen der IEC Norm in zwei Security Profile NOA BASIC und NOA EXTENDED zusammen: das Profil NOA BASIC (Tabelle 1, blauer Bereich) wird aus

den IEC Schutzmaßnahmen zu den Stufen SL1 und SL2 gebildet, das Profil NOA EXTENDED (roter Bereich) entsprechend aus den Stufen SL3 und SL4.

Die Zuordnung der verschiedenen Schutzmaßnahmen aus der IEC 62443-3-3 zu den Profilen NOA BASIC bzw. EXTENDED wurde hinsichtlich der spezifischen Anforderungen einer Prozessautomatisierung überprüft und in Einzelfällen angepasst. Eine vollständige Liste mit den NOA Schutzmaßnahmen kann der NAMUR Empfehlung, NE 177, entnommen werden.

Das NOA Securitykonzept fordert, dass auch für NOA Komponenten in vergleichbarer Weise zwei Profile NOA BASIC und NOA EXTENDED definiert werden.

3. Aufbau des NOA Security Gateways

Die Bild 4 zeigt das NOA Security Gateway als Verbindung zwischen der CPC-Security Zone und den M+O Security Zonen. Dabei werden Daten von ausgewählten CPC-Geräten (Messgeräte, Prozessanalysatoren, elektrische Equipment usw.) gesammelt und der M+O Zone zur Verfügung gestellt, ohne die Verfügbarkeit, Funktionale Sicherheit und Security zu kompromittieren.

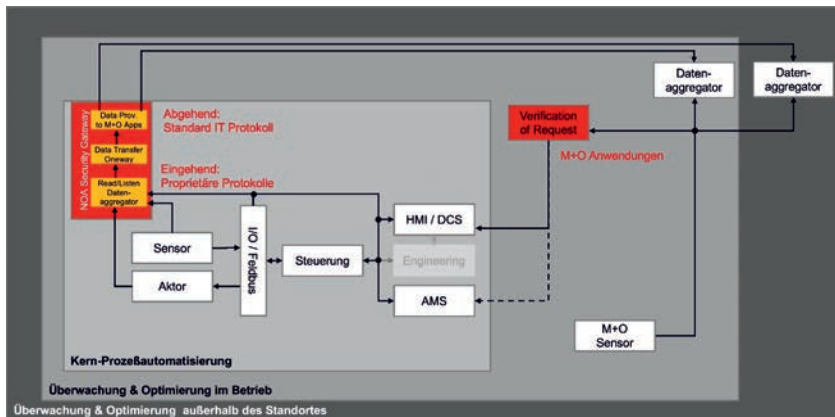


Bild 4: Der Informationsfluss durch ein NOA-Sicherheits-Gateway zwischen CPC und M+O Zonen [3]

Dem NOA Security Gateway obliegt es, die heute noch zahlreich verwendeten proprietären Protokolle aus der CPC Umgebung in ein verständliches, generisches Kommunikationsformat für die M+O Umgebung zu übersetzen.

3.1. Funktionale Anforderungen

Die Kernaufgaben und grundlegenden Security Anforderungen dieses Gateways lassen sich wie folgt zusammenfassen:

1. Das Gateway dient als Kommunikationsbarriere und stellt eine unidirektionale und rückwirkungsfreie Kommunikation vom CPC und in den M+O- Bereich sicher. Sämtliche Kommunikation läuft über dieses Security Gateway. Sobald die Kommunikationskette einmal aufgebaut wurde und in Betrieb ist, darf sie auch nicht mehr verändert oder umgeleitet werden.
2. Die Konfiguration der Feldgeräte sowie ihre Ausleseverhalten auf der CPC-Seite des Security Gateways ist vor direkte Eingriffe aus dem M+O Bereich heraus zu schützen.
3. Die Informationen, die in den M+O-Bereich übertragen werden, müssen dem NOA Informationsmodell [5] folgen.

Diese Kernaufgaben sind entweder in dem NOA Security Gateway integriert oder werden von einer systemischen, mehrkomponentigen Lösung erfüllt. Basierend auf den funktionalen Anforderungen lässt sich das Security Gateway Konzept wie in Bild 5 dargestellt in 3 Module beschreiben.

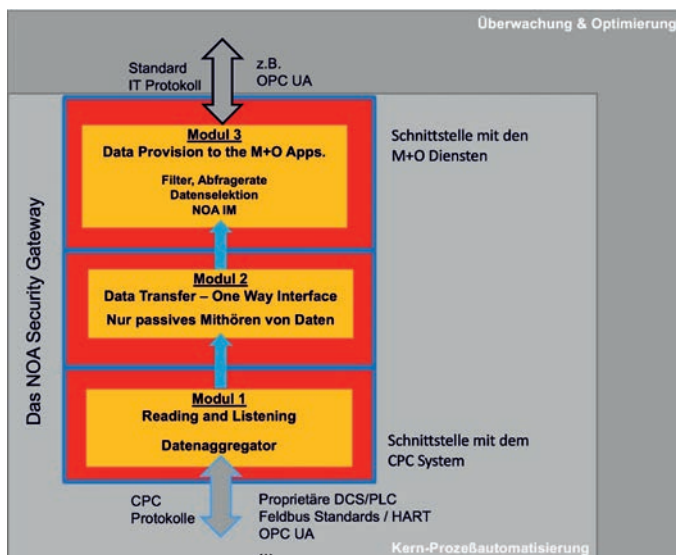


Bild 5: Der funktionelle Aufbau der Grundmodule eines NOA Security Gateways [3]

Das "Reading and Listening" Module 1 auf der CPC Seite sammelt die Daten auf der CPC Seite. Dies kann entweder im „passive listening mode“, Auslesen des Netzwerk-

Datenverkehrs, oder im „active reading mode“, Lesenfragen an das CPC-Netzwerk, erfolgen. Das „Data Transfer“ Module 2 fungiert als unidirektionale Schnittstelle für den reinen Datenverkehr zwischen Module 1, der CPC-Umgebung, und Module 3, der M+O Umgebung.

Das „Data Provisioning“ Module 3 auf der M+O Seite übersetzt die Daten aus der CPC-Umgebung in ein für die M+O Umgebung gebräuchliches IT-Protokoll wie OPC-UA. Dabei kann die Umgebung „On-Premise“ oder „Off-Premise“ sein. Die Datensemantik folgt dabei dem NOA Informationsmodell. Die Daten können direkt vom Gateway zur Verfügung gestellt oder vom NOA Aggregation Server (siehe Bild 5) abonniert werden.

3.2. Technische Anforderungen

Die technischen Anforderungen an das Security Gateway folgen der IEC 62443-4-2 [7] und teilen sich in zwei Kategorien NOA BASIC und NOA EXTENDED auf. Das BASIC Profil entspricht bis auf einigen Anpassungen SL 1 und 2. Wobei das NOA EXTENDED im wesentlichen SL 3 und 4 entspricht. Eine Liste der technischen Anforderungen für die Aspekte Identifikation und Authentifizierung ist in Tabelle 2 dargestellt.

Tabelle 2: Die technischen Anforderungen an das NOA Security Gateway für Identifikation und Authentifizierung [3,7]

Security Requirements and Requirement Enhancements for the CPC Gateway	SL NOA Basic Requirements	Additional Requirements for SL NOA Extended
FR 1 – Identification and authentication control (IAC)		
CR 1.1 – Human user identification and authentication	x	x
RE (1) Unique identification and authentication:	x	x
RE (2) Multifactor authentication for all interfaces		x
CR 1.2 – Software process and device identification and authentication	x	x
RE (1) Unique identification and authentication		x
CR 1.3 – Account management	x	x
CR 1.4 – Identifier management	x	x
CR 1.5 – Authenticator management	x	x
RE (1) Hardware security for authenticators	x	x
NDR 1.6 – Wireless access management	x	x
RE (1) Unique identification and authentication	x	x
CR 1.7 – Strength of password-based authentication	x	x
RE (1) Password generation and lifetime restrictions for human users		x
RE (2) Password lifetime restrictions for all users (human, software process, or device)		x
CR 1.8 – Public key infrastructure certificates		x
CR 1.9 – Strength of public key-based authentication		x
RE (1) Hardware security for public key-based authentication		x
CR 1.10 – Authenticator feedback	x	x
CR 1.11 – Unsuccessful login attempts	x	x
CR 1.12 – System use notification	x	x
NDR 1.13 – Access via untrusted networks	x	x
RE (1) Explicit access request approval		x
CR 1.14 – Strength of symmetric key-based authentication	x	x
RE (1) Hardware security for symmetric key-based authentication		x

4. Lifecycle-Betrachtung für NOA-Security

Allgemein werden Automatisierungssysteme gemäß eines Lifecycle-Modells, z.B. in der Norm IEC 24748, entwickelt und betrieben. Die Ausgangsgrundlage bietet die Spezifikation ihrer Schutzziele und Security Profile (siehe Bild 6). Dies erfolgt anhand eines bewerteten Risikos. Die in diesem Beitrag zuvor genannten Anforderungen zu den NOA Security Profilen basieren auf solch einer Risikoeinschätzung. In der darauffolgenden Design- und Implementationsphase werden ergänzende Rahmenbedingungen geschaffen, die den IT-sicheren Betrieb der Security Zonen aufrechterhalten. Dies setzt sich auch in der Betriebsphase und Außerbetriebnahme fort. Die Verantwortung für diese Vorgehensweise liegt beim Betreiber der Security Zonen bzw. eingebundenen Systemintegratoren.

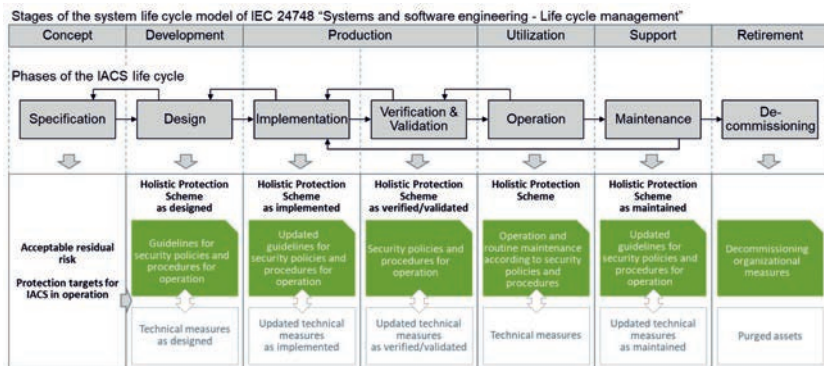


Bild 6: Stufen des Produktlebenszyklus nach IEC62443-2-2 [8]

Auch die Bereitstellung eines NOA Security Gateway durch die Hersteller folgt diesen typischen Stufen eines Lifecycles eines Automatisierungssystem: Spezifikation der im Kapitel 3 beschriebenen Security Profile, Design, Implementation und Verifikation der Security Funktionen des Gateways, sowie Aufrechterhaltung der Security-Funktionalität während der Betriebsphase eines Gateways.

Diese Vorgehensweise unterstreicht, dass sowohl Anwender als auch Hersteller und Integratoren eine gemeinsame Verantwortung für den IT-sicheren Betrieb eines Automatisierungssystems haben. Die Umsetzung der Securityanforderungen für die NOA Security Zonen allein zum Zeitpunkt der Inbetriebnahme einer Anlage führen zu keiner langfristig stabilen IT-sicheren Lösung.

Die Erfahrung aus der Praxis zeigt, dass ohne kontinuierlichen Informationsaustausch zwischen Betreibern, Integratoren und Herstellern bei der Implementation von Automatisierungs-

oder von Produktlösungen sich neue Security-Schwachstellen ergeben können. Bild 7 skizziert die Vorgehensweisen, diesem Problem entgegenzuwirken.

Im oberen Bereich wird der gegenseitige Austausch von Informationen während der verschiedenen Phasen der Produktlebenszyklen dargestellt. Die Betreiber und Integratoren von Automatisierungsanlagen formulieren ihre Systemanforderungen. Die Komponentenhersteller stellen detaillierte Dokumentationen bereit [9].

Im unteren Bereich ist ein vergleichbarer Informationsaustausch im Zuge des Security-(Asset-) Lifecycle nach VDI 2182 [10] dargestellt.

Die fast identische Darstellung beider Informationsprozesse zeigt, dass die Wichtigkeit abgestimmter Zusammenarbeit der verschiedenen Akteure erkannt ist, um effiziente, stabile und IT-sichere Produkte und Anlagen zu gewährleisten.

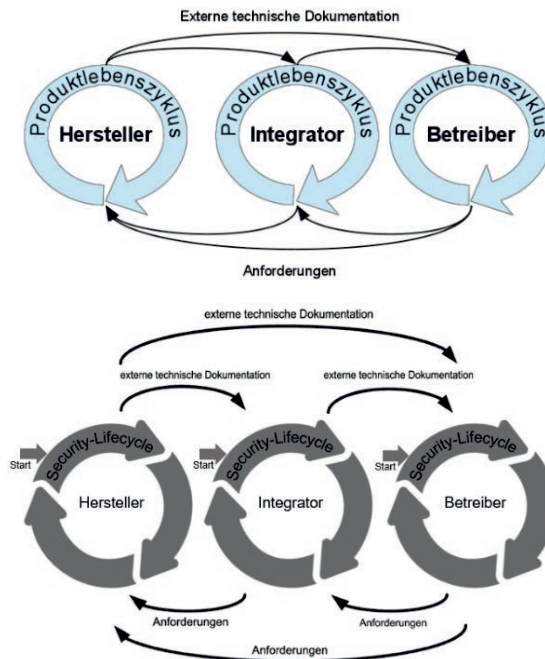


Bild 7: Gegenüberstellung | Security- und Produkt-Lifecycle

4.1 Information Security Management für NOA

Der Betrieb eines Cyber/Information Security Management System (CSMS/ISMS), wie er in der IEC 62443-2-1 [11] und der NE 169 [12] beschrieben ist, sieht zahlreiche Aufgaben im Security-Lifecycle vor. Hierzu zählen Themen wie Risk-Management und -Mitigation sowie Asset- und Vulnerability-Management. Ein Asset-Management umfasst die Sammlung-, Verwaltung und Auswertung aller Assets eines Automatisierungssystem. Das Vulnerability-Management nutzt, wie zahlreiche andere Prozesse auch, diese dokumentierten Assets für eine systematische Auswertung der extern bzw. durch die Hersteller bereitgestellter Vulnerability-Informationen und versetzt damit den Asset-Owner in die Lage, angemessene (Gegen-)Maßnahmen gezielt zu ergreifen.

Auch beim Asset Management ist ein wirksames Zusammenspiel zwischen Hersteller, Integratoren und Anwender nötig. So wird bspw. nach der Installation einer Komponente ein dauerhafter IT-sicherer Betrieb nur ermöglicht, wenn ein Hersteller entsprechende update-Dienste für „seine“ Komponenten als Serviceleistungen kontinuierlich vorsehen kann.

5. Zusammenfassung & Ausblick

Die Namur Open Architecture, kurz NOA, beschreibt ein Konzept zur Auslese und Verarbeitung von Feldgerätedaten. Der Kern von NOA ist eine Trennung des Prozessautomatisierungsbereiches (CPC) und dem Beobachtungs- und Optimierungsbereich (M+O). Damit wird es beispielsweise auch Bestandsanlagen ermöglicht, diese mit cloudbasierten Applikationen zu überwachen und zu optimieren. Die Berücksichtigung der damit einhergehenden Securityanforderungen, war von Anfang an ein fester Bestandteil der Ausarbeitungen.

Die zentralen Sicherheitsanforderungen werden in der NE177 in für den Anwender einfacher Weise dargelegt. Dabei orientieren sich die Forderungen an etablierten Standards wie der IEC 62443. Die beiden auf die Anforderungen der Prozessautomatisierung zugeschnittenen Profile: NOA Basic und NOA Extended definieren dabei die Mindestanforderungen, beziehungsweise Erweiterte Anforderungen für kritische Anlagen.

Die Kommunikationsschnittstelle zwischen CPC und M+O Bereichen wird mit dem NOA Security Gateway besonders geschützt. Neben funktionalen Anforderungen ist auch eine Beispielimplementierung skizziert. Die strengen Security Anforderungen an diesem Übergang erlauben es den CPC Bereich unbeeinflusst operieren zu lassen, und unterstützen damit eine schrittweise Digitalisierung der Anlage. NOA zeichnet sich dadurch als kosteneffizienter und nachhaltiger Ansatz aus.

Die Prozesse auf allen Seiten der genannten Parteien, vom Hersteller bis hin zum Anwender, müssen sicherstellen, dass die Anwendung des NOA-Konzepts diesen Vorgaben und Beispielen folgt.

Quellen

- [1] C. Klettner, T. Tauchnitz, U. Eppe, L. Nothdurft, C. Diedrich, T. Schröder, D. Großmann, S. Banerjee, L. Urbas and C. Iatrou, "NOA – NAMUR Open Architecture - Die NAMUR-Pyramide wird geöffnet für Industrie 4.0," atp edition 1-2 / 2017, pp. 20-37.
- [2] NAMUR-Recommendation, "NE 175 - NAMUR Open Architecture - NOA Concept", namur.net/en, 2020.
- [3] NAMUR-Recommendation, "NE 177 - NAMUR Open Architecture - NOA Security Zones and Security Gateway", namur.net/en, 2020.
- [4] IEC 62443, Normeneihe "Industrial communication networks – Networks and system security"
- [5] NAMUR-Recommendation, INTERNAL DRAFT "NE 176 - NAMUR Open Architecture – NOA Information Model", namur.net/en, 2020.
- [6] IEC 62443-3-3:2013 Industrial communication networks - Network and system security – Part 3-3: System security requirements and security levels," webstore.iec.ch, 2013
- [7] IEC 62443-4-2:2019 Security for industrial automation and control systems - Part 4-2: Technical security requirements for IACS components," webstore.iec.ch, 2019
- [8] IEC 62443-2-2:2020, "Security for industrial automation and control systems Part 2-2: IACS security program rating" CD, iec.com, 2020.
- [9] BSI, „Sicherheitsspezifische Empfehlungen für Maschinenbauer und Integratoren – Empfehlung: Integratoren, Maschinenbauer“, <https://www.allianz-fuer-cybersicherheit.de/>, 2020.
- [10] VDI-2182, Blatt 1, „Informationssicherheit in der industriellen Automatisierung - Allgemeines Vorgehensmodell", vdi.de, 2020-01.
- [11] IEC 62443-2-1:2020, "Security for industrial automation and control systems Part 2-1: Industrial automation and control system security management system" CD, iec.com, 2015.
- [12] NAMUR-Recommendation, „NE 169 - Automation Security Management in der Prozess-industrie“, namur.net/en, 2019.

Ontologiebasiertes Security Management

Ontology based Security Management

Christian Siegwart M.Sc., **Felix Scherhag** M.Sc.,

ZeMA gGmbH, Saarbrücken;

Kai Pervözl, Fraunhofer IAIS, Sankt Augustin,

Univ.-Prof. Dr.-Ing. **Georg Frey**,

Universität des Saarlandes, Saarbrücken

Kurzfassung

Die Digitalisierung und der stark zunehmende Einsatz digitaler Technologien liefern derzeit einen großen Beitrag bei der Umstrukturierung der Unternehmensprozesse. Die stark zunehmende Vernetzung stellt die Unternehmen jedoch wiederum vor zusätzliche Herausforderungen. Das Bewusstsein für Anforderungen und Maßnahmen im Kontext Informationssicherheit wächst zwar kontinuierlich, es besteht jedoch weiter Nachholbedarf bei den Unternehmen, insbesondere bei kleinen und mittleren Unternehmen (KMU). In diesem Beitrag wird ein ontologiebasiertes Security-Management vorgestellt, das es ermöglicht nicht nur die relevanten Informationen zu beschreiben, sondern auch die Beziehungen zwischen den Informationen maschinenlesbar zu repräsentieren. Für den Anwendungsfall Security Management für KMU wird auf diese Weise eine Wissensbasis für IT-Sicherheitsinformationen erstellt.

Abstract

Digitalisation and the rapidly increasing use of digital technologies are currently making a major contribution to the restructuring of business processes. However, the increasing networking leads to additional challenges for the companies. Although awareness of requirements and measures in the context of IT security is growing continuously, further improvement is still required for companies, especially small and medium-sized enterprises (SMEs). In this paper, an ontology-based security management is presented that allows not only to describe the relevant information, but also to represent the relationships between the information in a machine-readable way. In this way, a knowledge base for IT security information is created for the use case of security management for SMEs.

1. Einführung und Motivation

In der jüngsten Vergangenheit erleben wir einen deutlichen Anstieg der Bedrohungen durch Cyberkriminalität. Komplexe Angriffe entwickeln sich auch für Industrieunternehmen zu einem immer größeren Risikofaktor. Nach [1] werden die Risiken durch Cyberkriminalität weltweit als eines der größten Unternehmensrisiken gelistet. Neben der großen Gefahr durch zunehmende Angriffe stellt die Digitalisierung die Unternehmen zukünftig vor große Herausforderungen. Denkt man an den Bereich mobiles Arbeiten, so zeigt sich, dass die aktuelle Corona-Krise bereits heute zu einem deutlichen Digitalisierungsschub führt, der sich in Zukunft noch weiter verstärken wird. Betroffen sind dabei alle Unternehmen, jedoch haben kleine und mittlere Unternehmen besonders mit den Herausforderungen im Kontext IT-Security zu kämpfen. Die Studie [2] zeigt deutlich die Hindernisse für die Unternehmen auf dem Weg zur Erhöhung der Informationssicherheit im Unternehmen. Zeitmangel, zu hohe Kosten, ein Mangel an Knowhow oder gar das Fehlen des Bewusstseins für die Themen rund um IT-Sicherheit sind in diesem Zusammenhang zu nennen.

In [3] wurde bereits gezeigt wie ein ganzheitliches Reifegradmodell in Verbindung mit existierenden Standards (z.B. BSI IT-Grundschutz, IEC 62443) die Bewertung des Reifegrades im Kontext Security unterstützen kann. Die Informationen der Standards sind jedoch oft sehr umfangreich, komplex und treffen nicht die Bedürfnisse insbesondere der kleinen und mittleren Unternehmen. Daher ist es zwingend notwendig, alle relevanten Informationen im Kontext IT Security miteinander zu verknüpfen, um sie bestmöglich für die Unternehmen nutzbar zu machen. Für die Nutzung im Forschungs- bzw. Transferprojekt Grundschutz^{PLUS} Aktivator [4] werden neben den IT-Security-Informationen auch die Unternehmensinformationen betrachtet, um mit der entwickelten Wissensbasis bestmöglich das Informationssicherheitsniveau der KMU zu verbessern.

2. Grundlagen

Dieser Abschnitt zeigt zunächst die verwendeten Grundlagen, die für diesen Beitrag relevant sind. Dazu zählen eine kurze Darstellung des Forschungs- und Transferprojektes Grundschutz^{PLUS} Aktivator, die verwendeten anerkannten Standards im Bereich Informationssicherheit und eine kurze Darstellung, warum Ontologien eine vielversprechende Möglichkeit zur Wissensrepräsentation des Security Managements mit Fokus auf KMU bieten.

2.1 Grundschutz^{PLUS} Aktivator

Das vom BMWi geförderte Projekt Grundschutz^{PLUS} Aktivator [4] wird in Kooperation zwischen dem Fraunhofer Institut für intelligente Analyse- und Informationssysteme (IAIS) und

dem Zentrum für Mechatronik und Automatisierungstechnik (ZeMA) durchgeführt. Das Projekt hat zum Ziel, das IT-Sicherheitsniveau produzierender KMU zu erhöhen und Hürden in Bezug auf die Umsetzung von IT-Sicherheitsmaßnahmen abzubauen. Dazu wird eine interaktive Plattform geschaffen, die KMUs in die Lage versetzt, kompetente IT-sicherheitsrelevante Entscheidungen selbst zu treffen und die das Verständnis für Sicherheitsfragestellungen stärkt. KMUs erhalten speziell auf ihr Unternehmen zugeschnittene IT-Sicherheitskonzepte und Hilfestellung bei der Umsetzung.

2.2 Verwendete Standards

Für den Aufbau der Wissensbasis sind Informationssicherheitsinformationen von zentraler Bedeutung. Für den Kontext Informationssicherheit produktionsnaher KMU stehen in diesem Beitrag die Standards BSI IT-Grundschutz und die internationale Normenreihe IEC 62443 für den Produktionsbereich im Vordergrund.

BSI IT-Grundschutz

Der IT-Grundschutz des Bundesamtes für Sicherheit in der Informationstechnik (BSI) hat das Ziel mit einem standardisierten, ganzheitlichen Ansatz ein angemessenes Maß an Informationssicherheit in Unternehmen zu gewährleisten. Er besteht zum einen aus den BSI-Standards und zum anderen aus dem IT-Grundschutz Kompendium ([5]). Die Standards liefern Informationen zum methodischen Vorgehen. Das IT-Grundschutz-Kompendium stellt, als modular aufgebautes Nachschlagewerk, das zentrale Werkzeug des IT-Grundschutzes dar. Derzeit (Stand: Februar 2021) beinhaltet das Kompendium 97 IT-Grundschutz-Bausteine, die in zehn Bereiche aufgeteilt sind und sich mit unterschiedlichsten Aspekten der Informationssicherheit beschäftigen.

Jeder Baustein enthält jeweils eine Beschreibung der betrachteten Thematik, Vorgehensweisen und IT-Systeme, gibt einen Überblick über spezifische Gefährdungen und beschreibt detailliert standardisierte Sicherheitsanforderungen, die es zu beachten gilt.

Einen Sicherheitsprozess nach IT-Grundschutz im Unternehmen zu etablieren ist vollständig kompatibel zum Standard IEC/ISO 27001 [6]. Daher dient das IT-Grundschutz-Kompendium als Prüfgrundlage für Zertifizierungen nach ISO/IEC 27001 auf Basis von IT-Grundschutz.

IEC 62443

Die Normenreihe IEC 62443 „Industrial communication networks – Network and system security“ umfasst 14 Teile, die in die vier Bereiche General, Policies & Procedures, System und Component/Product unterteilt sind. Die IEC 62443 hat sich als Branchenstandard im

Bereich der industriellen Automatisierung etabliert und wird in Deutschland durch das Spiegelgremium DKE UK 931.1 „IT Sicherheit in der Automatisierungstechnik“ betreut. Die Norm hat die klassischen Security Schutzziele Vertraulichkeit, Integrität sowie insbesondere die Verfügbarkeit von Industrial Automation and Control Systems, kurz IACS, im Fokus. Darüber hinaus stellt die Einhaltung von Anforderungen der funktionalen Sicherheit (functional safety) ein weiteres Schutzziel der IEC 62443 dar. Wichtig ist dabei, dass diese Ziele nur mit einem ganzheitlichen Security-Ansatz, der sowohl Hersteller als auch Integrator und Betreiber umfasst, erreicht werden können. Besonders relevant für die Wissensbasis und damit auch für diesen Beitrag sind die Anforderungen aus Integrator- und Betreibersicht an die Automatisierungssysteme aus dem Normenteil 3-3 ([7]).

2.3 Ontologien

Ontologien bieten eine Methode der Wissensorganisation, die zunehmend den Einsatz in Anwendungen findet. Ursprünglich stammt der Begriff der Ontologie aus der Philosophie und ist dort eng mit der „Lehre des Sein“ und dem Bereich der Metaphysik verbunden. Neben den Anwendungen in Philosophie und Psychologie ([8]) werden Ontologien (ontologies) häufig in der Informatik eingesetzt, um Wissensbereiche abzubilden. Die bekannteste und am meisten zitierte Definition stammt von Tom Gruber aus dem Jahr 1993 ([9]) und beschreibt eine Ontologie folgendermaßen:

„an ontology is an explicit specification of a conceptualization“.

Die Definition wurde in zahlreichen weiteren Publikationen weiter ergänzt (z.B. [10], [11]), sodass bis heute keine endgültige und eindeutige Definition existiert. Ziel ist jedoch über alle Definitionen hinweg das Wissen eines bestimmten Anwendungsbereichs zu modellieren (conceptualization) und in maschinenlesbarer Form zu repräsentieren (explicit specification). Um ein Wissensmodell für eine bestimmte Anwendung (Application ontology) zu erstellen, müssen zum einen die Terminologie und auch die Zusammenhänge des Anwendungsgebietes modelliert werden. Zum anderen muss dieses Wissen über Ontologiesprachen (z.B. Resource Description Framework RDF, Web Ontology Language OWL) formal beschrieben werden, um die Maschinenlesbarkeit, Austauschbarkeit und letztendlich Verwendung als intelligentes Wissensmodell zu gewährleisten. Im Rahmen dieses Beitrags besteht eine Ontologie mit Fokus auf den Anwendungsbereich aus der Definition von Klassen (engl. classes, manchmal auch Konzepten (engl. concepts)), Eigenschaften (engl. properties) der Klassen und Einschränkungen (engl. restrictions). Werden zusätzlich zur Ontologie konkrete Instanzen der Klassen gebildet, so spricht man insgesamt von einer Wissensbasis.

Besonders im Kontext wissensbasierter Anwendungen entwickelt sich das Thema Wissensmanagement zu einem zentralen Erfolgsfaktor für die Unternehmen. Daher sind Wissensbasen unverzichtbar, da sie trotz des großen Aufwandes in der Erstellung durch die zusätzlichen semantischen Informationen des Anwendungsbereichs und logikbasierten Ansätze große Vorteile gegenüber der Nutzung klassischer Datenbanken aufweisen.

2.4 Ontologien im Kontext Security Management

Ontologien in der Informationstechnik kommen stets in abgegrenzten Bereichen (domain) zum Einsatz. Dieser Beitrag hat zum Ziel die Informationen im Kontext Informationssicherheit kleiner und mittlerer Unternehmen (siehe Projekt Grundschatz^{PLUS} Aktivator) abzubilden. Ein besonderer Fokus liegt auf der Abbildung der vier folgenden Bereiche und der Zusammenhänge zwischen ihnen:

- Unternehmen (z.B. Geschäftsprozesse, Assets, Kritikalität, ...)
- IT-Security-Anforderungen (z.B. Anforderungen aus BSI IT-Grundschatz, Anforderungen aus IEC62443, ...)
- IT-Security-Maßnahmen (z.B. Umsetzungshinweise, Best Practices, Umsetzungsdokumente, ...)
- IT-Security-Quellen (Bedrohungen, Standards, ...)

Ein großer Vorteil beim Einsatz von Ontologien ist die Austauschbarkeit und Wiederverwendung von bereits existierenden Modellen. Daher sollen im Folgenden bestehende Ontologien im Kontext Security Management näher betrachtet werden. In [12] werden 29 Ontologien im Kontext Security näher betrachtet und nach Einsatzzweck klassifiziert. Besonders interessant für diesen Beitrag sind die Ontologien, die sich auf Security-Anforderungen fokussieren. In [13] wird eine Taxonomie für Security Anforderungen mit Definitionen diskutiert. [14] zeigt die Nutzung einer Ontologie mit dem Fokus der Entwicklung sicherer Online-Behördenanwendungen. Etwas näher am Fokus dieses Beitrags sind die erstellten Security Ontologien in [15] und [16]. Bedrohungen, Assets, Anforderungen und Maßnahmen sind Bestandteil dieser beiden Ontologien, jedoch wird die Abbildung der Security-Informationen aus den beiden Standards BSI IT-Grundschatz und IEC 62443 sowie die Abbildung der IT-Services in den Unternehmen nicht berücksichtigt. Diese Verknüpfung ist für die Zielgruppe der kleinen und mittleren Unternehmen jedoch unerlässlich, um eine möglichst einfache Zuordnung der relevanten IT-Security-Anforderungen zu gewährleisten. [17] modelliert die Inhalte des BSI IT-Grundschatzes, jedoch gehen die Inhalte auf die alte, stark veränderte, Version des BSI IT-Grundschatzes zurück und die Abbildung der Unternehmensinformationen wird nicht betrachtet.

Mit Blick auf die Prioritäten der zu entwickelnden Anwendung liefert keine der betrachteten Security Ontologien eine passende Grundlage für die Wissensbasis. Die analysierten Informationen geben aber wertvolle Hinweise zur Erstellung einer passenden Ontologie, die in den folgenden Abschnitten genauer beschrieben wird.

3. Vorgehensweise

Die Entwicklung und Erstellung einer Ontologie ist mit großem Aufwand verbunden. Dies liegt daran, dass die Abbildung eines begrenzten Wissensbereichs von der Erfahrung und dem Wissen des Entwicklers der Ontologie abhängig ist. Jeder Entwickler beschreibt Terminologien anders und modelliert die strukturellen Eigenschaften des Anwendungsbereichs unterschiedlich. Die Herausforderung liegt darin, eine für die Anwendung sinnvolle und möglichst allgemeine und somit wiederverwendbare Wissensrepräsentation zu schaffen.

Für die Ontologieerstellung existiert keine fest definierte Vorgehensweise, jedoch gibt es zahlreiche Empfehlungen. Für die Entwicklung der Security Management Ontology (SMO), die in diesem Beitrag beschrieben wird, sind drei Schritte von besonderer Bedeutung (siehe Bild 1).

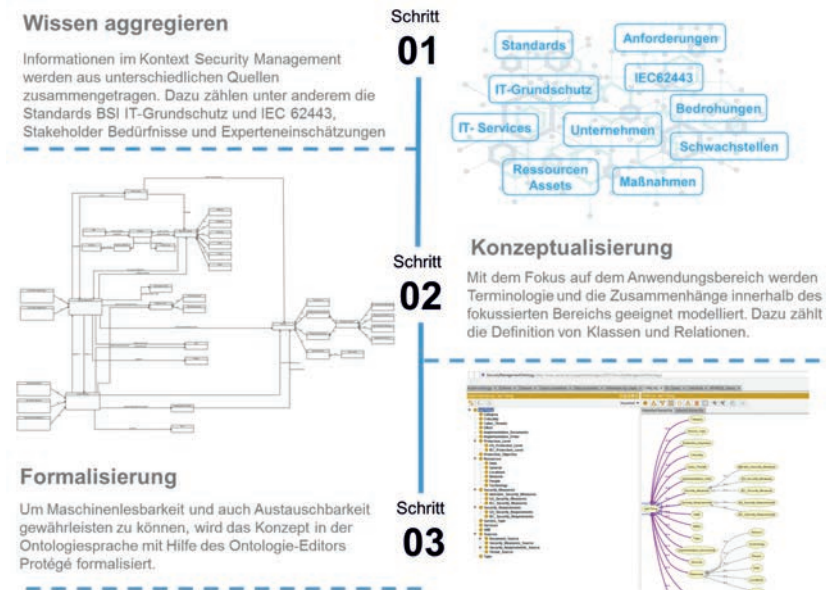


Bild 1 Vorgehensweise zur Realisierung der Security Management Ontology (SMO)

Zunächst ist es essentiell die Informationen des Anwendungsbereichs zu aggregieren. Dazu zählen im Besonderen die Informationen, die im Abschnitt „Grundlagen“ kurz beschrieben sind. Im folgenden Schritt werden die zentralen Konzepte und die strukturellen Zusammenhänge beschrieben, ehe diese mit Hilfe einer Ontologiesprache und einem Ontologieeditor formalisiert werden.

Die Konzeptualisierung und Formalisierung sind in der Vorgehensweise angelehnt an die Prozesse in [18] und [19]. Dazu zählen die im Folgenden beschriebenen Prozessschritte, die iterativ durchgeführt werden. Der iterative Prozess ermöglicht die kontinuierliche Verbesserung, Anpassung und Erweiterung der Ontologie.

3.1. Schritt 1: Wissen aggregieren

a) Scope

Wie bereits beschrieben, ist es aufgrund der Komplexität notwendig den Anwendungsbereich der zu entwickelnden Ontologie einzuschränken und zu fokussieren. Dabei stellt sich die Frage wie ein bestimmter Anwendungsbereich abzubilden ist und welche Aspekte nicht abgebildet werden sollen, um die Komplexität zu reduzieren. In diesem Beitrag steht das Security Management kleiner und mittlerer Unternehmen im Fokus. Alle wesentlichen Aspekte dieses Anwendungsbereichs werden daher zunächst im Scope definiert. Im bisherigen Verlauf des Grundschatz^{PLUS} Aktivator Projektes wurden bereits KMU befragt, welche Erwartungen sie an ein entsprechendes Tool stellen. Der wichtigste Punkt in diesem Zusammenhang ist das Reduzieren der Komplexität und auch die Unterstützung der KMU in der Nutzung. Auch diese wichtigen Informationen fließen in das Design und die Auswahl des Scope für die Security Management Ontologie ein.

Ein oft verwendetes Hilfsmittel zur Fokussierung auf den entsprechenden Anwendungsbereich stellen Competency Questions dar. Das heißt es werden Fragestellungen formuliert, die später mit Hilfe der Ontologie beantwortet werden sollen. Dieses Vorgehen hilft dabei den Fokus auf das Ziel zu legen und die Erwartungen an das Informationsobjekt zu beschreiben. Nach erfolgreicher Erstellung der Ontologie sind die Competency Questions ebenfalls sehr gut dazu geeignet, die Ontologie zu überprüfen. Im Folgenden ist ein Auszug der Competency Questions dargestellt, die für die Fokussierung auf den Bereich Security Management von KMU erstellt wurden:

- Q1: Welche Security Anforderungen sind relevant für die Cyberbedrohung Schadprogramme?
- Q2: Welche Ressourcen gehören zum Service Buchhaltung?
- Q3: Welche Anforderungen aus den IT-Grundschatz Bausteinen sind relevant für die Bedrohung des unbefugten Eindringens in Räumlichkeiten?

b) Terminologie

In diesem Schritt werden aus dem aggregierten Wissen die für die spätere Anwendung zentralen Begrifflichkeiten ermittelt. Im Fokus dieses Beitrages sind daher die Informationen aus den Standards BSI IT-Grundschutz und IEC 62443, Informationen aus bereits bestehenden Ontologien, Stakeholder Bedürfnisse, Experteneinschätzungen und die Inhalte der bereits erstellten Competency Questions.

3.2 Schritt 2: Konzeptualisierung

a) Klassenhierarchie

Um weiter in Richtung Formalisierung des Wissensbereichs zu gelangen, ist nun ein zentraler Schritt, aus der Terminologie Klassen und eine Klassenhierarchie zu bilden. Im Folgenden wird ein Ausschnitt der zentralen Klassen für die vier Dimensionen (Unternehmen, Anforderungen, Maßnahmen und IT-Security-Quellen) des Anwendungsbereichs kurz beschrieben. Die Klassen und auch Relationen werden in englischer Sprache verfasst, um eine bessere Vergleichbarkeit zu schaffen. Die späteren Inhalte (Instanzen) werden in deutscher Sprache verfasst, da sie für die Nutzung der KMU relevant sind.

1) Unternehmen:

SME: die zentralen Unternehmensinformationen mit Bezug IT-Security

Services: die zentralen IT-Services des KMU

Resources: die mit den Services zusammenhängenden Ressourcen/Assets

Protection Objective: Schutzziele der IT-Services

Criticality: Kritikalität der Schutzziele

2) Anforderungen

Security Requirements: Security-Anforderungen der Standards

Protection Level: Schutzlevel der Security Anforderungen

Implementation Order: Umsetzungsreihenfolge der Anforderungen

3) Maßnahmen

Security Measures: passende Maßnahmen zu den definierten Anforderungen

Implementation Documents: relevante Umsetzungsdokumente zu den Maßnahmen

Effort: Aufwandsabschätzung zu den Maßnahmen

4) IT-Security-Quellen

Cyber Threats: Relevante Security-Bedrohungen

BSI Grundschutz Building Blocks: Inhalte aus den BSI IT-Grundschutz Bausteinen

Requirements IEC 62443 3-3: Anforderungen aus dem Standard IEC 62443 Teil 3-3

b) Relationen

In diesem Schritt werden die Eigenschaften (properties) zu den erstellten Klassen definiert. Dies sind zum einen die Zusammenhänge die Klassen miteinander verbinden (object properties) und zum anderen Eigenschaften, die die einzelnen Klassen/Konzepte näher beschreiben (datatype properties).

Mit Abschluss dieses Schrittes sind die zentralen Begriffe und Zusammenhänge des Anwendungsbereiches definiert und die Konzeptualisierung ist abgeschlossen (Lightweight Ontologie).

3.3 Schritt 3: Formalisierung

Um aus der erstellten Lightweight Ontologie eine Heavyweight Ontologie zu entwickeln, werden die erstellten Inhalte nun formalisiert. Das heißt die Klassen, Subsumptionsbeziehungen der Klassen, Relationen und auch mögliche Einschränkungen werden mit Hilfe von Ontologiesprachen formalisiert. Detaillierte Informationen zu den Ontologiesprachen sowie zur Verwendung von Beschreibungsmustern sind in [18] genauer beschrieben.

Um der Komplexität der Ontologieerstellung gerecht zu werden, wird für die Formalisierung auf den bekanntesten Ontologie-Editor Protégé ([20]) zurückgegriffen. Bild 2 zeigt einen Ausschnitt des Tools, indem die erstellte Klassenhierarchie grafisch dargestellt ist. Protégé ermöglicht es, die Formalisierung der Konzeptualisierung mit einer grafischen Benutzeroberfläche zu realisieren. Der Export und Import der Ontologien ist für unterschiedliche Ontologiesprachen (z.B. OWL/XML, RDF/XML) möglich.

4. Wissensbasis und Validierung

Neben der Formalisierung wird ebenfalls für das Hinzufügen der Instanzen, also konkreter Elemente der erstellten Klassen, das Tool Protégé genutzt. Durch diesen Schritt wird aus der erstellten Security Management Ontology eine Wissensbasis geschaffen, die alle relevanten Inhalte inklusive semantischer Informationen des Anwendungsbereiches enthält.

Für eine Validierung der Wissensbasis wird in einem ersten Schritt die automatisierte Konsistenzprüfung in Protégé durchgeführt. Mit Hilfe eines Reasoners ist es möglich neues Wissen über die in der Ontologie hinterlegten Inferenzregeln zu generieren und somit kann sichergestellt werden, dass die hinzugefügten Inhalte auf formaler Ebene korrekt sind. Im folgenden Schritt werden Abfragen der Competency Questions auf die Wissensbasis erstellt, um zu überprüfen, ob die strukturellen Zusammenhänge stimmen und die benötigten Informationen bereitgestellt werden können. Auch dieser Schritt wird in Protégé mit Hilfe einer Abfragesprache für Ontologien (Simple Protocol and RDF Query Language, SPARQL) durchgeführt. In Bild 2 ist ein Beispiel einer Competency Question mit dem Ergebnis der Abfrage dargestellt.

Literaturangaben

- [1] Allianz Global Corporate & Specialty SE: Allianz Risk Barometer. Identifying the major business risks for 2020, 2020. <https://agcs.wufoo.com/forms/download-the-allianz-risk-barometer-2020-report/>, abgerufen am: 29.04.2020
- [2] Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste - WIK: Aktuelle Lage der IT-Sicherheit in KMU. Langfassung, 2017.
https://www.wik.org/fileadmin/Sonstige_Dateien/IT-Sicherheit_in_KMU/WIK-Studie_Aktuelle_Lage_der_IT-Sicherheit_in_KMU_Langfassung__2_.pdf, abgerufen am: 22.04.2020
- [3] Siegwart, C., Scherhag, F., Krammel, M. u. Frey, G.: Ganzheitliche IT-Security Reifegradbestimmung. In: Proceedings of the Kongress Automation 2020
- [4] Zentrum für Mechatronik und Automatisierungstechnik gGmbH, Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme: GrundschatzPLUS Aktivator.
<https://www.security4kmu.de>, abgerufen am: 11.12.2020
- [5] IT-Grundschatz-Kompendium Edition 2021; Februar 2021.
https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschatz/Kompendium/IT_Grundschatz_Kompendium_Edition2021.html
- [6] DIN EN ISO/IEC 27001. Information technology - Security techniques - Information security management systems - Requirements; Juni 2017
- [7] IEC 62443-3-3. Industrial communication networks - Network and system security - Part 3-3: System security requirements and security levels
- [8] Busse, J., Humm, B., Lübbert, C., Moelter, F., Reibold, A., Rewald, M., Schlüter, V., Seiler, B., Tegtmeier, E. u. Zeh, T.: Was bedeutet eigentlich Ontologie? In: Informatik Spektrum, August 2014
- [9] Gruber, T. R.: A Translation Approach to Portable Ontology Specifications. In: Knowledge Acquisition, 1993
- [10] Borst, W. N.: Construction of engineering ontologies for knowledge sharing and reuse. CTIT Ph D-thesis series, no. 97-14. Enschede, NL: Centre for Telematics and Information Technology 1997
- [11] Studer, R., Benjamins, V. u. Fensel, D.: Knowledge engineering: Principles and methods. Data & Knowledge Engineering 25 (1998) 1-2, S. 161–197

- [12] Souag, A., Salinesi, C. u. Comyn-Wattiau, I.: Ontologies for Security Requirements: A Literature Survey and Classification. In: Bajec, M. u. Eder, J. (Hrsg.): Advanced Information Systems Engineering Workshops. CAiSE 2012 International Workshops, Gdańsk, Poland, June 25-26, 2012. Proceedings. Lecture Notes in Business Information Processing, Bd. 112. Berlin, Heidelberg: Imprint Springer 2012, S. 61–69
- [13] A Taxonomy of Security-Related Requirements, Firesmith, D. G., 2015
- [14] Karyda, M., Balopoulos, T., Dritsas, S., Gymnopoulos, L., Kokolakis, S., Lambrinou-dakis, C. u. Gritzalis, S.: An ontology for secure e-government applications. Proceedings / the First International Conference on Availability, Reliability and Security. April 20 - 22, 2006, Vienna, Austria. Los Alamitos, Calif.: IEEE Computer Society 2006, 5 pp-1037
- [15] Tsoumas, B. u. Gritzalis, D.: Towards an Ontology-based Security Management. Proceedings / 20th International Conference on Advanced Information Networking and Applications. 18 - 20 April 2006, Vienna, Austria. Los Alamitos, Calif.: IEEE Computer Society 2006, S. 985–992
- [16] Souag, A., Salinesi, C., Mazo, R. u. Comyn-Wattiau, I.: A Security Ontology for Security Requirements Elicitation. In: Piessens, F., Caballero, J. u. Bielova, N. (Hrsg.): Engineering secure software and systems. 7th international symposium, ESSoS 2015, Milan, Italy, March 4 - 6, 2015 ; proceedings. Lecture notes in computer science, Bd. 8978. Cham: Springer 2015, S. 157–177
- [17] Fenz, S., Pruckner, T. u. Manutscheri, A.: Ontological Mapping of Information Security Best-Practice Guidelines. In: Abramowicz, W. (Hrsg.): Business information systems. 12th international conference, BIS 2009, Poznań, Poland, April 27-29, 2009 ; proceedings. Lecture Notes in Business Information Processing, Bd. 21. Berlin: Springer 2009, S. 49–60
- [18] Stuckenschmidt, H.: Ontologien. Konzepte, Technologien und Anwendungen. Informatik im Fokus. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg 2009
- [19] Noy, N. F. u. McGuinness, D. L.: Ontology development 101: A guide to creating your first ontology. 2001
- [20] Musen, M. A.: The Protégé Project: A Look Back and a Look Forward. AI matters 1 (2015) 4, S. 4–12

Dezentrale Authentifizierung und Autorisierung proaktiver Industrie 4.0-Verwaltungsschalen am Beispiel des IOTA-basierten Industrie-Marktplatzes

A. Dogan, A. Schnakenbeck, M. O. Rögner,
Univ.-Prof. Dr.-Ing. **A. Fay**, Helmut-Schmidt-Universität, Hamburg

Kurzfassung

Die Autoren haben ein Konzept entwickelt, wie die Feststellung der Identität von Industrie 4.0-Komponenten in dezentralen Systemen durch die Bereitstellung eines dezentralen Identitätsmanagements auf Basis der Distributed Ledger Technologie (DLT) und insbesondere dezentraler Identifikatoren (engl. *Decentralized Identifier*, DID) den beteiligten Akteuren in der Wertschöpfungskette mehr Sicherheit und Zuverlässigkeit garantiert. Auf diese Weise kann die Authentifizierung und Autorisierung für alle beteiligten Akteure in der Wertschöpfungskette ohne den Einsatz von zentralisierten Identity Providern oder Zertifizierungsstellen erfolgen.

1. Motivation

Die Vernetzung von Maschinen, Anlagen und Geräten über das und mit dem Internet eröffnet neue Chancen und Potenziale. Dabei werden sie nicht nur von Personen gesteuert, sondern sie kommunizieren selbstständig im Sinne einer Machine-to-Machine-Kommunikation (M2M) direkt miteinander. Eine wesentliche Herausforderung besteht darin, die Daten in einer sicheren, standardisierten und maschineninterpretierbaren Form automatisiert und über Unternehmensgrenzen hinweg allen beteiligten Akteuren in der Wertschöpfungskette zur Verfügung zu stellen. Die dazu notwendigen Voraussetzungen sind die Standardisierung der Kommunikationsstrukturen, der Systemsicherheit sowie der Sprache (Semantik). Die Plattform Industrie 4.0 hat sich dieser Aufgabe angenommen und das Konzept der Industrie 4.0-Komponente erarbeitet und damit eine Lösung zur Implementierung der verschiedenen Industrie 4.0-Anwendungsszenarien bereitgestellt [1, 2].

Der Sicherheitsaspekt (im Sinne von security) ist ein elementarer Bestandteil dieses Konzeptes. Kriterien wie beispielsweise die Vertraulichkeit, Datenintegrität und Zuverlässigkeit von Daten sind dabei essenziell für eine erfolgreiche Umsetzung von Industrie 4.0-Anwendungen. Eine vielversprechende Schlüsseltechnologie in diesem Zusammenhang stellt die DLT dar, die ihren Ursprung im Finanzsektor hat, jedoch für die Implementierung in nahezu allen Bereichen von Industrie 4.0 eine hohe Relevanz besitzt [3]. Insbesondere für dezentrale Systeme

ohne zentrale Autorität bietet die DLT, durch ein unter den Systemteilnehmern verteiltes und lückenloses Register, die notwendige Sicherheit, alle Datentransaktionen manipulationssicher und rückverfolgbar festzuhalten.

Die Autoren haben ein Konzept entwickelt, wie die Feststellung der Identität von Industrie 4.0-Komponenten in dezentralen Systemen durch die Bereitstellung eines dezentralen Identitätsmanagements auf Basis dezentraler Identifikatoren den beteiligten Akteuren in der Wertschöpfungskette mehr Sicherheit und Zuverlässigkeit garantiert [4]. Darauf aufbauend können die Identitäten sich gegenseitig verifizierbare Aussagen (engl. *Verifiable Claims*; z.B. Qualifikationen, Leistungen, Qualitäten oder Hintergrundinformationen) in einer kryptographisch sicheren, datenschutzfreundlichen und maschinell prüfbaren Form (*Verifiable Credentials*, VC [5]) ausstellen, wodurch ein vermaschtes Vertrauensnetzwerk zwischen den Identitäten, auch bekannt als Web of Trust, geschaffen wird und die Basis für ein dezentrales Zugriffsmanagement bildet. Auf diese Weise kann die Authentifizierung und Autorisierung für alle beteiligten Akteure in der Wertschöpfungskette ohne den Einsatz von zentralisierten Identity Providern oder Zertifizierungsstellen erfolgen, was wiederum neue digitale Geschäftsmodelle ermöglicht.

Der Beitrag ist wie folgt gegliedert: Zunächst werden in Kapitel 2 bis 4 die thematischen Grundlagen Industrie 4.0-Komponente, Distributed Ledger Technologie und Decentralized Identifier eingeführt. Kapitel 5 beschreibt die Anforderungen an die Sicherheit der Verwaltungsschale am Beispiel des IOTA-basierten Industrie-Marktplatzes. Im Anschluss wird in Kapitel 6 das Lösungskonzept der Autoren vorgestellt. Der Beitrag schließt in Kapitel 7 mit einer Zusammenfassung und einem Ausblick ab.

2. Industrie 4.0-Verwaltungsschale

Industrie 4.0 (I4.0) verspricht eine neue Stufe der Organisation und der Steuerung von Wertschöpfungsketten. Durch die neuartige Vernetzung von Objekten, Daten und Diensten entstehen dynamische Netzwerke der Wertschöpfung über Unternehmensgrenzen hinweg, die eine erhöhte Fähigkeit zur Selbstoptimierung und Selbstorganisation besitzen. Voraussetzung hierfür ist eine digitale Bild der physischen Welt. Anhand des Referenzarchitekturmodells Industrie 4.0 (RAMI 4.0, [6]) kann ein beliebiges Asset in einer Informationswelt beschrieben werden. Den Zugang zu den Daten, die einen Gegenstand beschreiben, ermöglicht die I4.0-Komponente. Sie besteht aus Asset und Verwaltungsschale (VWS) mit digitaler Verbindung innerhalb eines I4.0-Systems. Zudem ist sie eindeutig identifizierbar und kommunikationsfähig im I4.0-System. Um Informationen verfügbar zu machen, ist neben der hinreichenden, digitalen Bild der Informationen der interoperable Austausch der Informationen notwendig. Die Industrie 4.0 bietet zur Bild der Daten die Verwaltungsschale mit ihren Teilmodellen. Darauf aufbauend

wurde in [7] und [8] eine Richtlinie zur I4.0-Sprache entworfen, um I4.0-Komponenten miteinander interagieren lassen zu können. Die I4.0-Sprache gliedert sich in die drei Schichten Vokabular (regelt, wie Datenelemente aufgebaut werden, die in den Nachrichten verwendet werden), Struktur der Nachricht (regelt den Aufbau der Nachrichtenstruktur) und Interaktionsprotokolle (regelt den Ablauf von Nachrichten in einem Dialog). Bei der Kommunikation kann zwischen horizontaler (Komponenten derselben Betriebs- oder Fabrikebene) und vertikaler Kommunikation (von der Produkt- und Sensorebene bis hin zur Geschäftsebene) unterschieden werden. Beide Formen können von einer Verwaltungsschale angeboten werden. Die Abwicklung von Interaktionen zwischen Teilnehmern des I4.0-Systems wird von einem Interaktionsmanager durchgeführt. Dieser setzt verschiedene Protokolle und Entscheidungsalgorithmen um.

3. Distributed Ledger Technologie

Eine DLT ermöglicht den Betrieb und die Verwendung eines dezentral geführten Transaktionsregisters (engl. *distributed ledger*) in einem DLT-System. Ein Transaktionsregister ist eine Art Datenbank, die eine elektronische Erfassung von Transaktionsdaten (finanz-, logistisch- oder arbeitsbezogen) gewährleistet. In einem DLT-System werden dezentral beliebig viele gleichgestellte Kopien eines Transaktionsregisters von unterschiedlichen Teilnehmern (z.B. Geräten oder Prozessen), auch bekannt als DLT-Knoten (engl. *DLT node*), verwaltet. Neu hinzugefügte Transaktionen werden im gesamten DLT-System verteilt, und unter Verwendung eines Konsensmechanismus wird ein Konsens über den aktuellen Stand des Transaktionsregisters im gesamten DLT-System erzielt. [9]

In der öffentlichen Diskussion wird die Blockchain-Technologie häufig als Synonym für DLT angesehen. Objektiv betrachtet stellt die Blockchain-Technologie jedoch eine Ausprägung von DLT dar. In einer Blockchain werden die Transaktionen durch sogenannte Miner mittels kryptographischer Verfahren (Proof-of-Work) validiert, in Datenblöcken zusammengefasst und miteinander verknüpft, sodass eine manipulationssichere Datenblockkette entsteht. Das Netzwerk legt dabei fest, in welchem Zeitintervall ein neuer Block der Datenblockkette angehängen wird [10]. Eine konkrete Implementierung der Blockchain ist die Kryptowährung Bitcoin [11]. Das Konzept, wie Bitcoin es nutzt, hat allerdings einige Nachteile, wie beispielsweise eine Gebühr für jede Transaktion zu zahlen. Für Anwendungsfelder wie das IoT würde die Transaktionsgebühr höher sein als der eigentliche Wert der Transaktion selbst. IOTA, eine alternative Umsetzung der DLT, wurde für das IoT entworfen. Anstatt auf einer globalen Kette aus Blöcken basiert IOTA auf einem gerichteten, azyklischen Graphen, der Tangle genannt wird.

Die Knoten des Graphen bilden Transaktionen, die von den Teilnehmern des Netzwerks veranlasst werden. Die Kanten entstehen dadurch, dass jede eintreffende Transaktion zwei vorherige Transaktionen bestätigen muss. Beim Erstellen einer Transaktion muss ebenfalls ein Proof-of-Work geleistet werden. Dieser wird allerdings von keinem Miner durchgeführt, sondern von dem Ersteller der Transaktion, wodurch die Gebühren wegfallen. Die Blöcke werden hier also auf einzelne Transaktionen herunter gebrochen, die sich gegenseitig direkt referenzieren. Ein zweiter Vorteil ist, dass durch die Architektur des Tangles das Netzwerk besser skalierbar ist. Bei einer Blockchain besteht immer das Problem, dass jede Transaktion in die eindimensionale Kette an Blöcke eingefügt werden muss (Transaktionsdurchsatz konstant). Im Tangle hingegen sind parallele Knoten möglich, dadurch, dass jede Transaktion zwei bestehende referenziert. Das bedeutet, IOTA skaliert proportional mit der Anzahl der Transaktionen. Je größer der Tangle mit anwachsenden Transaktionen, desto schneller und sicherer wird IOTA. Wie in einer Blockchain wird auch hier eine Transaktion akzeptiert, wenn sie vermehrt referenziert wurde. [12]

4. Decentralized Identifiers

Beim klassischen Identitätsmanagement ist eine vertrauenswürdige, dritte Instanz notwendig, die dafür bürgt, dass bestimmte Attribute im Zusammenhang mit einer Entität stehen. In einer Public-Key-Infrastructure (PKI) kann dies beispielsweise der Zusammenhang zwischen Internetdomain und Public Key sein. Mit dem Aufkommen von Distributed-Ledger-Technologien ergibt sich die Möglichkeit ein dezentrales Management von Identitäten und den ihnen zugehörigen Attributen zu etablieren. So können die einzelnen Entitäten ihre Identitäten selbst erzeugen und verwalten. Dafür werden die Informationen, die zu einer Identität gehören, in der DLT gespeichert, wo sie dann vertrauenswürdig und manipulationssicher vorliegen.

DIDs sind eine von dem World Wide Web Consortium (W3C) entwickelte Gruppe von Identifikatoren, die einen digitalen, dezentralen Identitätsnachweis der Entitäten ermöglichen. Die Kontrolle über eine DID kann mittels digitaler Signaturen nachgewiesen werden, ohne dass eine zentrale Institution notwendig ist. Bei DIDs handelt es sich um Uniform Resource Identifiers (URIs), bestehend aus den drei Teilen Schema (did), Identifikator für die DID-Methode und methodenspezifische-ID. Diese DID verweist auf ein DID-Dokument, welches Informationen über das DID-Subjekt enthält. Bei einem DID-Subjekt kann es sich beispielsweise um eine Verwaltungsschale handeln. Das DID-Dokument kann kryptographische Informationen, Verifikationsmethoden und Serviceendpunkte enthalten, die eine Interaktion mit dem DID-Subjekt sowie eine Authentifizierung und Autorisierung ermöglichen. Die DID-Methode beschreibt die Implementierung in einem bestimmten Distributed-Ledger. Dies beinhaltet, wie die DID und

das Dokument erstellt, gelesen, aktualisiert und deaktiviert werden. Die DID-Methode, die in dieser Arbeit betrachtet werden soll, ist im IOTA-Tangle implementiert. [4]

5. Anforderungen an die Sicherheit der Verwaltungsschale am Beispiel des IOTA-basierten Industrie-Marktplatzes

Die Entwicklung des Konzeptes dieser Forschungsarbeit erfolgte unter Berücksichtigung der Anforderungen an den IOTA-basierten Industrie-Marktplatz [13]. Der auf Basis von IOTA konzipierte dezentrale Industrie-Marktplatz ist open-source und ermöglicht den Kauf und Verkauf von Produkten, Daten und Dienstleistungen auf einer hersteller- und branchenneutralen Plattform, d.h. ohne einen Plattformbetreiber wie auf einem herkömmlichen zentralisierten Marktplatz, der Einfluss auf die Prozesse zwischen den Entitäten einnehmen könnte (z.B. Zusammenführen von Angebot und Nachfrage, Preisbildung). Primär wurde die Plattform für die direkte Interaktion (z.B. Verhandlungen, Bezahlvorgänge) zwischen Maschinen entwickelt und berücksichtigt Spezifikationen und Richtlinien der Plattform Industrie 4.0. So können die Entitäten (z.B. Maschinen und Automatismen) die Rollen Daten/Service Provider (SP) oder Daten/Service Requester (SR) einnehmen und an protokollbasierten Interaktionen, wie in der VDI/VDE Richtlinie 2193 [7, 8] definiert, teilnehmen. Die Durchführung von Geschäftstransaktionen in verteilten Systemen ohne zentrale Autorität erfordert die Sicherstellung einer vertrauenswürdigen Infrastruktur, die im beschriebenen Anwendungsfall mit Hilfe von IOTA nur teilweise erfüllt wird. Die Übertragung von Daten (z.B. relevante Informationen der Verhandlung) in den IOTA-Tangle ist abhängig von der Vertrauenswürdigkeit der Entitäten. Hier bietet der IOTA-Tangle keinen Schutz gegen ungültige oder sogar bösartige Daten (auch bekannt als Orakelproblem [14]).

Aus der Analyse des konkretisierten Anwendungsfalles wurden Anforderungen an die Sicherheit einer Verwaltungsschale erhoben. Diese stimmen mit den in [15] genannten Anforderungen überein. Für das Identitäts- und Zugriffsmanagement-System (engl. *Identity and access management system*, IAM-System) gelten die Anforderungen aus [16]. Die Anforderungen werden in Tabelle 1 zusammengefasst:

Tabelle 1: Beschreibung der Anforderungen

Ziel	Beschreibung
A1 Identitäten und Authentifizierung	Die Grundvoraussetzung einer sicheren Interaktion zwischen SR und SP im lokalen (innerhalb einer Maschine) als auch globalen (unternehmensübergreifend) Kontext ist die eindeutige und unmissverständliche Identifizierung und Authentifizierung.
A2 Zugriffssteuerung	Verwaltung von Subjekten und deren Berechtigungen bezüglich der Interaktion mit einem bestimmten Objekt.
A3 Kommunikationssicherheit	Verwendung geeigneter Sicherheitsprotokolle für den Austausch vertraulicher Informationen zwischen SR und SP. Die Integrität der Daten muss sichergestellt werden.
A4 Ereignisprotokollierung	Die Rückverfolgbarkeit der Interaktion zwischen SR und SP durch Protokollierung und Auditing ist zwingend erforderlich bei der Aufklärung von Sicherheitsvorfällen.
A5 Dezentralität	Ein Kernaspekt des Industrie-Marktplatzes ist die dezentrale Architektur der Plattform. Die Unabhängigkeit von einer zentralen Autorität soll durch die Einführung eines IAM-Systems berücksichtigt werden.
A6 Zuverlässigkeit	Die Überprüfung einer Identität soll für alle Entitäten jederzeit möglich sein (Vermeidung von einem Single Point of Failure).
A7 Kompatibilität	Der Industrie-Marktplatz ist eine von vielen möglichen Anwendungen, welches das IAM-System unterstützen soll.
A8 Interoperabilität	Der Industrie-Marktplatz wurde primär für Maschinen entwickelt, weshalb der Informationsaustausch auch in einer standardisierten, maschineninterpretierbaren Form stattfinden soll. Das IAM-System soll relevante Spezifikationen und Richtlinien der Plattform Industrie 4.0 einhalten.

6. Lösungskonzept

In diesem Konzept wird ein Protokoll vorgestellt, das die Authentifizierung auf Basis von DIDs in I4.0-Sprache ermöglicht und dabei eine symmetrisch verschlüsselte Kommunikation zwischen den I4.0-Komponenten aufbaut. Diese entstehende, verschlüsselte Kommunikation ist Basis für weitere Konzepte, wie Autorisierung oder Verhandlungen in I4.0-Sprache [7, 8]. Da die I4.0-Sprache keine Verschlüsselung der Nachrichten vorsieht, wird in diesem Konzept eine zusätzliche Schicht vorgestellt, die eine verschlüsselte Kommunikation in I4.0-Sprache erlaubt. Diese Schicht ist notwendig, um entsprechende Metadaten zur Interpretation in I4.0-Sprache bereitzustellen, die für eine Entschlüsselung notwendig sind. Authentifizierung und Autorisierung werden in das Ausschreibungsverfahren [8] eingebettet, welches am Beispiel des dezentralen Marktplatzes durchgeführt wird (siehe Bild 1).

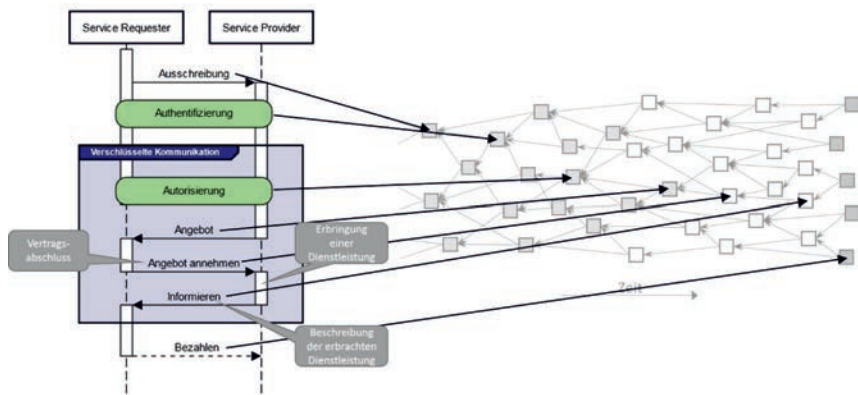


Bild 1: Authentifizierung und Autorisierung als Teilprozesse des Ausschreibungsverfahrens entsprechend der VDI/VDE Richtlinie 2193-2 am Beispiel einer Kommunikation im I-OTA-Tangle

Das Ausschreibungsverfahren findet zwischen zwei Akteuren in den Rollen des SR und des SP statt. Der Verhandlungsprozess startet mit dem Call for Proposal (CfP), der in Form eines Broadcast von einem SR über den dezentralen Marktplatz an alle SP gesendet wird. Nach dem CfP erfolgt zunächst die Authentifizierung anhand der Public Keys von SR und SP, die über DID-Dokumente veröffentlicht werden und mittels der Identifikatoren der VWS, den DIDs, abgerufen werden können. Zusätzlich zur Authentifizierung wird während des Protokolls ein symmetrischer Schlüssel ausgetauscht, mit denen SR und SP die restliche Kommunikation in dieser Sitzung verschlüsseln. Im Anschluss können beide mit einer Autorisierung fortfahren und dann das Ausschreibungsverfahren beenden. Für die verschlüsselten Nachrichten werden die Schutzziele Vertraulichkeit und Authentizität (und damit Integrität) erfüllt.

Authentifizierung

Grundlage für eine Authentifizierung ist, dass beide Entitäten ein DID-Dokument mit den entsprechenden Einträgen unter ihrer DID veröffentlicht haben. Dieser Prozess wird in [17] beschrieben. Inhalt dieser DID-Dokumente sind Public Keys und Serviceendpunkte, über die die Entitäten miteinander kommunizieren. Die Implementierung der DID-Methode stellt dabei einen authentischen Zusammenhang zwischen Identifikator (DID) und DID-Dokument her. Bei den Serviceendpunkten kann es sich um unterschiedliche Kommunikationsendpunkte, wie HTTP-Server, MQTT-Broker oder auch um eine Kommunikation über den IOTA-Tangle handeln. Des Weiteren sind im DID-Dokument zugehörige Metainformationen enthalten, sowie

eine Liste an unterstützten Authentifizierungsprotokollen und kryptographischen Algorithmen (Cipher Suite, vgl. Transport Layer Security (TLS) [18]). Bild 2 stellt den Interaktionsaustausch zwischen den beteiligten Akteuren für die Authentifizierung und die Erstellung einer verschlüsselten Kommunikation dar.

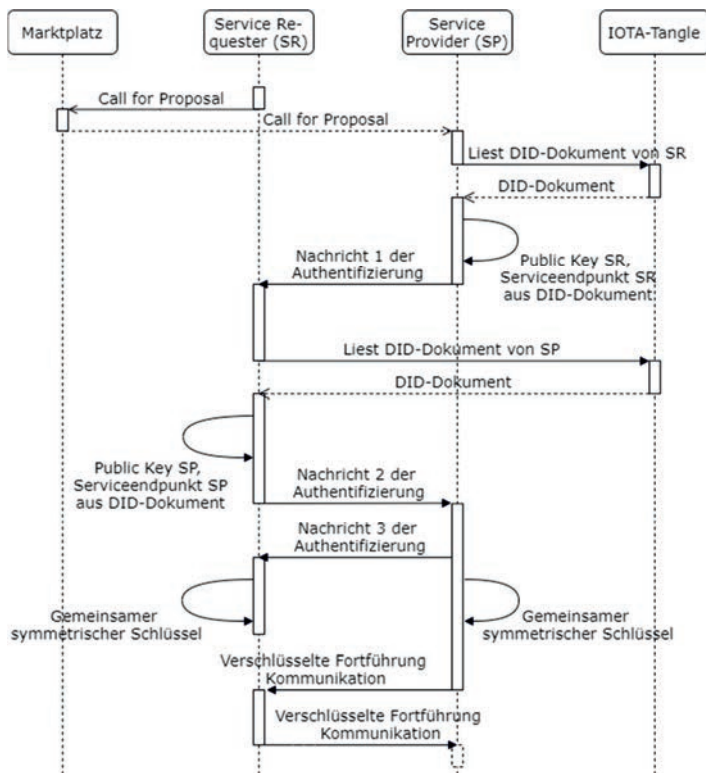


Bild 2: Authentifizierungsprotokoll

Der SP startet den Authentifizierungsprozess, indem er das DID-Dokument des SR abrufen und einen Serviceendpoint mit zugehörigem Public Key ausliest. Der dafür benötigte Identifikator der VWS des SR ist im Header des CfP angegeben. Das Schema der DID gibt dabei einen Hinweis auf die DID-Methode, nach der ein Abrufen durchgeführt wird [4]. Zudem liest der SP weitere Informationen zur Authentifizierung aus, beispielsweise auf welchen Algorithmen der

Public Key aufbaut. Anschließend erstellt der SP die erste Nachricht des Authentifizierungsprotokolls, verpackt in I4.0-Sprache (dargestellt in Bild 3), und sendet sie dem SR über dessen Serviceendpunkt. Das in dieser Arbeit verwendete Authentifizierungsprotokoll ist der ISO/IEC 11770-3 [19] entnommen. Es nutzt asymmetrische Kryptographie und schafft eine gegenseitige Authentifizierung über drei Nachrichten nach dem Challenge-Response-Prinzip. Zusätzlich wird über das Protokoll ein symmetrischer Schlüssel zwischen den Entitäten etabliert, mit dem die folgende Kommunikation verschlüsselt wird. Vergleichbar mit TLS besteht aber die Möglichkeit verschiedene Authentifizierungsprotokolle zu nutzen. Das Protokoll wird semantisch über seinen Identifikator im Header der I4.0-Nachricht identifiziert (Bild 3: semanticProtocol). Welche Protokolle von einer VWS unterstützt werden, wird im DID-Dokument angegeben. Erhält der SR die erste Nachricht des Authentifizierungsprotokolls, kann er dem Header der I4.0-Nachricht den Identifikator des SP entnehmen (Bild 3: sender – identification – id) sowie das verwendete Protokoll. Über den Identifikator des Protokolls wird die semantische Information für die Interpretation des Nutzwertes (Bild 3: interactionElements – keyToken) bereitgestellt. Der SR kann nun ebenfalls über den Identifikator des SP dessen DID-Dokument abrufen und Public Key und Serviceendpunkt auslesen. Es folgen die restlichen Nachrichten des Authentifizierungsprotokolls. Aus diesen geht der symmetrische Schlüssel hervor.

```

1 {
2   "frame": {
3     "semanticProtocol": {
4       "keys": [
5         {
6           "value": "www.iso.org/standard/11770-3/sktm6/did",
7           ...
8         }
9       ]
10    },
11    "sender": {
12      "identification": {
13        "id": "did:iota:987654321zyxwvutsr",
14        "idType": "URI"
15      },
16      ...
17    },
18    "receiver": {...},
19    ...
20  },
21  "interactionElements": [
22    {
23      "keyToken": "RmcWtblFVjRUS3RaU1...eWTMTbOY3CzsE7RX=="
24    }
25  ]
26 }

```

Bild 3: Nachricht 1 der Authentifizierung

Symmetrische Verschlüsselung

Die folgende Kommunikation wird nun verschlüsselt fortgesetzt. Zur Entschlüsselung der Nachricht benötigen die Entitäten (I) semantische Informationen, um die Nachricht zu interpretieren, und (II) den Identifikator des Absenders, um die zu verwendenden symmetrischen Schlüssel und kryptographischen Algorithmen identifizieren zu können. Daher wird lediglich der Body der I4.0-Nachricht verschlüsselt und der Header leicht modifiziert im Klartext gesendet. Diese Modifikation ist notwendig, um die semantischen Informationen zur Interpretation des Bodys bereitzustellen, in dem sich nun die verschlüsselte Nachricht befindet (Bild 4: `interactionElements`). Üblicherweise wird der Body dafür in Form von Teilmodellen der VWS modelliert [20]. Die damit verbundene Flexibilität des zu transportierenden Inhalts ist für kryptographische Protokolle hingegen unerwünscht [21]. Daher wird im Header der Nachricht ein zweiter Identifikator im Bereich `semanticProtocol - keys` (Bild 4) angegeben, der diese Semantik implizit liefert. Selbes gilt für das Authentifizierungsprotokoll. Im Body sind so nur die verschlüsselten Nutzdaten (Bild 4: `interactionElements - cipherData`) und ein Message Authentication Code (MAC) [22] enthalten. Über den Identifikator des Absenders können SP und SR intern die zu verwendenden symmetrischen Schlüssel und die kryptographischen Algorithmen abrufen. Die entsprechenden Informationen, die während des Authentifizierungsprotokolls ausgetauscht werden, können als Teilmodell in der VWS der Entitäten modelliert werden, wo sie über den Identifikator der Konversation (`conversationID`) nun abgerufen werden.

```

1 {
2   "frame": {
3     "semanticProtocol": {
4       "keys": [
5         {
6           "value": "https://w3c-ccg.github.io/security-vocab/#EncryptedMessage",
7           ...
8         },
9         {
10          "value": "www.admin-shell.io/interaction/bidding",
11          ...
12        }
13      ]
14    },
15    "type": "proposal",
16    "sender": {...},
17    "receiver": {...},
18    ...
19  },
20  "interactionElements": [
21    {
22      "cipherData": "VTJGc2RHVmtYMY3...ONjRUS3Rabb009Cg==",
23      "authenticationTag": "q25H1CzsE73lOmeyEle93w==",
24    }
25  ]
26 }

```

Bild 4: Verschlüsselte Fortführung der Kommunikation

Autorisierung

Autorisierung beschreibt den Prozess, der sicherstellt, dass die angeforderte Aktivität oder der angeforderte Zugriff auf ein Objekt mit den Berechtigungen der Identität (auch Subjekt genannt) möglich ist. Der Autorisierungsprozess setzt ein authentifiziertes Subjekt voraus. Das zugrundeliegende Konzept der Autorisierung ist die attributbasierte Zugriffssteuerung (engl. *Attribute Based Access Control*, ABAC [23]). Im Gegensatz zur häufig eingesetzten rollenbasierten Zugriffssteuerung (engl. *Role Based Access Control*, RBAC [24]), welches einen einfachen „Benutzer-Rollen-Rechte-Struktur-Ansatz“ (je nach Rolle in der Organisation hat der Benutzer unterschiedliche Zugriffsberechtigungen) verfolgt, stützt sich ABAC auf eine Reihe von Attributen der drei Kategorien Subjekt (z.B. Qualifizierungsniveau, Nachweis einer Sicherheitsprüfung), Umwelt (z.B. Zeitpunkt des Zugriffs, Standort der Daten) und Objekt (z.B. Betriebsmodi, Maschinenzustände) und dem wesentlichen Unterschied, dass das Subjekt dem Objekt zuvor nicht bekannt sein muss. ABAC ermöglicht die für die Industrie 4.0-Anwendungsszenarien erforderliche Granularität und Flexibilität bei der Zugriffssteuerung.

In [15] stellen die Autoren den prozessualen Zusammenhang im ABAC-Konzept am Beispiel einer I4.0-Entität beim Zugriff auf die CAE-Daten (Teilmodell der VWS) des Herstellers vor. Dieser Ansatz wird im Folgenden um die dezentralen Komponenten DID, VC und IOTA-Tangle erweitert und in Bild 5 dargestellt. Grundlage für die Autorisierung ist, dass das Subjekt die dafür nötigen Subjektattribute in Form von VC besitzt, welche von einer vertrauenswürdigen Instanz (z.B. Organisation, öffentliche Behörde) signiert sind. Des Weiteren müssen sowohl Subjekt als auch Aussteller der VC ein DID-Dokument mit den entsprechenden Eintragungen (Public Key) unter ihrer DID veröffentlicht haben. Der Aufbau eines VC sowie die notwendigen Schritte zur Erstellung und Signierung werden in [5] beschrieben. Die Erweiterung des Konzeptes sieht vor, dass der *Policy Decision Point* (PDP) eine Verifizierung der VC durchführt. Dazu wird das DID-Dokument des Erstellers benötigt, welches mit Hilfe der DID, eingetragen im VC, ermittelt werden kann. Für die Überprüfung muss der *Policy Information Point* (PIP) dem PDP die Information bereitstellen, welche Subjekte als vertrauenswürdig eingestuft werden. Bezogen auf den Industrie-Marktplatz können sowohl SR als auch SP Attribute in Form von VC besitzen, die für eine erfolgreiche Kommunikation mit dem Gegenüber notwendig sind.

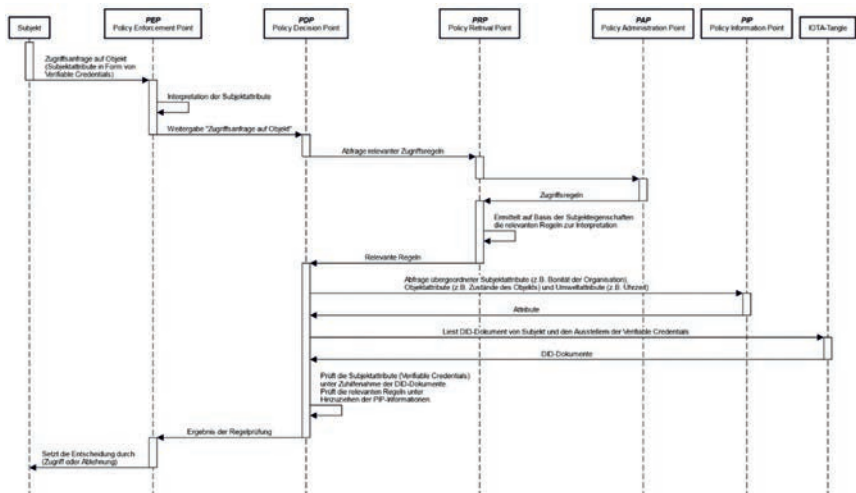


Bild 5: Erweiterter prozessualer Ablauf im ABAC-Konzept in Anlehnung an [15]

8. Zusammenfassung und Ausblick

In diesem Beitrag wurde von den Autoren ein Konzept vorgestellt, wie in einem dezentralen Netzwerk auf Basis von DIDs, DLT und der I4.0-Sprache eine Authentifizierung, eine symmetrisch verschlüsselte Kommunikation und eine Autorisierung zwischen I4.0-Komponenten durchgeführt werden kann. Hierzu wurden wesentliche Anforderungen an die Sicherheit der Verwaltungsschale in Anlehnung an den IOTA-basierten Industrie-Marktplatz erhoben und eine Lösung für den prozessualen Ablauf auf Basis bestehender, industrierelevanter Prozesse sowie das erweiterte Nachrichtenformat der I4.0-Sprache vorgestellt. Im Rahmen weiterer Forschungsvorhaben müssen aus Sicht der Autoren hybride Ansätze mit etablierten Konzepten (z.B. die Einbindung von X.509 Zertifikaten in VCs) aus der Industrie erarbeitet werden, um die Akzeptanz der verteilten Ansätze in der I4.0-Community zu stärken und die Transformation der aktuellen Konzepte zu dezentralen Lösungen zu ermöglichen.

Quellen

- [1] Plattform Industrie 4.0 (2016). AG Forschung und Innovation: Ergebnispapier - Aspekte der Forschungsroadmap in den Anwendungsprozessen. Berlin: Bundesministerium für Wirtschaft und Energie (BMWi).
- [2] Plattform Industrie 4.0 (2016). AG Forschung und Innovation: Ergebnispapier -Fort-schreibung der Anwendungsszenarien der Plattform Industrie 4.0. Berlin: Bundesminis-terium für Wirtschaft und Energie (BMWi).
- [3] Lewin, M., Dogan, A., Schwarz, J., & Fay, A. (2019). Distributed-Ledger-Technologien und Industrie 4.0-Eine Untersuchung der Relevanz für Industrie 4.0. Informatik Spek-trum: Vol. 42, No. 3.
- [4] W3C (2021). Decentralized Identifiers (DIDs) v1.0. Core architecture, data model, and representations. Online erhältlich unter: <https://www.w3.org/TR/did-core/>; abgerufen am 03. Mai 2021.
- [5] W3C (2019). Verifiable Credentials Data Model 1.0. Expressing verifiable information on the Web. Online erhältlich unter: <https://www.w3.org/TR/vc-data-model/>; abgerufen am 03. Mai 2021.
- [6] DIN SPEC 91345 (2016). Referenzarchitekturmodell Industrie 4.0 (RAMI4.0). Berlin: Beuth Verlag.
- [7] Verein Deutscher Ingenieure (2019). VDI/VDE 2193 Blatt 1. Sprache für I4.0-Komponen-ten - Struktur von Nachrichten.
- [8] Verein Deutscher Ingenieure (2020). VDI/VDE 2193 Blatt 2. Sprache für I4.0-Komponen-ten - Interaktionsprotokoll für Ausschreibungsverfahren.
- [9] International Organization for Standardization. (2020). ISO 22739:2020 - Blockchain and distributed ledger technologies - Vocabulary.
- [10] Skwarek, V. (2019). Eine kurze Geschichte der Blockchain. Informatik Spektrum, 42(3), 161-165.
- [11] Nakamoto, S., & Bitcoin, A. (2008). A peer-to-peer electronic cash system. Bitcoin. On-line erhältlich unter: <https://bitcoin.org/bitcoin.pdf>; abgerufen am 03. Mai 2021.
- [12] Popov, S. (2018). The Tangle. Online erhältlich unter: [https://assets.ctfassets-net/r1dr6vzfzhxhev/2t4uxvslqk0EUau6g2sw0g/45eae33637ca92f85dd9f4a3a218e1ec/iota1_4_3.pdf](https://assets.ctfassets.net/r1dr6vzfzhxhev/2t4uxvslqk0EUau6g2sw0g/45eae33637ca92f85dd9f4a3a218e1ec/iota1_4_3.pdf); abgerufen am 03. Mai 2021.
- [13] Belyaev, A., Diedrich, C., Köther, H., & Dogan, A. (2020). Dezentraler IOTA-basierter Industrie-Marktplatz. Industrie 4.0 Management 2/20, S. 36–40.

- [14] Caldarelli, G., Rossignoli, C., & Zardini, A. (2020). Overcoming the blockchain oracle problem in the traceability of non-fungible products. *Sustainability*, 12(6), 2391 Caldarelli, G., Rossignoli, C., & Zardini, A. (2020). Overcoming the blockchain oracle problem in the traceability of non-fungible products. *Sustainability*, 12(6), 2391.
- [15] Plattform Industrie 4.0 (2018). Zugriffssteuerung für Industrie 4.0-Komponenten zur Anwendung von Herstellern, Betreibern und Integratoren. Berlin: Bundesministerium für Wirtschaft und Energie (BMWi).
- [16] Dogan, A., Millenaar, J., & Fay, A. (2020). DID-basiertes Identitätsmanagement für Industrie-4.0-Komponenten in dezentralen Systemen. Tagung „Entwurf komplexer Automatisierungssysteme“ (EKA). Magdeburg.
- [17] Dogan, A., Schnakenbeck, A., Fay, A., Nagel, J. A., & Millenaar, J. (2020). DID-basiertes Identitätsmanagement für Industrie-4.0-Komponenten in dezentralen Systemen – Eine Realisierungsmöglichkeit für ein langlebiges Digitales Typenschild. In *Automation 2020* (pp. 983–998). VDI Verlag.
- [18] Rescorla, E., & Dierks, T. (2018). The transport layer security (TLS) protocol version 1.3.
- [19] International Organization for Standardization. (2015). ISO/IEC 11770-3:2015-08 - Informationstechnik - Sicherheitsverfahren - Schlüsselmanagement - Teil 3: Asymmetrische Techniken verwendende Mechanismen.
- [20] Plattform Industrie 4.0 (2020). Specification: Details of the Asset Administration Shell. Part 1 - The exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0RC01). Berlin: Bundesministerium für Wirtschaft und Energie (BMWi).
- [21] Schneider, B. (1996). Applied cryptography: protocols. ALGORITHMS, AND SOURCE CODE IN C (2d.
- [22] Krawczyk, H., Bellare, M., & Canetti, R. (1997). HMAC: Keyed-hashing for message authentication.
- [23] Priebe, T., Dobmeier, W., Muschall, B., & Pernul, G. (2005). ABAC—Ein Referenzmodell für attributbasierte Zugriffskontrolle. *Sicherheit 2005, Sicherheit-Schutz und Zuverlässigkeit*.
- [24] Sandhu, R., Ferraiolo, D., & Kuhn, R. (2000, July). The NIST model for role-based access control: towards a unified standard. In *ACM workshop on Role-based access control* (Vol. 10, No. 344287.344301).

ECLASS als semantische Basis eines dezentralen Registers für die Identifikation von Diensten von Industrie 4.0-Komponenten

Distributed Ledger-basierte Infrastruktur für Verwaltungsschalen

A. Dogan, Univ.-Prof. Dr.-Ing. **A. Fay**,
Helmut-Schmidt-Universität, Hamburg;
Dr.-Ing. **J. A. Nagel**, **G. Rijo**, Neoception GmbH, Mannheim;
A. Belyaev, Univ.-Prof. Dr. Ing **C. Diedrich**,
Otto-von-Guericke-Universität, Magdeburg;
Dr.-Ing. **C. Block**, ECLASS Head Office, Köln;
A. Bondza, Pepperl+Fuchs SE, Mannheim

Kurzfassung

Die Arbeit stellt ein Konzept einer Distributed Ledger-basierten Infrastruktur für das Identitätsmanagement für Industrie 4.0-Komponenten vor und führt das Konzept des dezentralen Registers für die Verwaltungsschalen ein. Anhand von drei Anwendungsfällen wird erklärt, wie diese Ansätze in der Praxis umgesetzt und angewendet werden können. Die Anwendungsfälle zeigen, wie die Verwaltungsschalen und die von diesen bereitgestellten Diensten in dem dezentralen Register angemeldet und von den Nutzern in verschiedenen Asset-Lebenszyklusphasen abgerufen werden können. Darüber hinaus zeigen die Anwendungsfälle, wie die Nutzer von Industrieanlagen mehrere Verwaltungsschalen mit der gleichen Anlage verknüpfen können und wie diese Verwaltungsschalen anhand einer eindeutigen Asset-ID gefunden werden können. Der dritte Anwendungsfall beschreibt die Möglichkeit der automatischen Aktualisierung der Zugriffsinformationen auf die Verwaltungsschalen bei Änderung der Eigentümer oder Betreiber eines Assets.

1. Motivation

Industrie 4.0 (I4.0) verfolgt das Ziel, die Grundbausteine für die Schaffung zukünftiger globaler digitaler Ökosysteme bereitzustellen. Die heutigen starren Wertschöpfungsketten sollen durch flexible, hochdynamische und global vernetzte Wertschöpfungsnetzwerke ersetzt werden [1]. In der vernetzten Welt von I4.0 wird nicht nur der Zugang zu Informationen immer wichtiger, sondern auch eine sichere, eindeutig standardisierte und maschineninterpretierbare Kommunikation zwischen allen Entitäten. Eine zentrale Rolle bei der Umsetzung von I4.0 übernehmen

Cyber-physische Systeme (CPS). CPS bestehen aus physischen sowie virtuellen Objekten und Prozessen, die miteinander, zum Teil global, vernetzt sind und ihr Handeln durch den Austausch von Informationen abstimmen. Für die Umsetzung müssen die Komponenten aus der physischen Welt in der Informationswelt abgebildet und die Interoperabilität durch standardisierte Daten- und Interaktionsmodelle sichergestellt werden [2]. Die Identifikation bildet die Grundlage für alle entstehenden Prozesse. Daraus ergeben sich neue Anforderungen, die am Beispiel des digitalen Typenschildes im Folgenden beschrieben werden.

Im Rahmen des vom ZVEI durchgeführten Projektes "Digitales Typenschild" [3] wurde eine digitale Umsetzung des Typenschildes am Beispiel einer I4.0-Verwaltungsschale (VWS) konzipiert. Neben der obligatorischen Kennzeichnung kann das digitale Typenschild auch weitere Informationen wie Handbücher, Zertifikate, technische Zeichnungen und Betriebsdaten in mehreren Sprachen enthalten. Entsprechend der DIN SPEC 91406 [4] wird ein global eindeutiger Identifikator benötigt, der am physischen Objekt angebracht wird und auf die digitale Repräsentation verweist. Die DIN SPEC 91406 verwendet eine Untermenge des URL-Schemas (Uniform Resource Locators) nach RFC 3986 [5], um einen global eindeutigen Bezeichner zu definieren. Die URL wird typischerweise für die Adressierung von Webseiten verwendet. Es liegt nahe, dass die URL, die nach aktuellen Konzepten der Plattform Industrie 4.0 als ein global eindeutiger Identifikator von VWS verwendet werden soll, von den meisten Unternehmen auch dazu verwendet wird, um unter dieser URL einen webbasierten Dienst zu hosten, der Informationen der I4.0-VWS anbietet. Die Verwendung eines solchen Identifikators zur Referenzierung einer Ressource im Internet kann problematisch sein, wenn es darum geht, die Informationen über den gesamten Lebenszyklus des Assets auf dem neuesten Stand zu halten. Zum Beispiel bedeutet eine Umstrukturierung der Website (z.B. die Änderung der Domain), dass der Identifikator nicht mehr aufgelöst werden kann. Daher ist es wünschenswert, einen Identifikator zu verwenden, der Änderungen der referenzierten Informationen erlaubt.

Damit sich die VWSen untereinander finden und vertrauen können, werden Konzepte benötigt, welche den Visionen von I4.0 entsprechen und die Registrierung von VWS in einem gemeinsamen Register ermöglichen, sodass ein vertrauenswürdiges Identitäts- und Zugriffsmanagement gewährleistet werden kann. Die derzeit in der I4.0-Community diskutierten Konzepte basieren auf dem klassischen zentralisierten Ansatz, bei dem zentralisierte Systeme die Rolle des Vertrauensvermittlers übernehmen und eine Registrierung, Authentifizierung, Autorisierung und Zertifizierung von VWSen ermöglichen. Die zentralisierten Systeme haben jedoch einige Eigenschaften, die im Zusammenhang mit der Entstehung globaler Unternehmens- und Länder-übergreifender Ökosysteme eher weniger geeignet sind. Zu diesen Nachteilen gehören unter anderem Abhängigkeiten vom Plattformanbieter, der als unausweichlicher Vermittler

zwischen den Plattformnutzern agieren würde, die Manipulation der Plattform durch den Anbieter zum eigenen Vorteil oder Risiken im Zusammenhang mit einem möglichen Missbrauch der Plattform durch Dritte.

Um dieses Problem zu lösen, wird im Rahmen dieses Beitrags von den Autoren ein Konzept mit Hilfe von drei relevanten Anwendungsfällen aus der Industrie vorgestellt. Das Konzept führt ein dezentrales Register für I4.0-Komponenten auf der Grundlage von dezentralen Identifikatoren (DIDs) [6] und der Distributed-Ledger-Technologie (DLT) ein. Das dezentrale Register erstreckt sich über alle Unternehmen und bildet eine betreiber- und herstellernerneutrale Plattform, einen sogenannten gemeinsamen Daten-Backbone, auf dem alle Unternehmen ihre VWSen registrieren können, ohne voneinander oder von Drittanbietern abhängig zu sein. Die DIDs dienen zur Identifizierung von Assets und den dazugehörigen I4.0-VWSen. Dies ermöglicht dem Benutzer den Zugriff auf das digitale Typenschild des Assets und weitere Dienste. Darüber hinaus können die Informationen, auf die sich der Identifikator bezieht, über den gesamten Lebenszyklus des Assets auf dem neuesten Stand gehalten werden. Die in dieser Arbeit beschriebenen Anwendungsfälle sind die Registrierung und das Auslesen eines Dienstes im dezentralen Register, die Referenzierung verschiedener I4.0-VWS unterschiedlicher Akteure mit nur einem Identifikator und die Eigentumsübertragung. ECLASS dient in dem vorgeschlagenen Konzept der semantisch eindeutigen Kennzeichnung von Diensten.

Der Beitrag ist wie folgt gegliedert: Zunächst werden in Kapitel 2 die thematischen Grundlagen DID und DLT vorgestellt. Im Anschluss wird das Konzept eines dezentralen Registers auf Basis von DID und DLT in Kapitel 3 und das Datenmodell zur Beschreibung eines I4.0-Dienstes in Kapitel 4 beschrieben. Kapitel 5 stellt die drei praxisrelevanten Anwendungsfälle vor. Der Beitrag schließt in Kapitel 6 mit einer Zusammenfassung und einem Ausblick ab.

2. Thematische Grundlagen

Im Folgenden werden die Grundlagen zum Thema DID und DLT beschrieben.

Dezentraler Identifikator (engl. *decentralized identifier*, kurz DID) - DIDs sind eine neue Art von Identifikatoren, entwickelt von der W3C Credential Community Group, die für die Identifizierung, Authentifizierung und Autorisierung von selbstverwalteten Identitäten (engl. *self-sovereign identities*) eingesetzt werden und dazu keine zentrale Autorität (z.B. *Identity Provider* oder Zertifizierungsstelle) benötigen [6]. DIDs sind URLs und somit auch URIs (Uniform Resource Identifier) gemäß RFC 3986 [5], die auf eine Ressource, das sogenannte DID-Dokument, in einem überprüfbareren Datenregister (z.B. DLT-Netzwerk) verweisen. Das DID-Dokument ent-

hält eine Beschreibung der kryptografischen Authentifizierungs- und Autorisierungsmechanismen sowie eine Liste der Dienste für jede Art der Interaktion mit der Identität. Der DID ist eine Zeichenkette und besteht grundsätzlich aus drei Teilen (Bild 1): dem Schema ("did"), dem Namen einer DID-Methode (engl. *method name*) und der methodenspezifischen ID (engl. *method-specific ID*). Eine DID-Methode definiert für ein bestimmtes überprüfbares Datenregister die Methoden zum Erstellen, Auflösen, Aktualisieren und Widerrufen von DIDs und den dazu gehörigen DID-Dokumenten. Darüber hinaus können auch Methoden zur Nutzung, wie z. B. der Authentifizierungs- oder Autorisierungsprozess, in einer DID-Methode beschrieben sein. Außerdem muss spezifiziert werden, wie eine global eindeutige methodenspezifische ID im jeweiligen überprüfbaren Datenregister ohne eine zentrale Instanz erzeugt werden kann. Darauf aufbauend können die Identitäten gegenseitig überprüfbare Aussagen (z. B. Qualifikationen oder Leistungen) in einer kryptografisch sicheren, datenschutzfreundlichen und maschinenlesbaren Form ausstellen (auch bekannt als *Verifiable Credentials*) [7].



Bild 1: Aufbau eines dezentralen Identifikators

Distributed Ledger Technologie (engl. *distributed ledger technology*, kurz DLT) - Ein Distributed Ledger ist ein dezentral organisiertes Transaktionsregister. In einem entsprechenden Netzwerk (DLT-Netzwerk) verwalten gleichberechtigte Netzwerkteilnehmer (Netzwerkknoten) exakte Kopien dieses Transaktionsregisters. Neue Transaktionen werden über das gesamte Netzwerk verteilt und unter Verwendung geeigneter Konsensmechanismen ein gemeinsamer Konsens über den aktuellen Stand des Transaktionsregisters erreicht. Die verteilte Haltung des Transaktionsregisters und der Einsatz von kryptographischen Verfahren ermöglichen eine transparente, manipulationssichere und ausfallsichere Architektur. [8]

Bei der Implementierung wird zwischen verschiedenen DLTs unterschieden. Die Unterscheidung hängt im Wesentlichen von der Art und Weise ab, wie eine Transaktion validiert und gespeichert wird. Die am weitesten verbreitete Technologie ist die Blockchain [9]. Eine alternative Technologie ist der Tangle [10], auch TDAG (transaction-based directed acyclic graph) genannt. Eine konkrete Implementierung ist der IOTA Tangle, welcher primär als Transaktions- und Zahlungsprotokoll für IoT-Geräte konzipiert wurde.

3. Dezentrales Register auf Basis von DID und DLT

In Bild 2 wird die Gesamtarchitektur des dezentralen Registers auf Basis von DID und DLT dargestellt. Unternehmen betreiben die VWSen verschiedener Assets in ihren privaten „*Trusted Networks*“. Die Partner in der Wertschöpfungskette, welche das Asset betreiben, wollen auf die VWS dieser Assets zugreifen. Die Betreiber kennen nur den DID der VWS, den sie aus einer eindeutigen Asset-ID (z. B. ID nach DIN SPEC 91406 [4]) ableiten können. Der Prozess der Überführung der Asset-ID in einen DID ist in [11] beschrieben. Die exakten Informationen über den Endpunkt für den Zugriff auf die VWS sind dem Asset-Betreiber nicht bekannt. Die Registrierung einer VWS (Schritt 1) erfolgt mit der Übermittlung eines DID-Dokuments, das als Teil einer Transaktion in einem Distributed Ledger gespeichert wird. Dieses Dokument wird durch die VWS-ID referenziert (Schritt 2), die in diesem System den DID darstellt und sich aus der eindeutigen Asset-ID ableiten lässt. Der Betreiber verbindet sich mit einem der Netzwerk-knoten des DLT-Systems und sucht die entsprechende Transaktion im dezentral organisierten Transaktionsregister. Aus dieser Transaktion kann das DID-Dokument extrahiert werden. Das DID-Dokument enthält eine maschinenlesbare Beschreibung, wie auf die VWS zugegriffen werden kann. Der Zugriff auf die VWS erfolgt von außerhalb des DLT-Netzwerks (Schritt 3).

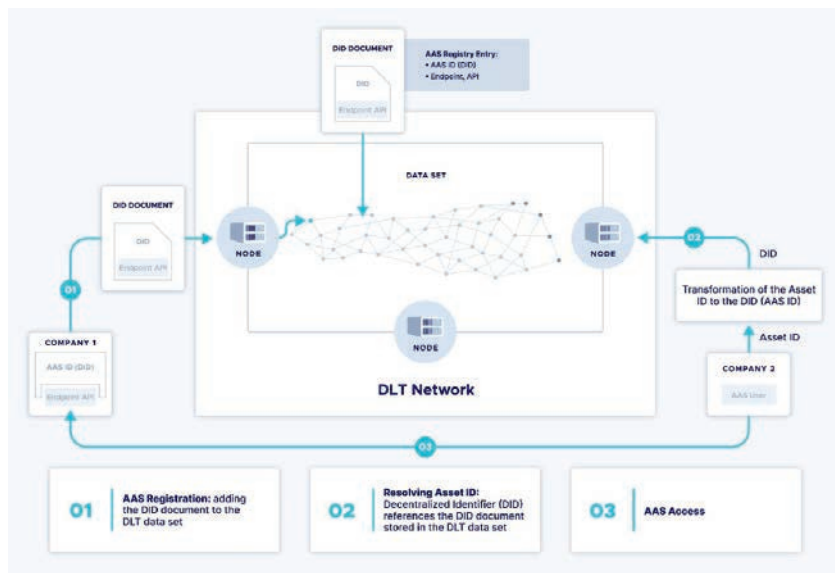


Bild 2: Dezentrales Register: Architektur und Interaktionen der beteiligten Akteure.

4. Definition eines Dienstes gemäß W3C DID-Spezifikation

Alle Dienste, die mit einer VWS in Verbindung stehen, werden innerhalb des dezentralen Registers in dem entsprechenden DID-Dokument aufgeführt. Für die Definition von Diensten ist der ECLASS-Katalog ein wichtiger Baustein, um sicherzustellen, dass die Dienste semantisch korrekt und maschinenlesbar beschrieben werden. Das DID-Dokument kann von einem Nutzer über die in Kapitel 3 beschriebenen Mechanismen abgerufen werden. Sobald der Nutzer das DID-Dokument abgerufen hat, sollte dieser in der Lage sein, alle von der VWS angebotenen Dienste aufzulisten und alle notwendigen Informationen für den Zugriff auf die Dienste vorzufinden. Das in [6] beschriebene Datenmodell gibt dem DID-Controller vor, wie das entsprechende DID-Dokument zu strukturieren und aufzubauen ist, um Kompatibilität und semantische Korrektheit sicherzustellen. Der DID-Controller ist in der Lage, Änderungen an einem DID-Dokument vorzunehmen. [6]

Die in diesem Beitrag vorgeschlagene Erweiterung zur Beschreibung von I4.0-Diensten ist in Bild 3 dargestellt. Da eine VWS mehrere verschiedene Dienste anbieten kann, muss das dezentrale Register eine Liste all dieser Dienste führen. Jeder Eintrag beschreibt einen einzelnen Dienst und ist über den Kontext definiert. Die Kontextinformationen stellen sicher, dass zwei Systeme, die mit demselben Dokument arbeiten, eine gemeinsam vereinbarte Terminologie verwenden. Der Kontext wird durch eine ECLASS-IRDI angegeben, die auf eine offene Standardspezifikation verweist, welche die Bedeutung zusätzlicher Eigenschaften definiert (z.B. ein ECLASS-Wert für die REST-API Spezifikation der VWS). Der Nutzer benötigt manchmal zusätzliche Informationen, um auf den Dienst zugreifen zu können. Diese zusätzlichen Informationen werden im DID-Dokument durch die Verwendung von Eigenschaften beschrieben (z.B. *protocol*, *authority*, *basePath*). Der Kontext gibt vor, welche Eigenschaften erlaubt sind. Im Falle eines digitalen Dienstes enthalten die Eigenschaften des Dienstes Informationen über den Endpunkt, an dem der Dienst abgerufen werden kann. Einige Eigenschaften sind obligatorisch, andere wiederum optional. Das dezentrale Register und die hier beschriebene Definition von Diensten im DID-Dokument sind Schlüsselkonzepte, welche von allen im Folgenden beschriebenen Anwendungsfällen verwendet werden.

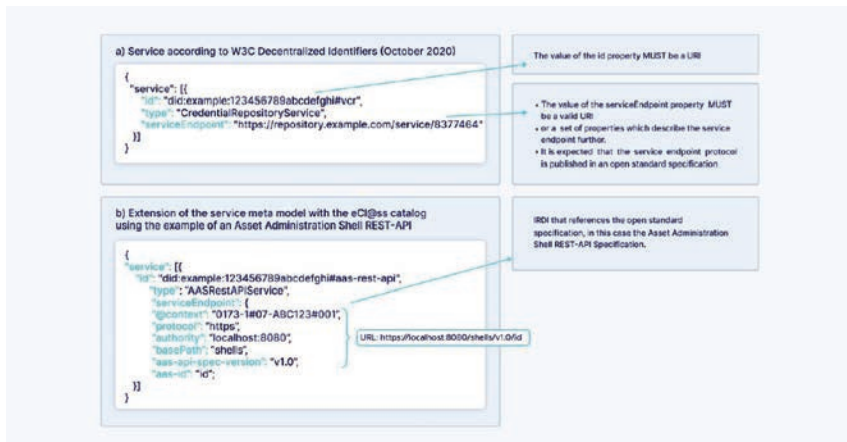


Bild 3: JSON-Serialisierung eines Dienstes im DID-Dokument

5. Anwendungsfälle

Anwendungsfall 1: Die Registrierung und das Auslesen eines Dienstes im dezentralen Register

In der täglichen Arbeitsroutine von Servicetechnikern nimmt das Erfassen von Informationen einen erheblichen Teil ihrer Arbeitszeit ein. Anstelle der zeitaufwändigen und mühsamen manuellen Suche nach relevanten Informationen in physischen und digitalen Datenbanken soll die Suche durch den Einsatz innovativer Technologien, wie dem digitalen Typenschild, z. B. zur Wartung einer Maschine, beschleunigt werden. Die Mitarbeiter müssen in der Lage sein, auf die aktuellsten Informationen eines Assets zugreifen zu können, ohne dass sie zeitaufwändig nach den Dokumenten suchen müssen. Das in [3] beschriebene Projekt „Digitales Typenschild“ verwendet einen Identifikator entsprechend der DIN SPEC 91406 [4]. Dieser global eindeutige Identifikator (Asset-ID) ermöglicht es, das physische Asset mit seinen digitalen Informationen (VWS) zu verknüpfen. Die Verknüpfung der VWS mit dem Asset kann dadurch erfolgen, dass die VWS unter der entsprechenden URL nach DIN SPEC 91406 bereitgestellt wird. Jedoch ergeben sich folgende Nachteile:

- Die Asset-ID ist statisch und kann nicht geändert werden. Die VWS kann nicht an einen anderen Ort (Domain) verschoben werden, da die Asset-ID, die auf das bereits im Feld eingesetzte Asset gedruckt ist, nicht mehr geändert werden kann.

- Das URL-Schema bindet das Asset an den Inhaber der Domain. Eine Übertragung ist mit Aufwand verbunden.

Dieser Anwendungsfall beschreibt, wie alle relevanten Informationen den Mitarbeitern sofort zur Verfügung gestellt werden können, wo auch immer diese gespeichert sind. Dazu registriert der Gerätehersteller zunächst sein Asset im dezentralen Register, indem dieser aus der Asset-ID eine DID ableitet und ein DID-Dokument hochlädt, das eine Liste aller Dienste enthält. Der Servicetechniker kann die Dienste des Maschinenherstellers in Anspruch nehmen, indem er mit einem Lesegerät (z. B. Smartphone) die Kennung (Asset-ID) ausliest, die auf dem Asset in Form eines 2D-Codes angebracht ist (siehe Bild 4). Der Resolver (z. B. eine Smartphone-Applikation) wandelt die Asset-ID in die entsprechende global eindeutige DID um und ruft das aktuellste, vom Hersteller bereitgestellte DID-Dokument aus dem dezentralen Register ab. In diesem Anwendungsfall werden ECLASS IRDIs verwendet, um sicherzustellen, dass die Dienste semantisch korrekt und maschinenlesbar beschrieben sind. Dazu werden ECLASS-Werte (Verweis auf offene Standards und Richtlinien), ECLASS-Dienstbeschreibungen (Identifizierung der Art eines Dienstes, z.B. Typenschild- oder Autorisierungsdienst) und ECLASS-Eigenschaften verwendet. Durch die Nutzung eines dezentralen Registers ergeben sich folgende Vorteile für alle Beteiligten:

- Der Registrierungsprozess ist unabhängig von einer zentralen Instanz, die die Möglichkeit hat, die Prozesse zu beeinflussen.
- Die Gültigkeit des referenzierten DID-Dokuments ist gewährleistet, da es auch dann aktualisiert werden kann, wenn sich die Domäne eines Dienstes ändern sollte.
- Das DID-Dokument ist immer verfügbar. Es gibt keinen Single Point of Failure.
- Die Registrierung muss nur in einem dezentralen Register erfolgen. Es besteht keine Notwendigkeit, das Asset in mehreren privaten Registern zu registrieren, um es für möglichst viele potenzielle Nutzer sichtbar zu machen. Dementsprechend können die Verwaltungskosten auf ein Minimum reduziert werden.
- Die Daten einer VWS müssen dort verfügbar sein, wo sie benötigt werden. Voraussetzung dafür sind die entsprechenden Zugriffsrechte. Mit firmeninternen Registern wäre dies kaum denkbar und praktisch unmöglich.

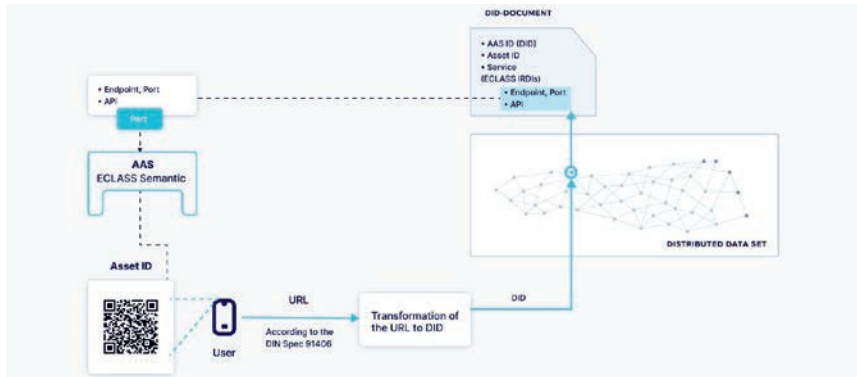


Bild 4: Auflösen einer DID

Anwendungsfall 2: Referenzierung verschiedener I4.0-VWS unterschiedlicher Akteure mit nur einem Identifikator

Für einen Betreiber ist es wichtig, auf die vom Hersteller bereitgestellten Daten zuzugreifen. Basierend auf Anwendungsfall 1 kann ein Betreiber oder ein beliebiger Partner in der Wertschöpfungskette alle Dienste und die dazugehörigen Informationen, die vom Hersteller bereitgestellt werden (z.B. Informationen im digitalen Typenschild), problemlos in einer strukturierten und klar definierten Weise finden. Sobald der Betreiber das Asset betreibt, werden neue Arten von Informationen erzeugt (z.B. Einsatzort, Verknüpfung zu weiteren Assets oder Betriebsdaten). Somit entsteht eine neue betreiberspezifische VWS. Diese soll ebenfalls im Diensteregister gespeichert und aufgelistet werden, damit z.B. Service-Dienstleister darauf zugreifen können. Im Folgenden wird ein Konzept vorgeschlagen (siehe Bild 5), wie eine neue VWS mit dem Asset verknüpft werden kann, ohne dem Asset einen zweiten betreiberspezifische Identifikator hinzuzufügen.

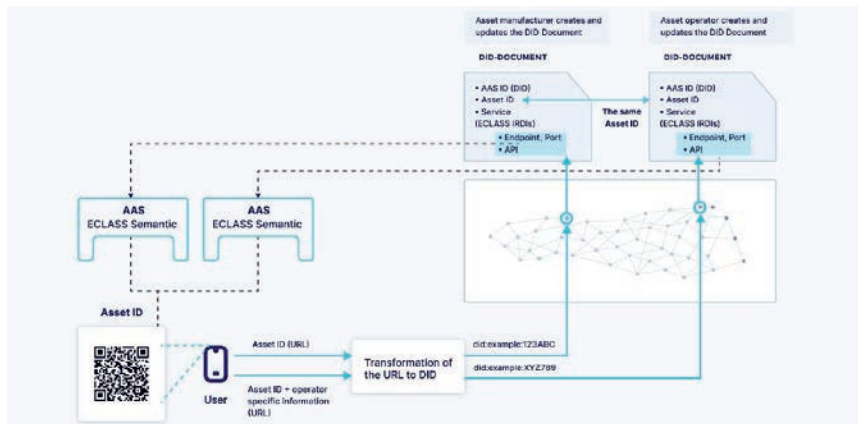


Bild 5: Mehrere VWS, die durch einen Asset-ID referenziert werden

Ein Identifikator, der auf einem Asset angebracht ist, wird immer global und eindeutig in einen DID umgewandelt und referenziert genau ein DID-Dokument im dezentralen Register. Dieses DID-Dokument kann nur vom DID-Controller des DIDs aktualisiert werden. Dies führt dazu, dass der Betreiber nicht in der Lage ist, das vom Hersteller kontrollierte DID-Dokument auf die eigenen Bedürfnisse zu erweitern. In Bild 6 wird eine mögliche Vorgehensweise für die Verknüpfung von unterschiedlichen VWS mit nur einer Asset-ID dargestellt. Der Resolver liest die ursprüngliche Asset-ID aus, erweitert diese um betreiberspezifische Informationen („operator“ in Bild 6) und ruft das entsprechende DID-Dokument auf. Die betreiberspezifischen Informationen müssen der Resolver-Applikation bekannt sein.



Bild 6: Verknüpfung der Asset-ID mit den Informationen der verschiedenen Akteure über den Lebenszyklus des Assets

Durch die Erweiterung des Identifikators ergeben sich folgende Vorteile:

- Alle Informationen sind über einen einzigen Identifikator abrufbar.
- Der Betreiber kann die vom Hersteller bereitgestellten Informationen leicht erweitern.
- Hersteller und Betreiber können unabhängig voneinander ihre VWS-Dienste durch Änderung des DID-Dokuments aktualisieren.
- Diese Erweiterung ist nicht auf einen einzelnen Betreiber beschränkt, sondern kann verwendet werden, um für jeden Partner in einer Wertschöpfungskette eine VWS anzuhängen (mögliche Standardisierung durch eine ECLASS Werteliste).

Anwendungsfall 3: Eigentumsübertragung

Die Eigentümerschaft eines Assets kann sich während seines Lebenszyklus ändern. Zum Beispiel könnte die Gerätesparte eines Herstellers an einen anderen Hersteller verkauft werden. Alle Informationen innerhalb des DID-Dokumentes aller bisher hergestellten Geräte müssen nun auf neue (Service-) Endpunkte aktualisiert werden. Der ursprüngliche Hersteller muss das DID-Dokument aktualisieren oder das Recht zur Aktualisierung an den neuen Besitzer übergeben, ohne den auf dem bereits in Betrieb befindlichen Gerät aufgedruckten Identifikator zu ändern. Auf diese Weise hat der neue DID-Controller die Möglichkeit, alle Dienste und Eigenschaften auf seine Bedürfnisse zu aktualisieren, ohne auf den früheren Controller angewiesen zu sein. Der auf einem Asset angebrachte Identifikator kann immer global und eindeutig in ein einziges DID umgewandelt werden und referenziert somit genau ein DID-Dokument im dezentralen Register. Dieses DID-Dokument wird von einer Entität in der Wertschöpfungskette kontrolliert. Derjenige, der die Registrierung zuerst durchführt, ist der Controller des DID. In der Regel wird dies der Asset-Hersteller sein. Wenn in irgendeinem Fall die Kontrolle über den DID auf einen neuen Eigentümer übertragen werden muss, so kann eine Eigentumsübertragung ausgeführt werden. Es gibt zwei Möglichkeiten, dies zu erreichen: (a) vom Hersteller verwaltet und (b) vom neuen Besitzer verwaltet. In (a) registriert der Hersteller das Asset im dezentralen Register und aktualisiert das DID-Dokument bei Bedarf auf Wunsch des Eigentümers. Dies setzt voraus, dass der Hersteller auch nach längerer Zeit noch bereit ist, das DID-Dokument zu aktualisieren. In (b) aktualisiert der Hersteller das DID-Dokument und überträgt gleichzeitig die Kontrolle über den Aktualisierungsvorgang (siehe Bild 7). Die Kontrolle über den Aktualisierungsvorgang wird ermöglicht, indem die kryptographischen Informationen (öffentlicher Schlüssel, Signatur) im DID-Dokument an den neuen Eigentümer angepasst werden.

Akzeptanz der verteilten Ansätzen in der I4.0-Community zu steigern und die Transformation der aktuellen Konzepte zu dezentralen Lösungen zu ermöglichen, würde es sich aus Sicht der Autoren anbieten, im nächsten Schritt einen hybriden Ansatz auszuarbeiten. Dieser würde es ermöglichen, von in einem dezentralen Register abgelegten DID-Dokumenten aus die Dienste eines zentralen Authentifizierungsservers in Anspruch zu nehmen.

Quellen

- [1] Plattform Industrie 4.0 (2019). Leitbild 2030 für Industrie 4.0. Digitale Ökosysteme global gestalten. Berlin: Bundesministerium für Wirtschaft und Energie (BMWi).
- [2] Colombo, A. W., Karnouskos, S., Kaynak, O., Shi, Y., & Yin, S. (2017). Industrial cyber-physical systems: A backbone of the fourth industrial revolution. *IEEE Industrial Electronics Magazine*, 11(1), 6-16.
- [3] Dogan, A., Zhang, Y., & Fay, A. (2019). Das vereinheitlichte digitale Typenschild: Anforderungen und Konzeption des Informationsmodells. *atp magazin*, 61(11-12), 92-101.
- [4] DIN SPEC 91406 (2019). Automatische Identifikation von physischen Objekten und Informationen zum physischen Objekt in IT-Systemen, insbesondere IoT-Systemen. Berlin: Beuth Verlag.
- [5] Berners-Lee, T., Fielding, R., & Masinter, L. (1998). Uniform resource identifiers (URI): Generic syntax.
- [6] W3C (2021). Decentralized Identifiers (DIDs) v1.0. Core architecture, data model, and representations. Online erhältlich unter: <https://www.w3.org/TR/did-core/>; abgerufen am 03. Mai 2021.
- [7] W3C (2019). Verifiable Credentials Data Model 1.0. Expressing verifiable information on the Web. Online erhältlich unter: <https://www.w3.org/TR/vc-data-model/>; abgerufen am 03. Mai 2021.
- [8] International Organization for Standardization. (2020). ISO 22739:2020 - Blockchain and distributed ledger technologies - Vocabulary.
- [9] Swan, M. (2015). *Blockchain: Blueprint for a new economy*. "O'Reilly Media, Inc."
- [10] Popov, S. (2018). The Tangle. Online erhältlich unter: https://assets.ctfassets.net/r1dr6vzfzhxhev/2t4uxvslqk0EUau6g2sw0g/45eae33637ca92f85dd9f4a3a218e1ec/iota1_4_3.pdf; abgerufen am 03. Mai 2021.
- [11] Dogan, A., Millenaar, J., & Fay, A. (2020). DID-basiertes Identitätsmanagement für Industrie-4.0-Komponenten in dezentralen Systemen. Tagung „Entwurf komplexer Automatisierungssysteme“ (EKA). Magdeburg

Ein Konzept für Produkt-Verwaltungsschalen zur Produktbezogenen Integration von Produktionsdaten für die Qualitätsoptimierung

A proposal of Asset Administration Shells for products, allowing product-related integration of production data for quality optimization purposes

M. Thies M.Sc., **T. Miny** M.Sc., Prof. Dr.-Ing. **T. Kleinert**,
RWTH Aachen University, Aachen

Kurzfassung

Für die Optimierung von Produktionsprozessen über mehrere Prozessschritte hinweg ist die Integration von Prozessdaten anhand der produzierten Produkteinheiten notwendig. In diesem Beitrag wird ein Konzept für die automatisierte Produkt-bezogene Integration der Daten auf Basis der Industrie-4.0-Verwaltungsschale entwickelt. Dazu werden die Modellierung der Produktidentität und mögliche System-Architekturen diskutiert und entsprechende Teilmodelle für die Verwaltungsschale modelliert.

Abstract

For the optimization of production processes across several process steps, the integration of process data based on the produced product units is required. In this paper, a concept for the automated product-based data integration using the Industrie 4.0 Asset Administration Shell is developed. For this purpose, the modeling of product identity and possible system architectures are discussed and appropriate submodels for the Asset Administration Shell are modeled.

1. Einleitung

Bereits seit einiger Zeit werden Methoden des Data Minings und Machine Learnings für die Daten-getriebene Optimierung von Produktionsprozessen erforscht [1]. Durch die Fortschritte in der Digitalisierung von Fertigungs- und prozesstechnischen Anlagen können zunehmend Daten aus Produktionsprozessen für diese Analysemethoden bereitgestellt werden. Eine Herausforderung stellt dabei die Zusammenführung der Daten aus verschiedenen Quellen dar, die in verschiedenen Systemen und Formaten vorliegen. Sollen mehrere Schritte des Produktionsprozesses gemeinsam analysiert und optimiert werden, müssen so etwa die von den

jeweiligen Maschinen und Anlagen erfassten Daten und Qualitätsmessungen anhand der produzierten Produkteinheiten zusammengeführt werden.

Für die Automatisierung dieser Produkt-bezogene Erfassung von Produktionsdaten sind verschiedene Voraussetzungen nötig: Zunächst ist ein Modell für die Identität der Produkte im betrachteten Produktionsprozess notwendig, das beschreibt, welche Teile im Materialfluss gemeinsam als *eine* Produkteinheit aufgefasst werden und wann ihr Lebenszyklus beginnt und – durch Übergang in eine neue Produkteinheit – endet. Dieses Modell erlaubt es, Produkteinheiten zu bezeichnen und so in Informationssystemen zu referenzieren, z.B. durch eine Identifikation-Zeichenkette (ID). Weiterhin ist es notwendig, die aktuelle Position der Produkteinheiten innerhalb der Produktionsanlage im Produktionsprozess zu verfolgen. Abhängig von Produktidentitäts-Modell, der Art des Produkts und der Sensorik in der Produktionsanlage kann dies anhand physischer Identifikationsmerkmale (z. B. RFID oder Barcode) oder (ggf. ergänzend) durch die Produktverfolgung anhand von Logistik-Modellen erfolgen.

Für die Erfassung der relevanten Produktionsdaten wird eine geeignete Infrastruktur benötigt, die es erlaubt, die Daten von verschiedenen Maschinen und Anlagen sowie ggf. zusätzlicher Sensorik zu übertragen, zusammenzuführen und dabei den Kontext der Datenentstehung, insb. Zeitpunkt und Ort der Datenaufnahme, zu erhalten. Mithilfe der Identifikation und Verfolgung von Produkteinheiten ist die automatisierte Verknüpfung dieser Daten mit den Produkteinheiten möglich. Abhängig von der Form der Daten und Produkteinheiten muss das Software-System und das zugrundeliegende Informationsmodell für diese Verknüpfung verschiedene Beziehungen zwischen Produkteinheiten und Datensätzen herstellen können. Schließlich müssen die Daten in einem Archivsystem unter Erhalt aller Kontext-Informationen, eingeschlossen der Verknüpfung mit Produkteinheiten, gespeichert und mit Bezug zum Kontext für weitere Analysen bereitgestellt werden.

Nach Kenntnis der Autoren wurde diese Zusammenführung der Daten in bisherigen Forschungsarbeiten manuell vorgenommen oder speziell für einen gegebenen Produktionsprozess realisiert. Ein Ansatz zur Verallgemeinerung und generischen Automatisierung der Produkt-bezogenen Datenintegration wurde bisher nicht veröffentlicht.

In diesem Beitrag wird daher ein verallgemeinertes Konzept zur Integration der Daten aus dem gesamten Produktionsprozess einer identifizierbaren Produkteinheit und ggf. ihrer Vorprodukte vorgestellt. Ziel ist es, ohne manuellen Konfigurationsaufwand Daten für die Optimierung des Prozesses bereitzustellen. Insbesondere werden die Aspekte der Modellierung der Identität von Produkteinheiten sowie der automatisierten Zuordnung und Bereitstellung von Produkt-bezogenen Daten näher untersucht. Bei der Modellierung der Produktidentität stehen Möglichkeiten zur Beschreibung der Identität von Einzelstücken bei Füge- und

Mischungsvorgängen und die daraus resultierenden Konsequenzen für die Datenhaltung im Fokus. Für die Produkt-bezogene Zusammenführung und Bereitstellung der Daten wird die Industrie-4.0-Verwaltungsschale genutzt, wobei die Produkte als Assets aufgefasst werden, sodass Verwaltungsschalen für jede Produkteinheit und jeden Produkttyp erstellt werden. Im Lebenszyklus der Produktdaten steht dabei die Entstehungsphase im Vordergrund.

Die weiteren Inhalte dieses Beitrags gliedern sich wie folgt: In Abschnitt 2 werden zunächst bekannte Teillösungen für die oben beschriebenen Funktionen sowie die konzeptuellen Grundlagen der Verwaltungsschale vorgestellt. Die neu entwickelte Modellierung der Identität von Produkteinheiten im Materialfluss und der Produktionsprozesskette wird in Abschnitt 3 diskutiert. Darauf aufbauend beschreibt Abschnitt 4 den neuen Ansatz zur Anwendung der Verwaltungsschalen zur Produktionsdatenerfassung, mit den dafür notwendigen Verwaltungsschalen-Teilmodellen sowie einer darauf ausgerichteten Systemarchitektur zur Datenerfassung. Nach einer Diskussion des Ansatzes in Abschnitt 5 fasst Abschnitt 6 die vorgestellten Konzepte zusammen und gibt einen Ausblick.

2. Stand der Technik

Dieser Abschnitt diskutiert bekannte Konzepte und Teillösungen für die zuvor erwähnten Funktionen, die für die Umsetzung einer Produkt-bezogenen Erfassung von Produktionsdaten erforderlich sind. Zudem wird das Konzept der Verwaltungsschale vorgestellt, das die Basis für die in den folgenden Abschnitten vorgestellte Lösung bildet.

2.1. Identifikation von Produkteinheiten

Für die Abgrenzung und Bezeichnung bzw. Identifikation von Produkteinheiten sind bereits zahlreiche Konzepte bekannt, insbesondere die Verwaltung von Chargen- und Seriennummern aus der Materialwirtschaft. Als Charge wird dabei üblicherweise eine Menge von produzierten Teilen oder Rohmaterialien bezeichnet, die durch die einmalige Durchführung eines zusammenhängenden Produktionsprozesses gefertigt wurde bzw. gemeinsam bestellt wurden [6]. Marktübliche ERP-Systeme erlauben es, Chargen durch Chargennummern zu identifizieren, Lagerbestände Chargen-bezogen zu verwalten und die eingesetzten Materialchargen von Rohmaterial und Vorprodukten als Materialeinsatz für einen einzelnen Fertigungsauftrag zu verbuchen, der zur Produktion einer neuen Produktcharge führt (z. B. [7]). So ist eine Chargenrückverfolgung möglich, die in einigen Branchen gesetzlich vorgeschrieben ist.

Über die Identifikation von Chargen hinaus können einzelne Produkte durch eine Seriennummer identifiziert werden. Dies ist vor allem für komplexere Baugruppen und Geräte üblich, die ein fertiges Produkt oder Teil eines fertigen Produktes darstellen. Marktübliche ERP- und

MES-Systeme unterstützen die Vergabe von Seriennummern, Verfolgung von dadurch identifizierten Produkteinheiten und die Speicherung grundlegender Informationen über die Produkteinheit. Durch die Hinterlegung von Materialstücklisten können Teil- bzw. Vorprodukte über ihre Seriennummern mit einer Produkteinheit verknüpft werden.

Für die eindeutige Identifikation von Produkten über Unternehmensgrenzen hinweg, z. B. auch beim Nutzer des Produktes oder beim Recycling, muss ein einheitliches Schema zur Erzeugung der Identifikationsmerkmale verwendet werden. Ein solches Schema zur weltweit eindeutigen Bezeichnung von physischen Gegenständen durch Zeichenketten, die z. B. in 2D-Barcodes für Auto-ID-Verfahren (vgl. Abschnitt 2.2) kodiert werden können, wird in [9] definiert. Die vorgestellten Konzepte bieten jedoch nur eine grobe Granularität bei der Abgrenzung der Produkteinheiten, insbesondere bei Produkten, die nicht als Einzelstücke gehandhabt werden. Zudem mangelt es an Konzepten für die konsistente Modellierung der Produktidentität in Füge- bzw. Mischungs- und Trennvorgängen.

2.2. Produktverfolgung

Die Verfolgung von Produkten wird bislang hauptsächlich mit Verfahren zur automatischen Identifikation (Auto-ID) gelöst. Diese ermöglichen es, die Identität einer Produkteinheit anhand von physischen Identifikationsmerkmalen wie z. B. aufgebrachten Barcodes, oder RFID-Transpondern automatisch festzustellen. Dazu existieren verschiedene Verfahren mit unterschiedlichen Vor- und Nachteilen [8].

2.3. Daten-Zuordnung und -Kontextualisierung

Die Erfassung und Zusammenführung von Daten aus verschiedenen Quellen sind an verschiedenen Stellen Gegenstand aktueller Forschung. Auf Chargen-Ebene wird in der Prozesstechnik bereits durch die ISA 88/DIN EN 61512 [6] eine Chargenhistorie vorgesehen, die neben den eingesetzten Teilanlagen und Materialien auch während der Produktion angefallene Ereignisse und Daten für jede produzierte Charge erfasst.

Unter Kontextualisierung wird die Einordnung der Daten in den Kontext, der bei ihrer Entstehung gegolten hat, verstanden, wobei der Kontext insbesondere zeitliche Zusammenhänge, Umgebungs- und Strukturzusammenhänge (z. B. Anlagenteil/Maschine, Prozessschritt), semantische Zusammenhänge (Einordnung in Ontologie-Modelle) und Situationszusammenhänge (z. B. Anlagen-Betriebszustand) umfasst [2]. Die Zuordnung der Produktionsdaten zu den zugehörigen Produkteinheiten kann als Spezialfall der Kontextualisierung von Prozess- bzw. Produktionsdaten bezüglich des Situationszusammenhangs betrachtet werden. Verschiedene Anwendungsbeispiele für Daten-Kontextualisierung in prozesstechnischen Anlagen

sowie Strategien zur Kontextualisierung werden in [2] vorgestellt. [10] zeigt ein Konzept zur Kontextualisierung und Verknüpfung von Prozessdaten auf Basis einer Graph-Datenbank. Für den Spezialfall der automatisierten Zuordnung von Daten zu Produkteinheiten sowie Speicherung und Bereitstellung der Daten mit Produktbezug sind bislang wenige Konzepte verfügbar. Einzelne Arbeiten zur Online-Erstellung von „Digitalen Zwillingen“ berücksichtigen bereits Informationen über Produkteinheiten (z. B. [12]); eine generalisierte Lösung zur Datenzuordnung ist den Autoren jedoch nicht bekannt.

2.4. Verwaltungsschale und Teilmodelle

Die Verwaltungsschale wurde als Konzept zur digitalen Repräsentanz von Assets entwickelt, um einen einheitlichen Zugriff auf die Informationen und Dienste eines Assets zu ermöglichen. Als Asset werden dabei grundsätzlich alle werthaltigen Betrachtungsgegenstände bezeichnet und somit neben physischen Gegenständen (z. B. Maschinen, Geräte) insbesondere auch Gegenstände der Informationswelt, wie z. B. Planungsdaten und Steuerungsprogramme [3]. Durch die Spezifikation „Verwaltungsschale im Detail“ der Plattform Industrie 4.0 [4], wird die Verwaltungsschale als einheitliches Informationsmodell für verschiedenartige Daten zu jeweils einem Asset ausgeprägt, ergänzt um Dateiformate zum Austausch bzw. Schnittstellen zur Bereitstellung der Daten. Das Informationsmodell sieht die semantische Unterteilung der Daten in „Teilmodelle“ vor, die jeweils einen Aspekt bzw. einen zusammenhängenden Datensatz zur Beschreibung des Assets enthalten. Teilmodelle können modular ergänzt und zur eindeutigen Interpretierbarkeit von standardisierten „Templates“ abgeleitet werden.

Innerhalb der Teilmodelle werden Daten und Informationen in verschiedenen Datenobjekten gespeichert, die neben den Nutzdaten verschiedene Attribute für Metadaten enthalten, wie z.B. die „semanticId“ für eine maschinenlesbare Referenz und die „description“ für eine menschenlesbare Beschreibung der Bedeutung des jeweiligen Datums. Zusätzliche Informationen über das Datenobjekt können in Form von „Qualifier“-Objekten angehängt werden.

Bisherige Forschungsarbeiten, wie z. B. [5], beschäftigen sich hauptsächlich mit der Repräsentation von Produktionsanlagen durch Verwaltungsschalen während der Nutzungsphase. In diesem Beitrag wird demgegenüber die Erstellung von Verwaltungsschalen mit Bezug auf in der Produktion befindliche Produkte sowie deren Produkttypen betrachtet.

3. Produktidentität

In diesem Abschnitt wird ein Modell für die Produktidentität vorgestellt, das beschreibt, wie Produkteinheiten in einem Materialfluss voneinander abgegrenzt werden und ob bzw. wann sich ihre Identität durch die durchlaufenen Produktionsschritte verändert. Ziel ist es, während

der Produktion für jede damit beschriebene Produkt(id)entität eine eindeutige Identifikations-Zeichenkette zu vergeben und einen Datensatz in Form einer Verwaltungsschale anzulegen.

3.1. Abgrenzung von Produkteinheiten im Materialfluss

Im Materialfluss durch eine fertigungs- oder prozesstechnische Anlage, können je nach Beschaffenheit des produzierten Produktes verschiedene Produkteinheiten mit unterschiedlicher Granularität abgegrenzt werden. Typisch ist die Abgrenzung und informationstechnische Erfassung von einzelnen Produktionschargen, teilweise auch von Verpackungseinheiten oder physisch getrennten Einzelstücken (vgl. auch Abschnitt 2.1).

In diesem Beitrag wird vereinfachend von Einzelstückfertigung ausgegangen, bei der sich ein Fertigungsschritt nur auf einzelne Stücke auswirkt (vgl. auch Klassifikation der „Bekanntheit im Informationssystem“ nach [3]). Einzelnen Produkteinheiten sind somit physisch voneinander abgegrenzt und können separat gehandhabt und verarbeitet werden. Um für Schüttgüter, Flüssigkeiten, Gase oder kleine Einzelstücke, die nicht unabhängig gehandhabt oder bearbeitet werden, eine ähnlich feingranulare Abgrenzung innerhalb einer Charge zu erreichen, wird ein komplexeres Materialflussmodell zur Verfolgung der Produkteinheiten benötigt. Dieses ist jedoch nicht Bestandteil dieser Arbeit, sondern erfordert weitere Forschung.

Grundsätzlich ist es möglich, verschiedene Stufen der Produktidentifikation parallel zu nutzen und die anfallenden Produktionsdaten „so genau wie möglich“ zuzuordnen. So kann etwa neben Verwaltungsschalen für jedes produzierte Einzelstück auch eine Verwaltungsschale für jede Charge angelegt werden. Werden Einzelstücke durch eine Referenz auf die Chargen-ID mit der Charge verknüpft, so können Chargen-bezogene Informationen gemeinsam mit Einzelstück-bezogenen Informationen ausgewertet werden.

3.2. Modellierung der Produktidentität über Produktionsschritte hinweg

In diesem Abschnitt soll diskutiert werden, ob und wie sich die Identität einer Produkteinheit innerhalb einer Kette von Produktionsschritten durch die aus den Produktionsschritten resultierenden Merkmalsänderungen sowie das Hinzufügen oder Abtrennen von Produkteinheiten verändert. Diese Fragestellung kann auf das bereits in der Antike diskutierte philosophische Problem der Identität von Gegenständen zurückgeführt werden. Hier wird der Fokus allerdings auf die praktischen Auswirkungen der Produktidentität auf die Informationshaltung gelegt.

Die Konzepte werden im Folgenden an einem einfachen Fertigungsprozess veranschaulicht, der im Forschungsprojekt BaSys 4.2 in Form einer dynamischen 3D-Simulation als Demonstrator für die im Projekt entwickelte Industrie-4.0-Middleware verwendet wird. In diesem



Bild 3.1: Verwaltung einer Produkteinheit unter Fortführung einer vorhandenen Identität in der andere Vorprodukte aufgehen

vollautomatisierten Fertigungsprozess werden zylindrische Kunststoffdosen mit einem Deckel verschraubt. Dazu sind verschiedene Transportmechanismen für die Dosen (Roboter, Förderbänder, Rundschalttisch), zwei Fertigungsstationen für das Einsetzen des Deckels und der Schraube sowie eine Schraubstation notwendig. Die Automatisierung ist in Form einer dienstbasierten Produktion umgesetzt (vgl. [11]). Es sei angenommen, dass jeder Dosenrumpf bereits zu Beginn des Prozesses eine Identifikations-Zeichenkette (ID) besitzt, ebenso wie Deckel und die Schrauben (in Bild 3.1 beispielhaft X01, Y01, Z01).

Zunächst wird angenommen, dass sich die Identität eines Produktes während des Produktionsprozesses nicht ändert und Fügeprozesse als „Ergänzung“ zu einem bestehenden Produkt betrachtet werden können. Das heißt, der Dosenrumpf wird als Ausgangsstadium für die fertige Dose gesehen; seine ID gilt somit nach dem Zusammenschrauben auch für das Endprodukt. Deckel und Schraube „gehen in der Dose auf“; ihr Lebenszyklus als identifizierbares Einzelteil endet mit dem Zusammenfügen. Die Produktidentitäten im Verlauf der Fertigung sind in Bild 3.1 dargestellt. Zusätzlich zeigt die Bild alle Produkt-Verwaltungsschalen, zugeordnet zu dem Fertigungsstand, den die enthaltenen Informationen nach Abschluss des Prozesses repräsentieren, sowie die Referenzen zwischen den Verwaltungsschalen.

In dieser Form der Modellierung können die Lebenszyklen der Einzelteile nicht mehr separat weiterverfolgt werden, falls sie z. B. später getauscht und anderweitig weiterverwendet oder recycelt werden. Auch beim Trennen des Produktes in andere als die ursprünglichen Komponenten, die weiterverfolgt werden sollen, ist es nur sehr umständlich möglich, den Informationsstand des Produktes zum Zeitpunkt des Abtrennens zu referenzieren und so die für das abgetrennte Material relevanten Prozessdaten zu rekonstruieren.

Um den Lebenszyklus der Teilprodukte weiter zu verfolgen, können diese als identifizierbarer Bestandteil des Produktes verwaltet werden, sodass das Produkt ein sogenanntes „Composite Asset“ darstellt, das weitere Assets enthält. Das Metamodell der Verwaltungsschale sieht dazu

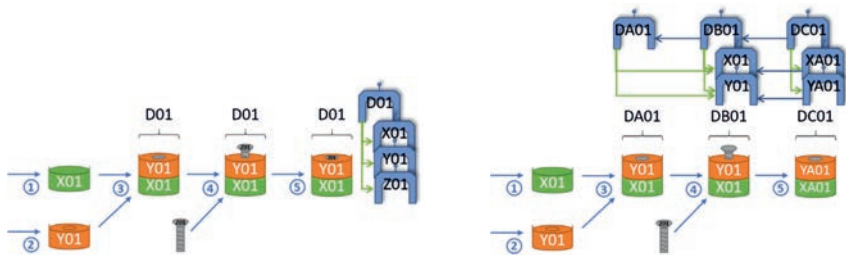


Bild 3.2: (a) Verwaltung von Vorprodukten als Teil-Asset, (b) mit zusätzlichen Identitäten nach jedem Prozessschritt

spezielle Elemente („Entity“-Objekte) vor [4]. Die Daten zu nachfolgenden Schritten im Produktionsprozess, die auf das Gesamtprodukt und damit auch auf die Teilprodukte wirken, müssen folglich in mehreren Verwaltungsschalen erfasst oder mindestens referenziert werden. Um die Teil-Asset-Beziehung auch für das ursprüngliche Ausgangsprodukt (im Beispiel: den Dosenrumpf) zu ermöglichen, ist es sinnvoll, dem Produkt nach Füge- bzw. Mischungsprozessen eine völlig neue Identität zuzuweisen. Für diese wird eine eigene Verwaltungsschale angelegt, worin die Ausgangsprodukte als Teil-Assets referenziert werden können. Dies ist in Bild 3.2 (a) dargestellt, wobei die grünen Pfeile die Teil-Asset-Referenzen darstellen.

Durch die Vergabe neuer Produktidentitäten nach einzelnen Prozessschritten kann das Problem des korrekten Referenzierens bei Trennvorgängen gelöst werden. Die vorherige Produktidentität wird dabei als Vorprodukt für diesen Schritt modelliert. Ihr Lebenszyklus endet in diesem Schritt, d.h. die zugehörige Verwaltungsschale verbleibt auf dem Informationsstand vor Ausführung des Prozessschritts. Ggf. enthaltene Teil-Assets können entweder ihre Identität behalten und in die neue Produkteinheit übernommen werden oder sie gehen ebenfalls in einem neuen Teil-Asset der neuen Produkteinheit auf. Siehe dazu Bild 3.2 (b). Im Extremfall könnte man somit für jede Produkteinheit nach jedem Prozessschritt eine neue Identität einführen. Neben der komplexeren Datenhaltung verursachen derart häufige Wechsel der Produktidentität jedoch Probleme bei der Verwendung von Auto-ID-Systemen.

Zusammenfassend können folgende Modellierungsvarianten in jedem Produktionsschritt unterschieden werden:

- Variante 1: Übernahme der Produktidentität des Vorproduktes
 - o Variante 1a: Weiterführen des Produkttyps (vorherige Schritte sind Teil des gleichen Workflows für die Produktion des Produkttyps)
 - o Variante 1b: Zuordnung zu einem neuen Produkttyp

- Variante 2: Einführen einer neuen Produktidentität (und eines neuen Produkttyps) und Auflösen des Vorproduktes (Ende des Lebenszyklus)
- Variante 3: Einführen einer neuen Produktidentität (und eines neuen Produkttyps) und weiterführen der Identität des Vorproduktes als Teil-Asset

In verschiedenen Prozessschritten können verschiedene Varianten genutzt werden, wobei innerhalb eines Füge- bzw. Mischungsprozessschrittes für verschiedenen Vorprodukte verschiedene Varianten verwendet werden können. So kann z. B. beim Zusammensetzen der Dose mit dem Deckel und der Schraube die Identität der Dose jeweils übernommen werden (Variante 1), die Schraube als Vorprodukt im Produkt aufgehen (Variante 2) und der Deckel als Teil-Asset der Dose weiterverwaltet werden (Variante 3). Welche Modellierungsvariante in jedem Prozessschritt für jedes Vorprodukt gewählt werden sollte, hängt von verschiedenen Kriterien ab, z. B. erlaubte Komplexität der Informationsmodellierung, kaufmännische und materialwirtschaftliche Kriterien, Anforderungen an die Datenhaltung im weiteren Lebenszyklus der Produkte (z. B. Wartung oder Recycling), steuerungstechnische Kriterien, usw.

4. Nutzung der Verwaltungsschale für die Produkt-bezogene Datenintegration

Die Verwaltungsschale ermöglicht es, verschiedene Daten und Informationen mit Bezug zu einem einzelnen Asset in einem gemeinsamen Informationsmodell zusammenzufassen und als Datei oder über eine Netzwerk-Schnittstelle bereitzustellen (vgl. Abschnitt 2.4). Im Rahmen dieses Beitrags werden zwei verschiedene Arten von Assets betrachtet: Zum einen Produkteinheiten gemäß Abschnitt 3, deren Produktionsprozess mit den dabei anfallenden Daten erfasst werden soll, und zum anderen die Produkttypen, zu denen die Produkteinheiten gehören. Der Produkttyp definiert die gemeinsamen Eigenschaften aller „gleichartigen“ Produkteinheiten (Produktinstanzen). Dies umfasst vor allem Planungsdaten, wie sie typischerweise in einem PDM/PLM-System verwaltet werden, z. B. Engineering-Dokumente und Fertigungsanweisungen. Für die Produktdatenerfassung ist vor allem der geplante Fertigungsablauf mit abstrakten Prozessschritten relevant, der es ermöglicht, die Prozessschritte der Produktinstanzen semantisch einzuordnen und in Beziehung zu setzen. Die von den Instanzdaten separierte Speicherung der Produkttyp-spezifischen Daten vermeidet Redundanz und vereinfacht spätere Ergänzungen und Fehlerkorrekturen.

Grundsätzlich sieht das Konzept der Verwaltungsschale die modulare Ergänzung von Daten durch einzelne Teilmodelle vor. In Abschnitt 4.1 wird jeweils ein Teilmodell für die Produkttyp- und Produktinstanz-Verwaltungsschalen für die Erfassung und Zuordnung von Produktionsdaten vorgeschlagen. Andere Teilmodelle können jederzeit in den Verwaltungsschalen ergänzt werden, z. B. das „Digitale Typenschild“ für Instanz-Verwaltungsschalen.

4.1. Verwaltungsschalen-Teilmodelle für die Produkt-bezogene Datenerfassung

Wie oben beschrieben, wird das Informationsmodell zur Produkt-bezogenen Datenerfassung in Typ- und Instanz-Daten aufgeteilt. Die erfassten Prozessdaten selbst stellen Instanz-spezifische Daten dar. Um sie jedoch semantisch einem Produktionsschritt zuzuordnen, und diese Zuordnung zwischen verschiedenen Instanzen abzugleichen, ist eine Definition der Produktionsschritte im Fertigungsablauf für den Produkttyp erforderlich.

Zu diesem Zweck wurde ein Teilmodell-Template „Workflow“ für Produkttyp-Verwaltungsschalen nach dem Metamodell der Verwaltungsschale [4] entworfen. Das entsprechend ausgefüllte Teilmodell für den Dosen-Produkttyp aus dem Beispielprozess ist in Bild 4.1 (a) dargestellt. Dabei wurde als Modellierungsansatz für die Produkteinheiten eine eigene Identität für die Gesamtdose gewählt, während alle Vorprodukte (Dosenrumpf, Deckel, Schraube) darin aufgehen und nicht weiter als Entität verfolgt werden (Variante 2 gemäß Abschnitt 3.2).

Das Workflow-Teilmodell enthält eine SubmodelElementCollection „workflow_steps“ als Liste aller Prozessschritte für die Produktion einer Produkteinheit dieses Typs. Jeder Schritt ist als SubmodelElementCollection mit einem Namen, einer Referenz auf eine semantische Definition des benötigten Maschinentyps und einer Liste der benötigten Vorprodukte und Hilfsmittel („BOM“ = Bill Of Materials) in Form von Referenzen auf die entsprechenden Typ-Assets gespeichert. In diesen Referenzen kann mithilfe der durch das Metamodell vorgesehenen Objektattribute wie description und semanticId die Semantik der einzelnen Vorprodukte maschinen- und menschenlesebar definiert werden. Für den Bezug zum Steuerungsprogramm ist eine Referenz auf einen BPMN-Workflow des Software-Frameworks Camunda vorgesehen.

Das Instanz-Teilmodell „Production Log“ enthält eine vergleichbare Liste von Prozessschritten, in der die Daten zu den tatsächlich durchgeführten oder noch bevorstehenden Prozessschritten für die jeweilige Produkteinheit erfasst werden. Ein Beispiel für die Darstellung des Production-Log-Teilmodells einer in Produktion befindlichen Dose im AASX Package Explorer ist in Bild 4.1 (b) zu sehen. Zu jedem Prozessschritt werden einige Daten einheitlich erfasst: Start- und Endzeitpunkt, eine Referenz zur abstrakten Definition des Schrittes im Workflow-Teilmodell, sowie die Maschine bzw. Anlage, die den Schritt durchgeführt hat. Weiterhin werden mit Bezug auf die Bill of Materials aus dem Workflow-Teilmodell die tatsächlich eingesetzten Vorprodukte und Hilfsmaterialien referenziert.

Die eigentliche Erfassung der Prozessdaten zum jeweiligen Schritt im Produktionsprozess der Produkteinheit erfolgt in den SubmodelElementCollections „events“ und „datapoints“. Ereignisse und Meldungen, die während des Prozessschritts aufgetreten sind, werden dabei in der Events-Collection gespeichert, erfasste Datenwerte in der Datapoints-Collection. Für die Ereignisse und Datenwerte ist kein festes Schema vorgesehen. Stattdessen werden die vorhandenen Werte mithilfe des semanticID-Attributs und zusätzlicher Qualifier-Objekte semantisch ausgezeichnet und so ihrem Entstehungskontext – über die Zuordnung zur Produkteinheit hinaus – zugeordnet, sodass eine beliebige Erweiterung um neue Datenwerte möglich ist. In der Praxis kann es sinnvoll sein, die Daten nicht wie hier gezeigt direkt im Verwaltungsschalen-Informationsmodell abzulegen, sondern in spezialisierten Datenarchiven aufzubewahren und aus den verschiedenen Verwaltungsschalen zu referenzieren. Zusätzlich zur Liste der

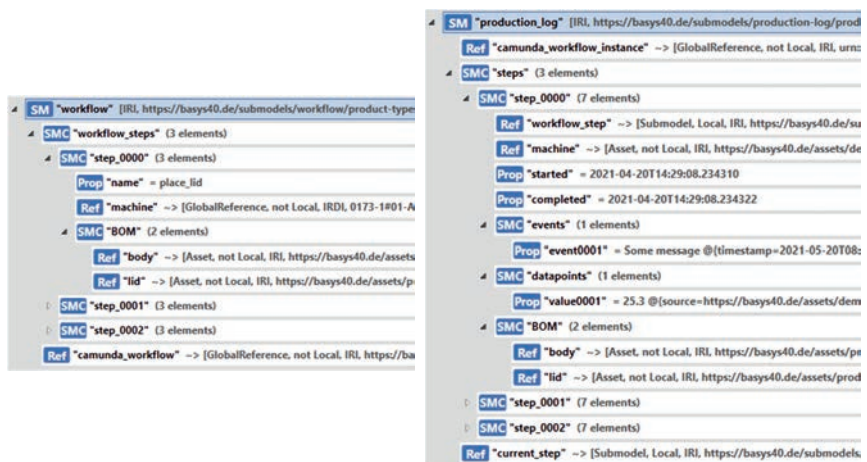


Bild 4.1: (a) Workflow- und (b) Production-Log-Teilmodell für den Beispielprozess im AASX Package Explorer

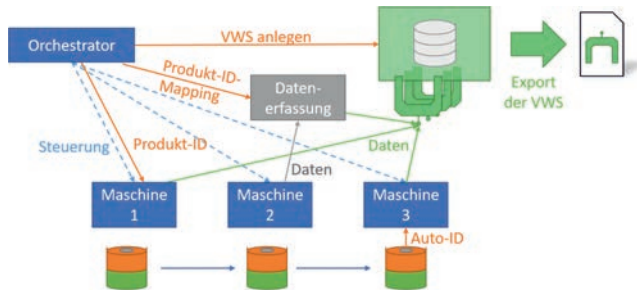


Bild 4.2: Übersicht über mögliche Systemarchitekturen zur Speicherung von Produktionsdaten in Produkt-Verwaltungsschalen

Prozessschritte enthält das Teilmodell eine Referenz „current_step“, die auf die SubmodelElementCollection des aktuellen Schrittes verweist und den Eintrag von Prozessdaten erleichtert.

4.2. Systemarchitektur & Datenerfassungsvorgang

In diesem Abschnitt wird diskutiert, wie die zuvor entwickelten Verwaltungsschalen-Teilmodelle eingesetzt werden können, um während des Produktionsprozesses die erfassten Produktionsdaten zu Produkteinheiten zuzuordnen. Dabei werden die Verwaltungsschalen als einheitliche Schnittstelle zum Abruf und Abspeichern von Asset-bezogenen Daten verwendet, d. h. konkret zum Abspeichern von Produkt-bezogenen Daten. Eine Komponente im Automatisierungssystem bietet dazu eine Schnittstelle an, über die für jede produzierte Produkteinheit eine Verwaltungsschale mit dem Production-Log-Teilmodell verfügbar ist (vgl. Bild 4.2).

Erfasste Produktionsdaten können von Steuerungskomponenten der Maschinen, vorhandenen Datenerfassungssystemen oder dem Orchestrator in die jeweils relevante Produktverwaltungsschale übertragen werden. Dafür ist es notwendig, dass die jeweilige Einheit die Identität der gerade bearbeiteten Produkteinheit kennt, um die korrekte Verwaltungsschale zu adressieren. Dies kann entweder durch am Produkt angebrachte Auto-ID-Merkmale (vgl. Abschnitt 2.2, Maschine 3 in der Bild) oder durch eine Produktverfolgung in der überlagerten Ablaufsteuerung („Orchestrator“) bzw. einem damit verbundenen Materialflussrechner erfolgen (Maschine 1 & 2 in der Bild) erreicht werden. Die Zuordnung kann dabei ggf. auch durch ein vorhandenes Datenerfassungssystem erfolgen. (Maschine 2 in der Bild).

Zunächst müssen die relevanten Verwaltungsschalen angelegt werden, d. h. die Objekte des Metamodells für eine Verwaltungsschale und die benötigten Teilmodelle erstellt und mit den grundlegenden Informationen ausgefüllt werden. Für die Verwaltungsschalen der Produkttypen kann dies manuell oder teil-automatisch durchgeführt werden, wobei das Workflow-

Teilmodell ggf. aus vorhandenen Engineering-Dokumenten generiert werden kann. Die Verwaltungsschalen der Produktinstanzen müssen dagegen vor oder bei Beginn des Produktionsvorgangs jeder einzelnen Produkteinheit angelegt werden, womit die Identität der Produkteinheit initial geschaffen wird. Dies kann – wie in Bild 4.2 dargestellt – durch den Orchestrator oder bereits vor Produktionsbeginn, z. B. beim Erfassen des Produktionsauftrags, geschehen. Während des Produktionsprozesses hält der Orchestrator die allgemeinen Daten des Produktionsablaufs im Production-Log-Teilmodell nach, insbesondere den Start- und Fertigstellungszeitpunkt der einzelnen Prozessschritte und die ausführende Maschine bzw. Anlage. Zudem aktualisiert er die „current_step“-Referenz. Erfasste Prozessdaten und Ereignisse können somit von den Maschinen bzw. Datenerfassungssystemen unabhängig in die Verwaltungsschale eingetragen werden. Für eine Produkteinheit-basierte Auswertung der erfassten Produktionsdaten müssen die in den Verwaltungsschalen hinterlegten Daten nach Abschluss des Produktionsprozesses in geeigneter Form zusammengeführt und ggf. vereinfacht werden.

5. Diskussion

Das vorgestellte Konzept der Datenintegration auf Basis von Produktverwaltungsschalen ermöglicht es, Daten und Ereignisse aus den Schritten eines Produktionsprozesses ohne manuellen Aufwand den produzierten Produkteinheiten zuzuordnen. Sofern ein Teilmodell für die Erfassung von Produktionsdaten in der beschriebenen oder einer ähnlichen Form standardisiert wird, ist zudem keine manuelle Konfiguration der Datenerfassung und -integration erforderlich. Durch die Verwendung der Schnittstelle der Verwaltungsschale ist eine einheitliche Architektur zum Aufbau des Datenerfassungssystems gegeben.

Andererseits erfordert das Konzept jedoch die Unterstützung durch die Steuerungsgeräte der einzelnen Maschinen, die für das Übermitteln der Produktionsdaten und Ereignisse eine Protokollunterstützung für die Kommunikationsschnittstelle und das Informationsmodell der Verwaltungsschale aufweisen müssen. Eine Nachrüstung bestehender Produktionsanlagen ist prinzipiell möglich, wenn Datenerfassungssysteme eingesetzt werden können, die zwischen den Steuerungssystemen und Verwaltungsschalen vermitteln. Die Verwendung der Verwaltungsschalen für die Datenerfassung erfordert bei der Auswertung der Daten zudem das Verfolgen vieler Referenzen und die Interpretation des komplexen Informationsmodells.

6. Zusammenfassung und Ausblick

In diesem Beitrag wird, nach Kenntnis der Autoren, erstmals ein Konzept für die Verwendung von Verwaltungsschalen zur Produkt-bezogenen Integration von Produktionsdaten vorgestellt. Dazu werden Produkteinheiten als Assets eingeführt und die Abgrenzung und Identität der

Produkteinheiten diskutiert. Verschiedene Modellierungsvarianten für diese Problemstellung werden entwickelt und ihre Auswirkungen auf die Informationsmodellierung mit Verwaltungsschalen analysiert. Weiterhin werden Teilmodelle für die Verwaltungsschale vorgestellt, die zur semantischen Definition des Produktionsprozesses und produktbezogenen Erfassung der Produktionsdaten und -ereignisse dienen. Es wird zudem ein Ansatz für die Umsetzung dieses Konzeptes in Form möglicher Systemarchitekturen und Abläufe der Datenerfassung gegeben. Eine weitere Forschungsaufgabe ist die Erweiterung des Konzeptes auf nicht unabhängig gehandhabte Einzelteile (wie z.B. Muttern) sowie andere Produktformen wie Flüssigkeiten und Schüttgüter. Mögliche Erweiterungen sind zudem die Berücksichtigung von gleichbleibenden Auto-ID-Merkmalen und ein Informationsmodell für die Erfassung der Veränderung von Produkteigenschaften. Durch ein zusätzliches Konzept zur einheitlichen Repräsentation von semantischen Verknüpfungen zwischen Assets, wie z.B. Einzelteilen, Vor- und Nachfolgeprodukten von Produkteinheiten, sowie durch ein Modell für die Vererbung von Informationen an andere Assets könnte zudem die Datenauswertung weiter vereinfacht werden.

Danksagung

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS19022 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin/beim Autor.

Referenzen

- [1] D. Weichert, P. Link, A. Stoll, S. Rüping, S. Ihlenfeldt, und S. Wrobel, „A review of machine learning for the optimization of production processes“, *Int J Adv Manuf Technol*, Bd. 104, Nr. 5–8, S. 1889–1902, Okt. 2019, doi: [10.1007/s00170-019-03988-5](https://doi.org/10.1007/s00170-019-03988-5).
- [2] U. Enste, A. Hartwich, C. Benthack, P. Wiemer, I. Weiß, und M. Henter, „Gib den Daten einen Kontext“, *atp*, Bd. 62, Nr. 9, S. 44–52, Sep. 2020, doi: <https://doi.org/10.17560/atp.v62i9.2478>.
- [3] „DIN SPEC 91345: Referenzarchitekturmodell Industrie 4.0 (RAMI4.0)“, DIN Deutsches Institut für Normung e. V., Berlin, Apr. 2016.
- [4] „Details of the Asset Administration Shell - Part 1 - The exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0RC01)“, Plattform Industrie 4.0, Berlin, Nov. 2020.
- [5] „Verwaltungsschale in der Praxis – Wie definiere ich Teilmodelle, beispielhafte Teilmodelle und Interaktion zwischen Verwaltungsschalen (Version 1.0)“, Plattform Industrie 4.0, Berlin, Diskussionspapier, Apr. 2019.
- [6] „DIN EN 61512-1: Chargenorientierte Fahrweise – Teil 1: Modelle und Terminologie“, DIN Deutsches Institut für Normung e.V., Deutsche Norm, 2000.
- [7] SAP SE, „SAP Help Portal: Chargennummernverwaltung“, *Chargennummernverwaltung*. <https://help.sap.com/viewer/68a2e87fb29941b5bf959a184d9c6727/9.3/de-DE/1ae2446f5f434a3288a48edcf9079763.html> (zugegriffen 27. Apr. 2021).
- [8] M. ten Hompel, H. Büchter, und U. Franzke, *Identifikationssysteme und Automatisierung*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.
- [9] „DIN SPEC 91406: Automatische Identifikation von physischen Objekten und Informationen zu physischen Objekten in IT-Systemen, insbesondere IoT-Systemen“, DIN Deutsches Institut für Normung e. V., Berlin, Dez. 2019.
- [10] C. von Trotha, T. Kleinert, und U. Epple, „Ein Konzept zur Kontextualisierung von Prozessdaten: Der erste Schritt auf dem Weg zur datenzentrierten Handlung“, *atp*, Bd. 62, Nr. 10, S. 68–76, Okt. 2020, doi: [10.17560/atp.v62i10.2502](https://doi.org/10.17560/atp.v62i10.2502).
- [11] T. Kuhn, S. Sadikow, und P. Antonino, „A Service-Based Production Ecosystem Architecture for Industrie 4.0“, *KI - Künstliche Intelligenz*, Bd. 33, Nr. 2, S. 163–169, Juni 2019, doi: [10.1007/s13218-019-00589-y](https://doi.org/10.1007/s13218-019-00589-y).
- [12] C. Cimino, E. Negri, und L. Fumagalli, „Review of digital twin applications in manufacturing“, *Computers in Industry*, Bd. 113, S. 103130, Dez. 2019, doi: [10.1016/j.com-pind.2019.103130](https://doi.org/10.1016/j.com-pind.2019.103130).

Integration von Module Type Package und Industrie 4.0 Verwaltungsschale

Integration of Module Type Package and Industry 4.0 Asset Administration Shell

Sten Grüner, Mario Hoernicke,
ABB AG Forschungszentrum, Ladenburg;
Alexander Kehl, Christian Barth, Matthias Freund,
Michael Hoffmeister, Festo SE & Co. KG, Esslingen;
Tobias Klausmann, Lenze SE, Aenzen

Kurzfassung

Um die steigenden Anforderungen bzgl. der modularen Produktion in der Prozessindustrie zu erfüllen, erarbeitet die deutsche Interessengemeinschaft Automatisierungstechnik der Prozessindustrie (NAMUR) und der Zentralverband Elektrotechnik- und Elektronikindustrie (ZVEI) eine Spezifikation für Prozessmodule unter dem Namen Module Type Package (MTP) [1][2]. Gleichzeitig entwickelt die Plattform Industrie 4.0 die Verwaltungsschale – eine interoperable Schnittstelle für den unternehmensübergreifenden Datenaustausch für Industrie 4.0 [3][4]. In dieser Arbeit zeigen wir, wie beide Konzepte kombiniert werden können, um Lücken in existierender Standardisierung zu schließen und modulare Prozessautomatisierung in das Industrie 4.0 Ökosystem einzubetten. Das vorgeschlagene MTP-Teilmodell für die Verwaltungsschale eröffnet dabei neue Anwendungsfälle, die durch die gemeinsame Verwendung von MTP und Verwaltungsschale und deren Infrastruktur entstehen.

Abstract

To serve the upcoming demand of modularized production in process industries, several working groups of NAMUR (German user association of automation technology in process industries) and ZVEI (German association of automation technology vendors) define a specification of process module interfaces called Module Type Package [1][2]. Simultaneously, a concept for an interoperable cross-company exchange information model called Asset Administration Shell is being developed under the governance of the Plattform Industrie 4.0 [3][4] in Germany. In this work, we present how both concepts can be combined to cover gaps in existing specifications and to embed MTPs and hence module process automation into the Industry 4.0

ecosystem. The proposed MTP submodel for AAS is a first step for additional use cases derived from the combination of MTP and AAS.

1. Introduction

To serve the upcoming demand of modularized production in process industries, several working groups of NAMUR (German user association of automation technology in process industries) and ZVEI (German association of automation technology vendors) define a specification of process module interfaces called Module Type Package (MTP) [2]. An MTP file (based on the AutomationML standard [5]) describes a module type, i.e. shared properties of every physical instance (Process Equipment Assembly, PEA) of this type. The file includes a description of the module's services (e.g., heating or pumping), human machine interface (HMI) and the mapping to a specific communication technology, i.e., OPC UA [6]. The MTP description files are interpreted during the engineering phase by an orchestration component of the MTP specification called POL (Process Orchestration Layer), where the engineer configures the module instances (PEAs). The next step is to create a module topology reflecting the physical layout of the production line (e.g., electrical and piped connections between the modules). Finally, a product-specific recipe is executed by the POL. Typically, a recipe is encoded as a Sequential Function Chart (SFC) which triggers services (e.g., heating) in specific modules in its states and monitors data points included in the module (e.g., temperature sensors) to initiate transitions between the states. Runtime communication between POL and PEA is realized using OPC UA which allows the POL to communicate to controllers (PLCs) within the PEAs.

Simultaneously, a concept for an interoperable cross-company exchange information model called Asset Administration Shell (AAS) is being developed under the governance of Plattform Industrie 4.0, a German consortium consisting of representatives of various IT and OT companies including electrical engineering, manufacturing and computer business associations. A unique feature of AAS is a protocol-independent information model which can be mapped to various file formats and communication protocols based on the needs of the particular digitalization use case. Examples are passive AAS (Type 1) which are exchanged using files (AASX-Packages) and reactive AAS (Type 2) which provide an API for dynamic read and write operations on AAS [7]. Type 2 data access can be realized by different communication protocols, e.g., RESTful APIs or OPC UA based on an information model specified by an OPC UA companion specification. Along with the core information model, domain-specific specifications for specific aspects of assets within the AAS are being developed. Examples for those are the

technical specification of an asset [8], e.g. its physical dimensions, and technical documentation which can be shipped within an asset's AAS [9].

In this paper, we sketch out how the two separate concepts can be interconnected by providing an overview of use cases for MTP and AAS co-existence in Section 2. Based on the extracted requirements, we describe the mapping of MTP to AAS and discuss alternatives and design decisions met for the definition of an MTP submodel for the AAS in Section 3. Furthermore, a brief overview of existing MTP submodel prototypes and tools is presented in Section 4. Finally, a summary and a list of future work topics are presented in Section 5.

2. Use Cases and Requirements

In course of the analysis of both standardization activities, we have identified two initial use cases for the usage of AAS as enhancement of existing MTP packages: *cross-company module instance information exchange* and *module type and instance documentation*. These use cases are discussed in the following sub chapters.

2.1 Cross-company Module Instance Information Exchange

Currently, MTP files include only type-specific information. During the PEA integration into the POL, additional instance-specific information, like the IP address of the OPC UA server within the PEA, is needed for operation. This requires separate initialization procedures for type and instance information. Furthermore, the instance information can be changed during the lifecycle of a PEA. Up to now, the instance information cannot be retrieved in a standardized manner.

To-be situation: Classification of type and instance information as well as references thereto during the lifecycle is one of the cornerstones of Industry 4.0 architecture. This manifests in IEC 62830 and the RAMI 4.0 model. The meta model of the AAS follows this differentiation and can be directly used to model module-specific instance and type data. Consequently, the type-specific information included in the MTP file should be available as a submodel in the AAS [9]. This submodel allows access to the MTP-file to associate the type specific information with other submodels of the AAS, e.g. documentation or technical data [7]. In this way, the components included in a module can be referenced with component-specific documentation.

Advantages of this use case are:

- Embedding of MTP- and PEA-specific information into the Industrie 4.0 ecosystem with further potential extensions e.g., searching for MTP files using Industry 4.0 tools like registries.
- Clear conceptual foundation for PEA data and a clear separation from type information of a module.
- Interoperable exchange of module data between POLs of different vendors.
- Interoperable exchange of module data between module vendors and POL as well as further OT management systems.

2.2 Module Type and Instance Documentation

Currently, the MTP concept does not provide an explicit possibility to include documentation. Documentation-related submodels are under development by the Plattform Industrie 4.0 community. Those models are based on the VDI 2770 standard. Therefore, building upon an embedding of MTP and PEA information into AAS submodels, we can enrich the corresponding AAS with additional information. One example of such information is type-specific and instance-specific documentation related to module type and PEA. This information is essential for a successful module commissioning.

To-be situation: The MTP file and type-specific AAS aspect enables the visualization and operation of module components. The documentation of the module can be divided into type- and instance-specific parts. Those parts contain module descriptions as well as manuals for components. Module type-specific documentation such as manuals are stored in the type-specific AAS whereas instance-specific documents like the site map of the operation location are stored in the instance-specific AAS. The documentation aspects of the AAS should provide links towards the corresponding components included in the MTP.

Advantages of this use case are:

- Availability of type- (e.g., technical module specifications) and instance-specific documentation (e.g., PEA commissioning protocols) in an interoperable format (documentation submodel based on VDI 2770 specification).
- Association of documentation elements with specific components of a module (e.g. operation manual for an incorporated valve)

- Re-use of existing tooling like the AASX Package Explorer¹ plug-in to view and edit documentation data for the module and its components.
- The MTP file stays unchanged, existing MTP tooling can be reused.

Based on the listed use cases, we conclude the following functional requirements:

1. Embedding one MTP file into an AAS submodel.
Definition and embedding of PEA-specific data in an AAS submodel. This includes embedding constants and variables into PEA-specific AAS like serial number (constant) or OPC UA endpoint (variable).
2. Possibility to re-use further AAS submodels, e.g., Nameplate or Documentation submodel.
3. Possibility to cross-reference single MTP elements from defined submodels. Example: attaching documentation from documentation submodels to certain elements included in an MTP file.

3. Design of MTP-related AAS Submodels

In this section, we discuss the decisions which were made during the design of the AAS submodel and the possibility of links between contents of MTP files, e.g. tags, and elements of AAS, e.g., submodels and submodel elements.

3.1 Modeling Design Decisions

A paramount design decision, to make regarding the first requirement, is the depth of modeling MTP content within AAS information model [9]. MTP is built upon AutomationML and hence provides a rich object-oriented model including different MTP aspects, e.g. module's HMI, its services, etc. Those structures can offer a baseline for mapping of MTP content to the AAS. Advantages of this re-modeling approach is the compatibility to existing AAS-specific tools and SDKs. This would also possibly allow to re-use AAS-specific infrastructure like registries to register and list MTP files contained in registered AAS.

On the other hand, re-modeling has several disadvantages including the need of double-maintenance of content within the AAS and the MTP files and the need to adapt existing MTP tools like POL systems.

¹ <https://github.com/admin-shell/aasx-package-explorer>

Based on this evaluation, we have chosen to embed the MTP as a file submodel element into the AAS submodel. This allows to easily re-use existing MTP tools and to minimize the double maintenance efforts. MTP can be embedded as a file to be transferred via AASX packages in Type 1 AAS or via an HTTP reference which can be navigated by AAS beyond Type 1. The existing proposal can be further extended with AAS-modeling of specific MTP information based on additional use cases if needed.

Let us consider an example of the MTP submodel in Fig. 1 as browsed by AASX Package Explorer.

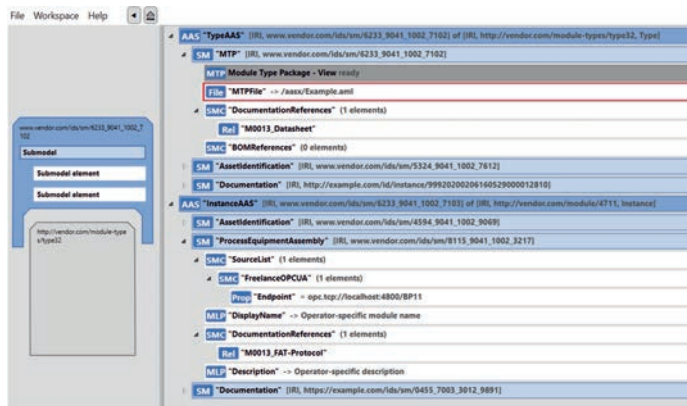


Fig. 1: MTP AAS and PEA AAS browsed by AASX Package Explorer tool.

Fig. 1 depicts two AASs – a *TypeAAS* embracing an MTP file and an *InstanceAAS* exposing PEA-specific information. Both AASs show how further standardized content can be added along with MTP-specific submodels: in both AAS, there is an *AssetIdentification* submodel containing identification properties of the respective asset, e.g., a vendor name for both MTP and PEA and a serial number for the PEA. In addition to the *AssetIdentification* submodel, a *Documentation* submodel containing semantically enriched PDF documents is contained in both AASs to differentiate between the MTP-specific documentation like datasheets or certificates and PEA-specific documentation like FAT or maintenance protocols. The PEA-AAS contains a *DerivedFrom* reference to the MTP-AAS allowing to locate the MTP file which matches the PEA.

Within the *TypeAAS* in Fig. 1, the red selection highlights the actual MTP file of the module type which is embedded as AAS submodel element of type *File*. This AAS element is equipped with a matching semantic ID and a mime type to make it uniquely identifiable by AAS consuming application.

The *InstanceAAS* contains a list of OPC UA endpoints which are specific to the PEA and are used to establish runtime communication with the PEA using the OPC UA communication protocol. Additional PEA properties to be shown in the POL are also embedded into the instance AAS: a *DisplayName* containing a human-readable name of the PEA and an operator-specific description of the PEA. Both descriptive properties are modeled as *Multi Language Property* elements within the submodel.

This lightweight AAS-model of the MTP and PEA is already able to cover the first use case of cross-company information exchange regarding MTPs and PEAs (cf. Section 2.1): PEA-specific information can be exchanged between POLs of different vendors. For example, a plant- or even application-specific OPC UA communication endpoint of the PEA can be exported from one POL and imported into another one.

3.2 MTP References within the Asset Administration Shell

Tackling the second use case – Module Type and Instance Documentation (cf. Section 2.2) – opens a need to create references between elements of the MTP or PEA submodel and elements of other AAS submodels. An example is a reference between a pump which is included within the module and a PDF containing a datasheet of the pump.

The information model of the AAS allows to create references between elements of different submodels and even AASs. However, the selected modeling approach for the MTP submodel does not duplicate the contents of the MTP file as elements in the AAS submodel (cf. Section 3.1). Therefore, the pump (represented by its tag name within the MTP file) does not exist in the AAS model and thus cannot be referenced from the AAS.

To close this gap, we decided to use the concept of references with *FragmentReference* keys allowing to encode MTP-specific (resp. AutomationML-specific) paths into references within an AAS. One example is shown in Fig. 2.

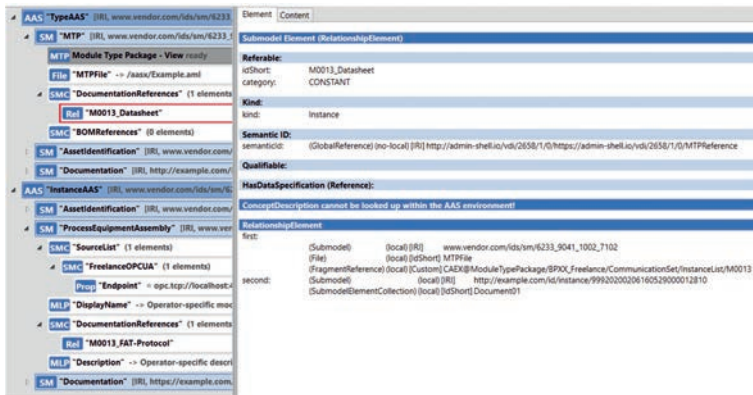


Fig. 2: MTP reference between a named element within the MTP file and a Documentation submodel element within the documentation submodel.

One can see a reference of type *MTPReference* called *M0013_Reference* which is specific to the MTP submodel pointing from a fragment reference “CAEX@ModuleTypePackage/BPXX_Freelance/CommunicationSet/InstanceList/M0013” within the MTP file to the “Document01” element of the documentation submodel which contains the datasheet of the motor. The textual encoding of the fragment reference is designed to make use of the CAEX²-structure (CAEX@” prefix) and points to a named element within the *ModuleTypePackage* hierarchy within the MTP file. Along with the demonstrated CAEX-based references to named elements, we also support references to unique IDs and *TagNames* within the MTP file.

The usage of fragment references is a method to reference nested elements of opaque elements in AAS submodels, e.g. files, to further submodel elements. On the one hand, a lot of engineering data is available in the AutomationML format and can now be easily linked to AAS elements. On the other hand, additional submodels can be referenced based on existing MTP or AutomationML files. One example of such references can be references to a *Bill of Material* (BOM) submodel within an AAS which is already indicated by an empty *BOMReferences* collection in Fig. 2.

² CAEX is the underlying metamodel of AutomationML and thus MTP

Further details on MTP-specific references will be included in the upcoming formal specification of the MTP submodel described in this paper.

4. Tools and Prototypes

To validate the submodel prototype, a plug-in to the open source tool AASX Package Explorer has been developed. This plug-in is activated as soon as an MTP submodel is detected within the AAS by using its semantic ID. A screenshot of the plugin can be found in Fig. 3. The plugin renders the Human Machine Interface which is embedded within the MTP. Furthermore, in case a PEA submodel is available, a connection to the contained OPC UA endpoint of the PEA is established and current values are read from the runtime OPC UA server.

This simple application illustrates the combination of both the type and the instance information and can be used along with an MTP-compliant POL.

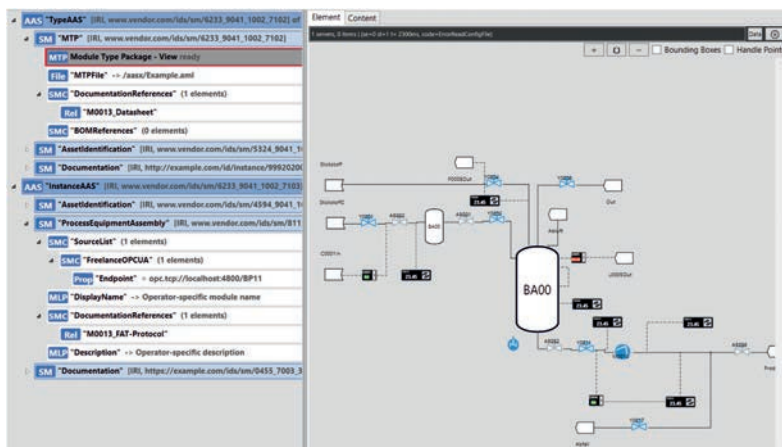


Fig. 3: MTP Plugin for AASX Package Explorer.

In parallel, there is ongoing work to enhance an existing MTP POL with the ability to export and import AAS-enhanced MTP files and PEA information.

5. Summary and Future Work

In this work, we have briefly discussed use cases and the design of MTP-related submodels to be embedded within module type- and instance- AASs. The proposed lightweight modeling

approach allows to enhance the current state of the MTP standardization by use cases like standardized exchange of instance specific PEA data and the embedding of documentation information to be supplied along with MTP files. This example shows how existing standards can be enhanced with the help of existing Industry 4.0 tools.

On the technical side, the introduced paradigm of "opaque" embedding of external artifacts within the AAS along with a set of fragment links can be re-used for other submodels or use cases, e.g., further AutomationML-based standards.

Topics of ongoing and future work are prototyping of documentation navigation using the AASX Package Explorer and a POL prototype as well as alignment, finalization, and publication of the MTP submodel via Plattform Industrie 4.0.

Acknowledgement

Authors were partially supported by German Federal Ministry of Education and Research in the scope of the BaSys 4.2 project (funding code 01IS19022).

References

- [1] J. Bernhausen et al. „Namur Modul Type Package–Definition.“ atp magazin 58. 01-02 (2016): 72-81.
- [2] VDI/VDE/NAMUR 2658 Blatt 1 „Automatisierungstechnisches Engineering modularer Anlagen in der Prozessindustrie - Allgemeines Konzept und Schnittstellen“ -Gesellschaft Mess- und Automatisierungstechnik. 10.2019.
- [3] R. Drath et al. „Die Rolle der Industrie 4.0 “Verwaltungsschale” und des “digitalen Zwilings” im Lebenszyklus einer Anlage: Navigationshilfe, Begriffsbestimmung und Abgrenzung“, Proceedings of Automation 2017, VDI-Berichte vol. 2293, VDI Verlag, Düsseldorf 2017, <https://doi.org/10.51202/9783181022931>.
- [4] Plattform Industrie 4.0 “Details of the Asset Administration Shell Part 1 - The exchange of information between partners in the value chain of Industrie 4.0”, 11.2020.
- [5] IEC 62714-1 “Engineering data exchange format for use in industrial automation systems engineering - Automation Markup Language - Part 1: Architecture and general requirements”. 04.2018.
- [6] IEC TR 62541-1 “OPC unified architecture - Part 1: Overview and concepts”, 11.2020.
- [7] Plattform Industrie 4.0 “Details of the Asset Administration Shell Part 2 - Interoperability at Runtime – Exchanging Information via Application Programming Interfaces”, 11.2020.
- [8] Plattform Industrie 4.0 “Submodel Templates of the Asset Administration Shell Generic Frame for Technical Data for Industrial Equipment in Manufacturing”, 11.2020.
- [9] M. Freund et al. „Das Module Type Package und die Verwaltungsschale – ein Ansatz zur Integration“, Proceedings of Automation 2020, VDI-Berichte vol. 2375, VDI Verlag, Düsseldorf 2020, <https://doi.org/10.51202/9783181023754-455>.

Das AutomationML Komponentenmodell

Dr.-Ing. **L. Hundt**, inpro, Berlin;
Dipl.-Ing. (DH) **M. Wiegand**, Festo SE & Co. KG, Esslingen;
Prof. Dr.-Ing. habil. **A. Lüder**, Otto von Guericke Universität, Magdeburg;
Dr. rer. pol. **T. Meyer**, Volkswagen AG, Wolfsburg

Kurzfassung

Im vorliegenden Beitrag wird das AutomationML Komponentenmodell vorgestellt. Dabei handelt es sich um einen Ansatz, der alle Engineering-relevanten Informationen zu einer Automatisierungskomponente und deren Zusammenhänge in einem konsistenten Datenmodell zusammenführt.

Es werden sowohl die zu Grunde liegenden Use Cases vorgestellt als auch das Modell beschreiben. Hierzu wird in dem Beitrag ein pneumatischer Zylinder als Beispiel für eine AutomationML Komponente erläutert. Abschließend werden Möglichkeiten der Nutzung, wie zum Beispiel der Aufbau von Digitalen Zwillingen diskutiert, sowie die Nähe des AutomationML Komponentenmodells zur Struktur der Verwaltungsschale der Industrie 4.0 Komponente und deren Nutzung im Engineering von Produktionssystemen thematisiert.

Abstract

This paper presents the AutomationML component model. This model is an approach that brings together all engineering-relevant information about an automation component and its interrelationships in a consistent data model.

The underlying use cases are presented, and the model is described. Furthermore, a pneumatic cylinder is explained as an example for an AutomationML component. Finally, possibilities of use, such as the construction of digital twins, are discussed and the proximity of the AutomationML component model to the structure of the management shell of the Industry 4.0 component and its use in the engineering of production systems are addressed.

1. Einführung Problemstellung

Automatisierte Produktionssysteme und -anlagen bestehen aus einer Vielzahl von Teilsystemen und -komponenten, die Hierarchien bilden können und auf unterschiedliche Weise miteinander interagieren. Beim Design dieser Systeme müssen verschiedene Aspekte berücksichtigt und kombiniert werden. Die einzelnen Engineering-Disziplinen, wie z.B. das funktionale Engineering, die mechanische Planung oder die Elektroplanung, benötigen unterschiedliche,

miteinander verknüpfte und abhängige Informationen auch über die einzelnen enthaltenen Automatisierungskomponenten.

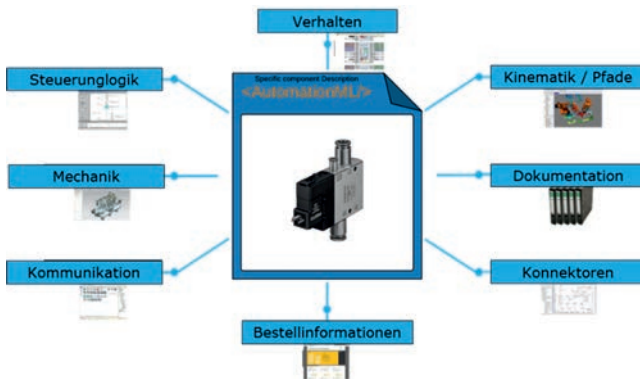


Bild 1: Aspekte einer Automatisierungskomponente

Bisher mussten diese Teilinformationen von den jeweiligen Planern und Konstrukteuren in den einzelnen Disziplinen meist manuell zusammengetragen werden, da es kein offenes standardisiertes Modell gab, welches die einzelnen Informationen zusammenführt und verknüpft. In diesem Beitrag wird das AutomationML Komponentenmodell vorgestellt, das in [1] definiert wurde und es erlaubt, eine Automatisierungskomponente als AutomationML Komponente zu beschreiben. Dieses ermöglicht die Aggregation nahezu aller Informationen in einem Modell, die für das komponentenbasierte Engineering eines Produktionssystems benötigt werden.

2. Anwendungsfälle für das AutomationML Komponentenmodell

Das AutomationML Komponentenmodell wurde vor dem Hintergrund folgender Anwendungsfälle entwickelt:

- Komponentenbeschreibung als Grundlage für CAX
- Gerätebeschreibungsdateien für Feldgeräte
- Materialwirtschaft und Lagerhaltung
- Simulation, insbesondere Virtuelle Inbetriebnahme
- Wartung und Dokumentation
- AutomationML Komponentenbeschreibungen als Enabler des Model-Based Systems Engineering (MBSE) und Erweiterung des Simultaneous Engineering

Beispielhaft wird an dieser Stelle der Anwendungsfall Komponentenbeschreibung als Grundlage für CAX im Detail erläutert.

Daten zu einer Automatisierungskomponente, die von Herstellern als AutomationML Komponente zur Verfügung gestellt werden, werden in den unterschiedlichen Engineering-Tools der am Planungsprozess beteiligten Gewerke verwendet. Bild 2 zeigt den Datenfluss von zwei unterschiedlichen AutomationML Komponenten von zwei Komponentenherstellern zu den Engineering-Tools, die in der CAX-Planung eines Komponentenanwenders eingesetzt werden (Import). Aus diesen und weiteren Engineering-Informationen bauen die Werkzeuge ein Gesamtsystem auf, das als AutomationML Modell bestehend aus Instanzmodellen der Komponenten besteht und das reale Automatisierungssystem, das realisiert werden soll, virtuell repräsentiert. Die daraus resultierenden Engineering-Daten innerhalb der AutomationML Komponentenbeschreibung sowie die Verbindungen zwischen den Modellen können ein inhaltsreicher Input für die virtuelle Inbetriebnahme sein, indem sie den Aufwand für die Datenerfassung reduzieren.

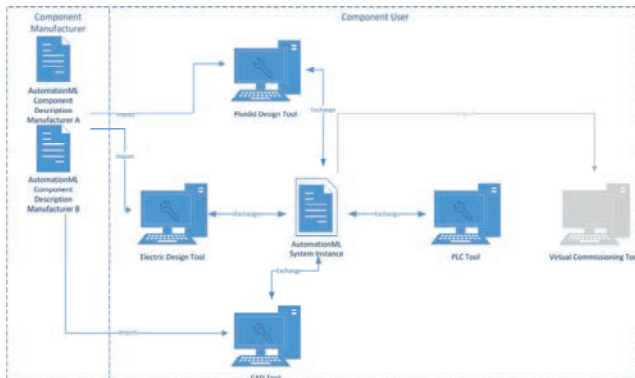


Bild 2: Anwendungsfall AutomationML Komponentenbeschreibung als Grundlage für CAX

Um dies zu realisieren werden die AutomationML Komponenten zwischen verschiedenen Werkzeugen ausgetauscht, die auf einer bestimmten Ebene die gleichen Informationen über eine Automatisierungskomponente und spezifische Informationen für ihren Zweck verwenden. Solche Informationen können z.B. werkzeugspezifische Icons oder Dokumente sein, die mittels der AutomationML Komponentenbeschreibung ausgetauscht werden. Zusätzlich werden die Verbindungen zwischen den Automatisierungskomponenten benötigt. Dies ist möglich, indem die Konnektoren der einzelnen Automatisierungskomponenten für die Engineering-Disziplinen abgebildet und deren Verbindung mit übertragen werden.

3. Das AutomationML Komponenten Datenmodell

3.1 Strukturierung von AutomationML Komponenten in CAEX

AutomationML Komponenten nutzen das Dateiformat AutomationML und dessen objektorientierte Modellierungsmechanismen. Alle Informationen zu Komponenten werden als AutomationML-Objekte gespeichert und können eine hierarchische Unterstruktur umfassen. Die Modellierungsarchitektur für AutomationML Komponenten ist eine verteilte Datei-Architektur unter Verwendung von Standard-AutomationML-Funktionen zur Speicherung von Daten. Hierzu werden mehrere bereits in AutomationML integrierte oder zusätzliche Engineering-Domänen durch Referenzieren bekannter externer Dateien genutzt. Dies wird durch bereits standardisierte oder neu definierte ExternalInterfaces realisiert.

Eine AutomationML Komponente, die eine Automatisierungskomponente beschreibt, wird als *SystemUnitClass* oder *InternalElement* mit der zugehörigen Rollenklasse *AutomationComponent* modelliert. Wenn die AutomationML Komponente ein Typmodell beschreibt, ist eine *SystemUnitClass* zu verwenden. Wenn ein Instanzmodell beschrieben wird, ist ein *InternalElement* zu verwenden.

Hinsichtlich der Strukturierung von AutomationML Komponenten sind verschiedene Prinzipien definiert.

- Jede AutomationML Komponente muss ein dediziertes Root-Element in der AutomationML-Datei haben, welches auf die in der Standard Rollenklassenbibliothek *AutomationMLComponentBaseRCL* definierte Rollenklasse *AutomationComponent* verweist.
- Die zu einer Automatisierungskomponente gehörenden Informationen müssen direkt im Wurzelement der AutomationML Komponente oder in einem Kindelement dieser Komponente gespeichert werden.
- Kindelemente, die einen Informationsaspekt der AutomationML Komponente beschreiben, müssen eine *SupportedRoleClass* haben, die in der Standard-Rollenklassenbibliothek *AutomationMLComponentBaseRCL* oder *AutomationMLComponentStandardRCL* definiert ist oder eine beliebige Rollenklasse, die von einer in diesen Bibliotheken definierten Rollenklasse abgeleitet ist.

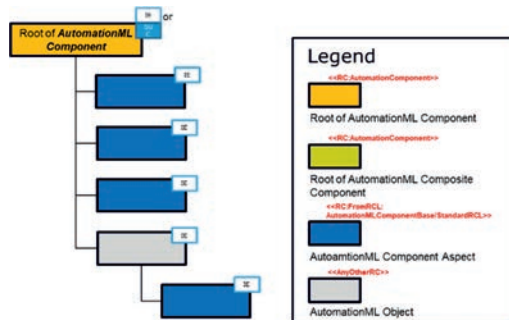


Bild 3: Strukturübersicht AutomationML Komponente

Bild 3 gibt einen Überblick über eine AutomationML Komponente, die ihr Wurzelement und vier Kindelemente enthält, die Informationen zur beschriebenen Automatisierungskomponente enthalten.

3.2 Abgedeckte Modellaspekte

Um die oben genannten Anforderungen zu realisieren, müssen verschiedene Informationen und Teilmodelle einer Automatisierungskomponente in der AutomationML Komponente abgebildet bzw. integriert werden.

Dies geschieht über zwei Mechanismen. Zum einen werden Informationen über die Automatisierungskomponente als Attribute der Komponente gespeichert und zum anderen werden Teilmodelle der Komponente als InternalElements der AutomationML Komponente modelliert bzw. über diese verknüpft.

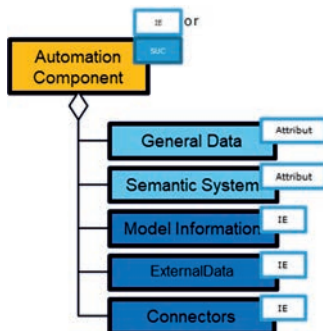


Bild 4: Modelaspekte der AutomationML Komponente

Es können die folgenden Kategorien an Information bzw. Modellen zur Komponente unterschieden werden.

1. Allgemeine Daten

Dieser Teil beinhaltet die Identifikations- und Klassifikationsdaten der Komponente, wie z.B. Produktidentifikationscode, Bestellinformationen, Seriennummer, Versionsnummer und andere technische Informationen. Alle diese Informationen werden auf Attribute im Roote-Element einer AutomationML Komponentenbeschreibung abgebildet.

2. Externe Daten

Dazu gehören z.B. die Dokumentation der Automatisierungskomponente oder die Symbole und Bilder der Komponente. Die externen Daten sind mit der AutomationML Komponente über ExternalInterfaces verknüpft, die zu einem untergeordneten InternalElement des Root-Elements gehören.

3. Modellinformationen

Die Modellinformationen umfassen verschiedene Arten von Modellen, die die Automationskomponente repräsentieren. Beispiele für solche Modelle sind: Funktionsdaten, Verhaltens- und Simulationsmodelle, 2D- und 3D-Modelle oder kinematische Modelle.

Auch diese Modelle werden über ein ExternalInterface, das zu einem untergeordneten InternalElement des Root-Elements gehört, als externe Daten mit der AutomationML Komponente verknüpft.

4. Konnektoren

Die Konnektoren beschreiben Informationen zu allen logischen, elektrischen, pneumatischen, hydraulischen und anderen Schnittstellen der Komponente. Diese Konnektoren werden als ExternalInterfaces modelliert, die zu einem InternalElement der AutomationML Komponente gehören.

3.3 AutomationML Komponenten Rollen- und Interfaceklassenbibliotheken

Die Grundlage für die Modellierung von AutomationML Komponenten sind die erforderlichen Rollenklassen, die angeben, dass eine SystemUnitClass oder ein InternalElement eine AutomationML Komponente ist oder zu einer solchen gehört. Im Hinblick auf die adressierten Aspekte von AutomationML Komponenten sind unterschiedliche Rollenklassen in drei AutomationML-Rollenklassenbibliotheken in [1] definiert. Der Name der ersten Rollenklassenbibliothek lautet *AutomationMLComponentBaseRCL*. Innerhalb dieser Rollenklassenbibliothek werden grundlegende abstrakte Rollenklassen spezifiziert, die im Allgemeinen nicht innerhalb eines AutomationML Komponentenmodells verwendet werden sollten, sondern eine abstrakte

Semantik liefern, von der detailliertere Rollenklassen abgeleitet werden können. Beispiele für eine solche Rollenklassen sind *Model* oder *GeometryModel*. Dabei ist *Model* die Elternklasse der spezifischeren Klasse *GeometryModel*. Bild 5 zeigt die Rollenklassenbibliothek *AutomationMLComponentBaseRCL*.



Bild 5: AutomationML Komponenten Rollen Klassen

Die zweite Rollenklassenbibliothek ist die *AutomationMLComponentStandardRCL*, vgl. Bild 5. Die Rollenklassen innerhalb dieser Bibliothek decken die konkreten Modell- und Informationaspekte einer AutomationML Komponente ab. Eine detaillierte Beschreibung aller Klassen ist in [1] zu finden.

3.4 Verbindungen zwischen und innerhalb von AutomationML Komponenten

Die volle Stärke von AutomationML als Format zur Beschreibung von Automatisierungskomponenten zeigt sich darin, dass die einzelnen Informationen und Modelle miteinander verknüpft und in Relation gesetzt werden können. Dies kann sowohl innerhalb einer AutomationML Komponente als auch zur Verknüpfung verschiedener AutomationML Komponenten erfolgen. Eine große Stärke von AutomationML ist die Darstellung von miteinander verknüpften interdisziplinären Informationen aus unterschiedlichen Domänen. So kann AutomationML beispielsweise Beziehungen zwischen Geometrie, Kinematik und Verhalten setzen, um einen vollständigen digitalen Zwilling für die virtuelle Inbetriebnahme darzustellen.

3.4.1 Beziehungen zwischen Konnektoren und Modellen innerhalb einer AutomationML Komponente

Die Konnektoren einer AutomationML Komponente stellen virtuelle oder reale externe Schnittstellen der realen Automatisierungskomponente dar. Meistens haben die Konnektoren einen Bezug zu den Datenmodellen innerhalb der AutomationML Komponente. So hat z.B. ein *MechanicConnector*, wie der Flansch an einer beweglichen Kolbenstange, eine

Beziehung/Einfluss auf das kinematische Modell der AutomationML Komponente, da sich bei einer Bewegung durch eine externe Kraft die Kolbenstange mitbewegen könnte.

Für die Modellierung der Beziehungen zwischen Konnektoren und Modellen gilt die allgemeine Regel, dass ein `InternalLink` zwischen einer Schnittstelle des Konnektors und einer Schnittstelle des Modells verwendet werden muss.

Es gibt viele verschiedene Möglichkeiten, Modelle innerhalb einer AutomationML Komponente zu verbinden, und nicht alle sind von praktischer Relevanz. In [1] sind sinnvolle Beispiele für mögliche Beziehungen zwischen Konnektoren und Modellen und deren Schnittstellen erläutert.

3.4.2 Beziehungen zwischen Modellen innerhalb einer AutomationML Komponente

Die Modelle einer AutomationML Komponente stellen Datenmodelle für eine bestimmte Domäne oder einen bestimmten Anwendungsfall der Automatisierungskomponente dar. In vielen Fällen innerhalb des Engineerings und des späteren Einsatzes einer AutomationML Komponente führt die kombinierte Verwendung einzelner Domänenmodelle zu den besten Ergebnissen. Daher müssen diese Modelle in Beziehung zueinander gesetzt werden.

Für die Modellierung von Beziehungen zwischen Modellen innerhalb einer AutomationML Komponente gelten die gleichen Regeln wie für die Modellierung von Beziehungen zwischen Konnektoren und Modellen. Es müssen also `InternalLinks` verwendet werden und Modellelemente, die eine entsprechende Rollenklasse haben, müssen miteinander verknüpft werden.

Ein Beispiel für eine Beziehung zwischen Modellen ist die Verknüpfung des Verhaltens und des kinematischen Modells einer AutomationML Komponente. Dazu muss eine Relation zwischen der Variablenschnittstelle des `Behaviour`-Modells mit dem `JointInterface` eines `COLLADAKinematicJoints` verbunden werden. Diese interdisziplinäre Verknüpfung von Logik und Kinematik stellt einen Mehrwert dar, der den Einsatz von AutomationML Komponenten z.B. in Software-Tools zur virtuellen Inbetriebnahme vereinfacht.

3.4.3 Beziehungen zwischen verschiedenen AutomationML Komponenten

Zwei oder mehr AutomationML Komponenten können mit Hilfe von Konnektoren in verschiedenen oder der gleichen Domäne miteinander verbunden werden, wodurch Composite Components entstehen.

Dabei werden unterschiedliche AutomationML Komponenten mit `InternalLinks` auf ihren Konnektoren verbunden. Diese bilden dann die resultierende AutomationML Composite Component, die einen eigenen Satz von Konnektoren definieren kann.

4. AMLX Container für AutomationML Komponenten

Ein weiterer Vorteil der AutomationML Komponentenbeschreibung ist, dass sie in einem Paket mit allen zugehörigen Dateien und Modellen zusammengefasst und ausgetauscht werden kann. Hierzu wird das AMLX-Format verwendet, vgl. [2].

Grundsätzlich können durch AMLX Container drei Anwendungsfälle unterstützt werden. Diese sind:

1. Austauschen einer einzelnen AutomationML Komponente
2. Austausch von Bibliotheken von AutomationML Komponenten
3. Austausch von detailliert spezifizierten Komponenten, die einem bestimmten Profil entsprechen

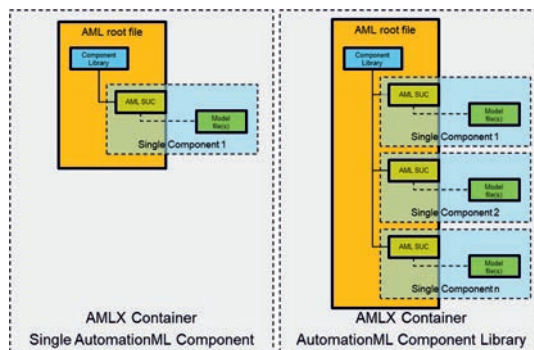


Bild 6: Der AMLX Container für AutomationML Komponenten

In Bild 6 ist der Aufbau eines solchen AMLX-Containers dargestellt. Auf der rechten Seite ist eine einzelne AutomationML Komponente dargestellt. Der Container enthält die Root-Datei, die eine AutomationML Komponente und die zugehörigen Modell- und Beschreibungsdateien enthält. Dies deckt den ersten Anwendungsfall ab. Die linke Seite der Abbildung zeigt die Struktur eines AMLX-Containers, mit dem eine komplette Bibliothek von AutomationML Komponenten in einen Container gepackt sind und ausgetauscht werden. Das Paket besteht ebenfalls aus einer AutomationML Root-Datei, die eine Bibliothek enthält. Diese Bibliothek kann mehrere Komponenten und deren zugehörige Dateien enthalten. Damit wird der zweite Anwendungsfall abgedeckt. In beiden Fällen gibt es keine zusätzlichen Regeln bezüglich Namenskonventionen innerhalb des Pakets, Regeln bezüglich obligatorischer oder optionaler Modelle. Die einzige Regel die existiert ist, dass die AutomationML Komponente innerhalb des Containers alle Regeln von [1] befolgen muss und dass AMLX gemäß den Regeln von [1] generiert werden muss.

Diese Serialisierungsmethode kann nun auch auf die AutomationML Komponente angewendet werden. Dabei werden, wie in Bild 8 rechts dargestellt, die einzelnen in einer AutomationML Komponente abgebildeten Informationsmengen als strukturierte *SubModels* (mit bzw. ohne *SubModelElementCollection*-Hierarchie) aufgefasst und entsprechend mit Rollen gekennzeichnet und deren Attributmengen in *InternalElements* mit den notwendigen verwaltungsschalenspezifischen Attributen überführt. Ein entsprechender Algorithmus dafür wurde in [4] und [6] vorgestellt.

6. Anwendungsbeispiel der Pneumatik Zylinder ADN-25-50-A-P-A als AutomationML Komponente

Im Folgenden wird an einem Beispiel dargestellt, wie ein Pneumatik Zylinder als AutomationML Komponente modelliert wird. Bei dem Zylinder handelt es sich um den ADN-25-50-A-P-A der Firma Festo, vgl. Bild 8.



Bild 8: Beispielkomponente Zylinder ADN-25-50-A-P-A

In der AutomationML Komponenten Beschreibung des Zylinders werden sind folgende Informationen bzw. Modelle integriert:

- Mehrere typische Datenblatt-Attribute zur Identifikation, technischen Beschreibung und kommerziellen Nutzung
- Bild, Icon und Pneumatiksymbol der Komponente und ein Herstellerlogo
- Dokumentation der Komponente
- Geometrie- und Kinematik Modelle mit Verlinkung auf das entsprechende COLLADA-Dokument
- Vereinfachtes logisches Verhaltensmodell mit Verlinkung auf das entsprechende AM-LLLogic-Dokument
- Mechanische Anschlüsse für Kolbenstange, Grundkörper und Sensorsteckplätze der Komponente

- Pneumatische Anschlüsse der Komponente
- Mehrere Relationen (interne Verknüpfungen) zwischen Kinematik- und Verhaltensmodell sowie den Konnektoren

Bild 9 zeigt die daraus resultierende AutomationML Struktur zur Beschreibung einer AutomationML Komponente. In dieser Struktur hat jeder einzelne Aspekt ein eigenes InternalElement als Unterelement, zum Referenzieren der einzelnen Dokumente und Modelle über ExternalInterfaces.



Bild 9: Modellierung Zylinder ADN-25-50-A-P-A als AutomationML Komponente

Dabei können nicht nur die Modelle an sich referenziert werden, sondern auch detailliertere Aspekte, wie zum Beispiel einzelne Variablen in Verhaltensmodellen. Diese Details können dann mit Aspekten weiterer Modelle, über InternalLinks, in Relation gesetzt werden. Alle direkten Informationen zum Zylinder werden als Attribute unter dem Root-Element „ADN-25-50-A-P-A“ der Komponente als Attribute gespeichert.

8. Zusammenführung und Ausblick

Im vorliegenden Beitrag wurde das AutomationML Komponentenmodell vorgestellt, welches die Möglichkeit bietet, Automatisierungskomponenten ganzheitlich zu beschreiben und diese Informationen inkl. der dazugehörigen Modelle und Dokumente auszutauschen. Das Modell basiert dabei auf AutomationML Rollenklassen und nutzt alle bisherigen Arbeiten des AutomationML e.V. und führt diese konsistent zusammen.

Aus Sicht der Autoren besteht das größte Nutzenpotential des AutomationML Komponentenmodells darin, dass es mit diesem möglich ist alle Informationen zu Automatisierungskomponenten einheitlich in einer standardisierten Beschreibung zu erhalten. und diese im Engineering zu nutzen. Damit erschließen sich vielfältige Anwendungsfälle, wie im Beitrag beschreiben.

Weiterhin wurde das anhand eines Zylinders beispielhaft gezeigt, wie ein solches Komponentenmodell in der praktischen Realisierung aussieht.

Der AutomationML e.V. arbeite daran, dass Modell für weitere Anwendungsfälle zu erweitern und Profile für bestimmt Anwendungsbereiche zu definieren.

- [1] AutomationML e.V.: Whitepaper AutomationML Part 6 - AutomationML Component, V 1.1.0, www.automationML.org, Jan 2020
- [2] AutomationML e.V.: BPR AutomationML Container, V 1.0.0, www.automationML.org, October 2017
- [3] Plattform Industrie 4.0: Details of the Asset Administration Shell. Part 1 - The exchange of information between partners in the value chain of Industrie 4.0 (Version 2.0). Federal Ministry for Economic Affairs and Energy (BMWi)(Hrsg.), ZVEI & Plattform Industrie 4.0, 2019, <https://www.plattform-i40.de/PI40/Navigation/EN/InPractice/Online-Library/online-library.html>.
- [4] AutomationML association: Application Recommendation - Asset Administration Shell Representation (AR AAS), 2019, https://www.automationml.org/0.red/uploads/dateien/1574258013-AR_004E - Assset Administration Shell Representation 20V1_0_0.zip
- [5] B. Boss, A. Graf Gatterburg, M. Hofmeister, A. Lüder, M. Müller, M. Schleipen, M. Wiegand: Serialisierung der Verwaltungsschale einer Industrie 4.0 Komponente mit AutomationML, Entwurf Komplexer Automatisierungssysteme (EKA), Mai 2020, Magdeburg, Germany, Proceedings.
- [6] A. Lüder, A. -K. Behnert, F. Rinker and S. Biffl, "Generating Industry 4.0 Asset Administration Shells with Data from Engineering Data Logistics," 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vienna, Austria, 2020, pp. 867-874, doi: 10.1109/ETFA46521.2020.9212149.
- [7] J. Wehrstedt, D. Rovere, P. Pedrazzoli, G. dal Maso, T. Meyer, V. Brandstetter, M. Ciavotta, M. Macchi, E. Negri: Open Semantic Meta-model as a Cornerstone for the Design and Simulation of CPS-based Factories In: J. Soldatos, O. Lazaro, & F. Cavadini: The digital shopfloor - industrial automation in the industry 4.0 era, April 2019
- [8] AutomationML e.V.: Whitepaper AutomationML Part 4 - AutomationML Logic Description, V 1.5.0, www.automationML.org, Jan 2017

Orchestration of Co-Simulations with FMU Integration as a Service

Orchestrierung von Co-Simulationen mit FMU-Integration as a Service

Dr. rer.-nat. **N. Schoch**, Dr. Ing. **J. Schlake**, Dr. Ing. **S. Grüner**,
Dr. Ing. **J. Heuschkel**, ABB Forschungszentrum, Ladenburg

Kurzfassung

Im Kontext von Industrie 4.0 werden Simulationen komplexer Systeme nicht nur immer wichtiger, sondern auch immer komplizierter, da sie oft mehrere Subsysteme mit teils hochgradig spezifischen mathematischen Eigenschaften beinhalten. Ein sehr gängiger Ansatz zur Kopplung und Integration solch spezifischer Subsysteme besteht darin, die einzelnen Systeme konform zum standardisierten Austauschformat des Functional Mockup Interface (FMI) zur Verfügung zu stellen und darauf basierend eine Co-Simulation aufzusetzen.

Diese Arbeit stellt eine prototypische Entwicklung eines Co-Simulations-Orchestrierungs-Services vor, welcher Simulations-Interoperabilität und Orchestrierung durch semantische Beschreibung, Kopplung und Kontrolle von FMI-konformen Sub-Simulationsmodellen als Bestandteile einer Gesamtsimulation ermöglicht und vereinfacht.

Zur semantischen Beschreibung und Kontrolle des Gesamtsimulationssystems wie auch der Teilsysteme baut der Ansatz dabei auf die Verwendung des Industrie 4.0 Standards der Verwaltungsschale/Asset Administration Shell (AAS) auf, in welche wiederum Tool- und Herstellerunabhängig die Functional Mockup Unit Blöcke eingebettet sind.

Der Simulations-Service folgt somit dem generellen Workflow (1) das via REST API empfangene AAS-basierte Simulationssystem zu parsen, (2) die spezifischen Subsystem-Blöcke in Inputs für die jeweils spezifischen Simulationstools zu übersetzen, um somit (3) eine Co-Simulation auszuführen, und schließlich (4) die strukturierten Simulationsergebnisse über die REST API des Simulationsservices wieder zurück zu geben.

Zur Verifikation des Ansatzes wird der Simulationsservice anhand eines Beispielszenarios einer Antriebs-Co-Simulation evaluiert, wo Drive-Experten exemplarisch die Antriebsdimensionierung manipulieren können, und Applikations-Experten die Master-Simulation kontrollieren.

Abstract

In the context of Industry 4.0 with many use cases including Digital Twins, Virtual Commissioning and Holistic Engineering, simulation of complex systems is getting more and more important, but also more and more challenging. Often, a cyber-physical production system consists of multiple sub-systems from different manufacturers or domains, having its own, separated models. On top, these models work in their own physical disciplines (e.g. electrics, mechanics or thermodynamics) and accordingly require different domain expertise as well as highly specific numerical solvers for integrating the underlying mathematical problem.

A common approach to couple such sub-system models into a combined system simulation is to require all sub-systems to implement the standardized Functional Mockup Interface (FMI). As such, the sub-system is able to participate as Functional Mockup Unit (FMU) in a so-called co-simulation.

This contribution presents a prototype of a service that facilitates and simplifies simulation interoperability through automated setup and orchestration of complex co-simulation systems, where all sub-systems are provided according to the FMI standard as FMUs. This system facilitates the standardized semantic description of simulation topologies as well as the automated execution and manipulation of the arising co-simulations, including the handling of simulation results and output data. Ultimately, the prototype allows to abstract the use of various simulation tools to the use of a single command line interface or RESTful API call.

To semantically describe and control the overall system simulation, we utilize Industry 4.0 Asset Administration Shell (AAS) interoperable exchange format for digital twins. Semantic-descriptions are considered future-proof, as they comply with AAS, OPC UA, and cloud-based simulation architectures. AAS-based models allow to represent both the overall system topology and the included sub-system models. Wrapped inside the AAS-based sub-system description, the respectively specific model and simulation of the sub-system are provided according to the tool- and manufacturer-independent FMI standard. This standard allows to represent the respective system component as a black-box container with defined interfaces that allow the efficient exchange and further processing of simulation data.

For the prototyping implementation of our simulation orchestration service, we support three commercial simulation tools that are compatible with the FMI standard: MATLAB/Simulink, SimulationX and Modelica Dymola.

The service workflow then goes as follows: (1) the simulation system AAS is parsed and (2) the respectively contained sub-systems are translated into the simulation tool specific inputs. Thus, a co-simulation is set up and executed (3), and the structured simulation results (4) are fed back to the user, who has access to the simulation service via a RESTful API.

To verify our approach, we evaluated our simulation orchestration service based on a power-train co-simulation. Here, drive experts can calibrate the drive dimension in the drive simulation, and application experts can run the master application simulation.

To summarize, our prototype implementation demonstrates the easy access, manipulation and execution of complex simulation systems with different simulation resources contained of interoperable subsystems. This is achieved by a common semantic control layer implemented in an AAS. The obtained results bring advantages to users in terms of easier access and setup of simulations for the initially mentioned purposes and lower entry-barriers to system simulations.

1. Introduction

In the context of Industry 4.0, with many use cases including Digital Twins, Virtual Commissioning, Simulation-based Calibration & Optimization and Holistic Engineering, simulation of complex systems is getting more important, but also increases complexity [1]. Often, the composed cyber-physical system consists of multiple sub-systems from different manufacturers or domains, with separated models each. On top, these models act in their own physical disciplines (e.g. electrics, mechanics, thermodynamics, etc.) and accordingly require different domain expertise as well as highly specific numerical solvers for integrating the underlying mathematical problems.

Today, a common approach to combine and couple such sub-system models into a composed system simulation, is to require all sub-systems to offer the standardized **Functional Mockup Interface** (FMI) [4,6]. As such, the sub-system is able to participate as Functional Mockup Unit (FMU) in a so-called **co-simulation** [2,3].

Among several large cross-industrial efforts in the context of bringing forward Industry 4.0, the **BaSys 4.2 project** [7] continues the success story of BaSys 4.0, with a particular focus on the continuous engineering of manufacturing processes and **Industry 4.0** for the process industry. The results will be incorporated as components into the BaSys middleware. Among others, BaSys 4.2 strongly focusses on (1) Edge Computing, to make decisions close to devices, (2) Model-based capability descriptions, and (3) Virtualization of production.

A key concept of Industry 4.0 is the **Asset Administration Shell (AAS)** [1,8,9], which is used to describe, represent and allow access to an industrial asset electronically and in a standardized manner. Its purpose is to exchange asset-related data among industrial assets and between assets and production orchestration systems or engineering tools. This logical

representation of the asset is called the AAS, and together with the asset, it forms the **Industry 4.0 Component** [10].

Enhancing model-based capability and functionality descriptions and fostering the virtualization of production, which also includes modeling and simulation, one of the major work packages that ABB AG is active in investigates on **simulation interoperability** and **simulation orchestration** [1].

In this context, a prototype for semantic orchestration of complex co-simulations with FMU integration was developed and shall be presented in this contribution. The prototype system facilitates and simplifies simulation interoperability through automated setup and orchestration of complex co-simulation systems, where all sub-systems are provided according to the FMI standard, and where the overall description of the simulation is provided in the AAS exchange format (AASX).

Currently, this system facilitates the standardized (yet extensible and exchangeable) semantic description of simulation topologies as well as the automated execution and manipulation of the arising co-simulations, including the handling of simulation results and output data. Eventually, the prototype is to abstract the usage of miscellaneous (possibly domain-specific) simulation tools to the use of a single command line interface or RESTful API call. So far, for the prototyping implementation of the simulation orchestration service, three commercial simulation tools that are compatible with the FMI standard (both for import and export) are provided: MATLAB/Simulink [11], SimulationX [12], and Modelica Dymola [13,14].

The rest of the paper is structured as follows: Section 2 presents an analysis of the user requirements and the thus-motivated solution approach. In Section 3, we evaluate the solution by means of a sample Industry 4.0 use case scenario. Lastly, we conclude and give an outlook in Section 4.

2. User Requirements and Solution Approach

In this section, we analyze the use case requirements, and thus motivate our solution approach. Then, the architecture and infrastructure setup will be outlined, and the overall workflow will be presented. A particular focus will be on facilitating simulation interoperability.

2.1. Requirements Analysis

Preceding the conceptualization and the architectural draft for the prototype of the simulation orchestration service, we gathered the requirements of potential service users in the above-mentioned Industry 4.0 use case contexts.

All considered use cases [15, 16] have the following feature requirements in common:

- REST API to the simulation orchestration service
- Standardized semantic topology & system description, sub-system blocks via FMU
- Orchestration and control of co-simulations
- What-if-scenario-simulation, possibly parallel, with parametrized scenario templates
- Simple user interface (UI) to upload/download simulation tasks and results

Non-functional requirements (such as simulation performance and stability) shall not be considered at this stage. Especially when dealing with multi-scale problems and problems from different application domains, these are typically important. Thus, in the future, we assume strong requirements, which is why the subsequently proposed solution and architecture is designed to extensively allow for and be able to guarantee scalability and stability.

2.2. Architecture and Infrastructure

Based on the above analyzed requirements, the following service architecture and infrastructure is proposed (Fig. 1).

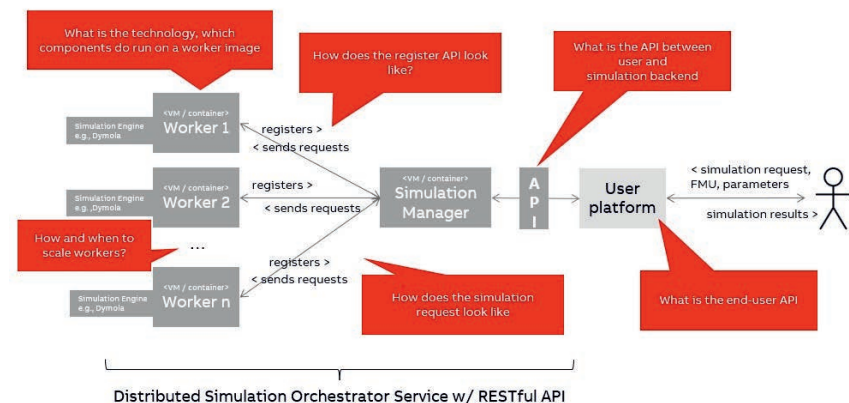


Fig. 1: Draft of the proposed goal architecture and infrastructure for providing the intended simulation functionality as a service.

The bright grey box marks the application system, that shall be connected to the simulation service in order to provide simulation functionalities to its clients. – The dark grey boxes mark the final vision and scope of work of this research paper. Important design and architecture as well as infrastructure questions are shown in red boxes. The components and functionalities implemented in the context of the here presented prototype include the Simulation

Orchestrator to process a Simulation Scenario that is specified by means of one FMU (e.g. to model one of the above intended use cases) and that runs on only one worker node using the simulation engine Dymola/Modelica.

2.3. Workflow and Use-case

Along with the Simulation Orchestrator, a RESTful (Web-)API is provided, such that the following generic workflow and use case is enabled for a client to interact with the Simulation Orchestrator Service (Fig. 2):

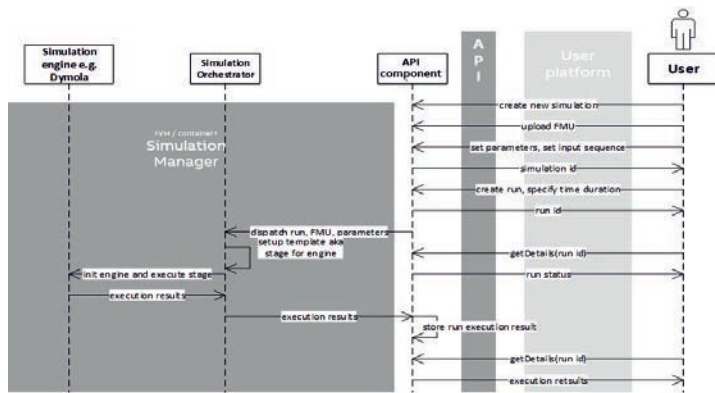


Fig. 2: Schema of the general workflow of the proposed MVP use case, showing the events and communication (data flow, requests and responses) for simulation orchestration.

This setup has been implemented in Python as a first proof-of-concept, which consists of two major (generally separate and independent) components, i.e., stand-alone Python package modules, that can be linked in order to provide the simulation orchestration service functionality:

1. The **Semantic Simulation Orchestrator for Co-Simulation using FMIs** to enable simulation interoperability and provide the orchestration service (with Dymola/Modelica as the simulation engine running in the backend), and
2. The **RESTful (Web-)API** to access the Simulation Orchestrator and to execute simulation tasks (along with a simple Demo Web App).

The general idea is to expose the functionalities of the simulation orchestrator (simulation, interoperability, co-simulation, FMU integration) via a RESTful API, and thus, to make it available to intended users that can make HTTPS requests to a specific URL and get back relevant

data (e.g. simulation results) in the response. The service expects that all data is sent and received in (or wrapped inside) JSON/AASX format.

2.4. The Functional Mockup Interface (FMI) Standard

The Functional Mockup Interface (FMI) is a standard that defines the interface to a Functional Mockup Unit (FMU) [4]. It is maintained as a Modelica Association Project [5] and supported by 100+ tools, hence achieving significant prevalence and acceptance. The FMU can be used for a) **model exchange**, where the FMU provides the simulation model which is executed by the simulation tool, or b) **co-simulation**, where the FMU contains the model and the solver as a compiled binary which can be called by the simulation host, or c) **scheduled execution**, where the importer triggers the execution of the model partitions. In essence, the FMU is zip file containing an XML file with the interface definition and meta data, together with the necessary files to execute the simulation [4,6].

FMUs for model exchange contain the simulation model; the model equations are exported and solved (i.e. integrated) by the importing tool. *FMUs for Co-Simulation* do not contain the simulation model in an importable format; the FMI only specifies the data transfer between host simulation (the tool where the FMU will be integrated) and the worker simulation which is embedded in the FMU. Hence, the FMU must contain a binary file, which is called by the master simulation and executes the embedded simulation model.

A disadvantage of FMUs for co-simulation, is the introduction of black boxes into the master simulation. Both simulations need to exchange not only state values but also derivatives. Since none of the simulators have access to the full state information of the overall system, it is known that simulation performance can suffer drastically. Testing different solvers in order to possibly improve performance is therefore recommended. Moreover, one needs to take special care of the time step size with respect to data exchange between all tools and their solvers, and of the initial values and parameters, particularly in the simulation start phase.

2.5. Nested Simulation System Topologies and Simulation Interoperability

The **semantic simulation orchestrator** establishes the core of the simulation service architecture (Fig. 3). It is accessible for users – both simulation experts and application experts (here, by the example of ABB Virtual Drive [15], which is a tool to simulate ABB drives) – who shall be able to describe, control and manipulate the simulation system (both simulation setup and execution) via a two-fold hierarchical approach:

On the outer level, we utilize Industry 4.0 Asset Administration Shell (**AAS**) interoperable exchange format for digital twins [8,9,10] and the related service infrastructure (AASX Services

and Platform Industry4.0 Services). This is, exchangeable, e.g. by means of an according meta-data-container or mapping process or a direct link to the user's system.

It shall be remarked we chose AASX format since complying semantic descriptions are not only considered future-proof for cloud-based simulation architectures, but also since AAS-based models allow to represent both the overall system topology and the included (possibly multi-level nested) sub-system models in a clearly defined hierarchical format.

Wrapped inside the **AAS**-based sub-system description, the respectively specific model and simulation of the sub-system are provided according to the tool- and manufacturer-independent **FMI** standard [4,6].

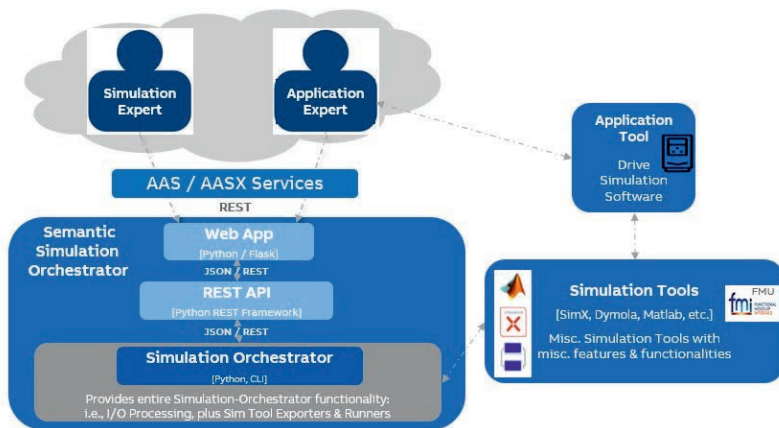


Fig. 3: Architecture Draft with focus on the context and container levels.

In total, this enables not only **interoperability** between simulation tools but also between users or domain experts (via common semantics and clearly defined sub-model interfaces and topologies). Moreover, standardized interfaces facilitate the platform-independent exchange of simulation models and thus increase the flexibility in the connection of external simulation tools and models. With this concept – the overall model topology, nested sub-models and model interfaces that are all represented semantically (through the AAS Industry 4.0 standard) and connected via FMUs within an overall master simulation. The resulting **Digital Twin** of the considered system is facilitated, and allows for representing the miscellaneous components and their interconnectivity and mutual impacts.

The **simulation orchestrator** itself finally manages translation, access and execution of simulation-tool-specific simulation scenarios and therefore is connected to a set of simulation

tools, such as Simulation X, Dymola Modelica or Matlab. Via the simulation tools, further connections to third-party tools can be established by means of the above described FMUs.

On the container level of the system architecture (Fig. 3), the system decomposes into inter-related containers that are themselves executable and deployable sub-systems.

As described above, our proof-of-concept consists of two major Python packages, that are linked together in order to provide the overall Simulation Orchestration Service functionality: (1) The Semantic Co-Simulation Orchestrator, and (2) the RESTful (Web-) API to access the Simulation Orchestrator (along with a simple Demo Web App). All communication and interaction with the additional components depicted in Fig. 4 follows the REST principles and takes place via AASX/JSON.

3. Evaluation – By means of a Sample Industry 4.0 Use Case Scenario

As has been motivated above, the proposed solution was built upon the analysis of the requirements and feature requests coming from the potential service users in the Industry 4.0 use case contexts [15,16]. To verify the proposed solution approach, the Simulation Orchestration service was evaluated based on a sample use case dealing with a powertrain co-simulation (e.g. as part of a Digital Twin of a motor with the goal of minimizing drive size and energy cost). Here, drive experts could calibrate the drive dimension in the drive simulation, and application experts could run the master application simulation and check the to-be-optimized KPIs, such as overall energy costs or performance.

3.1. Resulting Workflow and “Simulation-as-a-Service Functionality

The general service workflow of the proposed solution (see also Fig. 4) is as follows:

1. **Access via REST API:** The user has access to the Simulation Orchestration Service via RESTful API; The user can specify the simulation scenario: model topology, FMUs, input sequence, parameters, run time, etc. All data is expected to be sent and received in AASX/JSON format.
2. **Input Parsing:** The simulation system input file is being parsed and – depending on the intended goal simulation tool – send to the respectively suitable worker node(s).
3. **Tool-specific Translation & Template Instantiation:** the respectively contained sub-systems are translated into the simulation tool specific inputs; pre-defined generic templates are instantiated.
4. **Setup & Execution of Co-Simulation:** Thus, a co-simulation is set up and executed.

5. **Simulation Results Postprocessing:** The simulation results are structured and visualized in order to be fed back to the user via AASX/JSON format.
6. **Access via REST API:** The user has access to the simulation results via REST API.

3.2 Nested Simulation Topologies described through Combining AAS and FMI

As described above, the user is supposed to semantically describe and control the overall system simulation using a two-fold hierarchical / nested approach consisting of an AAS-conforming wrapping around possibly several FMU blocks. Fig. 4 shows an AASX/JSON instance file in the AAS Explorer [1], along with miscellaneous specifications (such as simulation parameters and simulation experiment controls) made in the according sub-models.

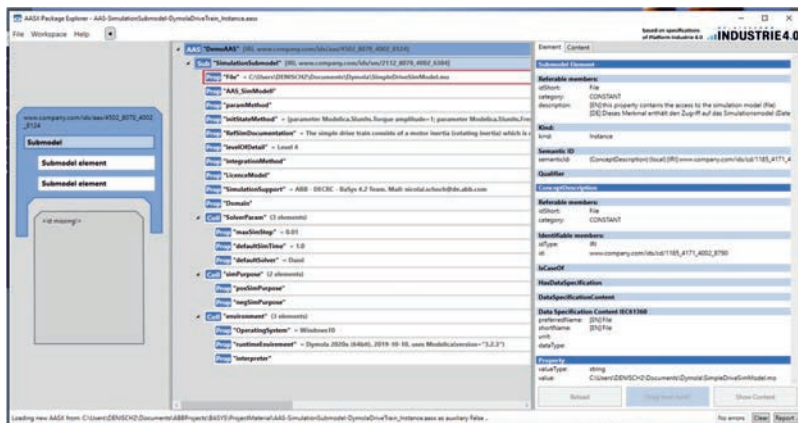


Fig. 4: Screenshot of a sample scenario in the AAS Explorer [1]. Please note, that an adapted preliminary version of the AAS Explorer has been used here.

Depending on the recognized set of goal simulation tools, the overall simulation system and its sub-systems are parsed and translated, and the according sub-simulation system input files are then created for subsequent execution through the respective simulation tools.

3.3. REST API and Simple Demo Web App for the Simulation Orchestrator

To provide a user-friendly alternative to the Command Line Interface access to the Simulation Orchestrator functionality, a rudimentary Demo Web App is attached to the REST API and offers a UI for simple control or manipulation of simulation scenarios. Fig. 5 shows the list of

basic endpoints, HTTP methods and expected formats, as well as the swagger.io-provided out-of-the-box playground [17] to test the API. This again emphasizes the service architecture.

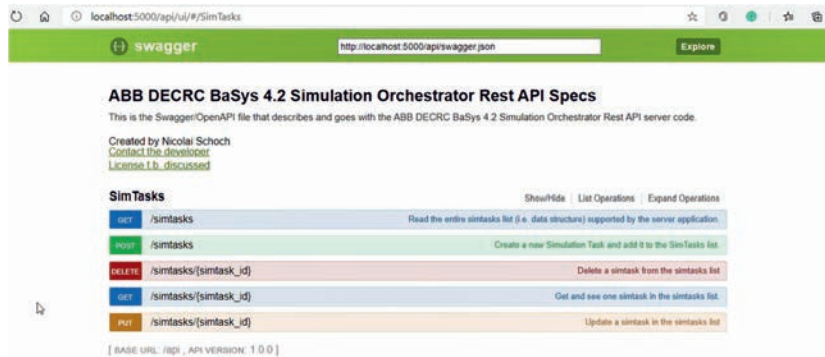


Fig. 5: Screenshot of the Demo Web App for easy access to the Simulation Orchestrator.

3.4. Orchestration of Co-Simulations by the example of a Powertrain Co-Simulation

To highlight the automated semantic setup, orchestration, and execution of a co-simulation, we defined and ran the exemplary powertrain co-simulation sample scenario (Fig. 6):

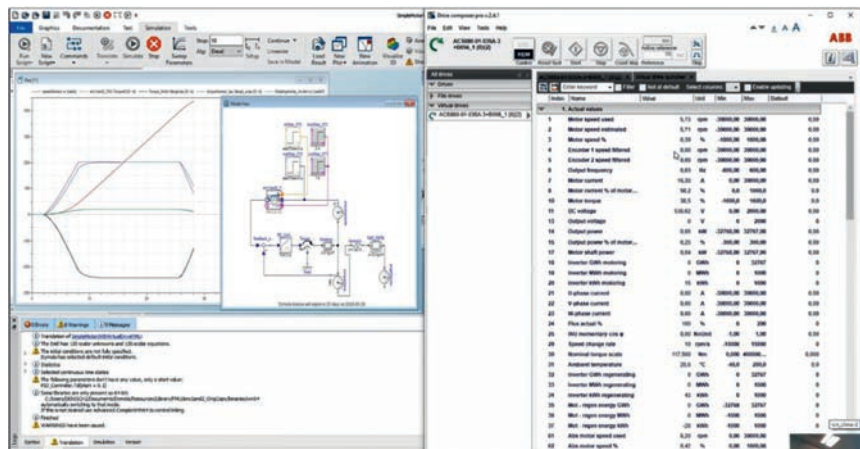


Fig. 6: Screenshot of the running Co-Simulation Sample Scenario with Dymola (master simulation) and Virtual Drive (sub-simulation) coupled via FMU (see block diagram).

Fig. 6 shows how the master simulation tool (here: Dymola [13]) is triggered and controls the overall system simulation setup and topology along with FMUs, through which sub-simulations and their respective tools (here: ABB Virtual Drive Simulation [15]) are controlled and exchange data with the master simulation via FMI. With this setup, e.g., the above-mentioned drive sizing experiments can be done easily and fully automatically.

4. Conclusion and Outlook

In this paper, we investigated on simulation interoperability in the context of Industry 4.0 use cases and analyzed typical problems as well as general user requirements. Thereon based, we presented the concept and prototype implementation of a **co-simulation orchestration service** to satisfy these requirements and to overcome these problems.

This service facilitates the **standardized semantic description of complex co-simulation topologies** via the Asset Administration Shell (AAS), as well as the **automated setup, manipulation, and execution as well as orchestration of co-simulation systems**, where all sub-systems are provided according to the FMI standard. Furthermore, the service allows to **abstract the usage of miscellaneous simulation tools and solvers** to the use of a single command line interface or **RESTful API call**.

As such, the proposed simulation orchestration service is an **enabler** for the typical portfolio of **simulation-related Industry 4.0 applications**, including Digital Twins, Virtual Commissioning, Calibration & Optimization and Holistic Engineering, which typically comprise complex cyber-physical overall systems. The service can bring advantages to users in terms of easier access to and setup of simulations for the given purposes and lower entry-barriers to system simulations in general.

To advance the prototype, we see, among others, the following major points for future work:

- Full virtualization and containerization, also with respect to scalability,
- Discussion and solutions for further non-functional requirements (such as robustness, scalability, performance, and efficiency),
- Easy access to functionalities for simulation-in-the-loop, e.g. in the context of optimization and calibration.

References

- [1] M. Kiele-Dunsche, J. Schlake, W. Engler, J. Wagner: "Verwaltungsschale: Interoperabilität für Simulationsmodelle", in Technologie & Trends (2020).
- [2] T. Jung, N. Jazdi, M. Weyrich: "Dynamische Co-Simulation von Automatisierungssystemen und ihren Komponenten im Internet der Dinge", VDI Berichte 2330 (2018).
- [3] S. Bittner, O. Oelsner, T. Neidhold: "Using FMI in a cloud-based Web Application for System Simulation", Proc. 11th Int. Modelica Conference (2015).
- [4] FMI Standard (last accessed on 2021-05-03): <http://fmi-standard.org/index.html>
- [5] Modelica Association (last accessed on 2021-05-03): <https://modelica.org/>
- [6] FMI Standard Specifications (last accessed on 2021-05-03): <https://fmi-standard.org/downloads/>
- [7] BaSys 4.0 Project Website (last accessed on 2021-05-03): <https://www.basys40.de/>
- [8] Industry 4.0 Asset Administration Shell (AAS) Website (last accessed on 2021-05-03): <http://www.i40-aas.de/>
- [9] BMWI: "Details of the Asset Administration Shell: Part 1 – The exchange of information between partners in the value chain of Industrie 4.0 (Version 2.0.1)", Plattform Industrie 4.0 & ZVEI (2020).
- [10] K. Garrels, M. Hoffmeister, J. Wagner: "Die Verwaltungsschale: Fundament für interoperable Industrie-4.0-Systeme", in Technologie & Trends (2019).
- [11] Matlab Webpage (last accessed on 2021-05-03): <https://www.mathworks.com/products/matlab.html>
- [12] SimulationX Webpage (last accessed on 2021-05-03): <https://www.simulationx.com/>
- [13] Dymola Webpage (last accessed on 2021-05-03): <https://www.3ds.com/de/produkte-und-services/catia/produkte/dymola/>
- [14] Modelica Language Webpage (last accessed on 2021-05-03): <https://www.modelica.org/modelicalanguage>
- [15] ABB Ability™ Virtual Commissioning for Drives Webpage (last accessed on 2021-05-03): <https://new.abb.com/drives/software-tools/virtual-commissioning-for-drives>
- [16] ABB Process Automation Digital: Genix Webpage (last accessed on 2021-05-03): <https://new.abb.com/process-automation/genix/the-twin-thats-key-to-decoding-asset-system-dna>
- [17] Open API Specification Website (last accessed on 2021-05-03): <https://swag-ger.io/specification/>

Assistenzsysteme der nächsten Generation für Produktionssysteme

Next Generation Assist Systems for Production Systems

Dr.-Ing. **Jan Fischer**, Dipl.-Math. **Roland Rosen**,
M.Sc. **Franz Georg Listl**, Dr. rer. nat. **Annelie Sohr**,
Dr. rer. nat. **Jan Christoph Wehrstedt**, Siemens AG, München

Kurzfassung

Der effiziente Betrieb automatisierter Produktionssysteme in der Prozess- und Fertigungsindustrie ist eine fortwährende Herausforderung. Neben der Berücksichtigung von Qualitäts-, Zeit- und Kostenzielen rückt immer mehr eine Erhöhung der Flexibilität, z.B. bezüglich kurzfristiger Auftrags- oder Zulieferungsänderungen, in den Vordergrund. Ein Ansatz diese Resilienz in der Produktion zu erreichen, stellen Entscheidungsunterstützungssysteme dar. Diese bieten unterschiedlichen Nutzern wie beispielsweise Produktionsplanern oder Maschinenbedienern eine Assistenz bei ihren Aufgaben. In diesem Beitrag wird ein Architekturansatz für die nächste Generation von Assistenzsystemen, sogenannte Digital Plant Companions (DPC), für diskrete und kontinuierliche Produktionssysteme beschrieben. Wichtige Hauptkomponenten dieser Architektur sind ein DPC-Koordinator, der die Unterfunktionen orchestriert, eine semantische Daten- und Modellintegration, Schnittstellenfunktionen zu Assets, wie z.B. dem Produktionsequipment oder IT/OT-Systemen und der Applikationsteil mit nutzerspezifischen Interaktionsmöglichkeiten. Der Ansatz wird an Beispielen einer Turbinenschaufelproduktion mit erforderlichen Make-or-Buy-Entscheidungen und der Produktionsplanung bei Bäckereien illustriert.

Abstract

The efficient operation of automated production systems in the process and manufacturing industry is an ongoing challenge. In addition to taking quality, time and cost targets into account, increasing flexibility e.g. with regard to short-term changes to orders or supplies, is increasingly being focused. Decision support systems represent one approach to achieving this resilience in production. These systems offer different users such as production planners or operators assistance for their specific tasks. This article describes an architectural approach for the next generation of assistance systems, so-called Digital Plant Companions

(DPC), for discrete and continuous production systems. Important main components of this architecture are a DPC coordinator who orchestrates various functions, a semantic data and model integration, an interface to assets such as production equipment or IT / OT systems and the application part with user-specific interaction options. The approach is illustrated using examples of a turbine blade production with necessary make-or-buy decisions and a production planning use case in bakeries.

1. Motivation

Ein effizienter und möglichst optimaler Betrieb von Produktionssystemen in der Prozess- und Fertigungsindustrie ist eine bekannte, aber fortwährende Herausforderung. Neben der Berücksichtigung von Qualitäts-, Zeit- und Kostenzielen rückt immer mehr die Flexibilität in den Vordergrund, da die Vielfalt herzustellender Produkttypen größer und die Losgrößen kleiner werden. Auch gilt es kurzfristige Auftragsänderungen aber auch veränderte Zulieferungen, z.B. zeitliche Änderungen aber auch bzgl. der Ausführung, Umfang und Qualität, in effektiver Weise zu berücksichtigen. Die zur Umsetzung dieser Punkte und damit zur Steigerung der Resilienz der Produktion erforderlichen Lösungen stützen sich auf die klassische Automatisierung und bilden allerdings eigene digitale Applikationen, die eine daten- und simulationsbasierte Entscheidungsunterstützung bieten (vgl. [1], [2]).

Aktuell wird die Nutzung von Simulation während des Betriebs von vielen Nutzern als noch sehr gering eingeschätzt, wie eine Umfrage aus 2020 unter Anwendern der Fertigungs- und Prozessindustrie ergeben hat [3]. Hingegen nehmen die Verwendung und Analyse von aktuellen und historischen Daten in der Einschätzung der Befragten zunehmend eine größere Rolle ein. Als Hauptgründe, warum die Simulation nicht „immer“ eingesetzt wird, wurde genannt, dass die Simulation „zu teuer“ sei und „zu viel Aufwand“ mache. Ein weiteres wichtiges Kriterium liegt in der fehlenden „Benutzerfreundlichkeit“. Diese genannten Hinderungsgründe motivieren die Autoren, in diesem Beitrag einen Ansatz für die nächste Generation von Assistenzsystemen für diskrete und kontinuierliche Produktionssysteme vorzustellen, der helfen kann, diese Barrieren zu überwinden. Wir bezeichnen diese Assistenzsysteme als Digital Plant Companions (DPC). Sie können verschiedenen Nutzergruppen (z.B. Planer, Operator und auch Maschinenbediener) bei ihren Aufgaben in der Produktion unterstützen. Dazu beziehen DPCs neben dem aktuellen Status des automatisierten Systems weitere Informationen aus verschiedenen Quellen wie Logistik, Instandhaltung und Produktionsplanung ein und verbinden diese Informationen mit daten- und simulationsbasierten Ansätzen. Durch den Begriff „digitaler Zwilling“ bzw. englisch Digital Twin sind in den letzten Jahren zum einen technologische, daten- und modellbasierte Applikationen, die parallel zum benutz-

ten Produkt oder im Betrieb befindlichen System genutzt werden, auf das steigende Interesse eines zugleich größer werdenden Nutzerkreises gestoßen [4]. Zum anderen haben sie zu einer Aufwertung und vermehrten Akzeptanz der Simulation bei der Entwicklung der Produkte und Systeme geführt. Eigene Begriffe, wie Digital Product Twin und Digital Production Twin belegen dies [5]. Aus der Lebenszyklusperspektive stehen einmal die Entwicklungsphasen im Vordergrund und ein anderes Mal die Betriebs- und Servicephasen. Die Entwicklung ist von Berechnungen und Simulationen geprägt, die Betriebsphase von den Daten, insbesondere Mess- und Betriebsdaten, die in immer größerem Umfang, erhoben, gesammelt, ggf. auf Edge-Geräte und in Clouds transferiert und ausgewertet werden. Diese Daten bilden ein Abbild des Produktes/Systems im jeweiligen Augenblick. Daher wird hierfür auch der Begriff „Digital Shadow“ verwendet. Damit stellen sich zwei Fragen:

- Wie passen diese unterschiedlichen Vorstellungen zusammen?
- Warum Assistenzsysteme entwickeln, die über daten- und simulationsbasierte Funktionen verfügen?

Der Begriff DZ integriert die Simulation und die von ihr benutzten Modelle und Daten sowie auch ihre Ergebnisdaten in einen erweiterten Systementwicklungs- und Betriebsprozess. Die Simulationsdaten und -modelle bilden zusammen mit allen digitalen Artefakten, die beginnend mit der Ideenfindung, beim Design, der Konstruktion, dem Engineering anfallen oder erhoben werden den digitalen Zwilling. Dieser entwickelt sich in den Phasen der (virtuellen) Inbetriebnahme sowie Betriebs- und Servicephasen weiter und stellt damit einen über alle Lebenszyklusphasen durchgängigen Begleiter des Produktes bzw. des Systems dar [5], [6], [7]. Diese visionäre Perspektive unterliegt in der Praxis pragmatischen Adaptionen, die z.B. durch begrenzte Entwicklungsaufwände, durch IP- und datenrechtliche Beschränkungen und auch technische Möglichkeiten (wie Datenverfügbarkeit) notwendig sind. Der klare Ansatz und sich in Zukunft weiter realisierende Trend bleibt, der sich in folgender Begriffsfassung des DZ widerspiegelt:

Der digitale Zwilling bezieht sich auf eine Beschreibung einer Komponente, eines Produkts, eines Systems, einer Infrastruktur oder eines Prozesses durch eine Menge zusammengehöriger beschreibender und ausführbarer Modelle. Er ist eine semantisch verknüpfte Sammlung aller relevanten digitalen Artefakte, einschließlich Design-, Konstruktions- und Engineering-Daten sowie Betriebsdaten als auch von Verhaltensbeschreibungen.

Ein verantwortungsvoller Umgang mit diesen digitalen Rohstoffen bedingt auch deren Wieder- und Weiterverwendung. Das bezieht sich auf wirtschaftliche Gründe, da vorhandene Informationen nicht ein weiteres Mal zeit- und kostenaufwändig erbracht werden müssen, aber auch auf inhaltliche Vorteile. Wie in [4] anhand von Virtuellen Inbetriebnahme-Modellen

(VIBN) aufgezeigt, werden nach Ansicht der Autoren in Zukunft auch Verhaltens- und Simulationsmodelle aus der Entwicklung und der VIBN in Applikationen im Betrieb und Service verstärkt Eingang finden. Ihr Vorteil liegt in der Verwendung von Wissen, das sich im Produktionsumfeld aus physikalischen und verfahrenstechnischen Kenntnissen ergibt. Dieses Wissen muss nicht erst erlernt werden, sondern steht sofort zur Verfügung. Diese Aussage setzt den Wert und die Wichtigkeit von datenbasierten und lernenden Verfahren nicht herab. Im Gegenteil, beide Ansätze ergänzen sich, und die Mess- und Betriebsdaten sind erforderlich, damit ein simulationsbasiertes Verfahren auf dem aktuellen Systemzustand aufsetzen kann. Materialflusssimulationen, die in diesem Beitrag im Anwendungsfokus stehen, bilden hier einen Anfang mit ihrer Modellierung von Vorgangs- und Transportzeiten. Komplexere Simulationen, die umfangreiche physikalische oder auch biochemische Berechnungsmodelle erfordern, werden in weiteren Schritten folgen. Einen Anfang setzen hier Diagnosemodelle, die bezogen auf eine Komponente dynamische Zustandsüberwachungen und Fehlerdiagnosen ermöglichen [8]. Für einen erfolgreichen Einsatz auf dem Shopfloor sind benutzerfreundliche Schnittstellen, die eine aufgaben- und rollenspezifische Verwendung ermöglichen, ein weiterer zu berücksichtigender Entwicklungsaspekt.

Mit dem DZ sind die prinzipiellen Voraussetzungen für die wirtschaftliche Erstellung von Assistenzsystemen in der Produktion gegeben. Zu deren effizienter Realisierung – in der nächsten Generation – wird darüber hinaus ein modularer Architekturansatz benötigt, der in diesem Beitrag in Abschnitt 3 nach einem Blick auf den Stand der Technik vorgestellt wird. Der Abschnitt 4 zeigt die Anwendungsszenarien Turbinenschaufelproduktion mit erforderlichen Make-or-Buy-Entscheidungen und eine Produktionsplanung in Bäckereien. Der Beitrag schließt mit einer zusammenfassenden Bewertung und einem Ausblick.

2. Stand der Technik

Digitale Assistenzsysteme treten in der Produktion in verschiedenen Weisen auf. So kann beispielsweise zwischen den unterschiedlichen Produktionsbereichen, wie Fertigung und Montage, dem Planungs- und Informationsmanagement sowie Schulungen und Inspektionen unterschieden werden. Auch die Art zwischen dem Einfluss und der Interaktion mit dem Nutzer des jeweiligen Systems oder bezüglich der genutzten Technologien kann differenziert werden [9]. Eine der meistgenutzten Arten von Assistenzsystemen innerhalb der Produktion sind Entscheidungsunterstützungssysteme. Kasie et al. diskutieren innerhalb einer Recherche sowohl simulationsbasierte Entscheidungsunterstützungssysteme in der Fertigung als auch Ansätze, die Simulation und KI-Methoden integrieren [10]. Auf dieser Grundlage wird ein theoretischer Rahmen für die Entwicklung von Entscheidungsunterstützungssystemen

vorgestellt, der hauptsächlich aus einem Simulationsmodell, einer Datenbank, einer KI-Komponente und einer Benutzeroberfläche besteht. Die Arbeit identifiziert dabei kritische Komponenten, die sich auch in der hier vorgestellten Arbeit teilweise wiederfinden. Es wird dabei auch bereits die Herausforderung adressiert, Methoden der künstlichen Intelligenz mit den modellbasierten Ansätzen eines Entscheidungsunterstützungssystems zu verknüpfen. Dennoch fehlen in dem Ansatz Technologien zur Extraktion und Integration von Produktionsdaten aus heterogenen Quellen innerhalb wissensbasierter Modelle sowie die Flexibilität in der Aufgabenerfüllung durch hybride Modelle aus einer Kombination von KI und Simulation.

Entscheidungsunterstützungssysteme sind üblicherweise auf einen genauen Use Case bzw. eine konkrete Aufgabe zugeschnitten, wie dies beispielsweise in [11], [12] auch zu sehen ist. In [12] wird ein Assistenzsystem auf Basis von Machine Learning Algorithmen, genauer gesagt Entscheidungsbäumen, und Fuzzy Logik vorgestellt. Das Assistenzsystem soll bei der Produktionsplanung unterstützen, um das Closed-Loop Supply Chain Management Problem zu lösen. In [11] werden Augmented Reality Methoden mit Simulation-basierter Optimierung kombiniert. Das konkrete Ziel ist dabei, die Modelle einer ereignisdiskreten Simulation über Augmented Reality anzusehen und damit greifbarer zu machen. Im Gegensatz zu den gerade genannten Ansätzen präsentieren Kunath und Winkler in [13] ein Framework für ein Entscheidungsunterstützungssystem im Auftragsmanagement, das auf einem Digitalen Zwilling des Fertigungssystems basiert. Wie auch in unserem Ansatz werden hier mehrere Subsysteme mit unterschiedlichen Datenquellen innerhalb des Digitalen Zwillings vereint. Ebenfalls ist eine Komponente zur automatischen Generierung von Simulationsmodellen vorhanden.

3. Ansatz

3.1 Gesamtarchitektur

Im folgenden Kapitel wird eine mögliche Architektur des DPC sowie dessen Hauptkomponenten erläutert. Ähnliche Ansätze wurden von uns bereits in [5] und [14] beschrieben. Daher wird hier, nach einer kurzen Überblicksdarstellung, der Fokus auf einige Teilaspekte des DPC gelegt. Abgrenzend gegenüber vielen anderen Assistenzsystemen richtet sich der DPC an systematischen Problemstellungen aus, stellt also Assistenzsystemfunktionen bereit, die eine Integrationen von mehreren Komponenten und Datenquellen erfordern und meist auf Produktionssystem- oder Fabrikebene einzuordnen sind. Typische Fragestellungen, die damit adressiert werden können, sind daher z.B. Planungs-, Reihenfolgebildungs- und Steuerungsprobleme. Zur Umsetzung von DPCs, wird ein modularer Lösungsansatz angewendet, der unterschiedliche Technologien (Ontologien, natürliche Spracherkennung, automatische

Modellgenerierung ...) in einer Gesamtarchitektur vereint. Strukturell besteht ein DPC aus drei bis vier Kernelementen, die in jeder Umsetzung leicht anders aussehen können, aber immer ähnliche Funktionen erfüllen (siehe auch Abbildung 1).

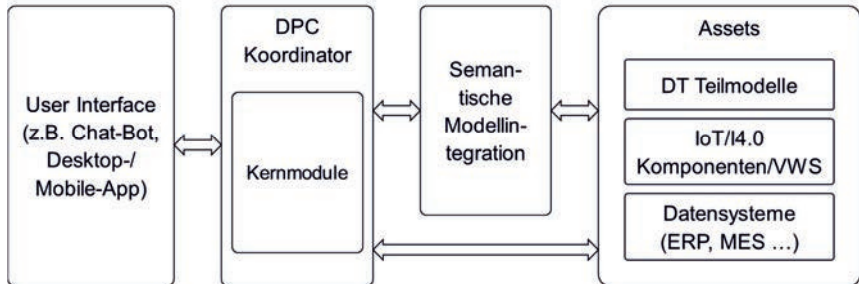


Bild 1: Struktureller Aufbau eines Digital Plant Companion

Dies sind (i) User Interfaces für Endanwender wie z.B. Chat-Bots, Web-Applikationen oder Mobile-Apps. Diese greifen auf Softwareschnittstellen des DPC zu, die dieser über das hier DPC Koordinator genannte Hauptmodul bereitstellt. (ii) Der DPC-Koordinator verwaltet alle ankommenden Anfragen, z.B. durch intelligente Interpretation oder durch das simple Bereitstellen einer funktionspezifischen Schnittstelle, auf die Endanwender zurückgreifen können. Er stellt auch wesentliche funktionale Bausteine bereit, die Kernmodule, die in vielen Assistenzsystemumgebungen notwendig sind. Dies sind z.B. Mechanismen zur Interpretation natürlichsprachlicher Eingaben (Intent-Erkennung), Funktionen zur Verkettung mehrerer Simulations- und Optimierungsläufe (Workflow-Erstellung) oder auch Services zur automatischen Generierung von Simulations- und Optimierungsmodellen. Einige der Kernmodule sind im folgenden Unterkapitel näher beschrieben (siehe auch Abbildung 2). (iii) Um auf die heterogenen und stark variierenden Datenquellen in Fabriken in der immer gleichen Weise zugreifen zu können, müssen diese entsprechend eines semantischen Datenmodells beschrieben sein. Diese bilden damit den modellbasierten Kern des Systems und beschreiben sowohl das Produktionssystem selbst als auch die Metainformationen von Simulations- und Optimierungsservices. Dies ermöglicht nicht nur das effiziente Verwalten von mehreren Simulationsläufen, sondern auch das dynamische Einbinden von Simulationen in (Graph-)Datenbankabfragen. Zur Modellierung der Ontologien wird weitgehend auf etablierte Standards zurückgegriffen (z.B. ISA88, ISA95), um eine produktionssystemübergreifende Wiederverwendung zu gewährleisten und damit den angesprochenen Aufwand für die Modellgenerierung zu minimieren. (iv) Schließlich greift der DPC auf unterschiedliche Assets einer

Produktion zurück, IT-Systeme, IoT-/Edge-Komponenten (ggf. als I4.0 Komponenten mittels Verwaltungsschale bereitgestellt) oder Cloud-Services. Diese sind daher nicht als Teil des DPC anzusehen, müssen jedoch ihre Daten an das semantische Datenmodell koppeln und über Programmierschnittstellen erreichbar sein. Hierzu ist in vielen Produktionen Retro-Fit Aufwand notwendig.

3.2 DPC Koordinator und Kernmodule

Der DPC Koordinator ist verantwortlich für den Gesamtablauf des DPC. Er steuert und verwaltet alle ankommenden Anfragen. Dabei bietet er als Server Schnittstellen an, mit denen Endanwender Applikationen interagieren können, z.B. zum Durchführen von Vorhersagesimulationen oder Optimierungsläufen. Gleichzeitig fungiert er als Client, um externe Assets wie z.B. Simulationsteilmodelle oder andere IoT Komponenten ansprechen zu können. Die Funktionen, die vom Koordinator bereitgestellt werden müssen, unterscheiden sich vom jeweiligen Anwendungsfall. Daher wird ein möglichst modularer und servicebasierter Ansatz angestrebt, der eher einem Baukastensystem ähnelt als einem monolithischen Funktionsblock. Dies ermöglicht auch, dass nicht für alle Funktionen neue Komponenten entwickelt werden, sondern dass weitgehend auf bestehende Ansätze zurückgegriffen werden kann. Abbildung 2 zeigt typische Kernfunktionen des Koordinators, von denen einige im Folgenden näher erläutert werden.

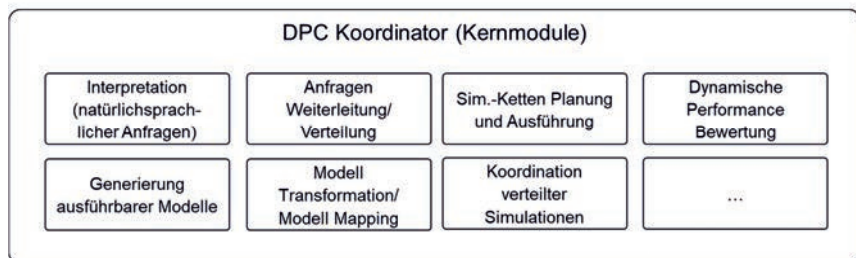


Bild 2: Kernmodule des DPC Koordinators

3.2.1 Modellgenerierung und -transformation

Eine wesentliche Funktion ist die Erzeugung ausführbarer Simulations- und Optimierungsmodelle auf Basis formal beschriebener Datenstrukturen. Es wird dabei in allen Fällen davon ausgegangen, dass die heterogenen Daten einer Produktion (ERP, MES, SCADA ...) bereits in einem generischen Datenmodell integriert worden sind, sodass produktionsübergreifend ein ähnlicher Zugriff darauf erfolgen kann (siehe Abschnitt 3.4). Darauf aufbauend müssen

diese Produktionsinformationen dann in die funktionsorientierten Modelle einer Simulation transformiert werden. Da dieser Schritt in großen Teilen das Expertenwissen von Simulations-/Modellierungs- und Applikationsexperten erfordert und es meist mehrere Varianten gibt, mit denen ein Sachverhalt in einer Simulation abgebildet werden kann, ist dieser Schritt auf absehbare Zeit vmtl. nicht vollautomatisierbar. Um trotzdem den Aufwand für die Modellierung zu reduzieren, wird eine Kombination aus folgenden Methoden und Techniken eingesetzt: (i) datenmodellbasierte Abbildungen (z.B. mittels XML Stylesheets), (ii) Graph basierte Methoden, z.B. Semantic Reasoning oder Abfragesprachen wie SPARQL, um Datenstrukturen direkt in den Wissensgraphen in das Zielformat zu transformieren, (iii) trainierte Modelle wie neuronale Netze, um z.B. Simulationsparameter aus historischen Produktionslogs zu ermitteln, und (iv) klassische imperative Programmieransätze mittels Skriptsprachen wie Python. Alle Ansätze bieten Vor- und Nachteile

3.2.2 Simulationsketten Planung und Ausführung

In vielen Fällen werden Simulations- und Optimierungsmodelle im laufenden Betrieb nicht als einzelne Services eingesetzt, sondern in Workflows in denen Datenvor- und nachverarbeitungsschritte oder auch Simulationen auf unterschiedlichen hierarchischen Ebenen nacheinander oder bedingt voneinander ausgeführt werden. Eine häufige Kombination ist auch die Durchführung von Optimierungsalgorithmen zur Produktionsplanung basierend auf einem stark vereinfachten Kapazitätsmodell einer Fertigung mit einer nachgelagerten Validierung mittels eines präziseren Modells, z.B. auf Basis einer ereignisdiskreten Materialflusssimulation.

In diesem Modul werden daher Funktionen bereitgestellt, um solche Workflows als Ketten von Services ausführen zu können. Hier wird unter anderem auf bereits existierende Ansätze zurückgegriffen (z.B. [15], [16]), um Simulationen als digitale Services verfügbar zu machen. In einfachen Fällen können die Workflows statisch beschrieben werden, sodass eine sequenzielle Abarbeitung der einzelnen Schritte a-priori festgelegt ist. In komplexeren Fällen hängt eine Serviceausführung von dem Ergebnis eines vorherigen Service ab oder von anderen externen Faktoren, und es muss dynamisch entschieden werden, wie und ob dieser Service ausgeführt wird.

Ferner ist es möglich, dass eine Funktion durch unterschiedliche Services ausgeführt wird. Zum Beispiel könnte eine Produktionsplanung mittels einer komplexen Optimierung durchgeführt werden, mittels einer Heuristik oder durch einfache Priorisierungsregeln. Dafür müssen Simulationsservices ihre Fähigkeiten wie z.B. Laufzeitverhalten und Ergebnislänge als Metainformation bereitstellen, um dann während der Abarbeitung des Workflows ad-hoc entschei-

den zu können, welcher Service gerade der am besten geeignete ist, um das gewünschte Ziel zu erreichen. Diese Metainformationen werden als semantische Modelle beschrieben (siehe Abschnitt 3.4).

Weitere Kernmodule, die hier nicht im Detail beschrieben werden, sind Module zur Interpretation natürlicher Spracheingaben zur Bereitstellung von Chat-Bot Applikationen. Hier besteht ein wesentlicher Aspekt zum einen in der Extraktion der Nutzerintentionen (z.B. Vorhersagesimulation, Optimierung, Zustandsschätzung) und zum anderen in der Extraktion der relevanten Parameter. Außerdem werden Methoden zur weitergehenden Analyse der Simulationsergebnisse sowie Funktionen zur dynamischen Laufzeitbewertung, z.B. bzgl. der Auslastung der zugrunde liegenden Rechnerarchitektur, bereitgestellt.

3.3 Assets

Die Zielsetzung des DPC liegt in der Entscheidungsunterstützung für den Menschen in seinen vielfältigen Aufgaben in der Produktion in einer jeweils spezifischen Situation. Um diese Assistenz zu geben, ist es erforderlich, dass der DPC die spezifische Situation kennt. In unserem Ansatz differenzieren wir hierzu zwischen drei verschiedenen Aspekten (vgl. Bild 1). Zum ersten die Produktion mit seinen Maschinen und dem Equipment, zum zweiten, die Informationen, die sich aus der Information Technology (IT) und Operation Technology (OT) ergeben, und zum dritten die Modelle – insbesondere auch Simulationsmodelle – die im jeweiligen Anwendungsfall das Verhalten der Produktionseinheit adäquat wiedergeben. Die Digitalisierung der Industrie – oder kurz Industrie 4.0 – gibt dazu Unterstützung in verschiedener Art. Durch ihre Technologien (u.a. Internet-of-Things, IoT) werden mittlerweile innerhalb der Produktion permanent große Datenmengen durch verschiedene Sensoren und auch von Maschinen erfasst und auf Edge-Computer oder in die Cloud transferiert und dort zur Verfügung gestellt. Mit dem Referenzarchitekturmodell Industrie 4.0 (RAMI 4.0) und der Definition der Verwaltungsschale (VWS) wird der herstellerübergreifende Zugriff und Austausch von Informationen durch entsprechende Standardisierungen vorangetrieben. Diese Informationen beziehen sich u.a. auf Maschinen und ihre Komponenten sowie Unterlagen wie z.B. Auftragsdaten. Damit unterstützen diese Industrie 4.0 Ansätze den hier skizzierten Architekturansatz und die ersten beiden oben genannten Aspekte.

Warum also nicht warten auf die „Fertigstellung Industrie 4.0“? Dagegen sprechen einige gute Gründe. Die Anforderungen an die Schnittstellen können nur durch die gleichzeitige Realisierung solcher fortschrittlichen Assistenzsysteme ermittelt werden. Außerdem sollen die Assistenzsysteme auch für existierende Anlagen („Brownfield“) entwickelt werden, die in

der Regel einen sehr unterschiedlichen Digitalisierungsgrad besitzen. Des Weiteren ist auch die Datenlandschaft sehr heterogen, die erforderlichen Daten stammen aus sehr unterschiedlichen Quellen, von verschiedenen Ebenen, wie z.B. MES, ERP, Maschinendaten, Datenbanken oder Data Warehouses und liegen oftmals in unterschiedlichen Formaten vor. In [17] wird gezeigt, dass der DZ in allen drei Dimensionen des RAMI 4.0 „schwebt“. Zur Konkretisierung wird aufbauend auf [7] eine mögliche Infrastruktur für DZ gezeigt, die in einer konkreten Situation die jeweils erforderlichen Modelle bereitstellen, die dann den DZ bilden. Bezieht man dieses Vorgehen auf ein Simulationsmodell der Produktion, so zeigt dies ein Vorgehen, das es ermöglicht, Simulationsmodelle von z.B. verschiedenen Maschinen auszuwählen und kombiniert in dem Assistenzsystem anzuwenden. Eine solche Kombinierbarkeit setzt eine Kompatibilität der Simulationsmodelle voraus. Hierzu gibt es Ansätze aus der Co-Simulation [18] und wiederum aus dem Ansatz der Industrie 4.0 VWS, bei dem ein VWS-Teilmodell ein Simulationsmodell bereitstellt und über eine „Simulation Model Description“ die Informationen zum Simulationsmodell vermittelt [19].

Die Industrie 4.0 Aktivitäten zur Standardisierung werden bei der Umsetzung des in diesem Beitrag gezeigten Architekturkonzeptes helfen und ihn auch beschleunigen. Für eine zeitnahe und praxisgerechte Realisierung eines DPC sind wegen des Umfangs und der Heterogenität der Daten und Modelle im jeweiligen Anwendungsfall konkrete Lösungsansätze zu entwickeln.

3.4. Wissensgraph

Eine wesentliche Komponente in der Architektur des DPCs ist der Wissensgraph der als Bindeglied zwischen den verschiedenen Datenquellen und -systemen, Teilen des Digitalen Zwillings und den Kernmodulen des DPCs fungiert. Er hat die Aufgabe Struktur in die genannten Teilsysteme zu bringen, diese miteinander in Verbindung zu setzen und schließlich eine semantische Schicht über diesen zu erzeugen. Über diese können die Kernmodule des DPCs automatisiert auf die Datenquellen und -systeme zugreifen und dadurch kann der gesamte Digitale Zwilling durch Zusammenführung aller Aspekte für das jeweilige System erzeugt werden.

Der Wissensgraph besteht dabei im Wesentlichen aus mehreren kleinen Wissensgraphen, die das Wissen einer spezifischen Domäne abdecken. Beispielsweise ist ein kleinerer Domänen-Graph dafür verantwortlich, die Daten und Datenerzeuger innerhalb der Fertigung zu beschreiben. Hierbei empfiehlt es sich auf produktionsnahe Standards wie ISA-95/ISA-88, AutomationML, OPC UA zurückzugreifen. So können wiederverwendbare und standardisierte Modelle, die in einer Produktionsontologie beschrieben sind, erzeugt werden. Ein weiterer

Aspekt, der durch einen Teilgraphen dargestellt werden muss, sind die Teilmodelle des Digitalen Zwillings und damit unter anderem auch Simulations- und Optimierungsmodelle. Um diese beschreiben, miteinander in Verbindung setzen und schließlich auch vergleichbar machen zu können, müssen zum einen die Komponenten als auch die Metadaten der jeweiligen Modelle enthalten sein. Darauf aufbauend werden die Modelle in einer übergeordneten hierarchischen Struktur eingeordnet, die die Granularität der Modelle widerspiegelt. Hierbei ist es sinnvoll sich für eine Einordnung an den Schichten von RAMI 4.0 zu orientieren. Die Instanzen dieses Teilgraphen sind in diesem Fall die implementierten Modelle. Die verschiedenen Teilgraphen müssen zusätzlich Informationen beinhalten, wie sie eine Abbildung bzw. Verbindung zueinander erzeugen, sodass je nach Aufgabe, die nötigen Teilgraphen hinzugezogen werden können.

4. Anwendungsfälle

Im Folgenden werden zwei Anwendungsfälle aus der industriellen Praxis beschrieben, in denen Teile des vorgestellten DPC Ansatzes umgesetzt wurden.

4.1 Optimierung einer Turbinenschaufelproduktion

Innerhalb des EU-geförderten Projektes ASSISTANT¹, wird ein Assistenzsystem für die Produktion auf Basis von digitalen Zwillingen erstellt. Die digitalen Zwillinge breiten sich über die Prozessplanung, Produktionsplanung und Steuerung, sowie Real-Time Control aus, um so eine flexible Entscheidungsunterstützung über mehrere zeitliche Planungsebenen in die Produktion zu bringen. Des Weiteren werden die digitalen Zwillinge durch die gemeinsame Nutzung von Machine Learning, Optimierung, Simulation sowie Domänen-Modellen umgesetzt. Die Daten innerhalb der Produktion werden durch eine sog. Data Fabric verwaltet, angereichert und für die Digitalen Zwillinge nutzbar gemacht. Validiert wird das Konzept unter anderem anhand einer Turbinenschaufelproduktion in Berlin. Das Ziel ist dabei die derzeitige Steuerung anhand der erwarteten Fertigstellungstermine der geplanten Produktionslast zunächst zu validieren und anschließend durch Berechnung der Starttermine für die einzelnen Aufträge zu optimieren. Darauf aufbauend soll eine Unterstützung für Make-or-Buy-Entscheidungen bezüglich einzelner Teile nach definierten Geschäftszielen erfolgen.

¹ <https://assistant-project.eu/>

4.2 Produktionsplanung für KMU-Bäckereien

Im Rahmen eines EU-geförderten Projektes wird analysiert, inwiefern der Assistenzsystemgedanke auch auf kleinere Organisationen übertragen werden kann – in diesem Fall kleine und mittelständische Bäckereien in Europa.²

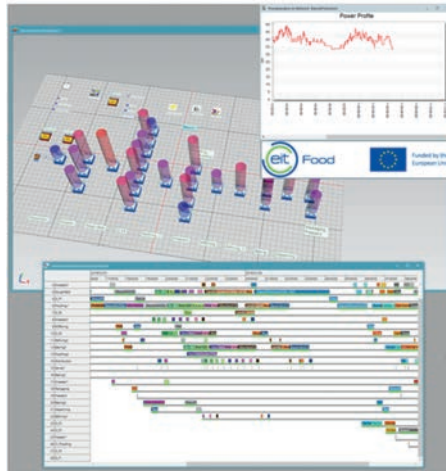


Bild 3: Screenshot einer Materialflusssimulation einer Bäckerei.³

Aufgrund des üblicherweise geringen Budgets für Simulation und Digitalisierung in KMUs, muss hier das Assistenzsystem weitgehend durch ein parametrierbares Produkt realisiert werden, um kostenintensive Beratungsdienstleistungen durch Experten, insbesondere für den Systemaufbau und die Modellpflege, zu minimieren. Prototypisch wurde hier eine Low-Code basierte Nutzeranwendung erstellt, die auf unterschiedliche Optimierungsservices zur Produktionsplanung zurückgreifen kann. Insbesondere sollen durch die Optimierung unnötige Energieverbräuche z.B. durch suboptimal sortierte Backprozesse sowie hohe energetische Lastspitzen minimiert werden. Eine Vorhersage eines energetischen Lastprofils kann z.B. durch eine Materialflusssimulation ermittelt werden (siehe Abbildung 3). Zur Beschreibung der notwendigen Bäckereidaten, wurde der Produktionsstandard ISA95/ISA88 umge-

² <https://www.eitfood.eu/projects/optimization-of-bakery-processes-by-a-computational-tool-together-with-consumer-feedback-to-minimize-ecological-footprint-and-food-waste-2020>

³ Planungsdaten entnommen aus [20]; Energielastzustände der Produktionsmaschinen fiktiv

setzt. Dieser erlaubt die Beschreibung der Produkte, Prozesse, Personal- und Maschinenbedarfe sowie die Ergebnisse von Produktionsplanungsläufen.

5. Zusammenfassung

In diesem Paper wurde ein Ansatz für ein Assistenzsystem zur Entscheidungsunterstützung im operativen Betrieb von diskreten Fertigungen und Prozessindustrieanlagen vorgestellt. Dabei wurden insbesondere die wesentlichen Komponenten sowie ihr Zusammenhang zu den Modellen der Fertigung, dem Digitalen Zwilling, beschrieben. Abschließend wurden zwei Anwendungsfälle skizziert, in denen Assistenzsysteme in der industriellen Praxis momentan umgesetzt werden.

Zur breiteren Umsetzung sind weitere Forschungs- und Entwicklungsarbeiten notwendig, die sich vor allem darum drehen, den Initialaufwand für den Aufbau des Systems sowie dessen Betrieb zu minimieren, also insbesondere Aktivitäten zur automatischen Generierung von Simulationsmodellen, zur Synchronisierung der Modelle mit dem aktuellen Anlagenzustand sowie der Kopplung von datengetriebenen Ansätzen (AI) mit Simulationsmodellen.

Diese Arbeit wurde von der Europäischen Kommission im Rahmen der Förderprojekte AS-SISTANT [101000165] und EIT Pro4Bake unterstützt.

Literatur

- [1] W. Kuehn, „Digital twins for decision making in complex production and logistic enterprises“, *Int. J. Des. Nat. Ecodynamics*, Bd. 13, Nr. 3, S. 260–271, Aug. 2018, doi: 10.2495/DNE-V13-N3-260-271.
- [2] R. Rosen, G. von Wichert, G. Lo, und K. D. Bettenhausen, „About the importance of autonomy and digital twins for the future of manufacturing“, *IFAC-Pap.*, Bd. 48, Nr. 3, S. 567–572, 2015.
- [3] M. Oppelt, L. Bruckner, R. Rosen, M. Barth, L. Urbas, und J. Jäkel, „Die aktuelle und zukünftige Nutzung von Simulation: Umfrageergebnisse für die Fertigungs- und Prozessindustrie“, *Atp Mag.*, Bd. 63, Nr. 04, S. 64–73, 2021.
- [4] R. Rosen u. a., „Die Rolle der Simulation im Kontext des Digitalen Zwillings: Virtuelle Inbetriebnahme-Modelle als Mittler zwischen Phasen und Anwendungen?“, *Atp Mag.*, Bd. 63, Nr. 04, S. 82–89, 2021.

- [5] R. Rosen, J. Fischer, und S. Boschert, „Next Generation Digital Twin: an Ecosystem for Mechatronic Systems?“, *IFAC-Pap.*, Bd. 52, Nr. 15, S. 265–270, 2019, doi: 10.1016/j.ifacol.2019.11.685.
- [6] R. Rosen, S. Boschert, und A. Sohr, „Next Generation Digital Twin: Öko-System für simulationsbasierte Lösungen im Betrieb“, *Atp Mag.*, Bd. 60, Nr. 10, S. 86–96, 2018.
- [7] S. Boschert und R. Rosen, „Digital Twin—The Simulation Aspect“, in *Mechatronic Futures*, P. Hehenberger und D. Bradley, Hrsg. Cham: Springer International Publishing, 2016, S. 59–74.
- [8] S. Boschert, C. Heinrich, und R. Rosen, „Next generation digital twin“, in *Proceedings of the 12th International Symposium on Tools and Methods of Competitive Engineering TMCE*, Delft University of Technology, 2018, S. 209–218.
- [9] X. Yang und D. A. Plewe, „Assistance Systems in Manufacturing: A Systematic Review“, in *Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future*, Bd. 490, C. Schlick und S. Trzcieliński, Hrsg. Cham: Springer International Publishing, 2016, S. 279–289.
- [10] F. M. Kasie, G. Bright, und A. Walker, „Decision support systems in manufacturing: a survey and future trends“, *J. Model. Manag.*, Bd. 12, Nr. 3, S. 432–454, Aug. 2017, doi: 10.1108/JM2-02-2016-0015.
- [11] I. Karlsson, J. Bernedixen, A. H. C. Ng, und L. Pehrsson, „Combining augmented reality and simulation-based optimization for decision support in manufacturing“, in *2017 Winter Simulation Conference (WSC)*, Las Vegas, NV, Dez. 2017, S. 3988–3999, doi: 10.1109/WSC.2017.8248108.
- [12] G. González Rodríguez, J. M. Gonzalez-Cava, und J. A. Méndez Pérez, „An intelligent decision support system for production planning based on machine learning“, *J. Intell. Manuf.*, Bd. 31, Nr. 5, S. 1257–1273, Juni 2020, doi: 10.1007/s10845-019-01510-y.
- [13] M. Kunath und H. Winkler, „Integrating the Digital Twin of the manufacturing system into a decision support system for improving the order management process“, *Procedia CIRP*, Bd. 72, S. 225–231, 2018, doi: 10.1016/j.procir.2018.03.192.
- [14] F. G. Listl, J. Fischer, R. Rosen, A. Sohr, J. C. Wehrstedt, und M. Weyrich, „Decision Support on the Shop Floor Using Digital Twins“, eingereicht für die APMS 2021 Conference, Nantes, France, Sep. 2021.
- [15] G. Höllthaler u. a., „Simulationsbasierte Produktionsassistenz“, *Wt Werkstattstech. Online*, Bd. 108, Nr. 3, S. 148–154, 2018.

- [16] J. Fischer, B. Obst, und B. Lee, „Integrating material flow simulation tools in a service-oriented industrial context“, in *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*, 2017, S. 1135–1140.
- [17] A. Deuter und F. Pethig, „The Digital Twin Theory - A New View on a Buzzword“, *Ind. 40 Manag.*, Bd. 2019, Nr. 1, S. 27–30, Feb. 2019, doi: 10.30844/I40M_19-1_S27-30.
- [18] C. Scheifele, C. Härle, R. Rosen, T. Heinzerling, und M. Röhler, „Co-Simulation als Realisierung digitaler Zwillinge: Austausch und Kopplung von (Teil-) Modellen für eine virtuelle Inbetriebnahme“, *Atp Mag.*, Bd. 63, Nr. 04, 2021.
- [19] Plattform Industrie 4.0, SG AAS, „Interoperabilität hoch zwei für Simulationsmodelle“. <https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/Interoperabilit%C3%A4t-hoch-zwei.html> (zugegriffen Apr. 30, 2021).
- [20] F. T. Hecker u. a., „Application of a modified GA, ACO and a random search procedure to solve the production scheduling of a case study bakery“, *Expert Systems with Applications.*, S. 5882–5891, 2014.

Automatic Detection of Start and End Times of Batch Phases for the Process Industry

Automatisierte Erkennung der Start- und Endzeitpunkte von Batchphasen in der Prozessindustrie

Dipl.-Ing. **Silke Merkelbach**, Prof. Dr.-Ing. **Leon Urbas**,
Josefine Gatter,
Professur für Prozessleittechnik/AG Systemverfahrenstechnik,
Technische Universität Dresden;
Dr. **Ruomu Tan**, Dr. **Marco Gärtler**, B.Sc. **Rebecca Gedda**,
ABB Corporate Research Center Germany;
Dr. **Franz David Böhner**, Bayer AG Leverkusen

Abstract

Data-driven and hybrid methods for optimising chemical and pharmaceutical processes are experiencing increasing interest in the process industry [1]. The potential for optimisation is particularly high for batch processes, where the largely transient process regime and the temporal segmentation into phases create a notable additional complexity compared to continuous processes. To analyse batch processes, knowledge of the beginning and end of these phases is necessary. However, data from the recipe sequences recorded by the data acquisition systems are error-prone, inaccurate, inconsistent with the purpose of the analysis, and in some cases do not exist at all. Given these circumstances, the labelling of the phases must sometimes be carried out retrospectively which, if done manually, is very time-consuming [2]. Automation is desirable, and solutions exist for specific applications, for example, for the classification of sequences [3] or the interactive segmentation of trends [4].

This paper compares the performance of different approaches towards automatic batch phase detection - change point detection [5], [6], classification directly on time series data [7] and classification using statistical features [8]. Statements, including maximum accuracy and minimum first- and second-order error rates, are collected from the simulation of a real process; without measurement errors, noise or imperfections. By applying them to "dirty" data from the real process, we demonstrate the robustness of the algorithms against typical artefacts of data recording in industrial practice.

Kurzfassung

Datengetriebene und hybride Methoden zur Optimierung von chemischen und pharmazeutischen Prozessen erfahren in der Prozessindustrie zunehmend größeres Interesse [1]. Besonders hoch ist das Optimierungspotenzial bei Batch-Prozessen, bei denen das weitgehend instationäre Prozessregime und die zeitliche Unterteilung in Phasen eine nennenswerte zusätzliche Komplexität im Vergleich zu kontinuierlichen Prozessen schaffen. Um Batch-Prozesse analysieren zu können, ist die Kenntnis von Beginn und Ende dieser Phasen notwendig. Die von den Datenakquisitionssystemen aufgenommenen Daten aus den Rezeptabläufen sind jedoch fehlerbehaftet, ungenau, im Detailgrad unstimmt mit dem Analysezweck oder in manchen Fällen gar nicht vorhanden. Daher muss die Auszeichnung der Phasen teilweise nachträglich durchgeführt werden und ist, wenn manuell durchgeführt, sehr zeitaufwändig [2]. Eine Automatisierung ist wünschenswert, für spezielle Anwendungen existieren Lösungen, beispielsweise zur Klassifizierung von Sequenzen [3] oder zur interaktiven Segmentierung von Trends [4].

Dieser Beitrag vergleicht die Leistung verschiedener Ansätze auf dem Weg zur automatischen Batchphasenerkennung - Change Point Detection [5], [6], Klassifikation mithilfe statistischer Features [8] und basierend auf Zeitreihendaten [7]. Aussagen, u. a. zur Genauigkeit und minimalen Fehlerraten erster und zweiter Ordnung, sind an der Simulation eines realen Prozesses erhoben; ohne Messfehler, Rauschen oder Fehlstellen. Über die Anwendung auf „schmutzige“ Daten aus dem realen Prozess zeigen wir die Robustheit der Algorithmen gegenüber typischen Artefakten der Datenaufzeichnung in der industriellen Praxis auf.

1. Introduction

Data enables enterprises to make objective decisions based on quantitative criteria. Usable information derived from data can take many forms such as statistical measures, custom calculations (Key Performance Indicators), data-driven models, but also parameters of mechanistic models.

The process industry, in particular the fine and specialty chemicals, tends to run processes in batches whenever flexibility and a high degree of freedom in the sequence and duration of single processing steps (batch phases) is desirable. The analysis of batch processes requires reliable information about the start and end points of the batch phases. Furthermore, information about utilised materials and the path of a batch through the plant can be relevant. Finally, to fully describe plant performance, unplanned events which upset the production schedule (such as equipment failures, malfunctions or maloperation) need to be classified. Without this informative context [9], collected operational data can be unusable. For manually operated

units, such event data is only known if it is manually collected. This raises doubts concerning their quality and creates expensive workload for plant employees. In automated plants, depending on installed process information technology and process control systems, archiving of production commands may also be difficult. In general, data quality issues may arise if operators interfere with the batch control system to enact manual scheduling decisions or react to unplanned events, which occur frequently in ageing plants. In general, batch plants lack behind continuous plants in their degree of automation [10].

Consequently, the availability and quality of event data which are relevant for basic and advanced process analytics can be limited. Measurements of levels, weights, pressures, temperatures and flows are relevant for a functioning regulatory control system. Values of actuators and sensors implicitly contain a lot of information about occurring batch phases, and whether they behave normally. Often, such data is stored on a process historian system as time series values. They pose a promising basis for evaluating the quality of existing batch phase statistics. Furthermore, if no batch statistics are archived, it may be possible to infer these statistics based on time series data.

This is a difficult task for a human expert. The potentially large number of batch phases, measured process values, types of process noise, and the occurrence of unplannable events create complexity. Given the inherent reproducibility of nominally executed batch phases, Machine Learning methods bear the potential to support this task. However, there exist several challenges for the application of Machine Learning (ML):

- ML has no intrinsic understanding of batch process constraints and dynamics
- Many methods of ML are not suitable for or not frequently used for time series analysis
- The noise-to-information ratio is limited by plant/process changes, installed sensors, operational discipline or level of automation, and equipment reliability

This article poses a first step to understanding how methods of ML can support a human expert in identifying, distinguishing, and understanding recipe events and extraordinary events - as well as recurrent behaviours and process noise. We applied and compared three promising approaches to the detection of phases using industrial data:

1. Classification of batch phases based on sensor measurements (e. g. [7])
2. Classification of batch phases based on statistical features of the time series data (Esmael et al. [8])
3. Change point detection based on the sensor measurements [11]

These methods have in common that they require only moderate amounts of training data and have shown good results in similar applications.

Specific focus is put on showing on simulated data how to properly pose problems using available process insight such that ML algorithms yield interpretable results. By applying the methods to real process data with measurement errors, noise and imperfections, we demonstrate the robustness of the algorithms against typical artefacts of data recording in industrial practice.

The general outline of the article is as follows: In section 2 we introduce the general framework containing the scope of the experiments, a description of the data and the applied evaluation metrics. In sections 3 to 5 we describe the test of the three chosen methods. Section 6 provides a comparison between the results and section 7 concludes our paper.

2. General Framework

In this paper we use simulated data and real data. Both the simulated and the real batch plant consist of multiple units each which have specific process steps running. Analysing the whole plant at once is highly complex and in some cases not feasible. To split this issue into smaller tasks, we start looking at single units of a process plant in this paper with the plan to track the batch through the whole plant later on.

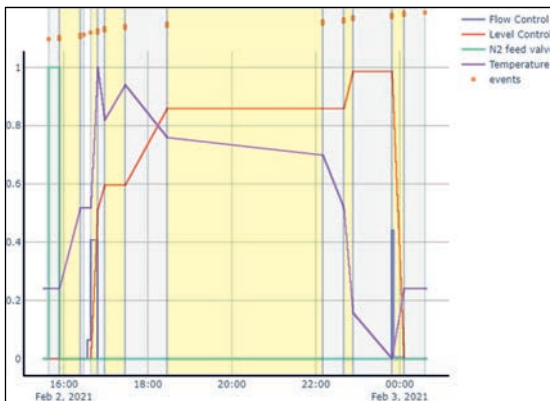


Fig. 1: Plot of a simulated batch run with events

The simulated data set consists of 370 individual runs, with four process variables and 13 unique phases. An optional sampling step occurs in 32 of the runs. A run takes on average 10 hours, the shortest is 7.2 hours and the longest 17.2 hours. One batch of the simulated data including its events and phases in between them is shown in Fig. 1.

Similarly, the real data set consists of 1190 individual runs collected over a period of one and a half years, with eight process variables and 43 different events and phases in between them. The median runtime is about

8.5 hours, while the extremes are 50 minutes and 36 days, respectively. Both extremes can't be ordinary batches and are therefore examples for imperfect batch data. Fig. 2 shows an exemplary batch run within the real data.

For both of the datasets we use the same pipeline, including pre-processing. The

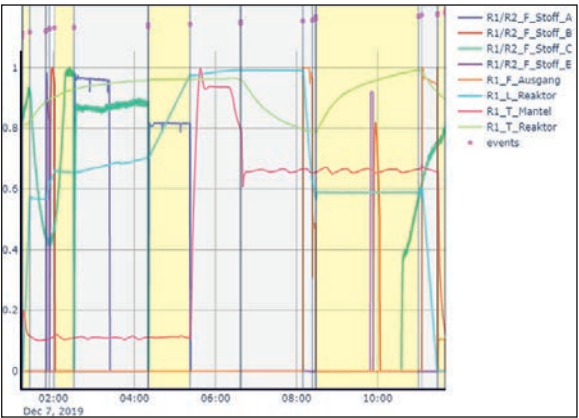


Fig. 2: Plot of a real batch run with events

quantitative quality metrics are aligned with the best practice suggested by the *RUPTURES* Python package [11]. To be able to compare all results, we transform the classification results from sections 3 and 4 to a series of change points.

Table 1 gives an overview over the chosen metrics. The Absolute Error metric (AE) is the absolute difference in number of predicted change points compared to the actual change points. The Meantime metric (MT) gives the average time difference in seconds between a predicted change point and the closest actual change point. The Precision metric indicates the percentage of accurate predictions and the Recall metric detects the actual change points being predicted within a tolerance. Over-segmentation causes Precision being close to zero, whereas Recall approaches 1 and vice versa for under-segmentation. The harmonic mean between the two gives the F1-score. For

satisfying segmentation performances, the F1-score is close to one. For the calculation of Precision, Recall and F1-score we use a tolerance of 5 minutes with regard to the shortest phase in the simulated data which is 5 minutes long. The Rand index (RI) is used to compare how well the partitions created by the actual and predicted change points agree and gives a percentage of agreement. A numeric value of 0 is associated with total

Table 1: Evaluation metrics with ranges and best ratings

Metric	Range	Best rating
Absolute Error (AE)	≥ 0	0
Meantime (MT)	≥ 0	0
Precision	0..1	1
Recall	0..1	1
F1-score	0..1	1
Rand Index (RI)	0..1	1
Hausdorff	≥ 0	0
Accuracy	0..1	1

disagreement, whereas 1 indicates two identical sets of segments. The greatest error of the prediction can be measured with the Hausdorff metric, which is the maximum temporal difference between a true change point and its estimation. In contrast to the Rand Index, a numeric value of 0 stands for two equal segmentations. In addition, we calculate the Accuracy of the classification in sections 3 and 4 to see which percentage of the phases was assigned correctly.

3. Classification Based on Time Series Data

The first investigated approach is using a supervised classifier directly on the time series data. By doing so, we use the fact, that each phase has a unique combination of variable values – at least the simulated data. A representative small dataset from a real plant can be annotated, meaning to assign the start and end times of phases and their names, by a process expert (e.g. operator, engineer). This dataset, generated at little effort, is used to train a classifier which afterwards is applied to the rest of the dataset for that specific unit of a batch plant. This minimises the amount of manual labour and still allows exploiting expert knowledge. For this experiment, the pre-processing is to firstly drop duplicates, to separate the phase labels from the sensor values and to interpolate all values in an interval of 10 seconds to have enough values even in the shortest phase which is 5 minutes in the simulated data. For the values of the temperature and level sensor we apply a linear interpolation, for the flow sensor we take the previous value and for the valve we fill with zero as constant value. For the phase name we apply a forward fill method to have the current phase name for every timestamp we have sensor values for. As classifier we use the random forest classifier provided by the Python package scikit-learn [12] as decision tree-based models are known to handle imbalanced datasets quite well. The sensor values are the input for our model, the phase name is the output. We used 15% of the simulated dataset for the training of the model which is a total of 55 batches. For the real plant we first used 10% of the data for training which contain 117 batches. The aim was to reduce the manual labelling effort to create the training dataset in the application later on. Since we reached an accuracy of only 34.07% we decided to increase the amount of training data. The best accuracy of 69.92% is reached with a training set of 351 batches (30%). The detailed results of these experiments are shown in Table 2.

Table 2: Results of the application of random forest classification on time series data

Dataset	Accuracy	AE (avg.)	MT (s)	Precision	Recall	F1-score	Hausdorff (s)
Simulated	0.9994	0.4966	234	0.9607	0.9958	0.9779	900
Real plant	0.6996	78.51	57,157	0.1260	0.5084	0.2020	1,250,850

4. Classification Based on Statistical Features

The second investigated approach follows the approach presented in [8]. It is based on the calculation of 22 statistical features, selecting the most relevant and applying a classifier to them. This shall improve the algorithm performance, eventually shorten the computation time and increase the explainability [8].

For the feature creation we utilize the same time series as for the training of the classifier in the previous section. For each variable 22 statistical features are calculated with a window size of 30 values to match the shortest phase, being 5 minutes. The features are mean, median, mode, variance, standard deviation, root mean square, interquartile range, range, skewness, kurtosis, second moment, percentiles (p10, p25, p50, p75, p90), min, max, sum, first, last and entropy.

By doing so we generate a total of 88 features from the 4 variables of the simulated data. The correlation matrix is used to identify pairs of highly correlated features having a correlation coefficient greater than 0.95 and remove one of them. This procedure reduces the set of 88 features to 53, which are then analysed by a principal component analysis, keeping at least 95% of the whole variance in the principal components. The back-transformation of the principal components yields a total of 14 features:

- 5 of *flow sensor*: mean, variance, kurtosis, last, skewness
- 4 of *temperature*: mean, kurtosis, entropy, standard deviation
- 3 of *N₂ feed valve*: mean, interquartile range, last
- 2 of *level sensor*: entropy, kurtosis

Applying the same strategy to the real data leads to 176 features from the 8 variables of which 60 are not highly correlated (correlation coefficient <0.95) and 33 remain after the principle component analysis. The remaining features are:

- 4 of *flow substance A*: mean, interquartile range, entropy, kurtosis
- 5 of *flow substance B*: mean, interquartile range, entropy, skewness, last
- 4 of *flow substance C*: mean, entropy, skewness, variance
- 5 of *flow substance E*: mean, interquartile range, entropy, skewness, last
- 4 of *flow product*: mean, interquartile range, entropy, skewness
- 4 of *temperature jacket*: entropy, skewness, standard deviation, variance
- 3 of *temperature reactor*: mean, skewness, standard deviation
- 4 of *level reactor*: mean, entropy, skewness, variance

Applying the random forest classifier to the features we achieve good results for the simulated data. For the real data the quality of the prediction is significantly reduced by artefacts in the signals. Table 3 provides the metrics of the experiment.

Table 3: Results of the application of random forest to selected features

Dataset	Accuracy	AE (avg.)	MT (s)	Precision	Recall	F1-score	Hausdorff (s)
Simulated	0.9994	0.3716	209	0.9674	0.9955	0.9813	900
Real plant	0.5964	106.5	288,499	0.0915	0.4681	0.1531	1,328,890

5. Unsupervised Change Point Detection

Start and end time of batch phases are often reflected by change points in the time trends of process variables, making change point detection (CPD) methods suitable for automatically detecting phase transitions. The simplest case is a change in a discrete variable, but also continuous variables convey information. For example, during heating of a reactor the temperature increases, whereas at the end of the heating phase the temperature remains unchanged or decreases. Hence the change points in the time trend of the temperature measurement will indicate the start and end of the heating phase. Nevertheless, unsupervised CPD needs to be coupled with a supervised method later on to achieve interpretable results.

Unsupervised CPD can be performed through formulating an optimisation problem.

This is implemented in the python package *RUPTURES* [8], which is used to make offline predictions of change points. If the number of change points is known, the problem is a best-fit problem. Here it is assumed that the number of change points is not known, and an additional term is added to control the number of change point predictions. The optimisation problem can be expressed as

$$\min_{\mathcal{T}} \sum_{j=1}^K c(y_{\tau_j:\tau_{j+1}}) + \beta |\mathcal{T}|$$

where $\mathcal{T} = \{\tau_0, \tau_1, \dots, \tau_{K+1}\}$ represents the set of K predicted change points τ_j , where $\tau_0 = t_0$ and $\tau_j = t_N$ are implicit change points at the first and last samples. The signal value at time t is represented y_t , and a segment from a to b is represented by $y_{a:b}$. The cost function for a segment is denoted as c and $\beta |\mathcal{T}|$ is a linear penalty function with a constant penalty of β . To solve the optimisation problem, there are two search direction methods: pruned exact linear time (PELT) and window-based approximation (WIN). Both approaches are used to make predictions and are presented in the tables below.

In this paper, two parametric cost functions are implemented. The first function assumes an underlying distribution, where the L2-norm is used to give the cost of a segment. The function is defined as

$$c_{L2}(y_{a:b}) := \sum_{t=a+1}^b \|y_t - \overline{y_{a:b}}\|_2^2$$

where $\overline{y_{a:b}}$ represents the segment mean from time index a to b . This enables the algorithm to identify sudden changes in the mean where the distribution of the data has changed. An alternative approach is to fit a model to the segments and calculate the cost as the residuals of the model, an example being linear regression. If previous samples are used as covariates, we get an autoregressive model, given by

$$c_{AR}(y_{a:b}) := \min_{\beta} \sum_{t=a+1}^b \|y_t - (\beta_0 + \beta x_t)\|^2, \quad x_t = [y_{t-1}, \dots, y_{t-k}]$$

where the autoregressive vector x_t comprises k of measurements of y_t in the past. Here we use five previous samples and a constant to construct the model.

Using the two search direction methods and the two cost functions, we can make predictions on selected batches of the simulated dataset and compare the performance. The results are seen in Table 4 where the most accurate predictions are given by WIN in combination with the autoregressive cost function (AR) of which the results are plotted in Fig. 3. When using PELT, the L2-norm gives better predictions than AR.

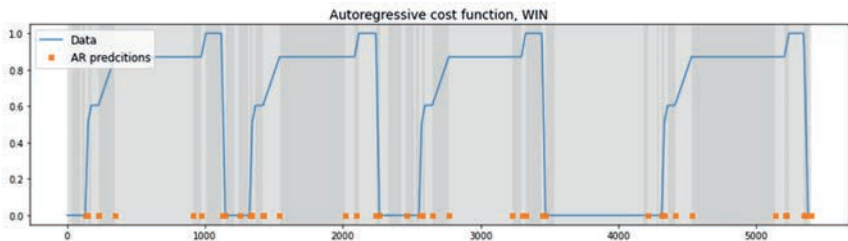


Fig. 3: Predictions using AR-model and WIN search direction for the simulated data. Alternating grey areas indicate segments between the true change points

Table 4: Results for predictions made on the simulated data

Search direction	Cost function	Penalty	AE (avg.)	MT (s)	Precision	Recall	F1-score	RI
PELT	L2	1	1	489	0.509	0.474	0.491	0.979
	AR	0.0001	3	105	0.289	0.228	0.255	0.926
WIN	L2	0.001	3.75	1899	0.548	0.404	0.465	0.956
	AR	0	4.25	105	0.750	0.526	0.619	0.992

Prediction on a few batches of the real data includes more process variables, and noise. As described in Section 3, the sampling interval for this data set is 10 seconds. In Table 5, the metrics for respective prediction approach are presented. If the window-based search direction is used, the L2 cost function has a higher F1-score and Rand Index than for the AR model. The Meantime for the L2 model cost function is 645 seconds, which is higher than the used margin of 300 seconds. When using PELT, the L2-norm is preferable with respect to all investigated metrics except Precision, where the Meantime is reduced to 519 seconds, which is lower than the Meantime for L2 with WIN. Similarly, the F1-score and Rand Index are higher than when WIN is used. The plotted results of PELT combined with the L2-norm is shown in Fig. 4.

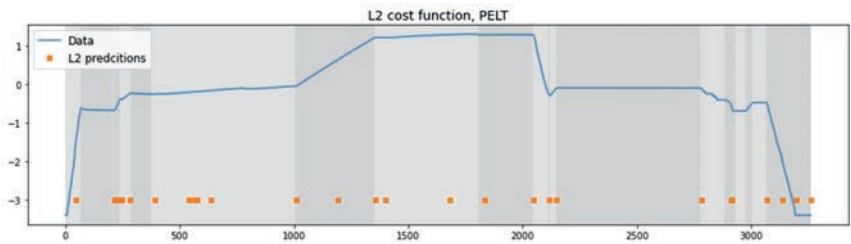


Fig. 4: Predictions using L2-model and WIN search direction for the original data. Alternating grey areas indicate segments between the true change points.

Table 5: Results for predictions made on the real data

Search direction	Cost function	Penalty	AE (avg.)	MT (s)	Precision	Recall	F1-score	RI
PELT	L2	30	0	519	0.577	0.577	0.577	0.955
	AR	20	2	695	0.333	0.308	0.320	0.894
WIN	L2	0	8	645	0.611	0.423	0.500	0.931
	AR	0	10	702	0.375	0.231	0.286	0.906

6. Comparison and Evaluation

As expected, for the simulated data the two classification algorithms perform nearly perfectly, leading to the conclusion, that such methods in the best case are providing good results for the automatic labelling of batch phases. However, for the real data the classification on time series yields an accuracy of 69.96%, the classification on statistical features 56.64%. Also, the other metrics are better when applying the classifier directly to the measurement data. The phase transitions cannot be identified clearly with both classification approaches which is shown by the rest of the metrics. Looking at the better performing classification on measurement data, especially the Meantime metric is quite high with nearly 16 hours. Due to the Hausdorff metric with more than 14 days, we assume one possible influencing factor could be the extremely long run lasting 36 days, which was not marked as maintenance in the data. We decided to leave it in because removing such a run from the data can be dangerous since needed information could be cleaned. With regard to the Annotation Error of about 78 per batch, the borders seem to not be clear in the data being smooth and therewith unclear transitions or having measurement errors. Applying a filter to the data could help providing better results. In addition, the dataset is imbalanced as some phases last a lot longer than other phases, which makes it harder for the algorithm to assign the correct phase because for the shorter phases there is not much data to train the model on.

Due to the higher complexity of the real dataset compared to the simulated dataset with 8 vs. 4 process values and 14 vs. 43 phases, the algorithm needs a lot more batches to train the classifier on than for the simulated dataset and still doesn't reach the same quality of the results. To address this issue, including process knowledge could be a valuable step towards automatic batch phase detection.

In addition, our research shows that the methods for unsupervised CPD implemented in *RUP-TURES* is applicable to data from batch processes, too. The choice of cost function influences

the performance of CPD. Moreover, the performance is also impacted by the heterogeneous characteristic of the real-life dataset, which may require additional handling. For the search method, WIN is more efficient than PELT in sense of computational time. For the settings compared in the paper, AR cost function achieved a better performance on the simulated data while L2-WIN is generally better for real-life data in sense of accuracy and computational efficiency. A future direction is to search for the cost function that can fit typical behaviour of batch data well.

7. Conclusion

Our experiments show that the tested approaches in general are suitable for the detection of phases in process data per unit of a process plant. On the simulated data they perform well, providing an accuracy of more than 99% using random forest classification. Applying the same methods to the real plant data leads to a maximum accuracy of around 69.96%. As expected, the higher complexity, noise and measurement errors complicate the classification.

Future work is aimed at teaching the model the sequence of the phases and to handle erroneous batch runs to increase the quality of the prediction. Also, different filters and feature generation methods will be tested. The improved models will be included in an interactive environment, where process experts can label batch phases with very much less effort compared to current labelling strategies.

Acknowledgements

The German Federal Ministry for Economic Affairs and Energy (BMWi) is acknowledged for funding this research as part of the KEEN platform (reference code 01MK20014T).

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

Literature

- [1] A. Bamberg, L. Urbas, S. Bröcker, N. Kockmann, and M. Bortz, "Was den Digitalen Zwilling zum genialen Kompagnon macht," *Chemie-Ingenieur-Technik*, vol. 92, 2020, doi: 10.1002/cite.201900168.
- [2] F. D. Böhner and J. K. Huusom, "A Debottlenecking Study of an Industrial Pharmaceutical Batch Plant," *Ind. Eng. Chem. Res.*, vol. 58, no. 43, pp. 20003–20013, 2019, doi: 10.1021/acs.iecr.9b03134.
- [3] Z. Xing, J. Pei, and E. Keogh, "A brief survey on sequence classification," *ACM SIGKDD Explor. Newsl.*, vol. 12, no. 1, pp. 40–48, 2010, doi: 10.1145/1882471.1882478.
- [4] J. Huss and L. Urbas, "Expert Oriented Design of an Engineering Tool for Data Base Investigation," 2006.
- [5] G. J. J. van den Burg and C. K. I. Williams, "An Evaluation of Change Point Detection Algorithms," *arXiv:2003.06222 [stat.ML]*. 2020, [Online]. Available: <http://arxiv.org/abs/2003.06222>.
- [6] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowl. Inf. Syst.*, vol. 51, no. 2, pp. 339–367, 2017, doi: 10.1007/s10115-016-0987-z.
- [7] A. B. S. Serapião, R. M. Tavares, J. R. P. Mendes, and I. R. Guilherme, "Classification of petroleum well drilling operations using Support Vector Machine (SVM)," 2006, doi: 10.1109/CIMCA.2006.66.
- [8] B. Esmael, R. K. Fruhwirth, and G. Thonhauser, "A Statistical Features Based Approach for Operations Recognition in Drilling Time Series," vol. 5, p. 4545, 2013.
- [9] C. Von Trotha, T. Kleinert, and U. Eppe, "Ein Konzept zur Kontextualisierung von Prozessdaten," *Atp Mag.*, vol. 62, no. 10, pp. 68–76, 2020, doi: 10.17560/atp.v62i10.2502.
- [10] U. Piechottka and V. Hagenmeyer, "A discussion of the actual status of process control in theory and practice: A personal view from German process industry," *At-Automatisierungstechnik*, vol. 62, no. 2, pp. 67–77, 2014, doi: 10.1515/auto-2014-1018.
- [11] C. Truong, L. Oudre, and N. Vayatis, "Selective review of offline change point detection methods," *Signal Processing*, vol. 167, 2020, doi: 10.1016/j.sigpro.2019.107299.
- [12] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011, doi: 10.1289/EHP4713.

AIQIIA: General AI Quality Inspector for Industrial Applications

M.Sc. **D. Díez Álvarez**, Mercedes-Benz AG, Stuttgart;
Prof. Dr. –Ing. **U. Berger**, BTU Cottbus-Senftenberg, Cottbus

Abstract

In this paper, we present a general solution for computer vision quality inspection tasks based on artificial intelligence for industrial environments with low data variability. The proposed solution improves the efficiency by eliminating any kind of programming, supervision, data gathering or training, which should be done just once, and enables the direct implementation of the inspector to any new process where a reference item is available. The currently high time and effort consuming task of developing a new inspection process is simplified to the selection of that reference. The method is based on meta-learning and on Siamese Neural Networks (SNN), where no information from concrete problems but the general behavior of comparison is learned, making the method suitable for any problem of this kind. The final accuracy is comparable to state-of-the-art algorithms even though the target is to optimize the implementation process not the performance of a specific network.

1. Introduction

Unpredictable constant changes in demand together with a growth of the desired degree of personalization of the products make smart factories a necessity if companies want to be efficient [1]. Smart factories are a growing tendency and have been designated as a differential factor for companies to remain competitive in future market scenarios [2]. According to the definition, smart factories are “flexible systems that can self-optimize performance across a border network, self-adapt to and learn from new conditions in real or near-real time, and autonomously run entire production processes.” [3]. Intelligent and autonomous systems (control, prediction or perception) are the enablers. Flexibility and adaptability are crucial factors and all the required technologies like AI, connectivity or automated decision making, are directly crucial as well. The technological state of the art (SoA) (plus regulations and standards) does not allow the implementation of smart factories yet and the steps the manufacturing industry is taking are not always correct.

The paradigm changes when considering smart factories in a holistic way and not smart specific processes only. The focus is therefore on building the machine that builds the machine, where the factory becomes the first product in the overall process [4], which (translated to automation) means the automation of the process of automating processes. We understand that the difference between smart factories and highly automated factories is the capability to self-adapt to unseen scenarios.

Artificial Intelligence (AI) has shown great potential and performance in multiple different industrial fields and the conversion of a given problem into an autonomous process has been proven. However, AI itself is not autonomous, meaning that it is able to perform the processes it has been programmed for automatically but it cannot adapt to unseen new scenarios due to the inherent limitations of the algorithms used. Autonomous factories require autonomous algorithms. Increasing and improving the flexibility of AI is a current research topic [5-8]. We propose and demonstrate the capability of AI to solve unseen scenarios conditioned to the way the AI algorithms are designed and trained.

In the example of computer vision (CV), AI can perform the task of quality inspection (QI) for a specific process and it has been proven that it can self-improve for that task [9] (with associated risks). However, it cannot self-adapt to a new process or product without an external interaction (human) that brings not only the decision of what to check but also the knowledge to adapt the algorithm and the data to train it. We present a QI system able to self-adapt to new products without any training or reprogramming by using Siamese Neural Networks (SNN) and a training strategy based on data variability instead of specific characteristics of the problems.

This document is structured as follows: section 2 is focused on the SoA of both QI and AI methodologies. Section 3 presents the main contribution of the paper by explaining the considered assumptions and describing the proposed solution. Section 4 explains the methodology and implementation details including network architecture, datasets, preprocessing and training. The final results and evaluation of the model is summarized in section 5 and the conclusions are described in section 6 together with the future work.

2. Related Work

Classical CV methods rely on the analysis of regions of pixels looking for specific gradients in both direction and magnitude. The extraction descriptors are defined through a mathematical mapping and are useful to detect shapes for image matching or quality inspection, however, these techniques are not invariant to object orientation or other kind of perturbations [10]. High-level feature extraction techniques such as deformable templates [11], active contours and snake [12,13], shape skeletonisation [14] or active shape models [15] address this problem

and are able to adapt their mathematical mapping to cover a wider range of problems. Even though, the flexibility is restricted to a certain degree.

Classical methods are limited to the mathematical target, thus to our knowledge and our ability to recognize and map the features. The definition of the target is not easy and the parametrization of possible perturbations is even harder. AI has shown proficient performance and potential in CV among different fields [16]. It has the ability to learn hidden features that we, as humans, are not able to learn, which implies the benefit of solving problems that cannot be solved currently at the cost of losing explainability of the solutions.

AI has gained interest by dealing directly with big amounts of bulk data though Deep Neural Networks (DNN). It is a current trend to reduce the size of the network to improve efficiency and flexibility [17] and, especially outside of the manufacturing sector, synergies between classical and intelligent solutions are under research [18].

While applying an already existing AI algorithm is as easy as opening a webpage [19], creating a proper AI solution involves many different specialists and is a time consuming and expensive task [20]. To overcome these problems, a change of focus from knowledge to behavior is investigated in meta-learning techniques. This approach proposes to learn general behaviors instead of specific characteristics of a specific problem. The first (and simplest) example is transfer learning, a SoA technique that allows the fast training of new algorithms from the already known information of similar problems. However, transfer learning still requires AI-experts to adapt the networks and algorithms to the new problem as well as data and time to retrain a new model.

SNN are a type of neural network architecture with two input channels instead of one. The networks learns a general behavior and the second channel can be used to provide a reference with the specific knowledge of the problem [21]. Convolutional Neural Networks (CNN) are invariant to transformations at high-level features, what has turned them into SoA NN for image processing and CV applications [22].

As shown, the main weakness of classical methods is the limitation to work with already known and predefine features, what eliminates the possibility of discovering new invariances and thus outperform current SoA. However, AI excels in this area in numerous fields such as physics and chemistry [23] while being used to create knowledge and not only to automate what is already known.

When properly designed and trained, the effort of creating a new classifier (under certain assumptions such as controlled environments and low data variability) is reduced to the

selection of a reference instead of recording new data, coding and training. The use of deep models with detailed information kept inside while limiting the generalizability can be narrowed down to really specific and unique problems. Industrial environments are standardized and both processes and products are known in detail as they have been predesigned. Thus most of the problems does not require complex AI solutions.

3. Contribution

We have shown that smart factories are necessary but the technology is not ready yet and regarding AI, the current approaches followed by industry are problematic. We have identified the flexibility and scalability of AI algorithms as a key factor limiting the adaptation of highly automated factories into smart and fully autonomous. We propose a new approach on how to apply AI that reduces the above mentioned limitations by facilitating the implementation of new AI solutions. In this paper we are going to validate the hypothesis of developing AI algorithms able to adapt to new and unseen scenarios opening the door to a change in how AI is applied by moving from specific use cases to general families of problems in the example of QI through CV. Our goal is to use AI to extract and analyze unidentified inherent invariances in problems with algorithms based on data distributions instead of on specific characteristics.

In a first stage, our method allows non-experts to create new solutions for specific problems with SoA accuracy through problem-agnostic AI algorithms able to solve the QI problem at the most general level. We work under the assumption of controlled environments and low data variability and we demonstrate that AI algorithms can learn behaviors and directly adapt from one problem to another inside the same family (QI through CV represents one family of problems that can be solved with one algorithm). As shown in Section 2, AI creates knowledge in other fields. In our case, AI finds the invariances that solve the problem at a general level as conservation laws solve general problems in physics. The main idea is to reduce the amount of information kept inside the AI making it agnostic to a specific problem and thus more general. Our novel approach proposes the use of SNN for low variability environments (industrial environments) and general QI problem solving.

4. Methodology

We consider that at a general abstraction level, QI is based on the comparison between the item to check and a reference to analyze the quality or correctness. As classical methods analyze the data through predefined maps able to generalize to different problems, we use AI in the same way but extracting more complex features. In the case of QI, we focus the problem

on data variability and how it changes between references of the same class and with regard to different classes.

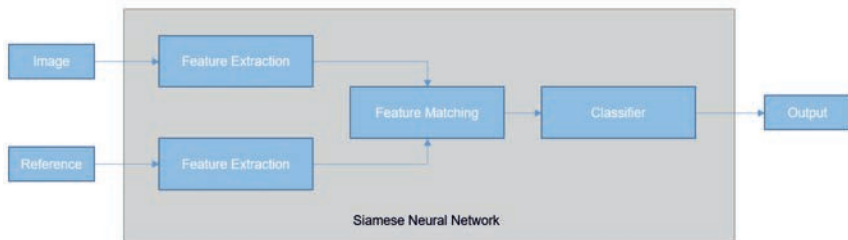


Fig. 1: Siamese Neural Network: modular architecture. Image and Reference are processed through the same feature extraction modules. Architecture (output channels, kernel size, stride):

(128, 7, 3) – (1024, 5, 3) – (1024, 1, 1) – MaxPool(2) – (1024, 3, 2) – (512, 3, 1) – (512, 1, 1).

The feature matching module divides the feature vectors (size 1x512) into 4 vectors (size 1x128) and performs 4 vector dot products: $\text{vector\_image}(i) \cdot \text{vector\_reference}(i)$ with (i) running from 1 to 4. The Classifier module is a fully connected NN with 4 layers with 4, 16, 8 and 1 neurons respectively.

Pairs of images are processed through a SNN where the image to analyze is supplied on channel 1 and channel 2 acts as a reference. Internally, SNN are composed of 3 modules (see Fig. 1): feature extraction, feature matching and classifier. The feature extraction module is the same for both input channels, thus the same kind of information is extracted from the image to analyze and from the reference. It is made up CNN, which are position-agnostic and can extract the information they are trained for with independence of where this information is located inside the data (position inside the image). The internal architecture of the CNN is a modification of NiN blocks [24] with a 1x1 kernel size CNN after two CNN layers (see Fig. 2). The feature matching module calculates the vector dot product of the extracted features to measure the divergence between them. This information is converted into the desired output at the classifier through a fully connected NN.

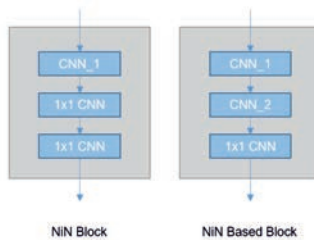


Fig. 2: Structure of NiN (network-in-networks) building blocks [24] at the left and our NiN-based building blocks at the right. 2 convolutional layers are followed by a convolution with kernel size = 1×1 .

We evaluate the capacity of SNN to learn the comparison behavior through different experiments. (1) Detect when the image is exactly the same as the reference. (2) An image is compared to a transformation of itself to understand the limits of AI and the limitations of CNN to find transformed features and the minimum detectable transformation. (3) A transformation of an image is compared to a different transformation of the same image and (4) real world test with real images from the same class.

We used 4 different datasets: 2 SoA benchmark open source datasets (CIFAR100 [25] and STL10 [26]) and 2 own datasets corresponding to different QI real use cases. The similarity between images of the same class in an industrial environment is much bigger than the similarity between images of the same class in open source datasets (see Fig. 3).



Fig. 3: Data variability in real world vs industrial dataset. At the left, 2 images of the same class from the CIFAR100 [25] dataset (also low quality) and at the right, 2 images of the same class from an industrial dataset. The difference between them and the data variability is lower in the industrial dataset.

To introduce data variability in the open source datasets with low divergence in a comparable range as in the industrial datasets, we used the same image as input and as reference and applied different kind of transformations to each of them. The transformation module performs random affine transformations (rotation, shift, perspective) and brightness changes to the input image keeping the reference without variation. Images are scaled and cropped to avoid empty areas or new edges while performing the transformations (see Fig. 4).



Fig. 4: Examples of the transformations applied to open source datasets: 3 images with 3 different transformations and a red square representing the cropped area. From left to right: original image, blurred and shifted image and rotated image.

Our industrial datasets correspond to a welding process where a nut is manually placed in the final position and in the next step, an automated machine welds it to the structure. The same kind of nut is located in 4 different positions and the perspective regarding the relative position between them and the camera makes it not possible to address the problem as only one. 4 samples are extracted at every iteration.

Our second dataset corresponds to a similar application of welded nuts but the nut itself and the position are different. We have now 3 nuts per picture that are extracted into different classes. The possible situations we wanted to analyze are: correct nut, incorrect nut and no nut. The real datasets are shown in Fig. 5.

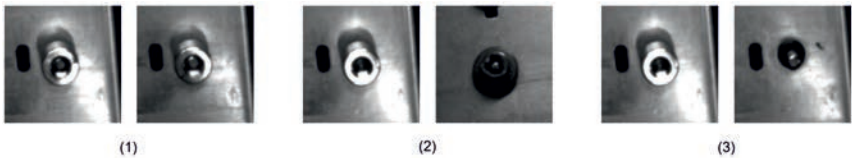


Fig. 5: Analyzed situations, from left to right: (1) correct nut – correct nut, (2) correct nut – incorrect nut (from real dataset 2) and (3) correct nut – no nut.

A summary of the experiments, datasets and transformations is shown in Table1.

Table 1: Datasets

*This table sums up the different datasets and transformations used in the 4 experiments. The indexes of the images inside the dataset are represented as (idx), thus image(i)-image(i) means that both images are the same, image(i)-image(ii) represents different images from the same class while image(i)-image(x) represents images from different classes. The transformations are: * = random rotation between -10° and +10°; ** = random rotation between -30° and -20° and between +20° and +30°; *** = transformation ** plus random shift in x and y direction between -10% and 10% of the image size.*

Experiment		Correct sample		Incorrect sample	
		Input	Reference	Input	Reference
(1)	CIFAR100 / STL10	Image(i)	Image(i)	Image(i)	Image(x)
(2)	CIFAR100 / STL10	Image(i)	Image(i)*	Image(i)	Image(i)**
(3)	CIFAR100 / STL10	Image(i)*	Image(i)**	Image(i)*	Image(i)***
(4)	Real	Image(i)	Image(i)	Image(i)	Image(x)

In terms of training, we used Leaky-ReLU as activation function for all layers of the fully connected network, except the last layer where we used the sigmoid activation function. As loss function, we found out that there is no significant difference between Softmax and Mean Square Error (MSE) in terms of accuracy but MSE converged faster during training. To check the models before final training we conducted a small test with 100 images and several epochs to check if the model was able to overfitting. Once demonstrated, we have trained the model with the whole dataset and a number of epochs running between 50 and 200 depending on the length of the datasets.

5. Experiments

The same image fed into channel 1 and 2 produces the same feature vector output. Slight changes in the input produce differences in the feature vectors and, after matching them, the classification module learns the threshold of acceptance. When analyzing images with a certain degree of divergence, the extracted features do not look exactly the same or the same information appears in a different position inside the extracted features vector. The comparison between this vectors is more complex producing therefore a decrease in accuracy of our network from 100% in same image pairs (NN are not even necessary as a simple subtraction would produce the same output) to a range between 70 and 80% depending on the applied transformations.

According to what we have explained before about CNN, we decided to use CNN to match the outputs of the feature extraction modules so they can find similarities between the outputted vectors regardless of the position inside the vector, where the relevant information is located. There was a problem in the definition of the input to this second CNN. We tried concatenation in both tensor dimensions and different architectures for the CNN but the improvement in the accuracy was not relevant (up to ~85% final accuracy).

Different functions were tested to measure the difference between both feature vectors and finally the vector dot product showed the best results. In the same way that the vector dot product gives the cosines of the relative angle of 2 vectors in the Euclidean space, it also gives a measurement of how different or how distant 2 tensors are in a higher vector space. The output of the vector dot product is just one value and only a threshold can be established. However, different kind of variations in different parts of the image introduce noise to the measurements and reduce the accuracy. We decided to divide each feature vector into 4 vectors and calculate 4 different vector dot products. This made it possible to create a more complex analysis layer with fully connected neural networks what seems to be crucial to improve accuracy by reducing the effects of noise to local areas.

With the new architecture the model is able to correctly classify images with regard to a reference and also between different degrees of transformations. This has been tested in the open source datasets. The training was carried out with CIFAR and some transformations for both the input image and the reference and was validated on STL10 dataset. The solution achieved SoA accuracy (>99%) in both the training and the validation dataset.

In our second experiment we tested this behavior on real datasets. Our dataset already contains some intrinsic variability between same-class references mainly due to changes in

position and light conditions. Therefore there are no transformations required now. We first needed to check if the difference in the images due to production failures are significantly different compared to the intrinsic variability of the dataset and if a network is able to overfitting the real dataset. We used the same networks as before. Transfer learning was possible and speeded up the training but does not eliminated it, as far as the specific variability of the industrial dataset was not properly introduced in the training dataset through the transformations. The algorithm converged to an accuracy of 87.5%, significantly worse as before. The NN was not deep enough so we modified the feature extraction module to include more layers going from 1 NiN block to 2. The network then reached an accuracy of 98.25%. Results are slightly worse compared to classical AI with 99.9% accuracy. However, the proposed solution is smaller in terms of number of parameters of the NN than a classical network solving the same problem (from 5.3 million taking EfficientNetB0 [27] as SoA to 80k). However, when applying the inspector to a new similar problem (similar degree of divergence between correct and incorrect sample) the classical method classifies all the samples as incorrect as they do not match the specific characteristics of the previous problem that has learned. A new training is required to achieve comparable accuracies with at least 100 new images. Our solution achieves 97.91% accuracy just by changing the reference. These results are summarized in Table 2.

Table 2: Results

Summary of experimental results: train and test datasets, test accuracy and architecture of the network. The architecture with 2 NiN blocks is presented in Fig. 1 and the architecture of the feature extraction module with 1 NiN block is as follows: (128, 7, 7) – (512, 8, 8) – MaxPool(4) - (512, 1, 1). Real 2 datasets represents a different type of nut (Fig. 5).

Train	Test	Test acc.	# NiN blocks
CIFAR100	CIFAR100	100%	1
CIFAR100*	CIFAR100 *	99,62%	1
CIFAR100*	STL10*	99,27%	1
CIFAR100*	Real	85,02%	1
Real	Real	98,25%	2
Real	Real 2	97,91%	2

To proof the flexibility of our solution, we directly used our model to classify metal pieces and to check possible manufacturing errors (Fig 6). This use case is less demanding as the differences between different pieces and between correct and incorrect pieces of the same

class are bigger than in the previous use cases. The network is able to classify all of them correctly with a 100% accuracy in a small test dataset (20 samples in total). The use of images from the CAD models of the corresponding pieces was also investigated and the results were the same as before: 100% for the small test dataset.



Fig. 6: CAD models as reference. (1) Real image and corresponding CAD model, (2) 2 real images, one correct and one incorrect (a hole is missing) and the corresponding CAD model. The network classifies these 3 situations correctly.

6. Discussion & Future work

We present a novel approach in the industrial field that focuses on the flexibility of the AI algorithms to enable the conversion of highly automated factories into fully autonomous or smart ones. We have demonstrated the flexibility of AI by training and testing the algorithms in completely different datasets. Our work changes the effort of developing new QI from the qualified work based on knowledge, understanding of AI and programming to the selection of a reference. The application and benefits of AI can grow exponentially improving efficiency. The proposed method is limited to highly standard environments and to a degree of accepted divergence between the references and the samples. When working out-of-domain the model fails, so different models should be trained according to the desired degree of divergence to detect. Though close to SoA, accuracies are slightly worse compared to tailored solutions thus there is a tradeoff between general algorithms and accuracy. The use of CAD images as references has only been tested on a less demanding use case and the behavior of the network in complex problems with this kind of references should be further tested.

We are aware of the limited applicability of our work to real life problems outside the industrial environment but we believe that the current tendencies of AI do not transfer properly from the outside world to industrial environments due to their specific conditions.

This work opens a path to new general approaches that take advantage of the standard conditions and the full knowledge of the environment, samples and processes to analyze as demonstrated with the use of CAD models.

AI has more potential than what we are currently exploiting in the manufacturing industry. It can be a game changer by creating new knowledge and automating tasks that we now think are not possible to automate. The current approaches will drive us to highly automated factories but not smart or autonomous ones in the terms we defined them.

Our effort was focused on QI through CV, but other kind of problems dealing with different type of signals (like sound or force) could be analyzed following this methodology. Other AI paradigms, as Reinforcement Learning, also have potential to develop general solutions.

7. Bibliography

- [1] Petit, J.P. (2019, November). *Smart factories @scale*. <https://www.capgemini.com/wp-content/uploads/2019/11/Report-%E2%80%93Smart-Factories.pdf>
- [2] Sjödin, D. R., Parida, V., Leksell, M., & Petrovic, A. (2018). *Smart Factory Implementation and Process Innovation: A Preliminary Maturity Model for Leveraging Digitalization in Manufacturing Moving to smart factories presents specific challenges that can be addressed through a structured approach focused on people, processes, and technologies*. *Research-Technology Management*, 61(5), 22-31.
- [3] Burke, R., Laaper, S., Mussomeli, A. (2017, August 31). *The smart factory*. <https://www2.deloitte.com/us/en/insights/focus/industry-4-0/smart-factory-connected-manufacturing.html>
- [4] Fehrenbacher, K. (June 6, 2016). *Tesla's Quest to Build the Machine behind the Machine*. <http://fortune.com/2016/06/06/teslas-the-machine-behind-the-machine/>
- [5] Munkhdalai, T., Yu, H. (2017). *Meta networks*. In: *Proceedings of machine learning research* 70, p. 2554.
- [6] Santoro, A. et al. (2016) *Meta-learning with memory-augmented neural net-works*. In: *International conference on machine learning*. pp. 1842–1850.

- [7] Sung, F. et al. (2018). *Learning to compare: Relation network for few-shot learning*. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1199–1208.
- [8] Vinyals, O. et al. (2016). *Matching networks for one shot learning*. In: *Advances in neural information processing systems*. pp. 3630–3638.
- [9] Ghamisi, P., Chen, Y., & Zhu, X. X. (2016). *A self-improving convolution neural network for the classification of hyperspectral data*. In: *IEEE Geoscience and Remote Sensing Letters*, 13(10), 1537-1541.
- [10] Nixon, M., Aguado, A. (2019). *Feature extraction and image processing for computer vision*. In: *Academic press*.
- [11] Yuille, A. L., Hallinan, P. W., & Cohen, D. S. (1992). *Feature extraction from faces using deformable templates*. In: *International journal of computer vision*, 8(2), 99-111.
- [12] Shah, S., Abaza, A., Ross, A., & Ammar, H. (2006, September). *Automatic tooth segmentation using active contour without edges*. In: *2006 biometrics symposium: special session on research at the biometric consortium conference* (pp. 1-6). IEEE.
- [13] Delmas, P., Coulon, P. Y., & Fristot, V. (1999, March). *Automatic snakes for robust lip boundaries extraction*. In: *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*.
- [14] Abu-Ain, W., Abdullah, S. N. H. S., Bataineh, B., Abu-Ain, T., & Omar, K. (2013). *Skeletonization algorithm for binary images*. In: *Procedia Technology*, 11, 704-709.
- [15] Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). *Active shape models-their training and application*. In: *Computer vision and image understanding*, 61(1), 38-59.
- [16] Khan, S., Rahmani, H., Shah, S. A. A., & Bennamoun, M. (2018). *A guide to convolutional neural networks for computer vision*. In: *Synthesis Lectures on Computer Vision*, 8(1), 1-207.
- [17] Chen, M. Y., Chiang, H. S., Lughofer, E., & Egrioglu, E. (2020). *Deep learning: emerging trends, applications and research challenges*.
- [18] Skoryk, A., Chyrka, Y., Gorovyi, I., Grechnev, O., & Vyplavin, P. (2020, August). *Comparative Analysis of Classic Computer Vision Methods and Deep Convolutional Neural Networks for Floor Segmentation*. In: *2020 IEEE Third International Conference on Data Stream Mining & Processing* (pp. 217-221).
- [19] MIT.edu (n.d.) *Personal image classifier*. From: <https://classifier.appinventor.mit.edu/>

- [20] Ng, A., Landing AI (2019) *AI Transformation Playbook*. From: https://landing.ai/wp-content/uploads/2020/05/LandingAI_Transformation_Playbook_11-19.pdf
- [21] Koch, G., Zemel, R., Salakhutdinov, R. (2015). *Siamese neural networks for one-shot image recognition*. In: *ICML deep learning workshop*. Vol. 2. Lille.
- [22] Shi, Q., Li, W., Zhang, F., Hu, W., Sun, X., & Gao, L. (2018). *Deep CNN with multi-scale rotation invariance features for ship classification*. In: *IEEE Access*, 6, 38656-38668.
- [23] Fleming, N. (2018). *How artificial intelligence is changing drug discovery*. In: *Nature*, 557(7706), S55-S55.
- [24] Lin, M., Chen, Q., & Yan, S. (2013). *Network in network*. In: *arXiv preprint arXiv:1312.4400*.
- [25] Xu, B., Wang, N., Chen, T., & Li, M. (2015). *Empirical evaluation of rectified activations in convolutional network*. In: *arXiv preprint arXiv:1505.00853*.
- [26] Coates, A., Lee H., Ng, A. (2011). *An Analysis of Single Layer Networks in Unsupervised Feature Learning*. In: *AISTATS*
- [27] Tan, M., & Le, Q. (2019, May). *Efficientnet: Rethinking model scaling for convolutional neural networks*. In: *International Conference on Machine Learning* (pp. 6105-6114). PMLR.

Scaling a Reinforcement Learning Approach for Online Job Shop Scheduling in Flexible Manufacturing Systems to Various Products

Schirin Baer, Amazon;

Sebastian Pol, **Danielle Turner**, **Felix Baer**, Siemens AG;

Tobias Meisen, Universität Wuppertal

Abstract

In this article, we present our results of using deep reinforcement learning to implement a reactive scheduling solution for flexible manufacturing systems. The number of product variants that can be produced in flexible manufacturing systems can be extremely large and results in a huge state space. Therefore, in this work, we focus on the generalization aspect of our solution to scale it to a variety of products. More specifically, we evaluate whether our trained agents are capable of handling any kind of job specification, which describes a product. For this purpose, we have developed and implemented a state design with an advanced job specification encoding that can deal with different shapes and sizes of job specifications for a fixed maximum number of machines within the plant. We use the multi-agent version of Deep Q-Networks to control the navigation of products through the plant, including transportation decisions and dispatching to machines. We trained a group of three agents on 600 randomly generated job specifications and proved their ability to generalize to jobs, which have not been seen during training. Furthermore, we demonstrate that the trained agents are able to deal with unseen job specification formats. With this design, we achieve a self-learning reactive scheduling system that is flexible and generalized to all kinds of product variants within a defined flexible manufacturing system.

1. Introduction

Flexible manufacturing systems (FMS) are modular and often intelligent plants that can produce a large variety of products using different machines with partly overlapping skill-sets. Each operation within a job specification can be carried out on multiple machines with emphasis on different properties, including processing time, energy consumption, and material costs. Other new technologies, such as Autonomous Guided Vehicles (AGVs) for transportation of products, are used which can also lead to both higher complexity and uncertainties in production control. AGVs can increase uncertainties as they plan their route based on dynamic obstacles and workers walking around. This adds to the chance that their arrival time will deviate from the plan. Future scheduling systems could therefore be hybrid, comprising of an offline planning to sort the order stack and release orders according to their

prioritization and an online scheduling responsible for machine allocation as soon as the order is released and the production starts. An online scheduling system that would fulfil the requirements of future production systems would be self-learning, has less engineering effort to set up and adapt, and would be scalable and generalizable, so that its usage in a physical setup is feasible.

The planning optimization problem is commonly formulated as a Job Shop Scheduling Problem (JSSP), where the optimal sequence of processing x operations of n job specifications on m machines needs to be found, while the order of operations is pre-defined [1]. The complexity of the JSSP is known as NP-hard and we even consider an extended version of JSSP, which makes it infeasible to learn every possible scenario. We consider various jobs that can be produced on different machines, called Flexible JSSP (FJSSP), and in addition constraints of the resources occurring from set-up-times and transportation, named as Multi-resource flexible JSSP (MrFJSP) by [15]. To handle this aspect, we train Deep Reinforcement Learning (RL) agents through which we can obtain a generalized solution that is able to reach near-optimal decision-making, even for unseen job specifications. We design a reactive scheduling system, which makes decisions online, allowing the system to react to unforeseen events without recalculating the complete schedule.

In this work, we focus on generalization to unseen job specifications, as we believe that this is a crucial component to prove the feasibility of our concept, which follows [2] and [3], and scales it to a large amount of product variantes. The use of the multi-agent variant of Deep Q-Networks (DQN) is demonstrated and its efficiency evaluated.

2. Related Work

RL approaches to solve the JSSP with online scheduling have been presented in [4], [5] and [6]. These approaches use decentralized Multi-Agent Reinforcement Learning (MARL) architectures, trained to find the correct dispatching rules for every machine. In contrast, we control products instead of machines to include transportation decisions in addition to dispatching. An approach recently proposed by [7] solves the complex JSSP in sheet-metal production with routing flexibility and highly heterogeneous jobs by combining RL with AlphaGo Zero, trained with schedules of a heuristic baseline to selectively search the solution space. [8] introduces a MARL approach, which includes the aspect of load management from the energy management domain, and [9] discusses RL agents that control machines and independently choose a product or stay idle. One common issue in MARL is the dynamics of learning, whereby actions made by agents interacting in an environment in parallel can influence the learning of others, leading to a non-stationary environment. [10] introduces an idea of how to overcome this problem in Multi-Agent Deep Reinforcement Learning. One agent

is trained by observing information about other agents, while their policies stay fixed. After a defined period of time the parameters are shared with the fixed agents and training with the new policies continues. We applied this method in [3], where the concept of generalization to multiple machine topologies was evaluated, whereas we focus in this work on generalization to job specifications with the goal to achieve a solution that is scalable to a large amount of product variants. As we consider only one optimization objective, we have changed our approach to use a single Q-Network for all agents as described in section 3, which makes the strategy of [10] and [3] obsolete.

3. Multi-Agent Reinforcement Learning Setup

To navigate multiple products through the FMS with the goal of achieving an efficient and proactive schedule, we use a decentralized MARL approach, where each product is controlled by one agent through a simulation including transportation decisions and dispatching to machines. Although this approach is decentralized, only one Q-network is used, which is trained as if in a single-agent environment. However, it is used n times within a time step for the n individual agent instances where each agent receives its own observation and selects its own action to reduce complexity and obtain flexibility.

This approach is feasible as we currently have homogeneous agents. Therefore, we obtain the same benefit as we would with parameter sharing, which has already been evaluated in [3]. All agents profit from the experiences of other agents and we overcome the issue of a transient environment by using only one network, which is updated after a defined period of time. Within this period the environment stays stationary. Although similar to joint action learner [11], we decided against this implementation, as the complexity of the action space grows exponentially with the number of agents, leading to a longer training time and inflexibility in terms of the number of agents. When training agents for different tasks or different optimization objectives, concurrent learning with distributed networks would be beneficial, as every agent could train within their local state-space, increasing learning duration [12]. However, we profit from using a single Q-network as long as we consider only one optimization objective for multiple agents.

The network architecture consists of an input layer with a size dependent on the agent's state-size (512), two hidden layers with 64 nodes and ReLU activations, and an output layer with 24 nodes corresponding to the action-size using a linear activation. As proposed in DQN [13], we use a replay memory buffer to collect experience tuples $[S_t, A_t, T_t, S_{t+1}]$ during training, a random sample of which is used in to update the Q-function after each epoch.

For training the agents, we use a simulation of our modular FMS with six different manufacturing modules and six positions where they can be located. The simulation is a Petri net that consists of 12 places and 24 transitions, including self-transitions for the opportunity

to stay in a manufacturing module for two successive tasks. With this concept, we define that the time steps and locations are both discrete, although we consider conveyor belts for the transportation. The agents are called to perform the action at each discrete decision making point using the state information, described in more detail in [2]. The action of one agent is to move the own token, which represent the controlled product, within the Petri net from A to B. This results into navigating the product through the FMS including machine dispatching. Each agent must therefore select one out of 24 actions at each time step. With this concept, we are flexible in the number of possible directions on each junction and profit from the efficiency and low engineering effort of using a Petri net as simulation.

We currently consider processing time as the only optimization objective, leading to a reward design that rewards choosing the faster module from the alternatives, depending on the respective processing times, and punishes invalid actions and steps on the conveyor belt. This work uses the same reward design as described in [3]. The characteristics of the Markov Decision Process are summarized in Table 1.

Table 1: Definition of state, action and local reward

State	Job specifications, position and progress of own and other agents; positions of available machines (machine topology)
Action	Direction from current position
Local reward	Time-dependent rewards favoring fast processing

A job specification describes the possible machines for each of the operations, and the relevant information for optimization. In this case, this information is a normalized integer value representing the processing time required for the given machine to complete the operation. Each operation within the job specification follows the format: $[M_{1n}, T_{1n}], [M_{2n}, T_{2n}], \dots, [M_{in}, T_{in}]$, where M_{in} represents the i_{th} machine able to perform the n_{th} operation, and T_{in} represents the time required to complete this operation on this machine. In the considered use case each job specification includes four consecutive operations, each having two possible machines on which the operation can be processed with different processing times.

4. Advanced state space to scale the concept to arbitrary products

To solve the MrFJSP using Deep RL, we aim for a trained agent that can handle an arbitrary job specification and schedule it optimally instead of being specialized to a single product. This requires the job specification to be part of the agent's state so that the agent is aware of which job specification it is currently scheduling and can distinguish between different job specifications. For this to work, the agent must be trained with several different job specifications and ultimately learn to interpret the job specification input by generalizing the

seen ones. This would allow the agent to correctly schedule job specifications not seen during training, as training for all job specifications is not feasible.

However, including the job specification in the agent's state significantly increases the state-size complexity and therefore complicates learning. Going even further, it would be worthwhile for the agent to not only know about the product it is handling, but also about other products which are currently being controlled by other agent instances. This is particularly important for the aspect of cooperation. The agent must be able to detect or even foresee possible conflicts between its job specification and the job specifications of other agents. Ultimately, it is necessary to avoid such conflicts and to work together to achieve a common goal within the production such as a minimal total makespan.

Integrating the job specification into the agent's state presents a challenge, as the job specification size can be variable but the state size must be fixed to be used as input for a neural network. For example, a product can require any number of operations depending on its complexity. Furthermore, each operation can be processed an arbitrary combination of the available modules with different properties, e.g. processing time. Such a job specification is demonstrated in Figure 1 on the right hand side, where the first value in the bracket indicates the machine and the second value the corresponding processing time. For example, the first operation of the job handled by Agent 1 could be either completed by module M3 in four time units or by module M1 in eight time units. The Gantt chart of the schedules of the three agents can be seen on the left hand side.

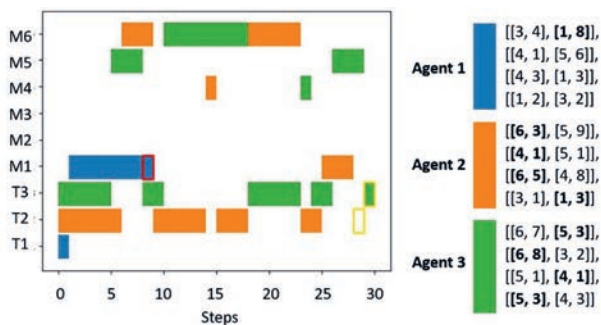


Figure 1: The Gantt chart on the left site shows one schedule of the job specification on the right site (with M1, ..., M6 indicating the modules and Tn the transport of agent n) created by the trained agents demonstrating the evaluation scheme. The yellow marker highlights whether a job specification is scheduled correctly. An empty marker shows that the agent failed to leave the module after finishing the last operation. A red marker shows an incorrect transition.

To encode the job specification with a constant size and consider the job specification of other agents, we propose an advanced job specification encoding. The idea of this encoding is illustrated in Figure 2, where $T_{Op,M}$ is the time it takes for module M to process operation Op .

The advanced job specification encoding allows the job specification to have any number of operations by introducing a lookahead. Therefore, the agent only sees n operations of the entire job specification. In Figure 2 the lookahead is set to two and is indicated by the gray rectangle.

Agent 1	Agent 2		Agent n
	$[T_{1,1}, T_{1,2}, T_{1,3}, T_{1,4}, T_{1,5}, T_{1,6}]$		$[T_{1,1}, T_{1,2}, T_{1,3}, T_{1,4}, T_{1,5}, T_{1,6}]$
$[T_{1,1}, T_{1,2}, T_{1,3}, T_{1,4}, T_{1,5}, T_{1,6}]$	$[T_{2,1}, T_{2,2}, T_{2,3}, T_{2,4}, T_{2,5}, T_{2,6}]$	...	$[T_{2,1}, T_{2,2}, T_{2,3}, T_{2,4}, T_{2,5}, T_{2,6}]$
$[T_{2,1}, T_{2,2}, T_{2,3}, T_{2,4}, T_{2,5}, T_{2,6}]$		...	$[T_{3,1}, T_{3,2}, T_{3,3}, T_{3,4}, T_{3,5}, T_{3,6}]$
$[T_{3,1}, T_{3,2}, T_{3,3}, T_{3,4}, T_{3,5}, T_{3,6}]$			

Figure 2: The advanced job specification encoding allows the inclusion of the job specification in the agent's state using a lookahead.

To combat the issue of a variable number of possible modules for each operation, the advanced job specification encoding considers every module in the manufacturing system. This does not introduce much overhead, as the system used contains a maximum of six modules. Therefore, the n^{th} operation of a job specification has six values $T_{n,1}, \dots, T_{n,6}$, stating the time each module requires to process the operation. To reduce complexity, the processing time of each module is normalized and discretized, in our case in the range of 1 to 9. Therefore, this range is used with the addition of zero for a module that cannot process the operation. The processing time is one-hot encoded, requiring 10 input nodes for each value of $T_{Op,M}$. If an agent has finished its job specification or is removed from the environment due to making a mistake, such as choosing an incorrect transition or module, the corresponding section is padded with zeros. The same applies if there are less operations remaining than the size of the lookahead. Another important aspect of this job specification encoding is that it considers the job specification of other agents with the same lookahead. Depending on the progress of other agents, an agent can see what their next n operations are, which may be useful for cooperative behaviour.

It should be noted that the advanced job specification encoding does not automatically provide a solution for a variable number of agents. In this work, a fixed number of three agents is used for all experiments, with each agent able to see the lookahead for its own job specification and that of the two other agents. With a variable number of agents in the system, it is possible to consider a subset of agents, for example three out of n , i.e. the agent itself and two others selected by some eligibility criterion. There are ways to incorporate this eligibility in the agents' tasks, as in coordinating Q-Learning proposed by [14].

Although the advanced job specification encoding is designed to keep the agent's state as compact as possible, it still requires substantial space. Considering three agents in the advanced job specification encoding with a lookahead of two, six possible modules and ten time units for each operation requires 360 additional input nodes for the neural network.

5. Experiments

Several experiments have been conducted to find out whether the state input is appropriately designed to enable the agent to generalize the job specification input and to correctly schedule job specifications that were not seen during training. In the following experiments, the agents have been trained for 3000 epochs using a total number of 600 randomly generated job specifications, comprising of four operations, each with two machine possibilities. During training, each agent was assigned a different job specification.

Regarding the use of different job specifications during training, multiple strategies have been tested. One possibility is to determine the optimal **job frequency**, i.e. the number of epochs in which the same combination of job specifications is used. For example, a job frequency of five would mean that the three agents are trained on the same three job specifications for five epochs before these are changed. Another possibility considers the order in which the 600 job specifications are sampled. One strategy would be to sample them in a **sequential order**, effectively separating the 600 job specifications into 200 combinations (1,2,3), (4,5,6), ... and training with them on loop. Another strategy is to sample the job specifications in a **random order** until all job specifications have been selected once and then to repeat the process. This would lead to the effect that the agent probably does not see the same combination of job specifications twice during training. Each time, a new random combination such as (67,212,31) or (517,84,403) is seen.

Regarding the lookahead of the advanced job specification encoding, a value of two is used for the following experiments. Keeping the lookahead small is crucial, as it also keeps the state size small. Furthermore, it is questionable whether a lookahead greater than two provides any notable benefits that outweigh the significant increase in learning complexity in our scenario. However this a hyperparameter that can be tuned to other production systems. In our experiments, a lookahead of two seemed sufficient for agents to foresee upcoming conflicts, such as blocking resources for each other, and effectively avoid them.

Experiments have been conducted to find the best hyperparameters. Through previous research [3] suitable hyperparameters were determined to be a gamma of 0.95, a learning rate of 0.0001, an epsilon of 0.99, an epsilon decay of 0.0097, a single network for all agents, a common replay memory and stochastic learning. In this work, experiments regarding replay memory size, batch size, network size and epsilon decay are considered, beside the training strategies *sequential order* and *random order* and a certain *job frequency* used to present the job specifications to the agents during training.

6. Results

Being able to evaluate and compare experiment results requires a form of quantitative assessment. For this purpose a basic rating scheme is introduced that evaluates the performance of the system based on the number of correct schedules. The validation set comprises ten scenarios in which three jobs must be scheduled. For each correctly scheduled job, one point is received. In total, 30 jobs can be scheduled correctly, leading to a maximum rating of 30 points. A job is scheduled correctly if the agent manages to complete each operation in the job specification. If an agent makes a mistake by entering an incorrect or empty module or by making an invalid transition no point is received. As agents sometimes finish the final operation of their job specification but then fail to leave the module, this behavior is evaluated with half a point. The requirement that a product must leave the module after completing the last operation is our definition of ending the production process in our training environment, as it is expected that the module is made available to other agents afterwards. An example of this rating scheme can be seen by the Gantt chart in Figure 1, which corresponds to a schedule created for one of the validation scenarios. The schedule in Figure 1 is rated with 1.5 points as Agent 1 made a mistake, Agent 2 completed its operations but failed to leave the module, and Agent 3 completed all operations correctly. Similarly, the schedule in Figure 3 is rated with 3.0 points, as it has completed all job specifications correctly.

Table 2 shows the comparison of job frequencies for the different training strategies. In earlier experiments it was found that training with the same job specifications for over 10 epochs lead to poor results. Therefore, this experiment only considers job frequencies 1, 3 and 5. Table 2 shows that there is no significant difference between the different job frequencies, meaning that the resulting validation schedules contain an equal number of correctly scheduled products. Comparing the training strategies "sequential order" and "random order" shows that training the agents with randomly selected job specification combinations results in a much better performance. While the sequential order separates the 600 training job specifications into 200 job specification combinations that are iterated through three times during training (with a job frequency of 5 and training for 3000 epochs), the random order makes a new combination of three job specifications each time. This makes better use of the 600 training job specifications leading to better generalization.

Table 2: Comparison of different training strategies

Training strategy	Job frequency		
	1	3	5
Sequential order	23	22.5	23
Random order	27.5	28	28

Additional experiments to fine-tune hyperparameters were conducted and achieved even better results. For this, variations in the epsilon decay rate, replay memory size, network size and batch size were explored. For our RL system, it was shown that the epsilon decay rate must be adjusted to the number of training epochs, so that epsilon reaches a lower bound of approximately 0.05 towards the end of the training. Changing the replay memory size did not lead to notable benefits in the training results. However, adjusting the network size and batch size resulted in a significant improvement. Table 3 shows the comparison of different network sizes, corresponding to the number of nodes per hidden layer, and batch sizes. The largest performance increase is achieved by training a larger network.

Table 3: Comparison of different hyperparameters

Network size	Batch size		
	2048	4096	8192
32	28.5	28	29.5
64	29	30	30

Integrating the job specification into the agent's state lead to a much higher state-size complexity and, therefore, an increased learning complexity, making a larger network size essential. Choosing a larger batch size also helped to increase performance by a small margin after 3000 epochs. This is because a larger batch size utilizes more training samples in each experience replay and allows the agents to learn using many more samples. This leads to longer training times for individual epochs, but may allow the network to converge sooner.

In conclusion, the best results were achieved using the "random order" training strategy, a job frequency of 5 epochs, a network with 64 nodes per hidden layer, a replay memory size of 4096 per agent (leading to a shared memory size of 12288), and a batch size of 8192.

The previous results were achieved with validation job specifications having the same format as those used during training, consisting of four operations whereby each operation could be processed on two possible modules. However, one of the main considerations of the advanced job specification encoding is that it must be able to handle a variable number of operations and number of options in each operation. To verify this functionality, an additional validation was carried out with a set of job specifications in which the total number of operations and the number of options for each operation varied. The network that performed best in the previous experiments was used to create schedules for these variably-sized job specifications. The analysis of the resulting schedules shows remarkable results in that most of the schedules are correct, despite the agent only having been trained using job specifications in the same format.

Figure 3 shows three examples of variably-sized job specifications and the corresponding schedule.

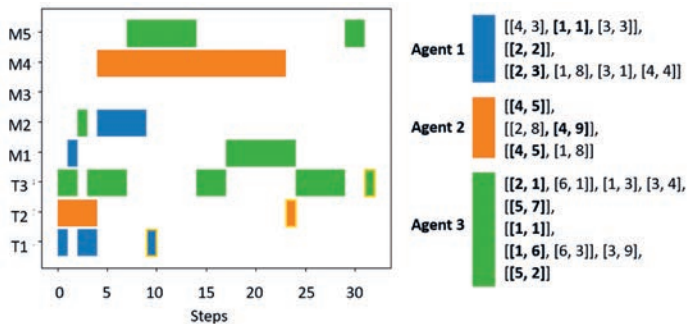


Figure 3: The Gantt chart of the schedule of the jobspecification on the right site (with $M1$, ..., $M6$ indicating the modules and Tn the transport of agent n) was created by the trained agents. The yellow markers highlight that all job specifications are scheduled correctly. This schedule shows a schedule for a set of job specifications with different formats of job specifications that were not used in the training. As this schedule is completely correct, it is rated with 3.0 points.

The results demonstrate that our advanced job specification encoding not only enables the agents to generalize the job specifications that have been seen during training and, therefore, correctly schedule unseen job specifications, but also to schedule job specifications in arbitrary formats. However, there is one major shortcoming. The schedules that are being created are often not optimal with regard to the makespan, which is very important for the schedules to be useful in practice. Analyzing the resulting schedules of many experiments has shown that the agents have a tendency to enter the first module able to process the current operation that they pass on the conveyor belt. However, this module is sometimes a sub-optimal choice with respect to a minimal makespan. For example, this behavior can be observed with Agent 1 in Figure 1, which chooses the slower module for the first operation as it is passed first. While there are cases in which an agent skips a valid module to then choose the faster module (see the first operation of Agent 3 in Figure 3), the tendency of entering the first valid module is clearly present. This issue will be addressed in future work.

6. Conclusion

In this article, we have discussed our latest results in creating a multi-agent RL system to solve the MrFJSP. An advanced job specification encoding was introduced to train agents that can schedule arbitrary products instead of being specialized to a single product. This encoding successfully allowed agents to handle job specifications of any length and shape. The concept enables the generalization of information within the job specifications with the possibility to

allow a flexible number of agents with the use of an eligibility criterion. In our experiments, we trained three agents on 600 randomly selected job specifications. The validation was done based on the number of correctly scheduled products. For this, we used ten validation scenarios, each consisting of three unseen job specifications. It was shown that the agents were able to create correct schedules for all of the validation scenarios. However, the resulting schedules were often not optimal with respect to the makespan. In future work, we will look into optimizing the created schedules by examining the aspect of cooperation.

References

- [1] M. Fera, F. Fruggiero, A. Lambiase, G. Martino, and M. Elena, "Production scheduling approaches for operations management," in *Operations Management* (M. Schiraldi, ed.), InTech, 2013.
- [2] S. Baer, Bakakeu Romuald Jupiter, R. Meyes, and T. Meisen, "Multi-agent reinforcement learning for job shop scheduling in flexible manufacturing systems," in *2019 Second International Conference on Artificial Intelligence for Industries (AI4I)*, IEEE, 2019.
- [3] S. Baer, Kumar Mohanty, Punit, D. Turner, and T. Meisen, "Multi agent deep q-network approach for online job shop scheduling in flexible manufacturing," in *International Conference on Manufacturing Systems and Multiple Machines (ICMSMM 2020)*, waset, 2020.
- [4] T. Gabel, M. Riedmiller, and N. Group, "M.: Adaptive reactive job-shop scheduling with learning agents," *International Journal of Information Technology and Intelligent Computing*, 2007.
- [5] Thomas Gabel, *Multi-Agent Reinforcement Learning Approaches for Distributed Job-Shop Scheduling Problems*. 2009.
- [6] B. Waschneck, A. Reichstaller, L. Belzner, T. Altenmüller, T. Bauernhansl, A. Knapp, and A. Kyek, "Optimization of global productionscheduling with deep reinforcement learning," *Procedia CIRP*, vol. 72, pp. 1264–1269, 2018.
- [7] A. Rinciog, C. Mieth, P. M. Scheikl, and A. Meyer, "Sheet-metal production scheduling using alphago zero," <https://doi.org/10.15488/9640>.
- [8] C. B. H. A. M. Roesch, C. Lindner and R. Gunther, "Industrial load management using multi-agent reinforcement learning for rescheduling," in *2019 Second International Conference on Artificial Intelligence for Industries (AI4I)*, IEEE, 2019.
- [9] S. Qu, J. Wang, S. Govil, and J. O. Leckie, "Optimized adaptive scheduling of a manufacturing process system with multi-skill workforce and multiple machine types: An ontology-based, multi-agent reinforcement learning approach," *Procedia CIRP*, vol. 57,

pp. 55–60, 2016.

[10] M. Egorov, "Multi-agent deep reinforcement learning." Stanford University, CS231n Course Report, 2016.

[11] D. Bloembergen, K. Tuyls, D. Hennes, and M. Kaisers, "Evolutionary dynamics of multi-agent learning: A survey," *Journal of Artificial Intelligence Research*, vol. 53, pp. 659–697, 2015.

[12] L. Panait and S. Luke, "Cooperative multi-agent learning: The state of the art," *Autonomous Agents and Multi-Agent Systems*, vol. 11, no. 3, pp. 387–434, 2005.

[13] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *International Conference on Machine Learning* 2016, pp. 1928–1937, 2016.

[14] M. Wiering and M. van Otterlo, *Reinforcement Learning*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012

[15] Jian Zhang, Guofu Ding, Yisheng Zou, Shengfeng Qin, and Jianlin Fu. Review of job shop scheduling research and its new perspectives under industry 4.0. *Journal of Intelligent Manufacturing*, 30(4):1809–1830, 2019.

Betriebsparallele Adaption einer Co-Simulation für diskrete Fertigungsanlagen

M. Sc. **C. Härle**, Prof. Dr.-Ing. **M. Barth**, Hochschule Pforzheim;
Prof. Dr.-Ing. **A. Fay**, Helmut-Schmidt-Universität Hamburg

Kurzfassung

Simulation nimmt eine zunehmend wichtige Rolle im Engineering und Betrieb modularer Maschinen und Anlagen ein. Nach wie vor erfordern sowohl die Erstellung als auch die Pflege der Modelle einen beträchtlichen manuellen Aufwand. Um diesem Problem zu begegnen wurde in Vorveröffentlichungen ein Assistenzsystem für die automatische Komposition und Konfiguration von Co-Simulationen eingeführt, dessen Fokus auf Modellen für die Inbetriebnahmephase liegt. In diesem Beitrag erweitern die Autoren das Konzept des Assistenzsystems auf die Betriebsphase. Ziel ist es, die Simulation hinsichtlich Genauigkeit und Laufzeit zu optimieren. Zur Generierung unterschiedlicher Modellkonfigurationen werden evolutionäre Algorithmen eingesetzt, welche die vorgegebenen Ziele in unterschiedlicher Ausprägung erfüllen. Anschließend wird dem Bediener, auf Basis des jeweiligen Use-Cases, eine geeignete Variante vorgeschlagen. Das Konzept wird anhand eines industriellen Anwendungsbeispiels prototypisch validiert.

1. Motivation und Problemstellung

Der Maschinen- und Anlagenbau setzt zunehmend auf modulare Systemlösungen [1]. Hierdurch wird gewährleistet, dass die Investitionsgüter trotz volatiler Kundenanforderungen, Produktportfolios und Produktionsrandbedingungen lang genug eingesetzt werden können, um rentabel zu sein [2]. Aus Sicht der Simulationstechnik stellt eine sich oftmals ändernde Systemkonfiguration jedoch neue Herausforderungen dar [3]. Neue Modulsimulationsmodelle (MSM) müssen integriert, bestehende ersetzt und Modellparameter angepasst werden. Sind mehrere MSM von einem realen Modul vorhanden, beispielsweise von einem Antrieb mit und ohne Betrachtung thermischer Einflüsse oder mit unterschiedlich detaillierten Reibungsmodellen, gilt es für verschiedene Simulationsszenarien den jeweils geeigneten Detaillierungsgrad (engl. *Level of Detail*, kurz LoD) zu wählen [4]. So kann, je nach Bedarf, das reale System hinreichend genau abgebildet und der Rechenaufwand minimal gehalten werden. In [4] wird konkretisierend gefordert, „dass der größte Teil eines Anlagensimulationsmodells stets im niedrigsten Detaillierungsgrad simuliert werden sollte. Nur einzelne Teile einer Anlage sollten, sofern dies erforderlich ist, mit höherer Detaillierung abgebildet werden“. Die Anlagenteile, bei

denen ein detaillierteres Modell erforderlich ist, gilt es möglichst automatisiert zu identifizieren. Zudem ist anzunehmen, dass verschiedene Disziplinen unterschiedliche Simulationsergebnisse benötigen. Der manuelle Aufwand für Aufbau und Pflege der Simulationen ist bislang jedoch zeit- und kostenintensiv sowie fehleranfällig [5]. Zudem beschränkt sich der Einsatz von Simulation nicht mehr ausschließlich auf das Engineering. Auch parallel zum Betrieb kann Simulation gewinnbringend eingesetzt werden [6, 7]. So eröffnet der kontinuierliche Vergleich von Simulationsergebnissen und Anlagenmesswerten neue Diagnosemöglichkeiten und kann zu einem robusten Anlagenbetrieb beitragen. Beispielhaft können bei einem Sensorausfall, wie z. B. einer defekten Lichtschranke, Systemzustände durch das Modell geschätzt werden [6]. Neben diesen betriebsparallelen Einsatzmöglichkeiten kann auch eine Voraussimulation für Prognose- und Optimierungszwecke, beispielsweise um Prozessparameter zu optimieren [7], durchgeführt werden.

In der Praxis sind die Ursachen für eine Abweichung zwischen Mess- und Simulationsdaten vielfältig [8]. Bild 1 zeigt eine Übersicht der aus der Literatur zusammengetragenen Ursachen und unterteilt diese in die Cluster „Anlage“ und „Simulation/ Modell“. Die Datenaufbereitung betrifft beide Cluster und ist grundsätzlich immer notwendig. Hierzu gehören u. a. Interpolation, wenn Abtastrate und Schrittweite nicht zusammenpassen, das Verschieben der Zeitreihen gegeneinander, wenn beispielsweise die Messung zu spät gestartet wurde, oder eine

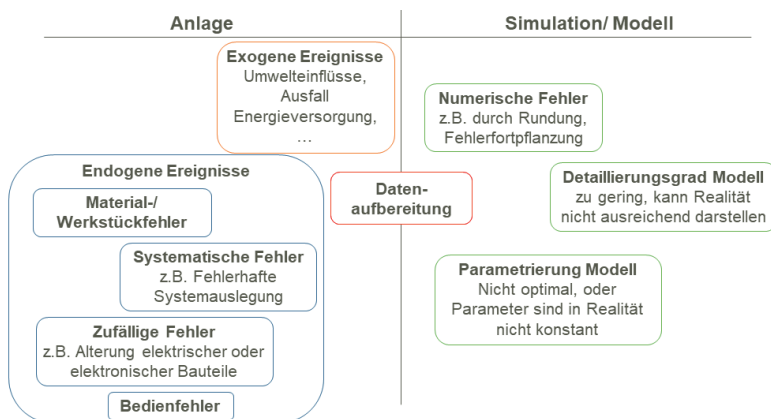


Bild 1: Ursachencluster für Abweichungen zwischen Mess- und Simulationsdaten.

Konsistenzprüfung der Einheiten. Wird dieser Schritt nicht durchgeführt, können Abweichungen auftreten, obwohl sich sowohl Anlage als auch Simulation im Soll-Zustand befinden. In

Vorveröffentlichungen der Autoren [9, 10] wurden verwandte Arbeiten im Bereich der automatischen Modellgenerierung und -adaption analysiert und der Bedarf für eine lebenszyklusübergreifende Lösung identifiziert. Eine Teillösung für die Inbetriebnahmephase wurde bereits erarbeitet, die im Rahmen dieses Beitrags auf die Betriebsphase erweitert wird. In Abschnitt 2 werden hierfür zunächst Ziele, Anforderungen und Randbedingungen, die teils aus den Vorveröffentlichungen resultieren, definiert. In Abschnitt 3 wird das Lösungskonzept erarbeitet. In Abschnitt 4 wird der entwickelte Prototyp des Assistenzsystems (AS) anhand eines industriellen Anwendungsbeispiels validiert.

2. Ziel, Anforderungen und Rahmenbedingungen

Wie in Abschnitt 1 beschrieben, bilden manuelle Modellierungstätigkeiten eine der wesentlichen Herausforderungen beim Einsatz von Simulation. Ziel ist es daher diese durch automatische Abläufe zu ersetzen, bzw. die Mitarbeiterinnen und Mitarbeiter bei deren Arbeit zu unterstützen. Hierzu muss das in diesem Beitrag erweiterte Assistenzsystem, neben den in [9, 10] definierten Funktionen (z. B. Modulmodelle laden, Initialwerte setzen und Modulmodelle koppeln), die folgenden Anforderungen umsetzen:

- A1. Simulations- und Messwerte, die zur Betriebsphase aufgenommen werden, miteinander vergleichen.
- A2. Feststellen, wenn eine Abweichung der in A1 beschriebenen Werte eine vorgegebene Toleranz überschreitet.
- A3. Mögliche Ursachen für die in A2 erwähnte Abweichung in den Simulationsmodellen lokalisieren.
- A4. Automatische Durchführung einer Adaption der betroffenen Simulationsmodelle.

Um den Vorgang der Komposition und Konfiguration zu automatisieren, müssen MSM für alle Module, die in der Gesamtsystemsimulation gekoppelt werden sollen, zur Verfügung stehen. Des Weiteren müssen die MSM über standardisierte Schnittstellen verfügen. Hier fiel die Wahl der Autoren auf das Functional Mock-Up Interface (FMI) [11], da dieser werkzeug- und herstellerunabhängige Standard weit verbreitet und akzeptiert ist. Auf die dadurch gewährleistete standardisierte Architektur aller MSM lässt sich zudem ein Optimierungsalgorithmus anpassen, um so die automatische Modelladaption zu realisieren. Darüber hinaus muss eine Beschreibung in Form eines Topologiemodells (TM) der geplanten Produktionsanlage existieren, für welches AutomationML (AML) [12] eingesetzt wird.

3. Lösungskonzept

3.1 Konzeptskizze und Möglichkeiten der Adaption

In diesem Abschnitt wird das Konzept zur automatischen Komposition und Konfiguration von Co-Simulationen um die betriebsparallele Adaption (blaue Pfade) erweitert, wie in Bild 2 dargestellt. Eine Basis-Aufgabe des AS ist es, die in der Anlage vorhandenen Module auf Grundlage des TM zu identifizieren, und die benötigten MSM aus FMU-Bibliotheken zu laden. Das AS generiert auf Basis des TM ein Skript zur Parametrierung und Steuerung der Simulation. Diese Funktionen werden im Folgenden auf die Betriebsphase erweitert. Neben den Kopp- lungsinformationen können aus dem TM weitere Informationen, wie z. B. Modellparameter und das Vorhandensein von MSM in unterschiedlichen Detailgraden, abgeleitet werden. Erkennt das AS, dass die Abweichung der Mess- und Simulationsdatensätze einen vorgegebenen Wert überschreitet, wird zunächst eine Ursachenanalyse durchgeführt. Es gilt die Frage zu beantworten, ob die Ursache anlagen- oder simulationsseitig einzuordnen ist (vgl. Bild 1). Hier- für wird zunächst angenommen, dass das Gesamtmodell korrekt ist und die Simulation ord- nungsgemäß läuft. Die Messdaten werden daraufhin mit einer Bibliothek aus Fehlerbildern abgeglichen, um Komponentenverschleiß oder -ausfall zu identifizieren. Wenn ein bekanntes Fehlerbild identifiziert wird, wird eine Meldung ausgegeben und eine Anlagenwartung muss eingeleitet werden (linker Pfad). Für diese Analysefunktionen existieren bereits Lösungen aus dem Bereich der prädiktiven Instandhaltung [13, 14]. Der Aufbau dieser Fehlerbilder-Bibliothek ist ein kontinuierlicher

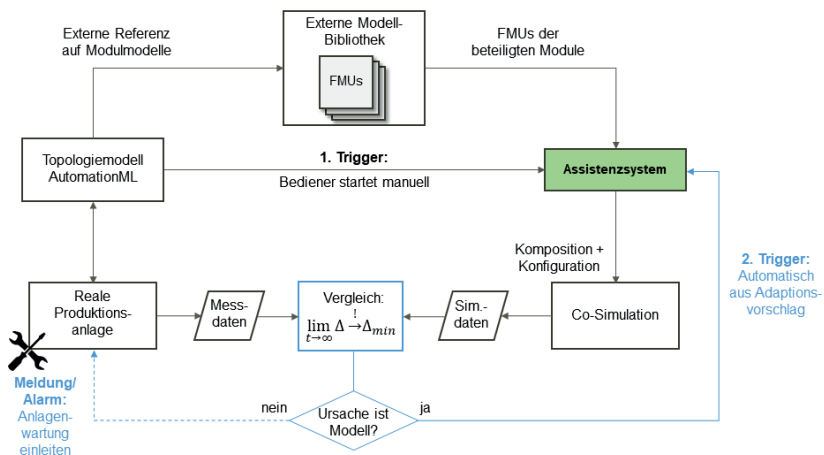


Bild 2: Konzept zur automatischen Modellgenerierung und betriebsbegleitenden -adaption.

anlagenspezifischer Prozess, welcher meist eine beträchtliche Menge an Messdaten voraussetzt. Da der Fokus auf der simulationsseitigen Adaption liegt, wird der linke Pfad in der Konzeptskizze im Rahmen der anschließenden Implementierung und Konzeptvalidierung nicht weiter betrachtet, sondern der rechte, d.h. die Ursache der Abweichung wird im Modell verortet. Im nächsten Abschnitt wird das Problem zunächst in eine maschinell interpretier- und lös-bare Form gebracht.

3.2 Formulierung der Adaptionaufgabe

Die Formulierung der Adaptionaufgabe führt zu einem Optimierungsproblem. Allgemein formulieren lässt sich ein Optimierungsproblem als

$$\min_{\underline{x} \in \mathbb{R}^n} f(\underline{x}) \quad , \quad f: \mathbb{R}^n \rightarrow \mathbb{R} \quad . \quad (1)$$

Die Funktion f wird als Güte- oder Zielfunktion, der Vektor $\underline{x}$ als Designvariablen bezeichnet. Möchte man beispielsweise einen Produktionsprozess hinsichtlich der benötigten Zeit optimieren, ist die Zeit die Zielfunktion. Designvariablen, über die man Einfluss auf die benötigte Zeit nehmen kann, sind beispielsweise Materialeigenschaften des Werkstücks, sowie die Prozessparameter. Das Optimierungsproblem (1) ist gelöst, wenn ein $\underline{x}^*$ gefunden wurde, für das gilt

$$f(\underline{x}^*) \leq f(\underline{x}) \quad \forall \underline{x} \in \mathbb{R}^n \quad . \quad (2)$$

Dabei müssen, insbesondere bei Optimierungen von realen Anlagen und Maschinen, unterschiedliche Rand-/ Nebenbedingungen erfüllt werden. Beispielsweise können Prozessparameter nicht beliebig variiert werden. Damit erweitert sich (1) zu

$$\begin{aligned} \min_{\underline{x} \in \mathbb{R}^n} \quad & f(\underline{x}) \quad , \quad f: \mathbb{R}^n \rightarrow \mathbb{R} \\ \text{s. t.} \quad & g_j(\underline{x}) \leq 0 \quad , \quad j = 1, \dots, e \quad , \quad (3) \\ & h_l(\underline{x}) = 0 \quad , \quad l = 1, \dots, r \end{aligned}$$

wobei die zusätzlichen Ungleichungen h_l und Gleichungen g_j den Lösungsraum einschränken. Im Rahmen der betriebsparallelen Modelladaption beschreibt eine offensichtliche Zielfunktion das Maß an Übereinstimmung zwischen Mess- und Simulationswerten $\underline{y}_{mess}$ und $\underline{y}_{sim}$, mit $\underline{y}_{mess}, \underline{y}_{sim} \in \mathbb{R}^{k \times v}$. Es wird angenommen, dass k Mess- und Simulationspunkte vorliegen und für die Anzahl an Sensorpaaren $v \geq 1$ gilt. Als Sensorpaare werden die Sensoren aufgefasst, die sowohl in der realen Anlage als auch in der Simulation enthalten sind. Ziel ist, wie beschrieben, die Minimierung der Abweichung zwischen den simulierten und real gemessenen Werten. Zunächst gilt es die Frage zu klären, wie das Maß an Übereinstimmung über alle Sensorpaaren und Mess-/Simulationspunkte bestimmt werden kann. Eine häufig eingesetzte

Methode ist das Bestimmtheitsmaß R^2 [15, 16]. Dieses nutzt die Residuenquadratsumme, welche beispielsweise beim Least-Square-Verfahren für lineare Regressionsrechnung verwendet wird, und betrachtet zusätzlich den empirischen Mittelwert $\bar{y}_{mess}$. Jedoch haben eigene Untersuchungen, die auf Basis der Untersuchungen in [17] durchgeführt wurden, für im Anlagenbau häufig eingesetzte nicht-lineare Modelle ergeben, dass sich Alternativen wie das *Akaike information criterion* (AIC), das *Corrected Akaike information criterion* (AIC_c) oder das *Bayesian information criterion* (BIC) besser eignen. Das AIC berechnet sich nach [17]:

$$AIC = 2n - 2 \ln(L) \quad , \text{ mit} \quad (4)$$

$$L = 0.5 \left(-N \left(\ln 2\pi + 1 - \ln N + \ln \sum_{i=1}^k z_i^2 \right) \right) .$$

Dabei beschreibt L die Likelihood-Funktion, n steht für die Anzahl an Modellparametern, da $\underline{x} \in \mathbb{R}^n$ angenommen wurde. Die Residuen aus dem nichtlinearen Least-Squares-Fit werden durch z_i beschrieben und N steht für die Anzahl der Residuen. Über die Likelihood-Funktion wird die Anpassungsgüte des Modells belohnt, die Anzahl an Modellparametern geht als sogenannter Strafterm ($2n$ in der ersten Zeile von (4)) in die Berechnung ein. Dieser verhindert ein Overfitting der Modelle. Durch mehr Modellparameter kann, bis zu einem Grenzwert, eine bessere Anpassungsgüte erreicht werden. Es ist allerdings nicht erstrebenswert diesen Grenzwert unter allen Umständen zu erreichen, da zusätzliche Parameter u. a. die Komplexität des Modells und den Aufwand für das Parametertuning erhöhen. Beim Vergleich mehrerer Modelle ist ein minimaler AIC-Wert zu bevorzugen. Die Berechnung der Modellgüte mittels BIC beinhaltet ebenfalls die Likelihood-Funktion, gewichtet bzw. bestraft die Anzahl an verwendeten Modellparametern jedoch unterschiedlich. Im Rahmen dieser Publikation wird exemplarisch AIC als Gütekriterium verwendet. Um mit den Definitionen aus ((-3) konsistent zu bleiben wird (4) umformuliert und in Abhängigkeit der Designvariablen $\underline{x}$ gesetzt

$$f(\underline{x}) = 2n - 2 \ln \left(0.5 \left(-N \left(\ln 2\pi + 1 - \ln N + \ln \sum_{i=1}^k \left(y_{i,mess} - y_{i,sim}(\underline{x}) \right)^2 \right) \right) \right) . \quad (5)$$

Im nächsten Schritt gilt es zu untersuchen, von welchen Einflussfaktoren die Simulationsergebnisse y_{sim} abhängen. Aus Sicht eines MSM stellt der individuelle Parametersatz $\underline{p}_{individ}$ den offensichtlichsten Einfluss auf die von diesem Modell generierten Simulationsergebnisse dar. Da bei einer Co-Simulation jedoch alle MSM über ihre Ein- und Ausgänge gekoppelt sind, müssen Wechselwirkungen zwischen den Modellen berücksichtigt werden.

selwirkungen zwischen den Modellen beachten werden. Damit haben auch die Parametersätze aller anderen Module $\underline{p}_{global}$ Einfluss auf die Simulationenwerte des individuell betrachteten Moduls. Die Parametersätze aller in der Co-Simulation enthaltenen Modulmodelle werden als $\underline{p}_{ges} = \begin{bmatrix} \underline{p}_{individ} \\ \underline{p}_{global} \end{bmatrix}$ zusammengefasst. Neben Modellparametern haben auch die Detaillierungsgrade aller Modelle $\underline{d}_{ges}$ einen Einfluss auf die Simulationsmodelle. Der Grund alle Detaillierungsgrade zu betrachten ist analog zum Wechselwirkungsargument bei den Parametersätzen. Im Rahmen dieses Beitrags wird der Detaillierungsgrad, in Anlehnung an [4], als $d \in \mathbb{N}_+$ angenommen, wobei Grad 1 die niedrigste Detaillierungsstufe eines Modells darstellt. Für das AS und den Optimierungsalgorithmus ist es wichtig, dass alle Modellquellen (z. B. die einzelnen Zulieferer) die gleichen Definitionen verwenden. Damit lässt sich (5) spezifizieren zu

$$f(\underline{p}_{ges}, \underline{d}_{ges}) = 2n - 2 \ln(L(\underline{p}_{ges}, \underline{d}_{ges})) . \quad (6)$$

Wie in (3) definiert, gibt es für Optimierungsprobleme häufig Randbedingungen zu erfüllen. In dem hier betrachteten Problem sind das zum einen die Grenzen der Modellparameter $\underline{p}_{ges}$. Im Allgemeinen ist davon auszugehen, dass die Modellparameter eine Unter- und eine Obergrenze haben, da es sich um physikalische Größen handelt. Die Grenzwerte werden in den beiden Vektoren $\underline{p}_{ges,min}$ und $\underline{p}_{ges,max}$ zusammengefasst. Mit dieser Erkenntnis und der in (6) formulierten Zielfunktion lässt sich das Optimierungsproblem nach (3) folgendermaßen aufstellen

$$\begin{aligned} \min_{\underline{p}_{ges}, \underline{d}_{ges}} \quad & f(\underline{p}_{ges}, \underline{d}_{ges}) \\ \text{s.t.} \quad & \underline{p}_{ges} - \underline{p}_{ges,max} \leq 0 \\ & -\underline{p}_{ges} + \underline{p}_{ges,min} \leq 0 \end{aligned} \quad (7)$$

In Bezug auf den Detaillierungsgrad gilt es ein geeignetes Maß zu finden. Detailreichere Modelle können zur Minimierung von (7) beitragen, aber auch die benötigte Rechenzeit erhöhen. Aus Sicht der Rechenzeit gilt es somit den Detaillierungsgrad des Gesamtmodells, welcher als Summe der Einzelmodelle angenommen wird, zu minimieren. Somit lässt sich eine zweite Zielfunktion f_2 formulieren

$$\min_{\underline{d}_{ges}} \quad f_2(\underline{d}_{ges}) = \sum_i^m d_m \quad , \quad (8)$$

wobei m der Anzahl an Modellen in der Co-Simulation entspricht. Das Optimierungsproblem F besteht somit aus zwei in Konflikt zueinanderstehenden Zielfunktionen.

$$\min_{\underline{p}_{ges}, \underline{d}_{ges}} F(\underline{p}_{ges}, \underline{d}_{ges}) = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} 2n - 2 \ln \left(L(\underline{p}_{ges}, \underline{d}_{ges}) \right) \\ \sum_m^m d_m \end{bmatrix} \quad (9)$$

$$\text{s.t. } \underline{p}_{ges} - \underline{p}_{ges, \max} \leq 0$$

$$-\underline{p}_{ges} + \underline{p}_{ges, \min} \leq 0$$

Diese Art von Optimierungsproblem wird in [18, 19] als Multi-Objekt-Optimierungsproblem (MOOP) oder *Pareto Optimierung* bezeichnet. Im nächsten Schritt gilt es geeignete Algorithmen zur Lösung dieser Problemklasse zu identifizieren.

3.3 Algorithmen-Wahl für Pareto-Optimierung

In [18] werden globale Such- und Optimierungsansätze in die drei Kategorien *enumerativ*, *deterministisch* und *stochastisch* eingeteilt. Obwohl eine enumerative Suche deterministisch ist, differenziert hier [18], da sie keine Heuristik verwendet. Die beschriebene Klassifizierung ist mit relevanten Vertretern der jeweiligen Kategorien in Bild 3 zusammengefasst.

Jeder dieser Ansätze besitzt nach [19] Vorteile aber auch grundlegende Einschränkungen. Der enumerative ist der am wenigsten effiziente, aber durch die Betrachtung aller möglichen Lösungen ein robuster Optimierungsansatz. Durch die Betrachtung aller möglicher Lösungen wird das globale Optimum garantiert gefunden. Hier wird jedoch schnell die Schwäche des Ansatzes deutlich. Eine Aufzählung oder Betrachtung aller Lösungen ist nur bei diskreten Problemen effektiv, da hier ein endlicher Lösungsraum vorliegt. Auch kontinuierliche Probleme

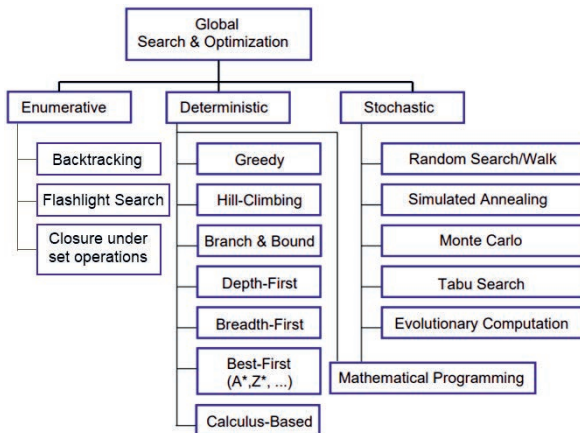


Bild 3: Einteilung globaler Such- und Optimierungsalgorithmen nach [18] mit populären Vertretern.

können enumerativ gelöst werden, hier ist jedoch zunächst eine Diskretisierung des Suchraums nötig. Nach der Umwandlung eines kontinuierlichen in ein diskretes Problem nimmt die Auflösung der Lösung umgekehrt proportional zur Anzahl der diskreten Punkte ab, die den (ursprünglich kontinuierlichen) Entwurfsraum repräsentieren. In der Praxis kann die Aufzählung aufgrund der schnell ansteigenden kombinatorischen Möglichkeiten und der daraus resultierenden Zeitkomplexität nur bei kleinen diskreten Designdomänen angewendet werden, d. h. bei wenigen Variablen, die lediglich mit einer geringen Auflösung diskretisiert werden [19]. In [20] wird dem enumerativen Ansatz die Eignung für realistische Optimierungsprobleme ab-erkannt.

Deterministische Optimierungsansätze sind nach [19] auf eine kleine Teilmenge der allgemeinen MOOP beschränkt. Die Verwendung von deterministischen Algorithmen basiert typischerweise auf Annahmen, beispielsweise hinsichtlich der Eigenschaften der Zielfunktion, des Optimierungsproblems oder des Suchraumes. Wenn diese Annahmen nicht erfüllt sind, ist ein Fehlschlagen des Optimierungsalgorithmus wahrscheinlich. Deterministische Algorithmen beruhen nach [19] des Weiteren auf einer Reihe von Annahmen, wie z. B. das Vorliegen eines linearen Problems in analytischer Form oder dem Ausbleiben von Wechselwirkungen zwischen Parametern. In Anbetracht realitätsnaher MSM gelten für das in (9) beschriebene Optimierungsproblem die meisten der formulierten Annahmen nicht. So ist f_2 beispielsweise unstetig und f_1 als nichtlinear anzunehmen. Wenn mindestens eine der Zielfunktionen oder der Randbedingungen nichtlinear ist, ist (9) ein nichtlineares Problem [21]. Zusammengefasst ist die Klasse der deterministischen Algorithmen für das in (9) beschriebene Optimierungsproblem ebenfalls nicht zu bevorzugen.

Im Bereich der stochastischen Verfahren werden evolutionäre Algorithmen routinemäßig bei der Lösung von Pareto-Optimierungsproblemen in verschiedenen Zweigen der Technik, Wissenschaft und Wirtschaft eingesetzt [23, 24]. Die zunehmende Akzeptanz dieser Algorithmen ist nach [25] auf ihre Fähigkeit zurückzuführen, mehrere Lösungen in einem einzigen Durchlauf zu finden, ohne Ableitungen zu arbeiten, zu Pareto-optimalen Lösungen zu konvergieren und sowohl für kontinuierliche Funktions- als auch kombinatorische Optimierungsprobleme genutzt werden zu können. Darüber hinaus sind evolutionäre Algorithmen weniger anfällig für die Form oder Kontinuität der Pareto-Front. Der Einsatz bei diskontinuierlichen oder konkaven Pareto-Fronten ist möglich, während diese beiden Aspekte für mathematische Programmieretechniken ein Problem darstellen [26]. Für die Implementierung der betriebsbegleitenden Adaption der Co-Simulation wählen die Autoren daher im Folgenden evolutionäre Algorithmen.

4. Konzeptvalidierung

4.1 Betrachtetes System als Validierungsszenario

Für die Konzeptvalidierung wird eine Gleitschleifanlage der Firma OTEC Präzisionsfinish GmbH betrachtet. Das CAD-Modell ist in Bild 4 (a) dargestellt, die 3D-Visualisierung des resultierenden Physik-Modells in Bild 4 (b). Mittels der Anlage können Werkstücke (WS) wie Bohrer und Fräser geschliffen und poliert werden. Hierzu wird das WS (I) in eine Halterung (II) eingespannt, in (b) durch den grauen Zylinder dargestellt, und in einen Bearbeitungsbehälter (III) abgesenkt, dargestellt in dunkelblau. Der Behälter ist mit Schleifmittel befüllt und wird in Rotation versetzt. Das Schleifmittel ist auf dieser Abstraktionsebene nicht dargestellt. Durch die Strömung des Schleifmittels und die Bewegung des WS wird der Materialabtrag erzielt. Die resultierende Kraft $f_{res} = \begin{pmatrix} f_x \\ f_y \end{pmatrix}$ auf das WS soll im Rahmen einer Simulation berechnet werden. Im Folgenden werden die Module „Bearbeitungscontainer“ und „Spannlanze mit WS“ betrachtet. Im nächsten Abschnitt werden die Physik-Modelle, deren Parameter sowie Detaillierungsgrade vorgestellt.

4.2 Physik-Modelle, Parameter und Detaillierungsgrade

Um den Modellierungsprozess zu beschleunigen und die Vorteile einer 3D-Visualisierung nutzen zu können, werden Elemente der Modelica-Standardbibliothek (*Mechanics* → *Multi-Body*) verwendet. Zuerst wird der Bearbeitungsbehälter betrachtet. Dieser besteht aus einer Masse mit Trägheit (*Body*), einem Drehgelenk, einem Achs-Antrieb sowie einem Sensor für die Drehgeschwindigkeit. Das Physik-Modell des Drehbehälters ist in Bild 5 dargestellt. Um

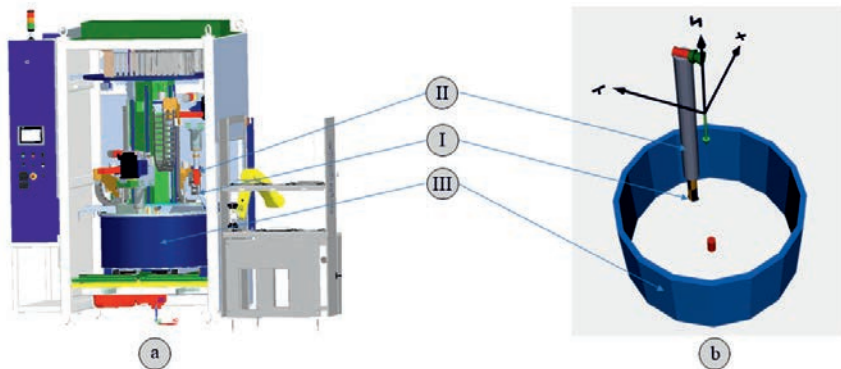


Bild 4: Modelle der betrachteten Anlage. CAD Modell (a), 3D-Visualisierung des Physik-Modells (b).

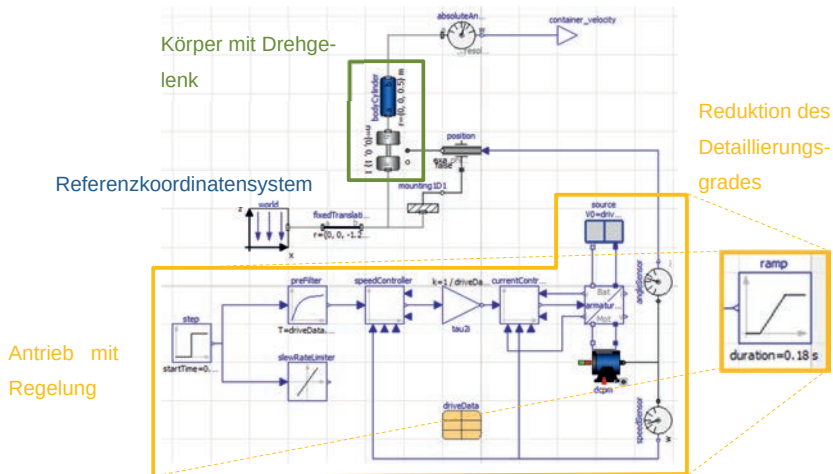


Bild 5: Physik-Modell des Bearbeitungsbehälters in hohem Detaillierungsgrad und gewählte Vereinfachung in orange.

die Komplexität des Modells zu reduzieren, wird in einem zweiten Modell der orange umrahmte Block (DC-Maschine mit Reibung, Trägheit und Kaskadenregler) durch die daneben abgebildete Rampen-Funktion ersetzt. Diese Vereinfachung reduziert das Gleichungssystem des Modells von 1721 auf 1186 Modellgleichungen. Die Drehgeschwindigkeit wird aus dem Modell geführt und dient als Eingang des Spannlansenmodells. Diese wird für die Berechnung von f_{res} benötigt, da sich hierüber und durch den Abstand des WS zum Behälterrand die Strömungsgeschwindigkeit des Schleifmittels annähern lässt.

Das Physik-Modell der Spannlanze mit WS ist in Bild 6 dargestellt. Die Spannlanze besitzt ein Drehgelenk mit Antrieb (grüner Block), um das WS in die Strömung zu stellen. Ein weiteres Drehgelenk ermöglicht dem WS eine Drehung um dessen Rotationsachse (blauer Block). Dies entspricht dem sogenannten Puls-Antrieb der Anlage. Die Drehrichtung des WS geht als Faktor in die Berechnung von f_{res} (roter Block) ein. Existierende Messdaten der auf das WS wirkenden Kräfte werden für die Plausibilisierung des Modells verwendet. Für die Validierung der Adaption durch das AS wird das Behälter- mit dem Spannlansenmodell gekoppelt. Dabei stellen der Detaillierungsgrad des Bearbeitungsbehälters sowie die frei gewählten Parameter *Schleifmittel-Dichte*, *Amplitude der Trapezoid-Funktion für Pulsantrieb*, *Drehgeschwindigkeit des Behälters* und die *effektive Eintauchtiefe WS in Schleifmittel* mögliche Designvariablen für die Adaption dar.

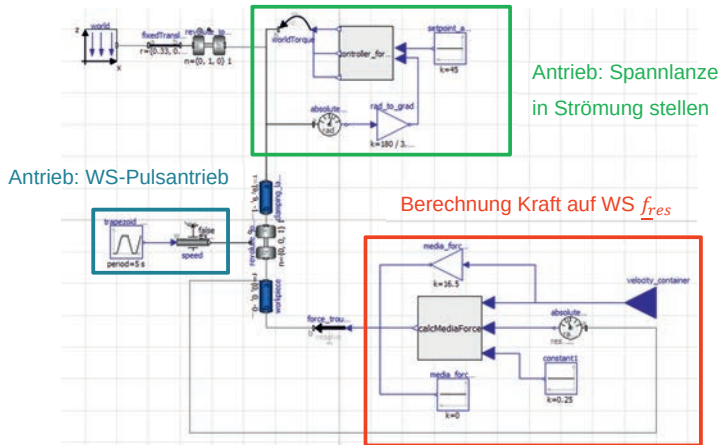


Bild 6: Physik-Modell der Spannlanze mit Werkstück inklusive Berechnung der resultierenden Kraft.

4.3 Ergebnisse der Adaption und Vergleich verschiedener Algorithmen

Um das AS mit der notwendigen Adaptionfunktionalität auszustatten, wird es durch die Open Source Bibliothek MOEA Framework [27] erweitert. Diese zeichnet sich u. a. durch eine kontinuierliche Entwicklungshistorie und eine umfangreiche API aus. Das MOEA Framework bietet mehr als 20 Algorithmen (u. a. NSGAII, SPEA2, VEGA) und die Möglichkeit diese problem-spezifisch zu konfigurieren (z. B. Populationsgröße, Mutationsrate). Zudem können die Abbruchkriterien für die Adaption eingestellt werden, z. B. nach vorgegebener Berechnungszeit oder Anzahl an Evolutionsschritten. Erfolg und Performance einer Adaption hängen, neben den stochastischen Einflüssen, von den Erfahrungen und Kenntnissen des Bedieners hinsichtlich Wahl und Setup des evolutionären Algorithmus ab. Für das betrachtete Beispiel-Szenario kann gezeigt werden, dass bereits mit den Standard-Einstellungen eine Verbesserung der Simulationsergebnisse hinsichtlich der beiden Gütefunktionen erreicht werden kann. In einem ersten Versuch wird exemplarisch der Algorithmus ϵ -NSGAII gewählt und nach 1000 Evolutionsschritten terminiert. Dieser kombiniert den Standard-Algorithmus NSGAII, welcher in der Literatur häufig als Referenz herangezogen wird [28], mit einem ϵ -dominanz Archiv [28], um Konvergenz sicherzustellen. In Bild 7 sind im oberen Teil die gemessenen und simulierten Kräfte f_x und f_y vor und nach der Adaption aufgezeigt. Vor der Adaption werden die geschätzten Parameter aus dem TM der Anlage und für alle Module der

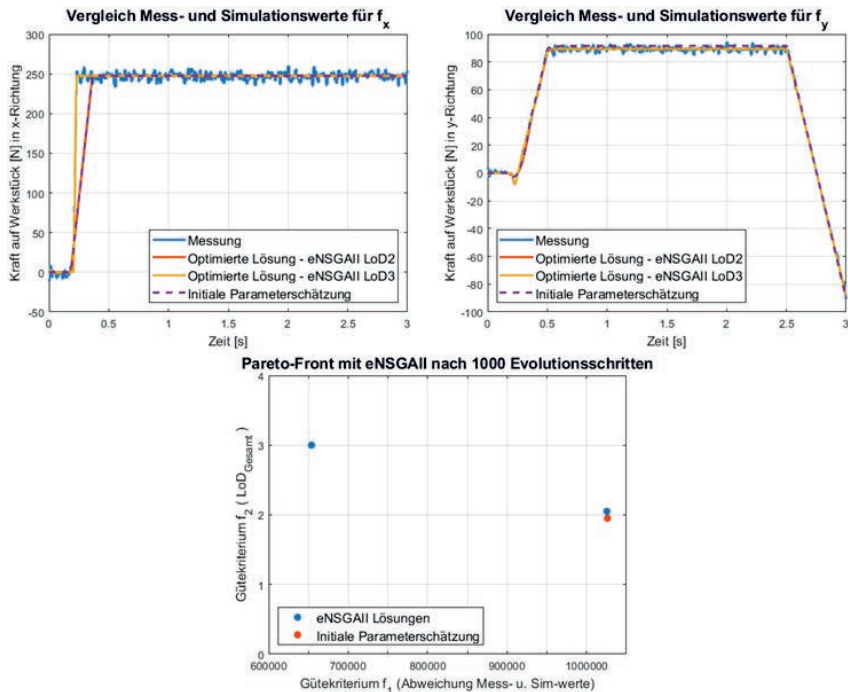


Bild 7: Plot der Kräfte auf das WS vor und nach der Adaption (oben) und Pareto-Front des eNSGAI nach 1000 Evolutionsschritten.

LoD 1 verwendet. Da die initialen Parameter bereits sehr nah an den korrekten Werten liegen ist die Differenz zu den Messwerten gering. In diesem Szenario sind die initialen Schätzwerte dem Algorithmus nicht bekannt. Dieser ist dennoch in der Lage mit dem einfach Behältermodell (LoD_B 1) unmittelbar gute Ergebnisse zu erzielen, und mit dem komplexeren Behältermodell (LoD_B 2) den Anstieg der Kraft f_1 bei Sekunde 0.2 genauer nachzubilden. Im unteren Teil von Abbildung 7 ist die Pareto-Front des MOOP dargestellt. Auf der x-Achse ist die Gütefunktion f_1 , auf der y-Achse f_2 aufgetragen. Die Pareto-Front beinhaltet lediglich zwei Punkte (blau), da auf der y-Achse nur zwei Werte angenommen werden können. Hier ist die Summe der LoD eingetragen, das Spannlanzenmodul besitzt nur ein Modell mit LoD_S 1, das Behältermodul zwei mit LoD_B 1 bzw. LoD_B 2. Sollte das AS das komplexere Behältermodell wählen, ist der Gesamt-LoD LoD_{Ges} 3, ansonsten LoD_{Ges} 2. Der rote Punkt beschreibt die Güte der initialen Parameterschätzung. Ein Evolutionsschritt (Vektor mit Designvariablen generieren, Simulati-

onsskript anpassen, Simulation mit drei Sekunden Simulationszeit durchführen, Gütefunktionen berechnen, ggfs. Pareto-Front aktualisieren) benötigt bei diesem Szenario im Schnitt etwa drei Sekunden auf einem durchschnittlichen Büro-Rechner. Bei 1000 Evolutionsschritten ist der komplette Adaptionsvorgang in unter einer Stunde abgeschlossen.

4.4 Bewertung hinsichtlich der Anforderungen

Über die in (6) formulierte Zielfunktion f_1 wird Anforderung A1, der Vergleich der Simulations- und Messwerte, erfüllt. Durch die mögliche Vorgabe einer maximalen Abweichung zwischen diesen Datensätzen wird Anforderung A2 abgedeckt. Die automatische Adaption betroffener Simulationsmodelle wird, wie in Abschnitt 4.3 gezeigt, ebenfalls abgedeckt. Anforderung A3, die Lokalisierung möglicher simulationsseitiger Ursachen für die Abweichung, teilweise erfüllt werden. So kann der Bediener einzelne Lösungen der angenäherten Pareto-Front betrachten beispielsweise nachvollziehen, dass

- ein Modulmodell nicht ausreichend detailliert dargestellt oder
- Modellparameter suboptimal eingestellt waren.

Für eine umfassende Ursachenidentifikation ist es allerdings notwendig, das bestehende Konzept um eine Sensitivitätsanalyse [29] zu erweitern. Bei dieser wird nach Zusammenhängen zwischen der Varianz von Eingangsfaktoren und Systemantworten gesucht.

Kritische Würdigung des Ansatzes: Industrielle Anwendbarkeit

Mit steigender Anzahl an Modulen, mit multiplen LoD und einer entsprechend größeren Anzahl an Parametern wächst der Suchraum nichtlinear an („Fluch der Dimensionalität“). Der Adaptionsvorgang wird entsprechend aufwändiger, es müssen entweder mehr Evolutionsschritte durchgeführt werden (negative Auswirkung auf Dauer der Adaption), oder es ist mit einer schlechteren Güte des Adaptionsergebnis zu rechnen. Studien zeigen, dass durch die Parallelisierung evolutionärer Algorithmen große Suchräume effizient untersucht, die Qualität der Lösung verbessert und somit diesem Problem begegnet werden kann [30].

Zusammenfassung und Ausblick

In diesem Beitrag wurde ein bestehendes Konzept für die automatische Modellgenerierung im Bereich der diskreten Fertigung um Funktionen für die betriebsparallele Modelladaption erweitert. Hierfür wurde ein Multi-Objekt-Optimierungsproblem formuliert, welches mittels evolutionärer Algorithmen gelöst werden kann. Das Ergebnis dieses Optimierungsproblems ist eine Menge an Lösungen, welche unterschiedliche Güteeigenschaften hinsichtlich Abweichung der Mess- und Simulationswerte sowie der Modelldetaillierungsgrade und damit der benötigten Rechenzeit haben. Der Bediener des AS kann letztlich entscheiden, welche Lösung(en) für

den weiteren Simulationsbetrieb sinnvoll sind. Das Konzept wurde anhand einer Anlage, bestehend aus zwei Modulen, validiert, und es wurde nachgewiesen, dass das AS alle formulierten Anforderungen erfüllt.

Soll das AS zukünftig für reale Applikationen eingesetzt werden, ist es sinnvoll Pareto-Fronten bzw. Lösungsmengen zu speichern, um eine Datenbank aufzubauen. So können Designvariablen (Parameter + Detaillierungsgrad) für weitere Projekte bereitgestellt werden und die Pareto-Front für neue Adaptionen als Teil der initialen Population gewählt werden, um die Adaption zu beschleunigen. Zudem kann die Pareto-Front als Referenz-Set gewählt werden. Dies ermöglicht eine detaillierte Analyse zukünftiger Ergebnisse, z. B. kann so die Konvergenzgeschwindigkeit von Algorithmen verglichen werden. Ein offenes Problem stellen kurzfristige, disruptive Änderungen an der realen Maschine (z. B. Arbeitspunktwechsel) dar. Hierbei wäre das Modell unter Umständen nur phasenweise „gut“. Ein evolutionärer Algorithmus optimiert jedoch über den kompletten Mess- und Simulationsdatensatz. Die vorgeschlagenen Lösungen wären somit Pareto-optimal für den kompletten Zeitraum. Es gilt daher zu prüfen, ob es sinnvoll wäre den Zeitbereich in Optimierungs-Intervalle aufzuteilen und wie diese Intervalle sinnvoll gewählt werden können.

In einem weiteren Arbeitspaket planen die Autoren das AS für die Lebenszyklusphase „Umbau“ zu erweitern. Hier gilt es, neue, entfernte oder ersetzte Module zu identifizieren und die Co-Simulation entsprechend neu aufzubauen.

Literaturverweise

- [1] Dispan, J.: „Modulare Bauweise und neue Produktionskonzepte im Werkzeugmaschinenbau,“ Working Paper Forschungsförderung, No. 118, Hans-Böckler-Stiftung, Düsseldorf, 2019.
- [2] Marks, P.; Hoang, X. L.; Weyrich, M.; Fay, A.: A systematic approach for supporting the adaptation process of discrete manufacturing machines,“ in Research in Engineering Design 29, 621–641, Springer-Verlag London Ltd., 2018.
- [3] Ladiges, J.: „Automatisierte Bestimmung von Eigenschaften industrieller Produktionssysteme unter Einfluss evolutionärer Änderungen,“ Dissertation, Helmut-Schmidt-Universität / Universität der Bundeswehr Hamburg, Institut für Automatisierungstechnik, 2018.
- [4] Puntel Schmidt, P.: „Methoden zur simulationsbasierten Absicherung von Steuerungscode fertigungstechnischer Anlagen,“ Dissertation, Helmut-Schmidt-Universität / Universität der Bundeswehr Hamburg, Institut für Automatisierungstechnik, 2018.
- [5] Rose O., März L.: „Simulation,“ in: März L., Krug W., Rose O., Weigert G. (eds) „Simulation und Optimierung in Produktion und Logistik.“ VDI-Buch. Springer, Berlin, Heidelberg.
- [6] Scaramellini, S. et al.: „Parallel Simulation Techniques for Real-Time Field Management: An Innovative Integrated Production Optimization Tool Application,“ in Offshore Mediterranean Conference and Exhibition, Ravenna, Italy, 2017.
- [7] Svensson, B.; Danielsson, F.; Lennartson, B.: „Time-synchronised hardware-in-the-loop simulation - Applied to sheet-metal press line optimisation,“ in Control Engineering Practice 20(8), 792-804, 2012.
- [8] Christie, M. et al.: „Error analysis and simulations of complex phenomena,“ in Los Alamos Science 29, 6-25, 2005.
- [9] Härle, C.; Barth, M.; Fay, A.: „Simulationsmodellgenerierung im modularen Maschinen- und Anlagenbau - Assistenzsystem zur automatischen Komposition und Konfiguration von Co-Simulationen,“ in atp magazin 62(9), 72-79, 2020.
- [10] Härle, C.; Barth, M.; Fay, A.: „Assistance system for the automated composition and configuration of a Co-Simulation,“ in Proceedings of 34th European Simulation and Modelling Conference, 155-162, 2020.
- [11] Blochwitz, T. et al.: „The Functional Mockup Interface for Tool independent Exchange of Simulation Models,“ in Proceedings of the 8th International Modelica Conference, 105-114, 2011.

- [12] Drath, R.; Lüder, A.; Peschke, J.; Hundt, L.: „AutomationML – the glue for seamless Automation Engineering,“ IEEE International Conference on Emerging Technologies and Factory Automation, 2008.
- [13] Susto, G. A. et al.: „Machine Learning for Predictive Maintenance: A Multiple Classifier Approach,“ in IEEE Transactions on Industrial Informatics 11(3), 812-820, 2015.
- [14] Thyago P. et al.: „A systematic literature review of machine learning methods applied to predictive maintenance,“ Computers & Industrial Engineering 137, 2019.
- [15] Fahrmeir, L.; Künstler, R.; Pigeot, I.; Tutz, G.: „Statistik. Der Weg zur Datenanalyse,“ Springer Spektrum, Berlin/ Heidelberg, 2016.
- [16] Lehmann, T.: „Methodenbaukasten zur Quantifizierung der statistischen Güte und deren Sensitivität von Last- und Verschleißanalysen mit einem Beispiel im Kontext alternativer Antriebskonzepte,“ Dissertation, Fakultät Statistik, Technischen Universität Dortmund, 2020.
- [17] Spiess, Andrej-Nikolai, Natalie Neumeyer: „An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach,“ in BMC Pharmacology. 10: 6, 2010.
- [18] Van Veldhuizen, D. A.: „Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations,“ Dissertation, Graduate School of Engineering of the Air Force Institute of Technology, 1999.
- [19] Lampinen, J.: „Multiobjective Nonlinear Pareto-Optimization - A Pre-Investigation Report,“ Lappeenranta University of Technology, 2000.
- [20] Kleijnen, J.P.C.; Wan, J.: „Optimization of simulated systems: OptQuest and alternatives,“ in Simulation Modelling Practice and Theory, Volume 15, Issue 3, 354-362, 2007.
- [21] Chinchuluun, A.; Pardalos, P.M.: „A survey of recent developments in multiobjective optimization,“ in Annals of Operations Research 154, 29–50, 2007.
- [22] Burkard R.; Zimmermann U.: „Konvexe Funktionen,“ in „Einführung in die Mathematische Optimierung“. Springer-Lehrbuch, vol 5045. Springer, Berlin, Heidelberg, 2012.
- [23] Zitzler, E.; Thiele, L.: „Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach,“ in IEEE Transactions on Evolutionary Computation, vol. 3, no. 4, 257-271, 1999.
- [24] Deb K.: „Multi-Objective Evolutionary Algorithms,“ in Kacprzyk J., Pedrycz W. (Hrsg.) Springer Handbook of Computational Intelligence. Springer Handbooks. Springer, Berlin, Heidelberg, 2015.

- [25] Suman, B.: „Study of simulated annealing based algorithms for multiobjective optimization of a constrained problem,“ in Computers & chemical engineering, 28(9), 1849-1871, 2004.
- [26] Coello Coello, C. A.: „A Short Tutorial on Evolutionary Multiobjective Optimization,“ in International Conference on Evolutionary Multi-Criterion Optimization, 21-40, 2001.
- [27] MOEA Framework: <http://moeaframework.org/index.html>, [Letzter Aufruf: 12.04.2021].
- [28] Kollat, J.B.; Reed, P.M.: „Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design,“ in Advances in Water Resources 29(6), 792-807, 2006.
- [29] Siebertz K.; Bebbler D.; Hochkirchen T.: „Sensitivitätsanalyse,“ in: Statistische Versuchsplanung. VDI-Buch. Springer Vieweg, Berlin, Heidelberg, 2017.
- [30] Umbarkar, A.J.; Joshi, M.S.: „Review of parallel genetic algorithm based on computing paradigm and diversity in search space,“ in ICTACT Journal on Soft Computing 3(4), 615-622, 2013.

Lokalisierung von intelligenten Werkstückträgern innerhalb einer vollautomatisierten Fertigungsanlage mittels Bluetooth 5.1

Bastian Schulte, IWN GmbH & Co. KG, Bielefeld;
Harry Fast, Holger Flatt, Fraunhofer IOSB-INA, Lemgo

Kurzfassung

In diesem Beitrag wird ein auf Bluetooth 5.1 basierendes Low-Cost-Lokalisierungssystem in einer industriellen Umgebung hinsichtlich der Lokalisierungsgenauigkeit und Zuverlässigkeit evaluiert. Dafür wird die Position eines mobilen Werkstückträgers, welcher mit einem entsprechenden Bluetooth-Sender ausgestattet ist, innerhalb einer vollautomatisierten Fertigungsanlage bestimmt. Die Lokalisierung erfolgt dabei, anders als bei dem Bluetooth 5.0 Standard, via Triangulation. Zwecks Vergleichbarkeit mit hochperformaten aber auch kostenintensiven Lokalisierungssystemen werden Referenzmessungen mit einem, auf Ultrabreitband (=UWB) basierendem System durchgeführt. Die Messungen im industriellen Umfeld zeigen, dass hier mit Bluetooth 5.1 Lokalisierungsgenauigkeiten von bis zu 1,1 m möglich sind. Für Anwendungen im industriellen Kontext mit entsprechender Anforderung können somit kostengünstig Lokalisierungslösungen auf Basis der Bluetooth 5.1 Technologie umgesetzt werden.

1. Einleitung

Das Forschungsgebiet der Indoor-Lokalisierung hat in den vergangenen Jahren an Relevanz zugenommen. Für eine moderne, vollautomatisierte, modulare und flexible Fertigung sind Positionsdaten funktionsentscheidend [1], wie an dem folgenden Beispiel gezeigt wird. Neue Ansätze für Analysesysteme von Fertigungsanlagen im Bereich „Condition Monitoring“ sind mobile Werkstückträger, die die Anlage selbständig durchlaufen und Störungen erfassen sollen [2]. Um eine Störung einem Anlagenabschnitt zuordnen zu können, ist eine dauerhafte Positionsbestimmung des Werkstückträgers essenziell. Für die Indoor-Lokalisierung stehen bereits einige, auf Funk basierende Technologien zur Verfügung, die in Tabelle 1 dargestellt sind. Die Problematik der bisherigen Lösungen ist, dass hochpreisige Systeme zwar eine hohe Positionierungsgenauigkeit erreichen, günstigere Systeme erreichen jedoch nur eine vergleichsweise geringe Genauigkeit [3]. Mit der Vorstellung des neuen Bluetooth 5.1 Standards steht eine vielversprechende Lösung im Bereich der Indoor-Lokalisierung bereit, welche eine präzise Lokalisierung für Low-Cost Anwendungen verspricht. Vorgängerversionen des Bluetooth

Standards, wie z. B. die Version 5.0 bieten zwar ebenfalls die Möglichkeit einer Lokalisierung, jedoch wird hier die Lokalisierung über die Signalstärke realisiert [4]. Da die Signalstärken stark umgebungsabhängig gedämpft werden, ist eine Lokalisierungsgenauigkeit von 2-5 m möglich [4]. Für den hier genannten Anwendungsfall genügt diese Genauigkeit nicht [5]. Bei dem weiterentwickelten Bluetooth 5.1 Standard wird, zusätzlich zu der empfangenen Signalstärke, der Einfallswinkel des Sendersignals am Empfänger gemessen (= Direction Finding). Dadurch kann eine höhere Genauigkeit von bis zu 1 m bei der Lokalisierung in einem nicht industriellen Umfeld erreicht werden [6].

Tabelle 1: Vergleich funkbasierter Indoor-Lokalisierungstechnologien [1], [4], [6]

Kriterium		WLAN	BLE V5.0	BLE V5.1 Direction Finding	UWB	RFID
Genauigkeit		5-15 m	2-5 m	< 1 m	<= 0,3 m	0,5-2 m
Reichweite		< 150 m	< 75 m	< 75 m	< 150 m	Aktiv < 50 m Passiv < 1 m
Störanfälligkeit (durch Wände, Personen, Umgebungsbedingungen etc.)		hoch	hoch	hoch	gering	mittel
Energieversorgung		Aktiv (Akku)	Aktiv (Akku)	Aktiv (Akku)	Aktiv (Akku)	Aktiv (Akku) Passiv
Energiebedarf		mittel	gering	gering	gering bis mittel	gering
Anschaffungskosten	€	gering (wenn WLAN schon vorhanden)	mittel	mittel	hoch	mittel
Kosten mit steigender Teilnehmerzahl	€ 	gering	gering	gering	mittel	gering

Somit steht eine potenzielle, kostengünstige Lokalisierungslösung für industrielle Anwendungen mit hohen Anforderungen an die Lokalisierungsgenauigkeit bereit [6]. In [3][4][7] wurden bereits Performance-Messungen für die Bluetooth 5.1 Lokalisierungstechnologie für eine Büroumgebung und eine Sporthalle umgesetzt. Da sich das industrielle Umfeld jedoch grundlegend von diesen unterscheidet (Abschirmung, Material) wird in diesem Beitrag das Potential einer Lokalisierungslösung im industriellen Kontext auf der Basis eines Bluetooth 5.1 Lokalisierungssystems untersucht. Parallel dazu werden Referenzmessungen mit einem UWB-Lokalisierungssystem durchgeführt. Dieser Beitrag ist wie folgt gegliedert: In Abschnitt 2 wird die Funktionsweise eines Bluetooth 5.1 Lokalisierungssystems mit Direction Finding beschrieben. Darauf folgend werden in Abschnitt 3 das Messkonzept und die Anforderung an das System beschrieben. In Abschnitt 4 wird die Messdurchführung an zwei Fallstudien beschrieben. Die Ergebnisse aus den Messungen werden in Abschnitt 5 vorgestellt und diskutiert. Abgeschlossen wird dieser Beitrag mit dem Abschnitt 6, in dem eine Zusammenfassung und ein Ausblick auf spätere Arbeiten gegeben wird.

2. Lokalisierung via Bluetooth 5.1

Die Vorgängerversionen von Bluetooth 5.1 messen Distanzen zwischen einem Sender (Beacon) und mehreren Empfängern über die Signalstärke, Received Signal Strength (= RSS), des Beacons. Durch die Messungen der eingehenden Signalstärken wird der Abstand zwischen dem Beacon und den Empfängern bestimmt. So können Kreise für die potenzielle Position um jeden Empfänger aufgespannt werden. Über die Berechnung der Schnittstellen der Kreise von mindestens drei Empfängern, kann eine Lokalisierung mit einer Genauigkeit von 2-5 m erreicht werden [4]. Durch die Weiterentwicklung des Bluetooth Standard (Bluetooth 5.1) wird der Winkel (= Angle of Arrival (AoA), siehe Bild 1 (a)) des eingehenden Signals bestimmt. Somit wird die Richtung der Position des Senders bezogen auf den Standort des Empfängers angegeben.

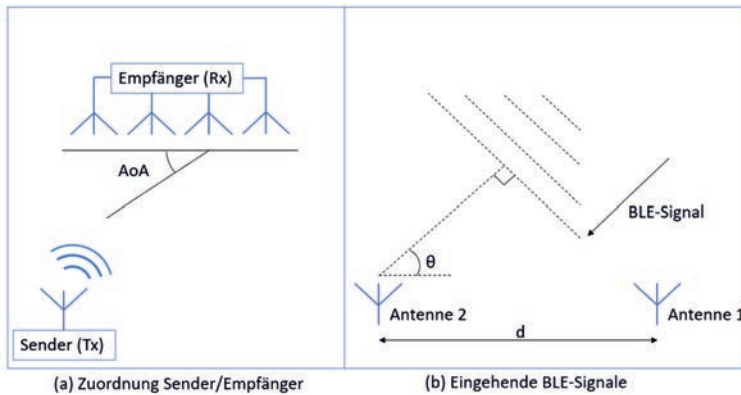


Bild 1: Schematische Darstellung der AoA Methode, adaptiert von [5]

Die Ermittlung des AoA wird über ein Antennenarray realisiert. Dafür sind auf dem Signalempfänger mehrere Antennen mit gleichem Abstand d , siehe Bild 2 (b), montiert. In ausreichendem Abstand zum Bluetooth-Sender kann die elektromagnetische Welle des Bluetooth-Signals als ebene Welle approximiert werden. Diese ebene elektromagnetische Welle trifft nun zeitversetzt auf die einzelnen Antennen des Antennenarrays (siehe Bild 2 (b)). Nach Auswertung der Phasendifferenz kann über den, in Gleichung (1), dargestellten Zusammenhang der AoA θ berechnet werden. In die Berechnung geht die Phasendifferenz ψ , der Antennenabstand d und die Wellenlänge λ ein [5].

$$\theta = \arcsin\left(\frac{\psi\lambda}{2\pi d}\right) \quad (1)$$

Unter Zuhilfenahme mehrerer Empfänger (ausgeführt als Antennenarrays) lassen sich aus den ermittelten Winkeln über Triangulation die Position des Senders im Verhältnis zu den einzelnen Empfängern berechnen. Für die Triangulation müssen zwischen drei verschiedenen Fällen unterschieden werden (siehe Bild 2) [3]. Dabei ist entscheidend, an welcher Position sich der Sender im Verhältnis zu den Empfängern befindet. So können (für den 2-dimensionalen Raum) aus zwei benachbarten Empfängern die entsprechenden Entfernungen zwischen Sender und Empfänger berechnet werden.

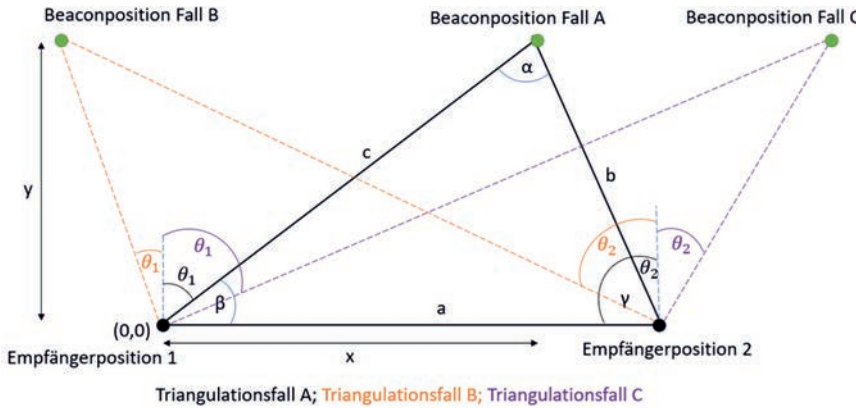


Bild 2: Schematische Darstellung der Triangulation, adaptiert von [3]

Bezogen auf den Nullpunkt (0,0), welcher der Empfängerposition 1 entspricht, kann der direkte Abstand c zwischen dem jeweiligen Sender (grünfarbener Punkt) und dem Nullpunkt für den Triangulationsfall A berechnet werden (siehe Gleichung 2).

$$c = \frac{a \cdot \sin(\gamma)}{\sin(\alpha)} = \frac{a \cdot \sin(90 - |\theta_2|)}{\sin(180 - (90 - |\theta_1|) - (90 - |\theta_2|))} \tag{2}$$

Die x - und y -Entfernungen im Verhältnis zu den Antennenpositionen (schwarzfarbene Punkte) werden daraus folgend mit den Gleichungen 3 und 4 berechnet.

$$x = c \cdot \cos(90 - |\theta_1|) \tag{3}$$

$$y = c \cdot \sin(90 - |\theta_1|) \tag{4}$$

Die Berechnung der beiden anderen Triangulationsfälle B und C sind in Tabelle 2 dargestellt.

Tabelle 2: Berechnung der Triangulationsfälle B und C

Triangulationsfall B	Triangulationsfall C
$c = \frac{a \cdot \sin(90 - \theta_2)}{\sin(180 - (90 + \theta_1) - (90 - \theta_2))}$	$b = \frac{a \cdot \sin(90 - \theta_2)}{\sin(180 - (90 - \theta_1) - (90 + \theta_2))}$
$x = -c \cdot \cos(90 - \theta_1)$	$x = b \cdot \cos(90 - \theta_2) + a$
$y = c \cdot \sin(90 - \theta_1)$	$y = b \cdot \sin(90 - \theta_2)$

3. Entwicklung eines Messkonzepts

Zur Umsetzung der Messungen gilt es im ersten Zuge ein Konzept zu entwickeln. Es werden mit zwei verschiedenen Lokalisierungssystemen, einem UWB-System und einem BLE 5.1-System, statische Versuche in einer industrieähnlichen Umgebung durchgeführt. So kann eine direkte Vergleichbarkeit in der Lokalisierungsgenauigkeit zwischen den beiden Systemen hergestellt werden. Parallel dazu werden weitere unbekannte Faktoren auf ihre Beeinflussung des Lokalisierungsergebnisses des BLE 5.1-Systems untersucht. Dazu zählen zum einen die Abstände zwischen den Antennenarraypositionen, die direkt das Positionierungsergebnis beeinflussen (siehe Gleichung 2). Zum anderen gilt es zu betrachten, inwiefern die Anzahl der möglichen Triangulationsfälle und somit die Anzahl der Antennenarraypositionen, an denen Winkelwerte aufgenommen werden, das Ergebnis des BLE 5.1-Systems beeinträchtigen. In weiteren Versuchen wird die Messumgebung variiert. Dazu werden die Messungen in einer realen Produktionsumgebung wiederholt, um den industriellen Einfluss auf die Messergebnisse des BLE 5.1-Systems zu analysieren.

4. Fallstudien

Als Umgebungen für die Lokalisierungsversuche werden die industrieähnliche Umgebung in der SmartFactoryOWL, in der ein UWB-System des Unternehmens „Kinexon“ als Referenz fest installiert ist, und die reale industrielle Umgebung bei dem Unternehmen IWN GmbH & Co. KG ausgewählt. Im Zuge der Untersuchungen in der SmartFactoryOWL wird das Beacon auf einen Werkstückträger in einer verketteten Produktionsanlage montiert und an mehreren bekannten Punkten mit Hilfe des UWB-Systems (= Versuch 1) und dem BLE 5.1-System statisch bestimmt. Der Messaufbau des UWB-Referenzsystems ist ähnlich wie der des BLE-Systems in Bild 3. Der UWB-Signalsender (= Tag) wird ebenfalls auf den Werkstückträger montiert. Der Tag sendet seine aktuellen Positionsdaten an die entsprechenden Empfänger (= Anker).

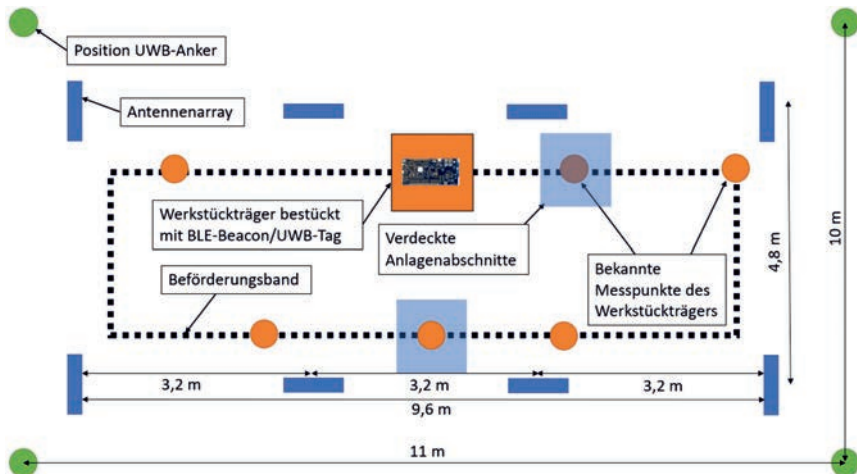


Bild 3: Schematische Darstellung des Messaufbaus

Da während den Versuchen nur ein BLE-Antennenarray zur Verfügung steht, werden an der jeweiligen Position des Arrays die Winkelwerte aufgenommen und anschließend wird das Antennenarray an eine neue Position (siehe Bild 3) gestellt. Dazu ist das Antennenarray auf ein Stativ montiert und somit mobil versetzbar (siehe Bild 4). Aufgrund der Mobilität des BLE-Antennenarrays werden verschiedene Messaufbauten gewählt.

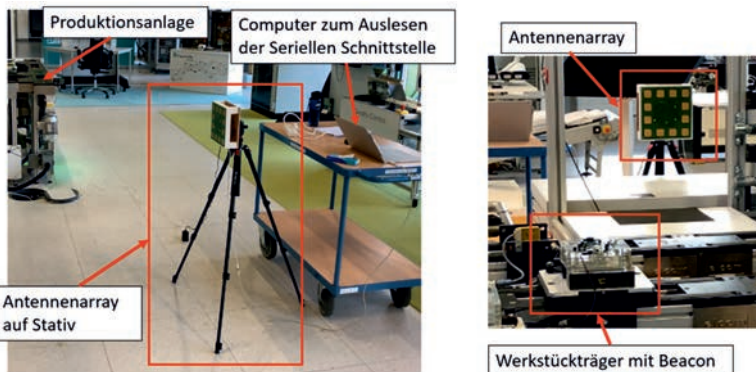


Bild 4 : Messaufbau des BLE-Systems

Der Unterschied in den Messaufbauten liegt zum einen in den Abständen zwischen den Antennenarraypositionen und zum anderen in der Anzahl der Antennenarraypositionen, an denen Winkelwerte aufgenommen werden. Für die ersten Messungen des BLE-Systems werden die Antennenarraypositionen direkt unterhalb der Kinexon-Anker platziert (= Versuch 2). Somit wird eine Vergleichbarkeit beider Systeme hergestellt. In einem zweiten Versuch wird der Abstand zwischen den Antennenarraypositionen des BLE-Systems reduziert (siehe Bild 3), da dieser die Positionsberechnung direkt beeinflusst (siehe Gleichung 2). Außerdem werden nicht nur an vier Antennenarraypositionen Winkelwerte aufgenommen (= Versuch 3), sondern auch an acht Antennenarraypositionen (= Versuch 4). So kann der Einfluss der Anzahl der Antennenarrays auf das Lokalisierungsergebnis analysiert werden. Für den Versuch 5 werden Messungen mit den entsprechenden Abständen und acht Antennenarraypositionen (siehe Bild 3) bei dem Unternehmen IWN GmbH & Co. KG durchgeführt. Aufgrund der festen Installation des UWB-Systems in der SmartFactoryOWL können nur mit dem BLE-System Messungen durchgeführt werden. Da sich hier keine vergleichbare Produktionsanlage wie in der SmartFactoryOWL befindet, wird das Beacon an einer festen Position neben einem Bearbeitungszentrum (= Zerspanungsmaschine) montiert. Die Bedingungen, die innerhalb der Maschine herrschen, erlauben keine Montage des Beacons innerhalb der Anlage. Es werden Messungen sowohl bei laufender, als auch bei Maschinenstillstand durchgeführt. Da in diesem Umfeld reale Produktionsbedingungen herrschen, kann der entsprechende Einfluss dieser Umgebung auf das Lokalisierungsergebnis analysiert werden.

5. Ergebnisse

Um eine bessere Vergleichbarkeit der jeweiligen Positionsabweichung zu gewährleisten, wird der direkte Abstand Δc zwischen Soll- und Ist-Position ermittelt (siehe Gleichung (5)).

$$\Delta c = \sqrt{\Delta x^2 + \Delta y^2} \quad (5)$$

Da die Position innerhalb einer Messreihe mehrmals bestimmt wird, wird der Median (robust gegenüber Ausreißern) über die jeweiligen Abweichungen ermittelt, um so eine Gesamtabweichung über die gesamte Messreihe bestimmen zu können. In Tabelle 3 sind die berechneten Abweichungen für die einzelnen Versuche in den verschiedenen Messumgebungen dargestellt. Dabei zeigt sich, dass die Genauigkeit des UWB-Systems bei 0,3 m liegt (= Versuch 1). Für den Versuch 2, bei dem sich die BLE 5.1 Empfänger unter den UWB-Ankern befinden, liegt die Genauigkeit bei 2,3 m. Wird der Abstand der Antennenarraypositionen zueinander reduziert und vier Empfängerpositionen hinzugezogen, kann eine Genauigkeit von 1,8 m erreicht werden (= Versuch 3). Unter Berücksichtigung von acht Empfängerpositionen verbessert sich die Genauigkeit auf 1,1 m (= Versuch 4). Die erzielten Genauigkeiten der Messreihen

im Unternehmen IWN GmbH & Co. KG (= Versuch 5) sind ebenfalls in der Tabelle 3 aufgezeigt. Hier bleibt das Lokalisierungsergebnis im Vergleich zu Versuch 4 bei Maschinenstillstand mit 1,1 m auf gleichem Niveau. Durch das Betreiben der Maschine wird das Lokalisierungsergebnis auf 1,9 m verschlechtert. Somit hat das Betreiben einer zerspanenden Fertigungsanlage einen negativen Einfluss auf das BLE-System.

Tabelle 3: Messergebnisse

Messumgebung SmartFactoryOWL				
	UWB-System Versuch 1	BLE-System		
		Unter Kinexon Anker Versuch 2	4-AP Versuch 3	8-AP Versuch 4
Durchschnittliche Genauigkeit [m]	0,3	2,3	1,8	1,1
Messumgebung IWN GmbH & Co. KG				
BLE-System-Versuch 5				
Durchschnittliche Genauigkeit [m]	Maschinenstillstand-8 AP		Laufende Maschine-8 AP	
	1,1		1,9	

Zusammenfassend zeigt sich, dass die Verwendung von acht Empfängerpositionen und die Verkleinerung der Abstände zwischen den Empfängerpositionen eine Verbesserung in der Positionsbestimmung ermöglichen. Weiterhin zeigt sich die hohe Störanfälligkeit des BLE-Systems gegenüber laufender Fertigungsmaschinen. Somit kann ein Lokalisierungssystem für Anwendungsfälle, bei denen die nötige Lokisierungs-genauigkeit im Bereich > 1 m liegt, mit Low-Cost-Komponenten und Bluetooth 5.1 Technologie in einer industriellen Umgebung eingesetzt werden. Eine mögliche Anwendung hierfür ist beispielsweise das Tracking von Warenströmen entlang der Produktion.

6. Zusammenfassung und Ausblick

In dem vorliegenden Beitrag wird ein auf Bluetooth V5.1 basierendes, prototypisches Low-Cost-Indoor-Lokalisierungssystem aufgebaut und in einer industriellen Umgebung evaluiert. Dafür werden Messungen in einer industrieähnlichen Umgebung und einer realen Industrieumgebung unter Variation von Abständen, Antennenanzahl und Betriebszustand von benachbarten Anlagen durchgeführt. Zwecks Vergleich mit einem kommerziellen Indoor-Lokalisierungssystem werden Messungen mit einem UWB-Indoor-Lokalisierungssystem, welches eine Genauigkeit von 0,3 m liefert, herangezogen. Die Auswertung der Messungen ergeben, dass die Positioniergenauigkeit des Bluetooth 5.1-Lokalisierungssystems in einer industriellen

Umgebung eine Lokalisierung von 1,1 m ermöglicht, wenn benachbarte Maschinen keine Signalstörungen erzeugen, ein geringer Abstand (siehe Bild 3) zwischen den Signalempfängern herrscht und an acht Empfängerpositionen Werte gemessen werden. Eine Verschlechterung der Positioniergenauigkeit auf 1,9 m wird bei Messungen nahe von laufenden Fertigungsmaschinen festgestellt. Eine Genauigkeit wie in büroähnlicher Umgebung von < 1 m sind in der Industrie nicht realisierbar. Unter den oben genannten Voraussetzungen kann somit ein Low-Cost-Lokalisierungssystem im industriellen Umfeld mit Genauigkeiten von > 1 m umgesetzt werden. In nachfolgenden Arbeiten gilt es die Performance des Bluetooth 5.1-Lokalisierungssystems für den dynamischen Fall zu untersuchen, da dieser Anwendungsfall häufig in der Industrie vorzufinden ist (z.B. bewegte Werkstückträger, autonome Fahrzeuge, etc.). Des Weiteren soll untersucht werden inwiefern Sensordatenfusion z.B. mit Inertialsensorik (Gyrometer, Beschleunigungsmessung) zu einer verbesserten Lokalisierungsgenauigkeit führen.

7. Literatur

- [1] Mautz, R. (2012): Indoor Positioning Technologies, ETH Zurich Habilitation, 2012
- [2] Fast, H., Schneider, D., Flatt, H., Jasperneite, J., Stübbe, O.: Drahtlose Sensordatenübertragung für die mobile Schwingungsanalyse in Produktionsanlagen, Kommunikation in der Automation (KommA 2019) Magdeburg, Nov 2019
- [3] Andersson, P., Persson, L. (2020): Evaluation of Bluetooth 5.1 as an Indoor Positioning System, KTH Sweden Master's Degree, 2020
- [4] Cominelli, Patras, Gringoli (2019): Dead on Arrival: An Empirical Study of the Bluetooth 5.1 Positioning System, 2019
- [5] Woolley, M. (2019): Bluetooth Direction Finding: A Technical Overview, 2019
- [6] Sambu, P., Won, M (2021): An Experimental Study on Direction Finding of Bluetooth 5.1: Indoor vs Outdoor, 06.2021, Connected Smart Sensor System Lab University of Memphis, TN
- [7] Sesma, V., Egorov, V. (2020): Improved Accuracy for Indoor Positioning with Bluetooth 5.1: From Theory to Measurements, LTH Sweden Master's Degree, 2020
- [8] Atteya, K., Amer, M. (2019): Indoor positioning: Bluetooth Angle of Arrival, LTH Sweden Master's Degree, 2019
- [9] Agatonovic, M., Stankovic, Z., Doncov, N., Sit, L., Milovanovic, B., Zwick, T. (2012): Application of artificial neural networks for efficient high-resolution 2d doa estimation, Radiengineering vol. 2012, S. 1178-1186

Compression and Fault Diagnosis of High Frequency Sensor Data

Dr. I. Amihai, Dr. R. Gitzel, Dr.-Ing. A. Boyaci, ABB, Ladenburg

Kurzfassung

Daten aus der echten Welt wie medizinische Scans, Audiosignale und industrielle Sensordaten sind häufig hochdimensional und müssen in eine komprimierte aussagekräftige Darstellung umgewandelt werden, bevor sie für maschinelles Lernen und statistische Anwendungen verwendet werden können. Ein solcher Prozess wird als Dimensionsreduktion (DR) bezeichnet und kann die Klassifizierung, Visualisierung und Komprimierung hochdimensionaler Daten erleichtern und den Fluch der Dimensionalität mildern. Es wird traditionell mit linearen Methoden wie der Principal Component Analysis (PCA) durchgeführt, die relativ einfach zu implementieren sind, für komplexe nichtlineare Daten jedoch unzureichend sein können.

In diesem Artikel haben wir verschiedene DR-Methoden für Beschleunigungsdaten untersucht, die von elektrischen Schaltgeräten erfasst wurden. Solche Daten werden häufig zur Überwachung des Zustands elektrischer Geräte verwendet, aber ihre hohe Dimensionalität macht es schwierig, sie für Anwendungen des maschinellen Lernens zu verwenden, vor allem wenn relativ wenige Beispiele verfügbar sind. Insbesondere haben wir lineare DR unter Verwendung von PCA mit mehreren nichtlinearen DR-Methoden verglichen: PCA mit einer radialen Basisfunktion (RBF) und einem Polynomkern und einem Faltungsautoencoder. Darüber hinaus haben wir eine SDR-Methode (Stepwise Dimensionality Reduction) getestet, bei der das Clustering bei einer anfänglichen Codierung der Daten durchgeführt wird, gefolgt von einer zweiten Codierung der Clusterdaten.

Anschließend haben wir das DR-Modell mit der besten Leistung verwendet, um eine Methode zur Erkennung von Fehlern in elektrischen Geräten zu entwickeln. Da Modelle für maschinelles Lernen natürlich eine bessere Leistung erbringen, wenn ihre Eingabe den Trainingsdaten ähnelt, haben wir das Modell auf gesunde Asset-Messungen trainiert und es verwendet, um gesunde und fehlerhafte Messungen zu komprimieren und zu rekonstruieren, die während des Trainings nicht verwendet wurden. Wie wir vermuteten, erzeugte unser Modell beim Komprimieren und Rekonstruieren zusätzlicher gesunder Sensordaten im Vergleich zu fehlerhaften Sensordaten viel kleinere Fehler. Diese Fehlerwerte wurden anschließend als Eingaben in ein unbeaufsichtigtes maschinelles Lernmodell verwendet, das gesunde und fehlerhafte Beispiele korrekt unterschied.

Abstract

Dimensionality reduction can provide many benefits for machine learning algorithms, can be applied to any kind of signal, and is especially relevant for high-dimensional data. In this paper, we tested several dimensionality reduction methods to encode accelerometer signals generated from circuit breaker drives sampled at a high temporal resolution. We compared several PCA methods, a convolutional-autoencoder, and a novel algorithm based on the combination of a clustering algorithm and autoencoders, which we termed stepwise dimensionality reduction. Autoencoder methods achieved a significantly lower reconstruction error in comparison to PCA methods, i.e. enabled better compression of the acceleration signals. Secondly, we used the best performing model to construct a highly accurate condition monitoring method which uses the reconstruction errors of the model to predict failures in electrical equipment.

1. Introduction

Real-world data, such as medical scans, audio signals, and industrial sensor data are often high-dimensional and need to be transformed into a compressed meaningful representation before they can be used in machine learning and statistical applications. Dimensionality reduction (DR) can facilitate classification, visualization, and compression of high dimensional data, and mitigates the curse of dimensionality. It is traditionally performed using linear methods like Principal Components Analysis (PCA), which are relatively easy to implement but can be inadequate for complex nonlinear data.

In this paper, we explored several DR methods for accelerometer data acquired from circuit breaker (CB) drives. Such data is often used for monitoring the health of CBs, but its high dimensionality makes it difficult to utilize for machine learning applications, particularly when there are relatively few examples available. Specifically, we compared linear dimensionality reduction using PCA to several non-linear methods: PCA with a radial basis function (RBF) and a polynomial kernel, and a convolutional autoencoder. Moreover, we developed a new method, which we termed "stepwise dimensionality reduction" (SDR), which is described in the text. Briefly, clustering is performed on an initial encoding of the data, followed by a second encoding of the clustered data (see Algorithm 1). We hypothesized that since our data consists of two types of breaker drive operations (opening and closing), it would be easy to generate well defined clusters of the encoded representations. We further hypothesized that this would lead to superior performance of SDR since it trains autoencoders that learn to represent specific clusters of data.

Secondly, we used the best performing DR algorithm to construct a model for detecting failures in electrical equipment. Specifically, we trained an autoencoder model on healthy asset measurements and used it to compress and reconstruct healthy and faulty measurements

not used for training. Since machine learning algorithms would naturally perform better when their input resembles their training data, we hypothesized that a model trained on healthy sensor data would generate lower errors when compressing and reconstructing additional healthy sensor data in comparison to faulty sensor data. An autoencoder was chosen for this task as it performed better than other methods on our first task.

2. Methods

A. Dimensionality Reduction

If we consider a dataset $X = \{x_1, x_2, \dots, x_N\}$, $x_n \in R^D$, we try to find a lower-dimensional compressed representation of x_n , $z_n = B^T x_n \in R^M$, so that $M < D$, and where we define the projection matrix: $B := [b_1, b_2, \dots, b_M]$. The task is to find projections $\tilde{x}_n$ that are as similar as possible to the original data x_n , e.g. by minimizing $\|x_n - \tilde{x}_n\|^2$.

B. Principal Component Analysis

In PCA we assume that the columns of B are orthonormal; i.e. $b_i^T b_j = 0$ if and only if $i \neq j$ and $b_i^T b_i = 1$, and we assume a linear relationship between the original data x_n and its low-dimensional code z_n so that $z_n = B^T x_n$ and $\tilde{x}_n = B z_n$. B^T can be thought of as an encoder, which encodes the original data x_n as a low-dimensional (compressed) code z_n , that minimizes the error $\|x_n - \tilde{x}_n\|^2$. PCA can be performed via an Eigendecomposition of the data covariance matrix and projecting the data onto the principal subspace. A reconstruction of the data is obtained by projecting the compressed data back onto the original data space. Kernel PCA is based on the "kernel-trick", which allows for the computation of inner products in the feature space instead of the data space, and thus enables performing PCA with non-linear transformations.

C. Autoencoders

An autoencoder is a type of neural network that consists of an encoder, which learns a lower dimensional or sparse representation of the data, and a decoder that reconstructs the original data from its encoded representation. As with PCA, an autoencoder learns to minimize the error between the original and reconstructed data. Autoencoders can be constructed using many types of neural network architectures. The simplest form is of a feed-forward neural network with one or more hidden layers between the input and output layers, and where the input and output layers have the same number of parameters. A convolutional autoencoder is an autoencoder that incorporates convolutional, and usually also pooling, operations.

D. Stepwise dimensionality reduction algorithm

The stepwise algorithm (see Algorithm 1) incorporates an initial "stage 1" encoder, followed by a K-Means algorithm that clusters the encoded data, and then additional "stage 2"

encoders are trained on each cluster of data. During inference, the K-Means algorithm is used to first assign cluster identities to the test data stage 1 encodings, and then the stage 2 auto-encoders are used to compress and reconstruct the test data from each cluster.

E. Algorithm 1:

1. Train a stage 1 encoder (e.g. an autoencoder or PCA) and generate encodings for the training and test datasets (and the validation set).
2. Fit a clustering algorithm (e.g. K-Means) on the encodings of the training dataset.
3. Train stage 2 encoders separately on training examples that were assigned to each cluster.
4. Compute the cluster assignments of the test data.
5. Use the stage 2 encoders to encode the test data in each cluster.

2. Experiment 1

A. Data

The data contained 118 accelerometer samples recorded during opening and closing breaker drive operations, where each sample contained 8750 acceleration samples that were down-sampled to 1750 samples using a Fourier resampling method (for specifications, see).

B. Model Specifications

We used the Scikit-Learn Python package to generate PCA models with a linear (i.e. no-kernel), RBF, and polynomial kernels of 2 and 3 degrees. Furthermore, we used the PyTorch Python package to build and train the autoencoders. The convolutional autoencoder was comprised of two convolutional “blocks” that consisted of a 1-dimensional convolutional layer with 32 3x1 filters with a stride of 1, followed by a ReLU activation function, batch-normalization, and a max-pooling operation with a 2x1 filter size and a stride of 2. Finally, a linear transformation layer was used to generate encodings of the appropriate dimension. The decoder consisted of a single convolutional block, followed by a max-unpool layer, which computes the partial inverse of a max-pool layer by setting all non-maximal values to zero, and a 1-dimensional convolutional layer with 32 input filters and a single output filter to generate the decoded reconstructions of the inputs.

The autoencoder was trained with a constant learning rate of 0.1, which was arrived at via manual experimentation using the validation set. The mean squared error (MSE) loss function was used. The autoencoder was trained for 1000 epochs, and the parameters during the epoch that generated the lowest loss on the validation set were saved for inference. The SDR algorithm was composed of a K-Means algorithm, where the number of clusters was set to 2 (equal to the number of breaker drive operations we had in the data), and three convolutional autoencoders, one for stage 1 and two for stage 2. These autoencoders were parameterized

in the same way as the convolutional autoencoder described above. Convolutional autoencoders were used for SDR since they achieved the lowest reconstruction error in our dimensionality reduction task (see Experiment 1 Results).

C. Task

The modeling task was to compare the reconstructions of the data using the models described in the previous section. Specifically, we reduced the dimension of the data to 1-4 dimensions and then back to the original data space and compared the reconstruction errors of the models. The reconstruction errors were compared after each model was used to compress the data using the same encoding size. 80% of the samples were used for training, and the remaining 20% were equally split between validation and testing data. The root MSE (RMSE) between the reconstructed and original samples was used as the dependent variable, and was analyzed using a 6 x 4 repeated measures ANOVA with Model (PCA, PCA with an RBF kernel, PCA with a 2-degree polynomial kernel, PCA with a 3-degree polynomial kernel, convolutional autoencoder, SDR) as a between samples factor and encoding dimension (1,2,3,4) as a within samples factor. Each condition was performed 10 times.

D. Reconstruction Performance

The RMSE of all the experimental conditions are displayed in Table 1. The ANOVA showed a statistically significant main effect of Model ($F(5,4) = 555.1$, $p < 0.0001$, see Table 1). The effect of Dimension was not significant ($F(3,12) = 0.809$, $p > 0.49$), and the Model x Dimension interaction was also not statistically significant ($F(15,12) = 0.625$, $p > 0.85$). Additional pairwise comparisons showed that the RMSE of the convolutional autoencoder and SDR were significantly lower than those of the PCA-based models ($p < 1e-28$ for all comparisons, see Fig. 1). Moreover, the RMSE of the convolutional autoencoder was significantly lower than that of SDR ($p < 1e-5$). The RMSE generated by the PCA-based models did not significantly differ.

E. Clustering Performance of SDR

In all conditions, there was a perfect correspondence between the cluster assignments and the breaker drive operation on the training and test datasets (see Fig. 2 for an example in 2 dimensions).

TABLE 1: MEAN RMSE OF ALL CONDITIONS, WITH STANDARD ERRORS DISPLAYED IN PARENTHESIS.

Algorithm Reconstruction Errors (RMSE)				
	1-dim	2-dim	3-dim	4-dim
PCA	10.01 (0.1)	9.88 (0.12)	9.74 (0.12)	9.65 (0.12)
PCA: RBF Kernel	9.85 (0.09)	9.82 (0.1)	9.79 (0.1)	9.76 (0.1)
PCA: Polynomial Kernel, degree = 2	10.05 (0.11)	9.89 (0.11)	9.8 (0.12)	9.76 (0.11)
PCA: Polynomial Kernel, degree = 3	10.02 (0.11)	10.01 (0.11)	10.09 (0.17)	9.99 (0.13)
Convolutional Autoencoder	4.02 (0.08)	4.22 (0.08)	4.24 (0.12)	4.25 (0.11)
SDR	5.85 (0.55)	4.98 (0.33)	5.15 (0.51)	5.42 (0.43)

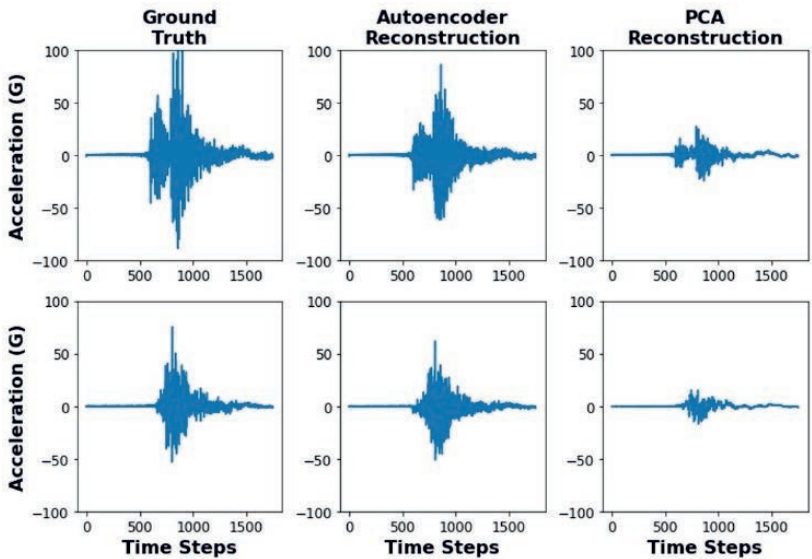


Fig. 1: Examples of autoencoder and PCA reconstructions.

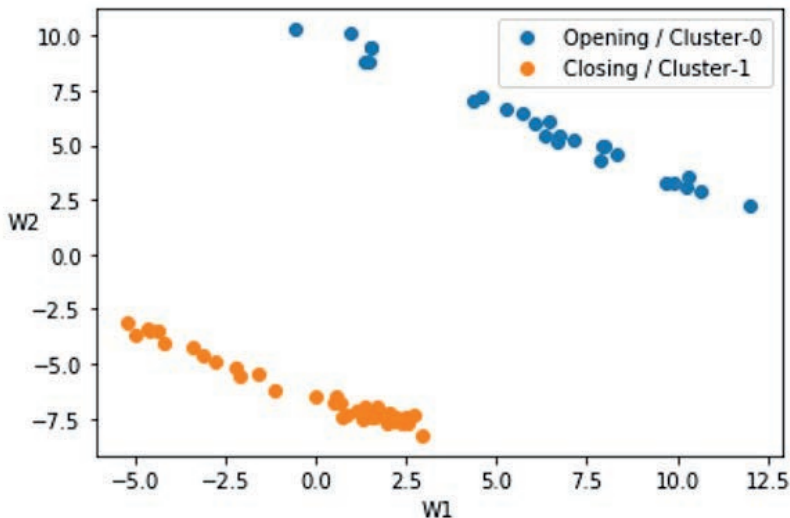


Fig. 2: K-Means clustering of 2-dimensional encodings of the test-dataset. W1 and W2 indicate the encoding parameter values. The opening and closing operation samples were always assigned to different clusters, regardless of the dimensionality of the encodings.

3. Experiment 2

A. Data

Data from an additional electrical asset were recorded and contained 160 accelerometer recordings generated during healthy operations and 30 samples recorded during operations in which the asset was faulty. Each recording contained 4000 acceleration samples. Of the 160 healthy samples, 100 were randomly selected for training, and the remaining 60 were equally split between the validation and test sets. All the 30 faulty samples were assigned to the test set.

B. Model Specifications and Task

The convolutional autoencoder was chosen for this experiment, as it was the best performing model in Experiment 1. The modeling task was to encode and reconstruct the training data. The same parameterization and training procedures used in Experiment 1 were used in Experiment 2.

C. Results

The reconstruction error, as measured by the RMSE, was always lower for healthy test samples relatively to faulty samples (see Fig. 3). Furthermore, using a K-Means clustering algorithm allowed us to completely distinguish healthy and faulty samples. When the RMSE of all samples were visualized, 3 distinct clusters were clearly visible, hence we set $k=3$. The algorithm successfully grouped all the healthy samples into one cluster and formed 2 clusters consisting entirely of faulty data (see Fig. 4).

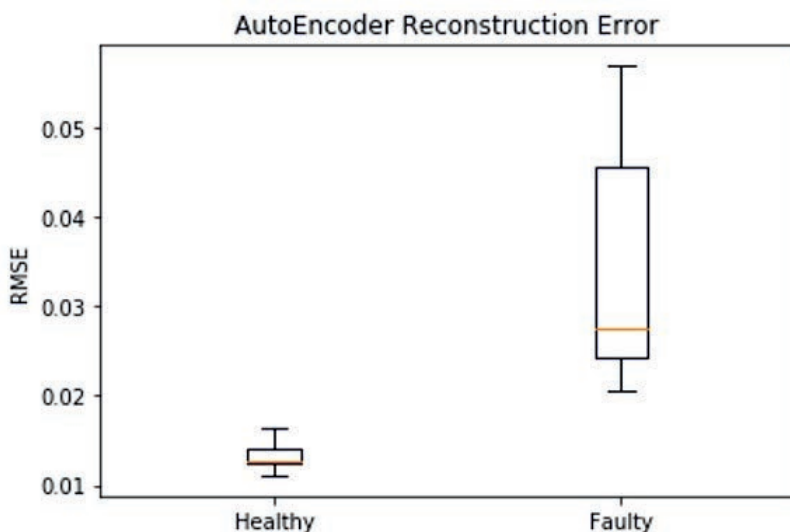


Fig 3: Reconstruction errors for healthy and faulty breaker drive data samples on the test set.

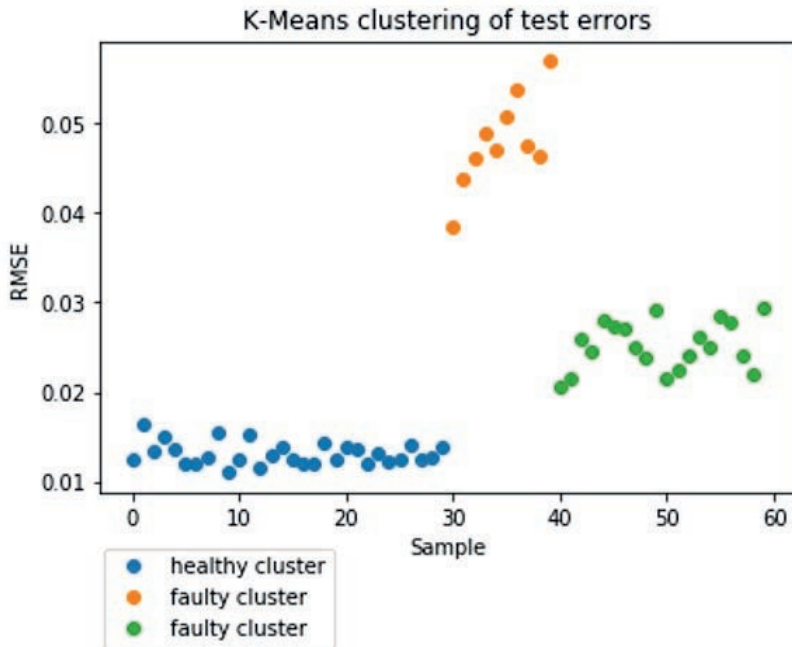


Fig 4: K-Means clustering of test sample RMSE.

4. Discussion

In this paper we first assessed the efficacy of different dimensionality reduction methods for high frequency acceleration signals. In this regard, autoencoder based methods, including the newly introduced SDR, demonstrated superior performance on this data in comparison to PCA based methods. However contrary to our hypothesis, the single convolutional autoencoder performed better than SDR, which suggests that our dataset may have been too small to benefit from the training of separate algorithms for different subsets of data. Consistent with this hypothesis, the variance of SDR was higher than that of the other models (see Table 1), which suggests that it was overfitting to the smaller data used for training each of its sub-models. It is also possible that even though the clusters used to train SDR were well defined, the encoded representations were largely based on features that were contained in both clusters of data, which allowed even a single convolutional autoencoder to generalize well. Thus, in future research it would be useful to compare the SDR to the single convolutional autoencoder approach on larger datasets.

Secondly, we tested the capability of using dimensionality reduction as a tool for diagnosing faults in electrical equipment. Specifically, we trained a convolutional autoencoder to

encode and reconstruct healthy signals, and then tested it using both healthy and faulty data. Subsequently, we used a K-Means algorithm to cluster the sample errors, and the algorithm was able to form clusters comprised entirely of healthy or faulty measurements. Hence, we may conclude that this method can indeed be used to classify certain electrical asset failures, although further research should compare different failure types to assess the generalizability of the method.

In previous studies, autoencoders have been successfully used for fault diagnosis in various industrial settings, primarily as algorithms for anomaly detection, i.e. the detection of rare events, or outliers, that are likely to represent faults. Like our current approach, such systems involve training the autoencoder mainly on healthy data, and high reconstruction errors were used as indicators of faults. Such autoencoder networks have been successfully applied to detect errors in gas turbine combustors and rotating machines. Here we used a similar approach to show that autoencoder networks could be used to detect faults using accelerometer data generated from electrical equipment, as faulty and healthy samples generated distinct clusters of data based on their reconstruction errors.

REFERENCES

- [1] L. J. van der Maaten, E. O. Postma and H. J. van den Herik, "Dimensionality reduction: A comparative review," *Journal of Machine Learning Research*, vol. 10, no. 1-41, pp. 66-71, 2009.
- [2] L. Jimenez and D. Landgrebe, "Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 28, no. 1, pp. 39-54, 1997.
- [3] H. K. Høidalen and M. Runde, "Continuous monitoring of circuit breakers using vibration analysis," *IEEE Transactions on Power Delivery*, vol. 20, no. 4, pp. 2458-2465, 2005.
- [4] M. Hoffmann, S. Wildermuth, R. Gitzel, A. Boyaci, J. Gebhardt, H. Kaul, I. Amihai, B. Forg, M. Suriyah, T. Leibfried, V. Stich, J. Hicking, M. Bremer, L. Kaminski, D. Beverungen, P. zur Heiden and T. Tornede, "Integration of Novel Sensors and Machine Learning for Predictive Maintenance in Medium Voltage Switchgear to Enable the Energy and Mobility Revolutions," *Sensors*, vol. 20, no. 2099, 2020.
- [5] M. Deisenroth, A. Faisal and C. Ong, *Mathematics for Machine Learning*, Cambridge University Press, 2020.
- [6] B. Schoelkopf and A. Smola, *Learning with Kernels – Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, 2002.
- [7] B. Schoelkopf, A. Smola and K. Mueller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Computation*, vol. 10, no. 5, pp. 1299-1319, 1998.
- [8] "scipy.signal.resample," [Online]. Available: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.resample.html>.
- [9] F. Pedregosa, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, p. 2825-2830, 2011.
- [10] A. Paszke and e. al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Proceedings of Advances in Neural Information Processing Systems*, 2019.
- [11] D. Lee, B. Lithgow and R. Morrison, "New fault diagnosis of circuit breakers," *IEEE Trans. Power Delivery*, vol. 18, pp. 454-459, 2003.
- [12] N. K. Verma, V. K. Gupta, M. Sharma and R. K. Sevakula, "Intelligent condition based monitoring of rotating machines using sparse auto-encoders," *IEEE International Conf. on Prognostics and Health Management (PHM)*, pp. 1-7, 2013.

Zustandsüberwachung und Klassifizierung manueller Montageprozesse mit selbstadaptiven Modellen

Philip Sehr, Natalia Moriz,
Institut für industrielle Informationstechnik (inIT),
Technische Hochschule Ostwestfalen-Lippe, Lemgo

Kurzfassung

Assistenzsysteme werden in der manuellen Montage angewendet, um angesichts hoher Variantenvielfalt Qualitätsstandards zu sichern. Dabei werden Werker schrittweise durch den Prozess geführt. Jedoch ist der Umfang der benötigten Assistenz stark von den Bedürfnissen des Werkers abhängig und resultiert in einer steigenden Nachfrage nach einer adaptiven Assistenz. In diesem Beitrag wird ein modellbasierter Ansatz vorgestellt, bei dem im ersten Schritt ein individuelles Leistungsniveau des Werkers automatisch bestimmt und anschließend ein Modell des Normalverhaltens erstellt wird. Diese Modelle werden im Betrieb eingesetzt, um den individuellen Assistenzbedarf eines Werkers abzuleiten und bei Bedarf eine direkte Änderung der angebotenen Hilfestellung durchzuführen. Der Ansatz verfolgt das Ziel, ein stabiles Leistungsniveau herbeizuführen und die Motivation der Mitarbeiter in der manuellen Montage zu fördern.

1. Einleitung

Durch Methoden moderner Informations- und Kommunikationstechnik wurden automatisierte Produktionsanlagen im Zuge der vierten industriellen Revolution zu Cyber-physischen Produktionssystemen (CPPS). Die Daten, die aus einem solchen System gewonnen werden, können genutzt werden, um Services wie Zustandsüberwachung, Diagnoseverfahren und vorausschauende Instandhaltung zu realisieren. Diese Verfahren helfen Stillstandzeiten zu minimieren und ermöglichen einen effizienten Betrieb in der Produktion.

Neben CPPS sind auch manuelle Montagesysteme weiterhin relevant für die Industrie. Eine steigende Nachfrage nach individuellen Produkten erfordert einen flexiblen Produktionsprozess, welcher – besonders für kleine und mittelständische Unternehmen – häufig nur mit manueller Montage wirtschaftlich zu realisieren ist. Angesichts der steigenden Variantenzahl wächst auch die Komplexität der manuellen Montage sowie die daraus resultierende Belastung für den Werker. Dies kann negative Auswirkungen auf die Produktivität haben. Um diesem Effekt entgegenzuwirken, werden vermehrt Montageassistenzsysteme verwendet, die den Werker mit schrittweisen Anleitungen durch den Montageprozess führen. Jedoch hängen

der Umfang und die Art der benötigten Assistenz stark von der Aufgabe und der ausführenden Kraft ab. Während unerfahrene Werker die schrittweise Unterstützung noch benötigen, fühlen sich Werker mit mehr Erfahrung durch das System häufig behindert oder bevormundet. Dies führt zu einer geringeren Akzeptanz des Systems, was sich in den meisten Fällen negativ auf die Produktivität auswirkt. Aus diesem Grund werden vermehrt adaptive Assistenzsysteme vorgeschlagen, die sich dem Erfahrungslevel des Werkers automatisch anpassen. In der Regel beruhen bisherige Ansätze auf vordefinierten Erfahrungsstufen, die durch bestimmte Ausführungszeiten oder Fehlerhäufigkeiten charakterisiert sind. Des Weiteren sind statische Kriterien für die Adaptivität eines Systems angesichts der hohen Abhängigkeit von der individuellen Erfahrung des Werkers und der Aufgabe häufig ungeeignet.

Um eine Lösung für diese Problemstellung zu bieten, wird in diesem Beitrag ein System vorgestellt, welches Werker in Abhängigkeit des bestehenden Leistungsniveaus unterstützt. Sowohl zur Ermittlung des Leistungsniveaus als auch zur Ermittlung des Bedarfs an Unterstützung soll ein gelerntes Modell verwendet werden, welches das Verhalten des Werkers während der Montagearbeit abbildet. Kommt es während der Montagearbeit zu Verzögerungen, können diese anhand des Modells identifiziert werden. Dieser theoretische Ansatz entspricht im Wesentlichen Methoden der modellbasierten Diagnose, die bei CPPS bereits Anwendung finden. Jedoch ist die direkte Übertragung dieser Methoden auf manuelle Montagesysteme mit Herausforderungen verbunden. Die Zeit, die ein Werker für dieselbe Montageaufgabe benötigt, bleibt langfristig nicht konstant, da er mit zunehmender Erfahrung seine Fähigkeiten verbessert, wodurch die benötigte Montagezeit sinkt. Um Störungen im Montageprozess anhand eines Modells identifizieren zu können, darf in dem Modell nur das Verhalten des Werkers repräsentiert werden, das seinem regulären Leistungsniveau entspricht. In diesem Beitrag wird ein Lösungsansatz vorgestellt, mit dem die Erstellung eines repräsentativen Modells des Normalverhaltens des Werkers während der Montage ermöglicht wird. Gelernte Modelle sollen weiterverwendet werden, um den individuellen Bedarf an Assistenz abzuleiten.

2. Stand der Technik

2.1 Adaptive Montageassistenz

Auf Grund des Bedarfs nach adaptiver Montageassistenz sind bereits einige Ansätze präsentiert worden, die eine adaptive Werkerassistenz bieten. Funk et.al [1] haben ein System entwickelt, welches aus der Anzahl korrekt und falsch ausgeführter Montageschritte einen Quotienten errechnet, der als Schwellwert für verschiedene Erfahrungsstufen verwendet wird. Sie unterscheiden drei Erfahrungsstufen (Anfänger, Fortgeschrittener, Profi) und ordnen jeder Er-

fahrungsstufe verschiedene Medien zwecks Assistenz zu. Während Anfänger Medien mit geringer Informationsdichte erhalten (Videos), bekommen Fortgeschrittene bzw. Profis nur noch Medien mit hoher Informationsdichte (Bilder) bzw. keine Unterstützung. Die Autoren stellen fest, dass die Messung der 3D-Kamera, die für die Erkennung falscher Greifbewegungen verwendet wird, im industriellen Umfeld sehr von Fehlern behaftet sein kann, weshalb mehrere Sensorsysteme verwendet werden sollten.

Kölz et al. haben ein System entwickelt, welches Menschen mit Behinderung durch adaptive Assistenz beim Lernprozess unterstützt [2]. Dieses System basiert auf vorab definierten Ausführungszeiten, die mit dem Methods-Time-Management-Standard ermittelt wurden: Abhängig vom Erfahrungslevel wird die Standardzeit um einen entsprechenden Faktor erhöht. Bleiben die Beschäftigten über mehrere Durchgänge in einem Zeitbereich, wird das Erfahrungslevel angepasst. Obwohl dieses System positive Ergebnisse innerhalb der Länge der Lernphase erzielte, wurde es nicht in einer realen industriellen Umgebung getestet. Dabei können reale Produktionsumgebungen mehrere Ablenkungsfaktoren enthalten, z. B. Licht, Lärm oder Absprachen mit Kollegen. Verzögerungen oder Fehler im Montageprozess, die durch einen dieser Faktoren verursacht werden, würden dann das Kompetenzniveau des Werkers fälschlicherweise beeinflussen.

Da sich die Montageschritte auch in Hinsicht auf ihren Schwierigkeitsgrad unterscheiden, beeinflusst dies auch den Bedarf an Montageunterstützung bei jedem Schritt. Bannat et al. haben ein System entwickelt, das eine adaptive Assistenz in Bezug auf die Montageschritte bietet [3]. Der Montageprozess wird hier als Graph modelliert. Diese Struktur erlaubt die Modellierung von alternativen Montagesequenzen, die sich in ihrem Detailgrad unterscheiden. Basierend auf der Montagehistorie des Werkers und den Sensoreingaben kann das System die am besten geeignete Montagehilfe für den Werker auswählen.

Ein anderer Ansatz basiert adaptive Assistenz auf individuellen Informationen über den Werker: Galakse und Anderl haben ein Profildatenmodell vorgeschlagen, das eine adaptive Werkerassistenz ermöglicht [4]. Dieses Modell beinhaltet neben den kompetenzbasierten Informationen auch ergonomische Profildaten, die zur Optimierung der angebotenen Montageassistenz genutzt werden können. Allerdings wurde das Modell noch nicht praktisch angewendet. Der in diesem Beitrag präsentierte Ansatz strebt eine automatische Erkennung des individuellen Assistenzbedarfs an und setzt auf ein individuell modelliertes Leistungsniveau eines Werkes. Dafür wird der Montageprozess mit Einsatz von Methoden des maschinellen Lernens und modellbasierter Diagnose abgebildet.

2.2 Modellierung von Montageprozessen

Ein Montageprozess besteht – sehr vereinfacht – aus einer Abfolge von Arbeitsschritten, bei denen Teile zusammengefügt werden. Dabei kommt es vor, bedingt durch den physikalischen Aufbau eines Produktes, dass Teile in unterschiedlicher Reihenfolge zusammengebaut werden können. Diese Beziehung kann als AND/OR Graph modelliert werden [5]. Dieses Modell kann alle zulässigen Montagesequenzen für ein Produkt gleichzeitig abbilden, weshalb es für die Planung von Montageabläufen verwendet wird [6], [7]. Die Knoten des Graphen repräsentieren dabei Zwischenzustände der Montage, bei denen nur ein Teil der Komponenten montiert wurde. Es existieren auch weitere Ansätze für die Modellierung der Zwischenzustände, wie beispielsweise durch einen Zustandsvektor [8], welche zwar die Form des resultierenden Graphen beeinflusst aber nicht die grundlegende Struktur des Modells.

Ein weiterer Ansatz zur Modellierung manueller Montageprozesse ist die Verwendung von Petri-Netzen [9], [10]. Dieser Formalismus erlaubt eine detailliertere Darstellung des Montageprozesses und kann z.B. Stückzahlen abbilden.

Die Analyse der hier beschriebenen Modellierungsansätze zeigt, dass die Struktur der Modelle immer einem Graphen entspricht, obwohl sich die Darstellungsformen unterscheiden kann. Auf Grund der Recherche werden diskrete Automaten als Basis für den im Folgenden präsentierten Ansatz gewählt.

2.3 Modellbasierte Diagnose

Bei jedem Produktionssystem kann es zu Anomalien im Produktionsablauf oder zu Systemausfällen kommen. Eine Diagnose auftretender Fehler ist bei komplexeren Produktionssystemen eine große Herausforderung, da komplexe Systeme große Datensätze produzieren, die für die Fehlerdiagnose ausgewertet werden müssen. Um diese Herausforderung zu adressieren, werden im Bereich von CPPS häufig modellbasierte Diagnoseverfahren verwendet. Bei diesen Methoden werden die Daten aus einer Produktionsanlage nicht direkt für die Fehlerdiagnose genutzt, sondern mit Daten eines Modells verglichen, welches das Verhalten des Systems beschreibt. Die im Rahmen des Vergleichs identifizierten Abweichungen können für die Erkennung von Fehlern verwendet werden [11]. Das Verhalten des Systems kann dabei entweder als Abfolge diskreter Events [12] oder als kontinuierlicher Verlauf modelliert werden [13]. Viele moderne Produktionssysteme sind hybrid und enthalten neben kontinuierlichen Verläufen auch diskrete Kontrolloperationen, wie etwa das Öffnen von Ventilen. Deshalb werden häufig diskrete Automaten verwendet, um das Systemverhalten zu modellieren [12], [14]. Um kontinuierliche Verläufe mit in die Fehleranalyse einzubeziehen, können diskrete Automaten

erweitert werden [15]. Die in diesem Beitrag vorgeschlagenen Ansätze für die adaptive Assistenz während der manuellen Montage basieren auf den Methoden der modellbasierten Anomalieerkennung.

3. Anwendungsfall

Der in diesem Beitrag vorgestellte Ansatz für die individuelle Assistenz wurde für den Anwendungsfall im Rahmen eines Inklusionsbetriebs konzipiert. In dem Betrieb werden verschiedene manuelle Tätigkeiten, wie z.B. Verpackung oder manuelle Montage im Kundenauftrag bearbeitet. In der Produktion des Betriebs werden Menschen mit und ohne Einschränkungen beschäftigt. Angesichts der häufig wechselnden Aufgaben werden Montageassistenzsysteme eingesetzt, um den Qualitätsstandard der Produktion zu gewährleisten. Aktuell wird im Betrieb ein in Kooperation zwischen dem Fraunhofer IOSB-INA und der technischen Hochschule Ostwestfalen-Lippe entstandene Assistenzsystem eingesetzt. Ein Aufbau der Hardware ist in Bild 1 dargestellt: Über dem Montagetisch ist ein Projektor angebracht, der Montageinformationen direkt auf den Arbeitsbereich projiziert. Gleichzeitig wird der Projektor für eine Pick-to-Light Funktion verwendet. Hierbei werden die Bauteilkästen angeleuchtet, aus denen im nächsten Montageschritt ein Bauteil entnommen werden muss. Bei Verwendung des Projektors für diese Funktion anstatt einer einfachen LED kann die Darstellungsform verändert und die Anzahl der benötigten Teile mit angezeigt werden. Der Bewegungen des Werkers während der Arbeit werden mit einer Tiefenbildkamera aufgezeichnet. Diese erlaubt es dem Assistenzsystem, bestimmte Handlungen des Werkers zu erkennen und darauf zu reagieren. So werden beispielsweise Greifbewegungen des Werkers in die Bauteilkästen erkannt. Auch die Interaktion mit virtuellen Buttons auf der Arbeitsfläche kann auf diese Weise realisiert werden.

Wegen der stark individuellen Bedürfnisse der Mitarbeiter soll ein bereits im Betrieb verwendetes System um adaptive Montageassistenz erweitert werden.



Bild 1: Aufbau des Assistenzsystems [22]

4. Verhaltensmodellierung bei der manuellen Montage

4.1 Prozessmodellierung

Da der zeitliche Ablauf der Montage individuell von Werker und Montageaufgabe abhängt, soll für jede Kombination ein eigenes Modell verwendet werden. Eine manuelle Erstellung dieser Modelle ist wegen zu hohem Modellierungsaufwand ungeeignet. Des Weiteren werden für die Umsetzung der individuellen Assistenz Methoden notwendig, welche im laufenden Betrieb eine Adaption der Modelle zulassen. Aus diesen Gründen werden in diesem Ansatz online-fähige Lernverfahren angewendet. Auf Basis der Daten, die durch das bereits vorhandene System während der laufenden Montagearbeit entstehen, werden Modelle maschinell gelernt. Grundsätzlich sollen zwei wesentliche Aspekte bei der Modellierung des Montageprozesses berücksichtigt werden.

Lernverhalten: Der Lernprozess eines Werkers findet zu Beginn jeder neuen Montageaufgabe statt. Nach mehreren Wiederholungen steigt die Erfahrung des Werkers. Dies wirkt sich auf die benötigte Ausführungszeit aus. Dabei hat jeder Werker ein charakteristisches Lernverhalten, welches von verschiedenen Aspekten der Montageausgabe, der Lernmethode und Persönlichkeit des Werkers beeinflusst werden kann [16]. Die während der Lernphase beendeten Montagedurchgänge sind daher nicht repräsentativ für das spätere Leistungsniveau des Werkers und sollten nicht im Modell des Normalverhaltens abgebildet werden.

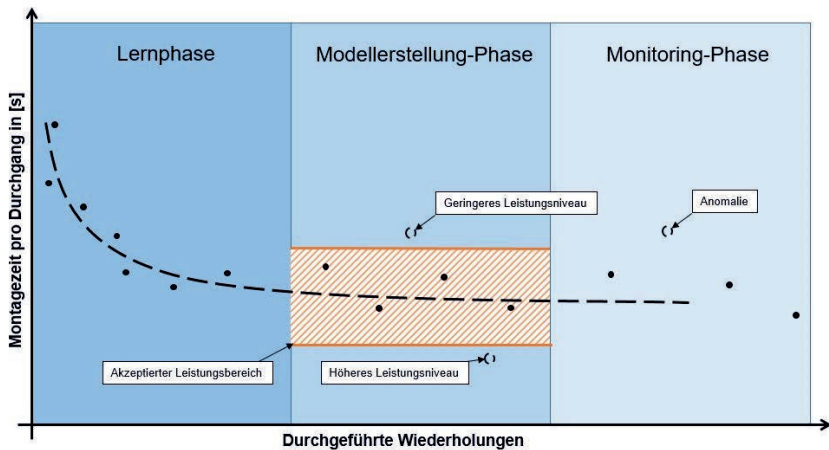


Bild 2: Phasenmodell zur Modellierung des Normalverhaltens eines Wokers während der Montage

Nicht prognostizierbare Abweichungen im Prozessablauf: Viele Situationen des betrieblichen Alltags können zu Verzögerungen im Montageablauf führen, wie etwa Absprachen mit Kollegen oder Vorgesetzten oder persönliche Bedürfnisse. Die Sensorik des Assistenzsystems ist nicht ausreichend, um die Gründe für Verzögerungen zu erkennen und zu evaluieren. Daher müssen die Montagedurchgänge, die solche Störungen beinhalten, ebenfalls nicht ins Modell des Normalverhaltens einfließen.

Um den manuellen Montageprozess zu modellieren, wurde ein Ansatz entwickelt, der die beiden genannten Aspekte behandelt. Der Ansatz besteht aus einem mehrstufigen Verfahren, welches in Bild 2 dargestellt ist und in den folgenden Unterkapiteln vorgestellt wird.

Phase 1 (Lernphase): Während dieser Phase soll der Lernprozess des Wokers beobachtet werden. Hierzu werden die Ausführungszeiten der ersten Montagedurchgänge verwendet, um mittels Regression ein individuelles Lernkurvenmodell zu erstellen. Sobald mit Hilfe des Lernkurvenmodells das Ende des Lernprozesses ermittelt werden kann, geht das System in die nächste Phase.

Phase 2 (Modellerstellung-Phase): Nachdem der Lernprozess des Wokers abgeschlossen ist und er sein stabiles Leistungsniveau erreicht hat, beginnt die Phase für die Erstellung des Modells des Normalverhaltens. Abgeschlossene Montagedurchgänge werden bezüglich der Durchführungszeit und Montagefehler analysiert. Datensätze mit einer zu großen Abweichung

vom ermittelten Leistungsniveau werden von Trainingsdaten ausgeschlossen bzw. gesondert behandelt. Vom Assistenzsystem erkannte Greifbewegungen und User-Interaktionen stehen in Form von Events zur Verfügung, weshalb sich die Modellierung des Montageablaufs als diskretes Eventsystem anbietet. Zur Modellierung des Normalverhaltens wird ein endlicher diskreter zeitbehafteter Automat verwendet.

Phase 3 (Monitoring-Phase): Nachdem das Modell des Normalverhaltens erstellt wurde, können zeitliche Abweichungen im Montageprozess erkannt werden. Hierzu können spezielle Algorithmen, wie z.B. ANODA verwendet werden [17]. Erkannte Abweichungen können dann genutzt werden, um Verbesserungen in der Montageassistentz abzuleiten.

4.2 Lernphase

Für die Modellierung einer Lernkurve stehen verschiedene Ansätze zur Verfügung. Nach dem typischen Verlauf einer Lernkurve benötigt ein Werker zu Beginn der Lernphase noch viel Zeit für seine Aufgabe. Die benötigte Montagezeit nähert sich mit steigenden Wiederholungen eines Montagedurchgangs asymptotisch einem bestimmten Wert an. Die Lernkurve kann durch drei Parametern charakterisiert werden: die Zeit der ersten Ausführung, die End-Ausführungszeit und die Lerngeschwindigkeit [18]. Einige Modelle berücksichtigen auch weitere Faktoren wie vorherige Erfahrung des Werkers oder den Effekt des Verlernens von Fähigkeiten. Anzanello und Fogliatto [19] haben mehrere Lernkurvenmodelle in einer detaillierten Recherche miteinander verglichen. Zwei Modelle stellen für den vorliegenden Anwendungsfall geeignete Kandidaten dar, da sie nur das Lernverhalten ohne zusätzliche Einflussfaktoren berücksichtigen (vgl. Gl. (1) und (2)).

$$\text{Lernkurvenmodell nach Wright} \quad t(x) = C x^{-b} \quad (1)$$

$$\text{Plateau-Modell} \quad t(x) = C x^{-b} + B \text{ mit } (0 < b < 1) \quad (2)$$

Der Funktionswert $t(x)$ beschreibt dabei die Montagezeit, die für die x -te Wiederholung des gleichen Montagevorgangs benötigt wird. Um ein geeignetes Lernkurvenmodell zu bestimmen, wurde ein Experiment in der Laborumgebung durchgeführt, welches die dem industriellen Anwendungsfall vergleichbaren Montageaufgaben nachbildet. Bei diesem Experiment mussten Probanden unter Verwendung des in Abschnitt 3 beschriebenen Assistenzsystems acht farbige Bausteine auf einer Platte anordnen. Die Aufgabe

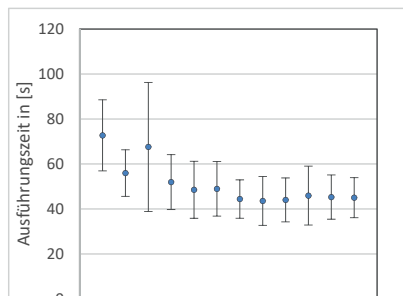


Bild 3: Mittlere Ausführungszeit während der Lernphase

wurde dabei bewusst mit geringer Schwierigkeit gestaltet, um mechanische Fähigkeiten aus dem Lernvorgang auszuschließen. Die Probanden mussten die Aufgabe insgesamt 12 mal wiederholen, wobei ihre Ausführungszeiten dokumentiert wurden. Die Ergebnisse des Experiments sind in Bild 3 dargestellt.

Auf Basis der Ausführungszeiten jedes Probanden wurde ein Regressionsverfahren unter Verwendung des Levenberg-Marquardt Algorithmus angewendet. Im nächsten Schritt wurden die R^2 -Fehler beider Modelle verglichen [20]. Auf Grund der Auswertung stellt das Plateau-Model bei den meisten Probanden eine besser geeignete Lernkurve dar und wird für den industriellen Anwendungsfall weiterverwendet.

Um das Ende der Lernphase für jeden Werker individuell bestimmen zu können, werden Auswertungskriterien für das Lernkurvenmodell benötigt. Die Ergebnisse des durchgeführten Experiments haben gezeigt, dass das individuelle Leistungsniveau – sogar bei Aufgaben mit geringem Schwierigkeitsgrad – stark vom Werker abhängt. Festgelegte Schwellwerte für das Leistungsniveau sind ungeeignet. Stattdessen werden zwei Kriterien definiert, die individuell für jede Lernkurve bestimmt werden:

Änderungsrate der Lernkurve

Während das absolute Leistungsniveau vom Werker abhängig ist, strebt die Änderungsrate der Lernkurve zum Ende des Lernprozesses gegen den Wert 0. Um ein anwendungsfallunabhängiges Abbruchkriterium zu definieren, wird jedoch nicht die absolute, sondern die relative Änderungsrate des Leistungsniveaus betrachtet. Anwendungsfallspezifisch wird mit p_{max} eine maximal akzeptierte prozentuelle Änderungsrate vorgegeben:

$$\left| \frac{t(x+1) - t(x)}{t(x)} \right| * 100 < p_{max} \quad (3)$$

Standardfehler der Lernkurve

Die Ausführungszeiten eines Werkers können signifikanten Schwankungen unterliegen, so dass die Änderungsrate der Lernkurve auf das Ende des Lernprozesses hindeutet, obwohl sich noch kein stabiles Leistungsniveau eingestellt hat. Für die Handhabung solcher Fälle wird ein zweites Kriterium definiert, welches die relative Standardabweichung zwischen des Lernkurvenmodells $t(x)$ und den realen Messungen $m(x)$ einbezieht. Dieses Kriterium wird nur über aktuellsten k Wiederholungen berechnet, damit einmalige deutlich längere Ausführungszeiten keinen zu nachhaltigen Effekt auf die berechnete Dauer des Lernprozesses haben. Eine Konstante e_{max} bestimmt dabei einen je nach Anwendungsfall maximal akzeptierten Wert für das Kriterium (4).

$$\sqrt{\frac{\sum_{i=0}^{k-1} (m(x-i) - t(x-i))^2}{k}} < e_{max} \quad (4)$$

Das Ende der Lernphase ist erreicht, wenn beide Kriterien erfüllt sind. Der Wert $t(x)$ am Ende dieser Phase entspricht dem Leistungsniveau des Werkers.

4.3 Modellerstellung-Phase

Nach dem Ende der Lernphase ist ein individuelles Leistungsniveau des Werkes für einen Montagedurchgang bestimmt worden. Dennoch ist es weiterhin möglich, dass der Werker während seiner Arbeit von diesem ermittelten Leistungsniveau abweicht. Während der Modellerstellung-Phase muss diese Situation berücksichtigt werden. Das Modell des Normalverhaltens wird in Form eines endlichen diskreten zeitbehafteten Automaten erstellt und beinhaltet jeden Montageschritt inklusive dem zulässigen Zeitintervall für seine Ausführung. Die Trainingsdaten werden analysiert, ob Messungen auf ein deutlich geringeres Leistungsniveau (längere Durchführungszeit) oder ein deutlich höheres Leistungsniveau (kürzere Durchführungszeit) hinweist.

Geringeres Leistungsniveau

Störungen im betrieblichen Ablauf oder Montagefehler können die für die Montage benötigte Zeit erhöhen. Montagedurchgänge, die von Fehlern beeinflusst sind, dürfen bei der Modellerstellung nicht genutzt werden, da sonst während der Monitoring-Phase ein ähnliches Verhalten des Werkers nicht als anomales erkannt werden kann. Es wird daher ein Toleranzbereich definiert, der eine maximale Abweichung von ermittelten Leistungsniveau erlaubt. Wird dieser Bereich verletzt, werden die entsprechenden Daten beim Modelllernen nicht berücksichtigt.

Höheres Leistungsniveau

Benötigt ein Werker weniger Zeit für die Montage als sein ermitteltes Leistungsniveau, deutet dies auf weitere Lernfortschritte des Werkers hin. In diesem Fall werden die Daten für das Modelllernen verwendet.

Das Ende der Modellerstellung-Phase ist von dem verwendeten Lernverfahren abhängig. In diesem Ansatz kommt der Algorithmus OTALA [21] zum Einsatz und bestimmt mit seinem Terminierungskriterium den Abschluss dieser Phase. Es kann sich nach dem Abschluss der Modellerstellungs-Phase ein neues individuelles Leistungsniveau für den Werker ergeben.

4.4. Monitoring Phase

Wenn die Erstellung des Verhaltensmodells abgeschlossen ist, kann es für ein Monitoring des Montageprozesses verwendet werden. Dies ermöglicht eine präzise Zustandsüberwachung

der Montage während des Betriebs. Der zeitliche Ablauf der Nutzerinteraktionen mit dem Assistenzsystem wird im laufenden Betrieb durchgehend mit dem gelernten Modell verglichen, um die Korrektheit der Ausführung des Montageprozesses zu ermitteln. Dabei können auch anomale Verhaltensweisen im Prozessablauf erkannt werden. Auf Basis dieser Informationen kann während des Betriebs die angebotene Assistenzfunktion bei Bedarf angepasst werden. Hierfür sind folgende zwei Ansätze vorgeschlagen: Es wird das Leistungsniveau eines Werkers analysiert, um einen negativen oder positiven Trend rechtzeitig zu erkennen. Die zunehmend steigenden Ausführungszeiten können ein Hinweis darauf sein, dass der Werker in einem bestimmten Montageschritt Hilfe benötigt. Anhand des Modells können diese Schritte genau identifiziert und eine weitere Hilfestellung vom Assistenzsystem angeboten werden. Im Fall der Tendenz zu sinkenden Ausführungszeiten können über das Assistenzsystem Veränderungen an den angebotenen Assistenzfunktionen vorgenommen werden, wie z.B. das Weglassen bestimmter Medien. So kann überprüft werden, ob der Werker mittlerweile die Aufgabe aus dem Gedächtnis beherrscht. Über das Verhaltensmodell kann danach festgestellt werden, ob sich die Veränderungen negativ auf das Leistungsniveau ausgewirkt haben. Diese Phase des Konzepts wird in den nächsten Schritten der Arbeiten angegangen.

5. Zusammenfassung

In diesem Beitrag wurde ein Ansatz für ein adaptives Montageassistenzsystem vorgestellt, bei dem über ein Modell das individuelle Verhalten und Leistungsniveau eines Werkers ermittelt und für die Optimierung der Assistenzfunktion verwendet wird. Dabei wurden besondere Herausforderungen herausgestellt, die bei der Modellierung eines manuellen Montageprozesses entstehen, und ein Lösungskonzept vorgestellt. Mit Hilfe des Regressionsverfahrens wird ein individuelles Lernkurvenmodell erstellt, mit dem das spezifische Leistungsniveau eines Werkers bestimmt werden kann. Im weiteren Schritt werden während der Montage gewonnene Daten für die Erstellung des Modells verwendet. Das gelernte Modell kann eingesetzt werden, um Anomalien im Ablauf des Montageprozesses zu erkennen um damit adaptive Werkerassistenz anzubieten.

Zukünftige Arbeiten evaluieren das Konzept im realen Betrieb. Außerdem werden weitere Anwendungsmöglichkeiten für die erstellten Modelle untersucht. Ein Modell, welches die Zustände und zeitlichen Abfolgen eines Montageprozesses enthält, könnte beispielsweise für die Klassifizierung verschiedener Prozesse verwendet werden. Eine Ähnlichkeit zwischen zwei Modellen verschiedener Montageprozesse kann ein Hinweis auf eine ähnliche Schwierigkeit dieser Montageaufgaben hindeuten. Diese Informationen können im vorliegenden Anwendungsfall verwendet werden, um Personal an neue Aufgaben heranzuführen.

Danksagung

Diese Arbeit entstand im Rahmen des Projekts „enableIT“ (FKZ 13FH110PX6), gefördert vom Bundesministeriums für Bildung und Forschung (BMBF).

Referenzen

- [1] M. Funk, T. Dingler, J. Cooper, and A. Schmidt, „Stop helping me-I'm bored! why assembly assistance needs to be adaptive“. In Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers, Sep. 2015, pp. 1269-1273.
- [2] M. Kölz, D. Jordon, P. Kurtz, and T. Hörz, „Adaptive assistance to support and promote performance-impaired people in manual assembly processes“ In Proceedings of the 17th international acm sigaccess conference on computers & accessibility, Oct. 2015, pp. 337-338.
- [3] A. Bannat, F. Wallhoff, G. Rigoll, F. Friesdorf, H. Bubb, S. Stork, H. J. Müller, A. Schubö, M. Wiesbeck, M. F. Zäh, „Towards optimal worker assistance: a framework for adaptive selection and presentation of assembly instructions“ In Proceedings of the 1st international workshop on cognition for technical systems, Oct. 2008, Cotesys.
- [4] N. Galaske, and R. Anderl, „Approach for the development of an adaptive worker assistance system based on an individualized profile data model“ In Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future, 2016, pp. 543-556.
- [5] L. H. De Mello & A. C. Sanderson (1990). AND/OR graph representation of assembly plans. *IEEE Transactions on robotics and automation*, 6(2), 188-199.
- [6] L. H. De Mello & A. C. Sanderson (1989, January). A correct and complete algorithm for the generation of mechanical assembly sequences. In *1989 IEEE International Conference on Robotics and Automation* (pp. 56-57). IEEE Computer Society.
- [7] U. Thomas, T. Stouraitis & M.A. Roa (2015, August). Flexible assembly through integrated assembly sequence planning and grasp planning. In *2015 IEEE International Conference on Automation Science and Engineering (CASE)* (pp. 586-592). IEEE.
- [8] J. Lanham & F. Dialami (2001, May). The assembly state vector: a new approach to the generation of assembly sequences. In Proceedings of the 2001 IEEE International Symposium on Assembly and Task Planning (ISATP2001). Assembly and Disassembly in the Twenty-first Century.(Cat. No. 01TH8560) (pp. 37-42). IEEE.

- [9] K.E. Moore, A. Güngör & S. M. Gupta (2001). Petri net approach to disassembly process planning for products with complex AND/OR precedence relationships. *European Journal of Operational Research*, 135(2), 428-449.
- [10] S. Casell & F. Zanichelli (1995, August). On assembly sequence planning using petri nets. In *Proceedings. IEEE International Symposium on Assembly and Task Planning* (pp. 239-244). IEEE.
- [11] S. Narasimhan & G. Biswas (2007). Model-based diagnosis of hybrid systems. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and humans*, 37(3), 348-361.
- [12] M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, & D. C. Teneketzis (1996). Failure diagnosis using discrete-event models. *IEEE transactions on control systems technology*, 4(2), 105-124.
- [13] P. J. Mosterman & G. Biswas (1999). Diagnosis of continuous valued systems in transient operating regions. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 29(6), 554-565.
- [14] R. Zemouri & J.M. Faure (2006, October). Diagnosis of discrete event system by stochastic timed automata. In *2006 IEEE Conference on Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control* (pp. 1861-1866). IEEE.
- [15] A. Vodenčarević, H. K. Büning, O. Niggemann & A. Maier (2011, October). Using behavior models for anomaly detection in hybrid systems. In *2011 XXIII International Symposium on Information, Communication and Automation Technologies* (pp. 1-8). IEEE.
- [16] M. De Greiff (2001). Die Prognose von Lernkurven in der manuellen Montage unter besonderer Berücksichtigung der Lernkurven von Grundbewegungen. VDI Verlag.
- [17] O. Niggemann, A. Vodencarevic, A. Maier, S. Windmann & H. K. Büning (2013, October). A learning anomaly detection algorithm for hybrid manufacturing systems. In *The 24th International Workshop on Principles of Diagnosis (DX-2013)*, Jerusalem, Israel.
- [18] T. Jeske & C. M. Schlick (2011). Influence of task complexity on learning times of sensorimotor tasks in assembly systems. In *ICPR (Vol. 21)*.
- [19] M.J. Anzanello & F. S. Fogliatto (2011). Learning curve models and applications: Literature review and research directions. *International Journal of Industrial Ergonomics*, 41(5), 573-583.
- [20] T. Jeske, C.M. Schlick & S. Mütze-Niewöhner (2014). Unterstützung von Lernprozessen bei Montageaufgaben. In *Flexible Produktionskapazität innovativ managen* (pp. 163-192). Springer Vieweg, Berlin, Heidelberg.

- [21] A. Maier (2014, July). Online passive learning of timed automata for cyber-physical production systems. In 2014 12th IEEE International Conference on Industrial Informatics (INDIN) (pp. 60-66). IEEE.
- [22] itsowl-TT-FlexInt: Flexible Inklusion in einem gemeinsamen Produktionsprozess durch adaptive Assistenz. Projektwebseite: <https://www.th-owl.de/init/forschung/projekte/b/filteroff/453/single.html>

Semantische Interoperabilität durch Natural Language Processing als Basis für Self-X-Fähigkeiten von

Verwaltungsschalen in semantisch heterogenen Asset-Netzwerken

Maximilian Both, Nicolai Maisch, Prof. Dr. Jochen Müller,
Technische Hochschule Köln, Köln

Kurzfassung

Dieses Paper stellt eine Methode vor, durch die Systeme eines semantisch heterogenen Industrie 4.0 Netzwerks befähigt werden ihre semantische Beschreibung selbstständig und konfigurationsfrei auf einen Standard abzubilden. Hierzu werden Methoden des Natural Language Processing (NLP), konkret Language Models, die auf allgemeinen Texten trainiert wurden und hierdurch semantische Beziehungen von Wörtern durch Word Embeddings repräsentieren, verwendet. Auf Basis dieser Language Models werden Sentence Embeddings trainiert, welche nicht nur Wörter, sondern Sätze in Form von Vektoren repräsentieren. Namen und Definition von Properties der Teilmodelle eines Assets durchlaufen diese Architektur und werden so durch Sentence Embeddings repräsentiert. Die Namen und Definitionen der Merkmale neuer Kommunikationsteilnehmer durchlaufen ebenfalls die Architektur und werden als Sentence Embeddings abgebildet. Auf deren Basis kann bestimmt werden, welches Merkmal der eigenen Teilmodelle dem Merkmal des Kommunikationsteilnehmers entspricht. Dies ermöglicht eine automatisierte Abbildung heterogener Semantik aufeinander und bildet die Basis für Systeme mit Self-X-Fähigkeiten. Der hierfür entwickelte Matching Dienst erweitert den aktuellen I4.0-Ansatz von homogener auf heterogene Semantik. Das Paper gibt einen Überblick über die erreichten Genauigkeiten der Methode und über die Erweiterung des Interaktionsmanagers von Verwaltungsschalen um den semantischen Matching Dienst.

1. Einleitung

Komponenten mit Self-X-Fähigkeiten werden die nächste Entwicklungsstufe der Automatisierungstechnik charakterisieren. Unter Self-X-Fähigkeiten werden Funktionalitäten eines Systems verstanden, welche intrinsische Automatismen zur Erkundung eines Netzwerks, Selbst-Konfiguration, -Diagnose und -Optimierung ermöglichen. Damit ein Netzwerk realisiert werden kann, in dem verschiedene Komponenten interoperabel im Sinne von I4.0 miteinander agieren können, müssen drei Grundsätze erfüllt werden: die interagierenden Komponenten benötigen eine einheitliche Struktur, ein gemeinsames Vokabular und eine standardisierte

Dialoggestaltung [1]. Bezüglich der einheitlichen Struktur hat sich im Bereich der I4.0 das Konzept der I4.0-Komponente, bestehend aus Verwaltungsschale und Asset, etabliert [2]. Um ein gemeinsames Vokabular zu etablieren, können Dictionaries wie ECLASS verwendet werden [3]. Für den Bereich der Dialoggestaltung wird eine einheitliche I4.0-Sprache definiert, welche die Struktur von Nachrichten und darüber hinausgehende Interaktionsprotokolle definiert [1]. Im weiteren Verlauf dieses Paper werden die Aspekte gemeinsame Struktur und gemeinsames Vokabular näher betrachtet. Bild 1 zeigt eine Monitoringanwendung, die mit verschiedenen Komponenten interagiert. Die Komponenten werden durch Verwaltungsschalen repräsentiert, besitzen also eine einheitliche Struktur und basieren auf einem gemeinsamen I4.0-Vokabular.

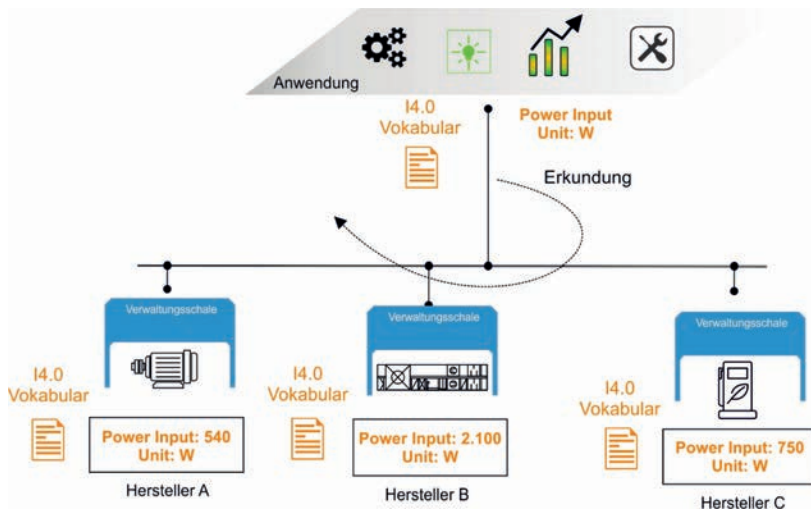


Bild 1: Erkundung von Verwaltungsschalen

Die Kombination aus einheitlicher Struktur und gemeinsamen Vokabular ermöglichen der Anwendung Informationen der Komponenten zu erkunden und zu integrieren. Verwenden die Komponenten jedoch kein standardisiertes Vokabular, z.B. den ECLASS Standard, sondern herstellerspezifische Beschreibungen, so ist ein interoperables Agieren nicht möglich [2]. Damit die Komponenten mit der Anwendung automatisiert interagieren können ist ein Matching der verschiedenen Vokabulare notwendig. Dies ist in der zweiten Abbildung dargestellt.

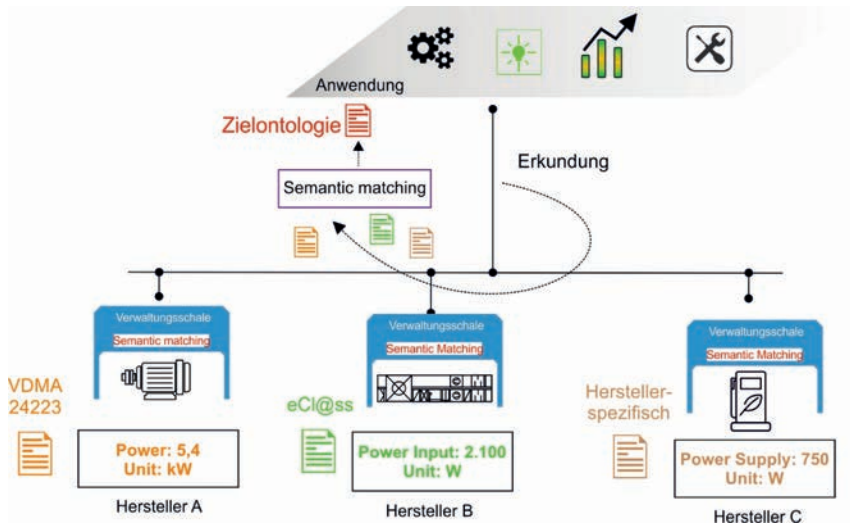


Bild 2: Matching heterogener Semantik auf eine Zielontologie

Die Komponenten basieren auf unterschiedlichen Vokabularen, welche der Anwendung unbekannt sind. Diese Informationen müssen auf die Zielontologie der Anwendung abgebildet werden, sodass diese die Informationen verarbeiten kann. Die Abbildung erzeugt initial einen hohen manuellen Aufwand, sodass Anwender den Einsatz von Monitoringanwendungen hinsichtlich Kosten und Nutzen abwägen. Im Folgenden wird eine Methode vorgestellt, welche ein automatisiertes Matching heterogener Semantik auf eine Zielontologie und somit semantische Interoperabilität ermöglicht. Ausgangslage des Matchings ist die Struktur der Verwaltungsschale. Die Informationen der Komponenten liegen also in einer einheitlichen Struktur vor, folgen jedoch keinem einheitlichen semantischen Standard. Zunächst wird untersucht, welche Informationen die Verwaltungsschale bereitstellt, die für ein Matching verwendet werden können. Für den Entwurf des Matchings werden Konzepte aus dem Bereich der Künstlichen Intelligenz, konkret des Natural Language Processing (NLP), verwendet.

Der betrachtete Use Case des Matching Dienstes sind nicht kritische Anwendungen (z.B. die Monitoringanwendungen der Abbildungen 1 und 2). In kritischen Automatisierungstechnischen Anwendungen, wie der Prozessautomatisierung, wird dieser nicht eingesetzt. Hierfür wäre eine 100 prozentige Genauigkeit des Dienstes notwendig, welche durch die angewandten Methoden nicht erreicht werden kann.

Dieses Paper stellt eine Weiterentwicklung bereits veröffentlichter Ergebnisse dar [4]. Durch die Erweiterung des verwendeten Datensatzes und der Analyse der Metainformationen von Verwaltungsschalen als weitere Trainingsparameter, konnten neue Erkenntnisse für die Weiterentwicklung des semantischen Matching Dienstes gewonnen werden.

2. Informationen der Teilmodelle von Industrie 4.0 Komponenten

Die Verwaltungsschale ist der digitale Repräsentant eines Assets in der digitalen Welt [2]. Ein Asset kann hierbei jede Entität sein, die einen Wert für eine Organisation hat [5]. Das Informationsmodell der Verwaltungsschale definiert die Struktur, wie ausgetauschte Informationen verschiedener Assets aufgebaut sein müssen [2]. Die Komposition aus Verwaltungsschale und Asset wird als I4.0-Komponente bezeichnet. Zentraler Baustein der Verwaltungsschale sind Teilmodelle, welche Eigenschaften und Funktionalitäten der Assets und deren Inhalte abbilden [2]. Teilmodelle stehen z.B. für die Themen Identifikation, Design oder Konfiguration. Teilmodelle wiederum enthalten *Submodel Elements* (SEs), welche die tatsächlichen Informationen der Komponenten enthalten. Beispiele für SEs sind *Properties* oder *Capabilities* [2]. Verschiedene Initiativen standardisieren Teilmodelle für Komponenten der Industrie, bspw. Pumpen oder Antriebe [6, 7]. Die Integration der Teilmodelle in Vokabulare wie den ECLASS-Standard [3], oder das IEC Common Data Dictionary [8], ermöglicht die eindeutige Identifikation der Teilmodelle, sowie deren Elemente und bereitet so den Weg für eine eindeutige Semantik. Sind die Teilmodelle und insbesondere deren SEs jedoch nicht in einem externen Standard verfügbar, sondern liegen in nicht standardisierter Form vor, erschwert dies Interoperabilität zwischen I4.0-Komponenten. Im Folgenden wird untersucht welche Informationen den SEs zugeordnet werden können, die dann wiederum für das Matching verwendet werden können.

Die Klasse SE ist die abstrakte Klasse aller Elemente, die ein Teilmodell enthalten kann. Zusammen mit ihren übergeordneten Klassen ist sie in der folgenden Abbildung 3 dargestellt.

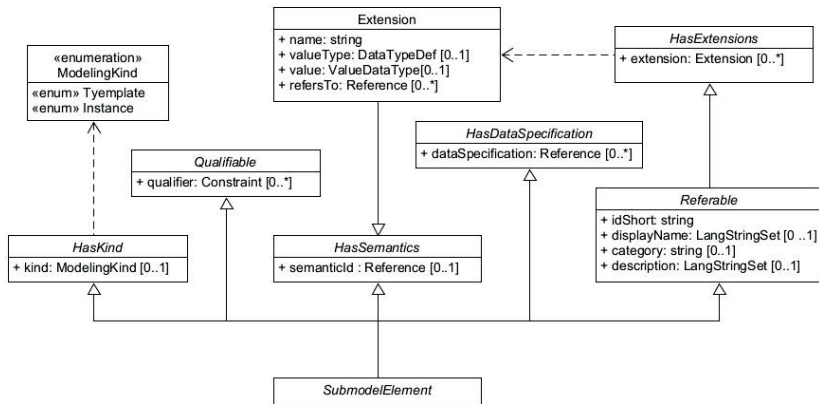


Bild 3: Klasse Submodel Element

Die für den Matching Dienst relevante Klasse ist *Has Semantics*. Jedes SE soll über eine *Semantic Id* verfügen, welche eine Referenz zur semantischen Definition des Elements enthält [2]. Mittels dieser Id kann entweder direkt auf einen externen Standard, wie ECLASS, verwiesen werden, oder auf eine *Concept Description* (CD). In dieser sind Attribute enthalten, welche das jeweilige Element semantisch definieren. Die jeweilige CD kann hierbei eine Kopie eines externen Standards sein. Ist sie dies jedoch nicht, kann nach dem aktuellen I4.0-Ansatz keine Interoperabilität mit anderen Verwaltungsschalen gewährleistet werden [2]. Grundsätzlich sollen CDs einem *Data Specification Template* folgen, in dem definiert ist welche Attribute die CD enthält. Für verschiedene SEs wie *Capabilities* oder *Properties* werden unterschiedliche Templates definiert, da für die jeweiligen SEs unterschiedliche Attribute relevant sind. In [2] ist das Template *DataSpecificationIEC61360* definiert, welches von der Plattform I4.0 für die Definition von *Properties* empfohlen wird.

Der entworfene Matching Dienst wird exemplarisch mit Teilmodellen des Projektes Pumpe 4.0 [7] umgesetzt. In den dort definierten Teilmodellen sind überwiegend *Properties* als SEs enthalten, sodass die Attribute des Template *DataSpecificationIEC61360* weiter analysiert werden. Die Klasse *DataSpecificationIEC61360* ist mit ihren Attributen in der folgenden Abbildung vier dargestellt.

«Template» DataSpecificationIEC61360
+ preferredName: LangStringSet + shortName: LangStringSet [0..1] + unit: string [0..1] + unitid: Reference [0..1] + sourceOfDefinition: string [0..1] + symbol: string [0..1] + dataType: DataTypeIEC61360 [0..1] + definition: LangStringSet [0..1] + valueFormat: string [0..1] + valueList: ValueList [0..1] + value: ValueDataType [0..1] + valueId: Reference [0..1] + levelType: LevelType [0..*]

Bild 4: Template für Data Specification IEC 61360 nach [2]

In der Klasse sind 13 Attribute definiert, welche ein *Property* zusätzlich zu den Informationen aus dem Metamodell der Verwaltungsschale weiter beschreiben. Für einen Matching Dienst könnten alle der Attribute verwendet werden, da sie unterschiedliche Aspekte eines Assets beschreiben. Beispielsweise kann mit dem Attribut *source of Definition* eingeschränkt werden, aus welchem Bereich das Merkmal stammt und so eine Eingrenzung vorgenommen werden. Problematisch für den Matching Dienst ist jedoch, dass die meisten Attribute des Template optional sind. Deshalb kann nicht gewährleistet werden, dass immer alle dieser Attribute für ein Matching zur Verfügung stehen. Nach [2] gibt es jedoch zwei Einschränkungen: Sowohl der *preferred Name* als auch die *definition* müssen für ein *Property* angegeben werden, wenn es nach diesem Template definiert wird. Da diese beiden Attribute immer angegeben werden müssen und alle anderen optional sind, werden für das Matching im ersten Entwurf nur diese beiden Attribute verwendet. Basis des Matching Dienstes sind Methoden aus dem Bereich NLP, konkret Language Models und Sentence Embeddings.

3. Pretrained language models und sentence embeddings

Das Ziel von NLP ist, dass Computer menschliche Sprache verstehen und darauf aufbauend Interaktionen realisieren [9]. In den letzten Jahren wurden die Methoden zur Extraktion von Wissen aus Text weiterentwickelt, so dass dieses in unterschiedlichen NLP Anwendungen, wie Text Classification oder Question Answering, verwendet werden kann. Die in diesem

Paper entwickelte Modellarchitektur setzt sich aus einem Pretrained Language Model (PLM) und darauf aufbauenden Sentence Embeddings (SEms) zusammen.

Aktuelle NLP Methoden basieren häufig auf PLMs [10]. Diese Modelle werden auf großen Mengen Text vortrainiert (Pretraining). State-of-the-art Modelle basieren auf der Transformer Architektur [11]. Das Transformer Modell setzt sich aus einem Encoder und Decoder zusammen und eignet sich besonders für sequence-to-sequence Aufgaben (z.B. Übersetzungen) [11]. Um die Transformer Architektur für andere Aufgaben als Übersetzungen zu verwenden, werden für die heutigen Modelle entweder der Encoder [12 bis 16] oder Decoder [17 bis 19] eingesetzt. Ergebnis dieser Trainings sind Vektoren, welche Wörter so repräsentieren, dass sie semantische Bedeutungen und Beziehungen der Wörter enthalten. Diese word embeddings (WE) können für die Erkennung von Ähnlichkeiten von Beziehungen verwendet werden, z.B. ob ein Wortpaar wie Berlin und Deutschland die gleiche Beziehung hat wie Madrid und Spanien. Vortrainierte WEs werden in einem zweiten Trainingsschritt mit spezifischen Datensätzen eines konkreten Anwendungsfalls (Text Classification etc.) verfeinert (Fine Tuning) [20]. Diese Art des Trainings, Pretraining auf allgemeinen Texten und Fine Tuning auf spezifischen Datensätzen, wird als Transfer Learning bezeichnet [21] und in diesem Paper angewandt. Im weiteren Verlauf wird das vortrainierte Modell DistilBERT [14] verwendet.

Der Anwendungsfall der semantischen Interoperabilität lässt sich dem NLP-Bereich Paraphrase Identification zuordnen. Paraphrase Identification (Detection) bezeichnet die Fähigkeit eines Modells zu erkennen, ob zwei Sätze die gleiche Bedeutung haben [22]. Ein Beispiel für eine solche Anwendung ist der Quora Question Pairs (QQP) Datensatz, bei dem vorhergesagt werden soll, ob Fragen die gleiche Bedeutung haben oder nicht [23]. Verschiedene LMs haben hohe Genauigkeiten von bis zu 90 Prozent für diesen Datensatz erzielt, unter anderem das Modell Sentence-BERT [24]. Bei diesem Modell werden die WEs von BERT zu einem SEM mit einer fixen Dimension zusammengefasst. Mittels der Kosinus-Ähnlichkeit kann im Anschluss überprüft werden, ob sich die SEms zweier Sätze ähneln. Dies ermöglicht den Abgleich, ob eine Frage des QQP Datensatzes ein Duplikat im Korpus hat oder nicht [24].

4. NLP-Modell zur Abbildung heterogener Semantik

Der Matching Dienst wird in die Verwaltungsschale integriert (Abbildung 5).

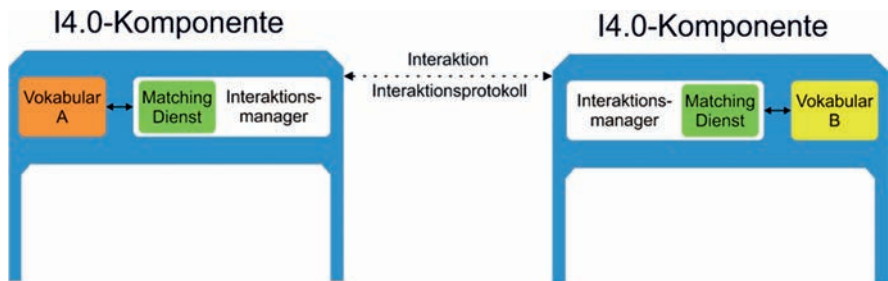


Bild 5: Der Matching Dienst als Teil der Verwaltungsschale

Die zwei Verwaltungsschalen der I4.0-Komponenten interagieren miteinander. Hierbei kennen sie das Vokabular der jeweils anderen nicht, basieren somit auf heterogener Semantik. Dieses Szenario würde nach dem heutigen I4.0-Ansatz Interoperabilität verhindern.

Damit Interoperabilität zwischen den beiden Verwaltungsschalen hergestellt werden kann, wird ein Matching der beiden Vokabulare benötigt. Der grundsätzliche Ablauf eines automatisierten Matchings ist Bestandteil dieses Papers und erweitert den aktuellen I4.0-Ansatz von homogener auf heterogene Semantik. Für die Abwicklung der Interaktionen ist der Interaktionsmanager verantwortlich, welcher die eingehenden Nachrichten verarbeitet [1]. Die Abbildung der unbekannten Semantik auf die Zielontologie wird durch den Matching Dienst realisiert. Hierdurch kann heterogene Semantik verarbeitet und der Ansatz des homogenen Sprachraums auf einen heterogenen Sprachraum erweitert werden.

Als konkreter Anwendungsfall werden unterschiedliche Semantiken auf die Teilmodelle des Projektes Pumpe 4.0 [7] abgebildet. Die Basis des Modells bildet das vortrainierte LM DistilBERT. Durch ein erweitertes Pretraining auf domänenspezifischer Literatur wird grundsätzlich eine Erhöhung der Genauigkeit erreicht [25].

$$\text{Genauigkeit} = \frac{\text{Anzahl korrekter Vorhersagen}}{\text{Gesamtzahl der Vorhersagen}}$$

Aus diesem Grund wird das Pretraining von DistilBERT mit ISO-Standards erweitert. Dieses Modell wird im Anschluss verwendet, um mit Datensätzen aus dem Bereich Paraphrase Identification verfeinert zu werden. Durch diesen Trainingsablauf erlernt das Modell die Semantik des Anwendungsfalls (Pretraining) und Methoden zur Lösung der Aufgabe des Paraphrase Identification (Fine Tuning). Zur Evaluation des Modells wird ein Datensatz erstellt, der unterschiedliche Definitionen und Namen für Merkmale der Teilmodelle von Pumpen enthält. Die Auswahl der ISO Standards umfassen die Bereiche Identifikation von Assets allgemein und spezifische Inhalte von Pumpen. Insgesamt bilden 303 Standards die Basis für das erweiterte Pretraining [26 bis 32]. Diese wurden in das XML-Format umgewandelt [33], und zu Textpassagen weiterverarbeitet, ohne Tabellen, Bildunterschriften oder Literaturangaben. Auf diesem Text wurde das durch [10] zur Verfügung gestellte DistilBERT Modell weitertrainiert.

Für den Anwendungsfall der semantischen Interoperabilität von I4.0-Verwaltungsschalen stehen aktuell keine Datensätze für ein Fine Tuning zur Verfügung. Deshalb wurden für das Training der SEms [34] die allgemeinen Datensätze MultiNLI [35], STS-Benchmark [36] und QQP [23] verwendet. Die drei Datensätze werden häufig eingesetzt, um im Bereich Paraphrase Identification neue Modelle zu trainieren. Hierdurch erlernt das Modell die Aufgabe des Paraphrase Identification und kann diese auf den konkreten Anwendungsfall adaptieren. Das Modell wurde nacheinander mit den drei Datensätzen verfeinert, sodass zum Abschluss das Modell ISO-DistilBERT-SE erstellt wurde.

5. Evaluation und Diskussion

Zur Evaluierung des Modells wurde ein Datensatz erstellt, der als Basis 250 *Properties* der Teilmodelle für Pumpen als Zielontologie verwendet [7]. Für jedes *Property* wurden mehrere Paraphrasen erstellt, sodass insgesamt 650 Paraphrasen der *Properties* zur Verfügung stehen. Ein Beispiel für ein *Property*-Paraphrase Paar ist in der folgenden Tabelle dargestellt.

Tabelle 1: Beispiel für ein Merkmal mit passender Paraphrase

Property aus der Zielontologie		Beispiel Paraphrase	
Name	Definition	Name	Definition
Name of product	Product designation under which the device is marketed	Product title	Title under which the product is known

Von den 250 *Properties* der Zielontologie werden zunächst 250 SEms gebildet. Hierbei durchlaufen Name und Definition der *Properties* getrennt das Modell, sodass sich zwei Vektoren bilden. Diese werden im Anschluss miteinander zu einem SE verkettet. Ebenso durchlaufen die 650 Paraphrasen das gleiche Modell. Für jede dieser SEms wird die Kosinus-Ähnlichkeit mit den 250 SEms des Korpus berechnet. Das jeweilige Paar mit der höchsten Ähnlichkeit wird von dem Modell als Paraphrase klassifiziert. Zur Bewertung von ISO-DistilBERT-SE wird die Bewertungsmetrik Genauigkeit verwendet. Bei dem auf Normen vortrainierten Modell ISO-DistilBERT-SE wurde eine Genauigkeit von 74,26 % erreicht. Ergebnisse mit einem kleineren Ausschnitt der Daten (36 *Properties* der Teilmodelle) haben jedoch deutlich bessere Ergebnisse mit einer Genauigkeit von 94 % erzielt [4]. Dies zeigt, dass durch die große Anzahl der *Properties* (250) des Modells in diesem Paper mehr Missklassifikationen stattfinden.

In weiteren Versuchen müssen Möglichkeiten untersucht werden, wie eine höhere Genauigkeit erzielt werden kann. Eine Möglichkeit besteht darin, das Pretraining des Modells auf einer größeren Anzahl an domänenspezifischen Texten durchzuführen, sodass die Semantik der technischen Domäne mehr Gewicht in dem Modell erhält. Gleichzeitig können weitere Attribute der Verwaltungsschale, sowohl aus dem Metamodell als auch Attribute der *Properties*, verwendet werden, um die Genauigkeit zu erhöhen. Durch ein vorab durchgeführtes Clustering auf Basis weiterer Attribute wie der *Unit* oder dem *Value type* der *Properties* kann die Anzahl der *Properties* reduziert werden, welche zu einer Paraphrase passen. Erst auf dieser reduzierten Anzahl wird dann das semantische Matching durchgeführt.

6. Fazit und Ausblick

Die Ergebnisse erweitern den aktuellen I4.0-Forschungsansatz zur Interaktion von semantisch homogenen auf semantisch heterogene Verwaltungsschalen. Hierzu wurden NLP-Methoden verwendet, um ein automatisiertes Semantic Matching auf eine Zielontologie zu realisieren. Die erreichte Genauigkeit von 74 % zeigt, dass die entwickelte Methodik ein Ansatz ist, um semantische Interoperabilität zu erreichen. Zur Akzeptanz in der Praxis muss die Genauigkeit des Modells jedoch weiter erhöht werden. Diese reicht bspw. zur Anwendung in automatisierungstechnischen Funktionen wie der Prozessführung nicht aus. Außerdem stellt der betrachtete Anwendungsfall nur einen kleinen Ausschnitt der etwaigen *Properties* von I4.0-Komponenten dar. Für konkrete Anwendungsszenarien, wie ein umfangreiches Plant Asset Management, müssen mehr *Properties* in der Zielontologie abgebildet werden. Zudem wurde das Modell nur auf einem kleinen Ausschnitt internationaler Normen trainiert. In weiteren Schritten wird das Pretraining auf mehr Normen durchgeführt, der Zielontologie werden mehr *Properties* hinzugefügt und der Evaluationsdatensatz wird erhöht. Die Ergebnisse werden in relevante Spezifikationsarbeiten eingespeist, um den Interaktionsmanager der Verwaltungsschale um einen Semantic Matching Dienst zu erweitern.

7. Literaturangaben

- [1] VDI/VDE 2193 Blatt 1:2020-04. *Sprache für I4.0-Komponenten. Struktur von Nachrichten*
- [2] Details of the Asset Administration Shell. Part 1 - The exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0RC01), Federal Ministry for Economic Affairs and Energy, Berlin 2020, Federal Ministry for Economic Affairs and Energy. Berlin 2020
- [3] ECLASS: ECLASS - Standard für Stammdaten und Semantik für die Digitalisierung. <https://www.eclass.eu/index.html>, abgerufen am: 17.11.2020
- [4] Both, M., Müller, J. u. Diedrich, C.: Automatisierte Abbildung semantisch heterogener I4.0-Verwaltungsschalen durch Methoden des Natural Language Processing. at - Automatisierungstechnik (in Druck)
- [5] Industrie 4.0 Begriffe/Terms. VDI Statusreport, VDI/VDE-Gesellschaft, 2019, VDI/VDE-Gesellschaft 2019
- [6] Antrieb 4.0 – Vision wird Realität. Merkmale, Daten und Funktionen elektrischer Antriebssysteme in Industrie 4.0 für Hersteller, Maschinenbauer und Betreiber, ZVEI - Zentralverband Elektrotechnik und Elektronikindustrie e. V., 2018, ZVEI - Zentralverband Elektrotechnik und Elektronikindustrie e. V. 2018
- [7] Both, M. u. Müller, J.: Digitization of pumps – Industry 4.0 submodels for liquid and vacuum pumps. 4th International Rotating Equipment Conference (2019)
- [8] IEC: IEC Common Data Dictionary. <https://cdd.iec.ch/cdd/iec61360/iec61360.nsf/TreeFrameset?OpenFrameSet>, abgerufen am: 17.11.2020
- [9] Jurafsky, D. u. Martin, J. H.: Speech and language processing. An introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall series in artificial intelligence. Upper Saddle River, NJ: Prentice Hall Pearson Education Internat 2009
- [10] HuggingFace's Transformers: State-of-the-art Natural Language Processing, Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., Platen, P. v., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q. u. Rush, A. M., 2019, 2019
- [11] Attention Is All You Need, Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u. Polosukhin, I., 2017, 2017
- [12] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Devlin, J., Chang, M.-W., Lee, K. u. Toutanova, K., 2018, 2018

- [13] RoBERTa: A Robustly Optimized BERT Pretraining Approach, Liu, Y., Ott, M., Goyal, N., Du Jingfei, Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. u. Stoyanov, V., 2019, 2019
- [14] DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, Sanh, V., Debut, L., Chaumond, J. u. Wolf, T., 2019, 2019
- [15] Cross-lingual Language Model Pretraining, Lample, G. u. Conneau, A., 2019, 2019
- [16] ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators, Clark, K., Luong, M.-T., Le V, Q. u. Manning, C. D., 2020, 2020
- [17] XLNet: Generalized Autoregressive Pretraining for Language Understanding, Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. u. Le V, Q., 2019, 2019
- [18] Improving language understanding by Generative Pre-Training, Radford, A., Narasimhan, K., Salimans, T. u. Sutskever, I., 2018, 2018
- [19] Language Models are Unsupervised Multitask Learners Alec, Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. u. Sutskever, I., 2019, 2019
- [20] To Tune or Not to Tune? Adapting Pretrained Representations to Diverse Tasks, Peters, M. E., Ruder, S. u. Smith, N. A., 2019, 2019
- [21] Pan, S. J. u. Yang, Q.: A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering 22 (2010) 10, S. 1345–1359
- [22] Socher, R., Huang, E. H., Pennington, J., Ng, A. Y. u. Manning, C. D.: Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection. Proceedings of the 24th International Conference on Neural Information Processing Systems. NIPS'11. Red Hook, NY, USA: Curran Associates Inc 2011, S. 801–809
- [23] Iyer, S., Dandekar, N. u. Csernai, K.: First Quora Dataset Release: Question Pairs. <https://www.quora.com/q/quotadata/First-Quora-Dataset-Release-Question-Pairs>, abgerufen am: 23.11.2020
- [24] Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, Reimers, N. u. Gurevych, I., 2019, 2019
- [25] Don't Stop Pretraining: Adapt Language Models to Domains and Tasks, Gururangan, S., Marasović, A., Swayamdipta, S., Lo Kyle, Beltagy, I., Downey, D. u. Smith, N. A., 2020, 2020
- [26] ISO: ISO/TC 184/SC 4. Industrial data. <https://www.iso.org/committee/54158.html>, abgerufen am: 11.12.2020
- [27] ISO: ISO TC 115. Pumps. <https://www.iso.org/committee/51766.html>, abgerufen am: 11.12.2020

- [28] ISO: ISO/IEC JTC 1/SC 17. Cards and security devices for personal identification. <https://www.iso.org/committee/45144.html>, abgerufen am: 11.12.2020
- [29] ISO: ISO/IEC JTC 1/SC 31. Automatic identification and data capture techniques. <https://www.iso.org/committee/45332.html>, abgerufen am: 11.12.2020
- [30] ISO: ISO/TC 46/SC 4. Technical interoperability. <https://www.iso.org/committee/48798.html>, abgerufen am: 11.12.2020
- [31] ISO: ISO/TC 46/SC 9. Identification and description. <https://www.iso.org/committee/48836.html>, abgerufen am: 11.12.20
- [32] ISO: ISO/TC 59/SC 2. Terminology and harmonization of languages. <https://www.iso.org/committee/49076.html>, abgerufen am: 11.12.20
- [33] GROBID, 2008-2020. <https://github.com/kermitt2/grobid>, abgerufen am: 14.04.2021
- [34] Reimers, N.: SentenceTransformers Documentation. <https://www.sbert.net/index.html>, abgerufen am: 23.11.2020
- [35] Williams, A., Nangia, N. u. Bowman, S.: A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). Association for Computational Linguistics 2018, S. 1112–1122
- [36] Cer, D., Diab, M., Agirre, E., Lopez-Gazpio, I. u. Specia, L.: SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation. Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). Vancouver, Canada: Association for Computational Linguistics 2017, S. 1–14

Automatisierte Datenintegration für den Fehleranalyseprozess von Halbleiterbauelementen mithilfe von Ontologien und Graphen

S. Kamm, M.Sc., Dr.-Ing. **N. Jazdi**, Prof. Dr.-Ing. Dr. h.c. **M. Weyrich**,
Institut für Automatisierungstechnik und Softwaresysteme (IAS) der
Universität Stuttgart

Kurzfassung

Im Sinne der Qualitätssicherung sollen zukünftig produzierte Halbleiterbauelemente den steigenden Anforderungen an Zuverlässigkeit, Konnektivität und Automatisierung standhalten. Dazu ist ein Fehleranalyseprozess notwendig. Um die Effizienz zu steigern und die Fehleranfälligkeit zu reduzieren, soll der Fehleranalyseprozesses mit dazugehöriger Datenanalyse automatisiert werden. Dafür muss zunächst ein Konzept für die automatisierte Integration heterogener Daten aus unterschiedlichen Datenquellen auf Basis aktueller Daten Engineering Methoden entwickelt werden. Die Daten aus unterschiedlichen Quellen (z.B. Produktionsdaten, Datenblätter, Analyseverfahren) müssen eindeutig einem Bauelement zugeordnet werden können und neue Daten oder Analyseergebnisse diesem hinzugefügt werden können. In diesem Beitrag wird für die automatisierte Datenintegration des Fehleranalyseprozesses von Halbleiterbauelementen ein Konzept mithilfe von Ontologien und Graphen vorgestellt und dessen Mehrwerte anhand eines Anwendungsfalls mit einer prototypischen Umsetzung diskutiert. Dafür wird ein hybrider Ontologie-Ansatz genutzt, um die heterogenen Daten in einem Knowledge Graphen zu verknüpfen. Der vorgestellte Ansatz ermöglicht sowohl eine lokale als auch eine globale Datenanalyse. Dadurch können sowohl bisher entwickelte Datenanalyse-Ansätze genutzt werden, als auch neue Algorithmen aufgebaut werden.

1. Einleitung und Grundlagen

Halbleiterbauelemente sind wesentliche Komponenten für die Digitalisierung und Automatisierung von vielen Bereichen, wie z.B. der Produktion (Smart Manufacturing), dem Automobilbereich (Automatisierte und Autonome Fahrzeuge) oder der intelligenten Erzeugung und Nutzung von erneuerbaren Energien. Aufgrund steigender Anforderungen an die Funktionalität und der Bedeutung von Halbleiterbauelementen wird die Zuverlässigkeit dieser Komponenten immer bedeutender [1]. Es ist jedoch unmöglich, alle auftretenden Fehler vorzeitig zu entdecken. Es können Fehler während der Herstellung, der Bestückung oder der Lebensdauer auftreten [2]. Um die Zuverlässigkeit von Halbleiterbauelementen zu erhöhen, ist die Analyse von

defekten Elementen zur Bestimmung der Fehlerursache unabdingbar. Dazu muss zunächst ein Fehler erkannt werden, z.B. durch eine fehlerhafte Funktion einer Komponente. In der folgenden Fehlerdiagnose wird der auftretende Fehler klassifiziert, lokalisiert und charakterisiert [3]. Daraufhin kann die konkrete Fehlerursache bestimmt werden. Zusammengefasst wird im weiteren Verlauf von der Fehleranalyse gesprochen

Das gewonnene Wissen aus der Fehleranalyse kann in den Entwicklungsprozess fließen, um aufgetretene Fehler in Zukunft zu verhindern und damit die Zuverlässigkeit zukünftiger Komponenten zu erhöhen. Um den Fehler eindeutig bestimmen und lokalisieren zu können, wird eine Vielzahl an Methoden benötigt. Einige beispielhafte nicht-destruktive (zerstörungsfreie) und destruktive (zerstörende) Methoden sind in Tabelle 1 gelistet [2]:

Tabelle 1: Beispielhafte destruktive und nicht-destruktive Fehleranalysemethoden

Nicht-destruktive Methoden		Destruktive Methoden	
Methodik	Anwendung	Methodik	Anwendung
Mikroskopie	Externe und interne visuelle Inspektion	Entkapselung	Öffnen des Packages
Kurventracer	Identifikation der Breakdown-Spannung, Leakage	Chromatographie	Chemische Analyse
Time-Domain Reflektometrie	Elektrisches Reflektionsmuster	Sectioning	Querschnitt durch die Probe

Generell kann in nicht-destruktive und destruktive Techniken unterschieden werden. Die unterschiedlichen Techniken können unterschiedliche Bereiche der Fehleranalyse unterstützen. So können mithilfe eines Kurventracers offene Verbindungen, Kurzschlüsse oder eine Spannungsdegradation detektiert und bestimmt werden [2]. Mithilfe der Time-Domain Reflektometrie (TDR) können ebenfalls unterschiedliche elektrische Fehlermuster detektiert werden und der Fehler lokalisiert werden [4]. Mikroskopische Techniken wie z.B. die Scanning Acoustic Microscopy (SAM) können helfen, mechanische und thermomechanische Fehler zu detektieren und lokalisieren [5]. Mithilfe von destruktiven Techniken wird das Bauteil zerstört, und weitere Untersuchungen sind nur noch sehr beschränkt möglich. Diese Techniken können jedoch relevant sein für eine eindeutige Fehleranalyse. In Bild 1 ist ein vereinfachter Fehleranalyse Prozess dargestellt. Dabei werden in den unterschiedlichen Analyseschritten unterschiedliche

Fehleranalysetechniken eingesetzt. Die gewonnenen Erkenntnisse aus der Fehleranalyse fließen zum einen in den Entwicklungsprozess und zum anderen in die zukünftige Fehleranalyse mit ein.

Je nach Bauelement und Fehlerart sind nicht alle Schritte des Prozesses notwendig, um eine eindeutige Fehleranalyse zu ermöglichen. Dennoch zeigt Bild 1 die Komplexität des Prozesses, in welchem eine Vielfalt an Daten entsteht, die miteinander in Verbindung stehen (unterschiedlichste Messungen an eventuell unterschiedlichen Orten eines Bauteils). Auf Basis dieser Datenvielfalt soll die finale Fehlerursache bestimmt werden. Um diese Datenvielfalt effizient analysieren zu können, bedarf es jedoch einer möglichst einheitlichen und automatisierten Datenintegration, -haltung und -bereitstellung.

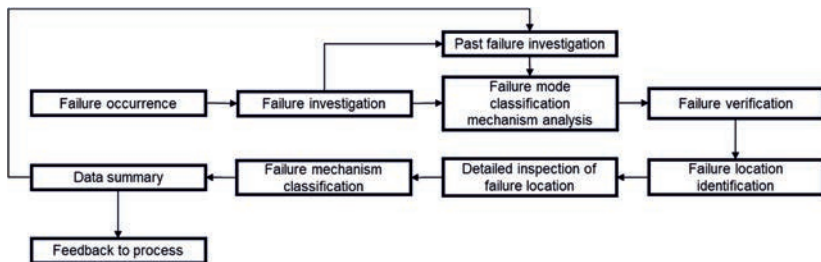


Bild 1: Beispielhafter Fehleranalyse Prozess in Anlehnung an [2]

Für die automatisierte Datenintegration und einheitliche formale und semantische Beschreibung der vorliegenden heterogenen Daten eignen sich Wissensbasen, da in diesen Wissensbasen die Daten und ihre Relationen formal und einheitlich beschrieben werden. In die Modellierung dieser Wissensbasen fließt vorhandenes Expertenwissen ein, was in der Fehleranalyse von Halbleiterbauelementen häufig über viele Jahre aufgebaut wurde. Eine weit verbreitete Realisierungsmöglichkeit für Wissensbasen sind Ontologien. Eine Ontologie wird als formale, explizite Spezifikation einer gemeinsamen Konzeptualisierung definiert, die sich durch eine hohe semantische Ausdruckskraft auszeichnet, die für eine erhöhte Komplexität erforderlich ist [6]. Ontologien können in sogenannte Lightweight Ontologien und Heavyweight Ontologien anhand ihrer Ausdruckskraft unterschieden werden. Lightweight Ontologien beinhalten Konzepte, Taxonomien, Beziehungen zwischen Konzepten und Eigenschaften die Konzepte beschreiben (z.B. ein UML Klassendiagramm). Heavyweight Ontologien fügen mathematische Axiome und Begrenzungen hinzu [7]. Ontologien versprechen dabei zwei Arten von Interoperabilität zu erreichen. Auf der einen Seite Interoperabilität zwischen mehreren Repräsentatio-

nen der Realität (z.B. unterschiedliche Datenmodelle), die sich in Computersystemen befinden. Auf der anderen Seite die Interoperabilität zwischen solchen Repräsentationen und der Realität [8]. Dadurch kann eine Ontologie die entstehende große semantische Heterogenität auflösen und eine einheitliche formale und semantische Datenbeschreibung realisieren. Ein wissensbasiertes System erweitert eine solche Wissensbasis um eine sogenannte *Reasoning Engine*, welche Logik und Folgerungsmechanismen einbringt, um neues Wissen zu generieren (z.B. Empfehlung fehlender Relationen oder Detektion von falschen Relationen). Darüber hinaus erweitert der Knowledge Graph ein wissensbasiertes System um die Datenintegration. Mithilfe eines Knowledge Graphen können ebenfalls mehrere Wissensbasen (z.B. Ontologien) miteinander verknüpft und gegebenenfalls erweitert werden. In Bild 2 (links) ist eine schematische graphische Darstellung des Knowledge Graphen mit seinen Bestandteilen gegeben. Zusammenfassend erfasst und integriert ein Knowledge Graph Informationen in einer Ontologie und wendet einen Reasoner an, um neues Wissen abzuleiten [9]. Ein daraus resultierender Graph (z.B. Bild 2 (rechts)) kann in einer Graph-Datenbank abgelegt werden. Knowledge Graphen haben seit der Einführung von Google's Knowledge Graph in 2012 [10] viel Aufmerksam-

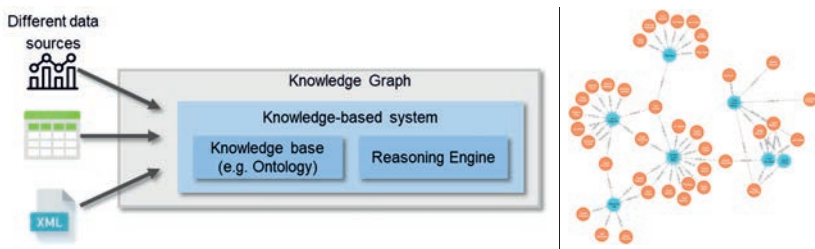


Bild 2: Schematische Darstellung eines Knowledge Graphen (links) und beispielhafter resultierender Graph (rechts)

keit gewonnen und werden als Herzstück der nächsten Generation von Enterprise Informationssystemen gesehen [11].

In diesem Beitrag wird gezeigt, wie eine automatisierte Datenintegration für den Fehleranalyseprozess von Halbleiterbauelementen mithilfe von Ontologien und Graphen aufgebaut und realisiert werden kann. Dafür werden zunächst in Kapitel 2 die Herausforderungen der automatisierten Datenintegration für den Fehleranalyseprozess von Halbleiterbauelementen sowie bestehende Arbeiten in diesem Gebiet aufgezeigt. Daraufhin wird in Kapitel 3 das Konzept vorgestellt. In Kapitel 4 wird eine Realisierung und prototypische Umsetzung des vorgestellten Konzeptes gezeigt. Abschließend gibt Kapitel 5 eine Zusammenfassung des Beitrags und Ausblick auf zukünftige Arbeiten.

2. Herausforderungen und bestehende Arbeiten

Durch die große Datenvielfalt im Fehleranalyseprozess von Halbleiterbauelementen entstehen Herausforderungen für die Datenintegration, -haltung und -nutzung. Die Daten entstehen an unterschiedlichen Messgeräten von unterschiedlichen Herstellern, was in einer großen semantischen Heterogenität resultiert [7]. Um die Daten möglichst automatisiert zu integrieren, zu verarbeiten und damit für den Fehleranalyseprozess nutzbar zu machen, bedarf es einheitliche, formale und semantische Beschreibungen der Daten, um diese semantische Heterogenität zu lösen [12]. Mithilfe einer semantischen Beschreibung kann die Interoperabilität von Daten unterschiedlicher Systeme (z.B. unterschiedliche Cyber-Physischer Systeme – CPS) erreicht werden. Zudem stellt eine semantische Beschreibung eine notwendige Voraussetzung für das Nutzen der Potenziale dieser heterogenen Systeme dar [13]. Diese semantische Beschreibung ermöglicht unter anderem die einheitliche und maschinenlesbare Bereitstellung der Daten, um im weiteren Verlauf die Fehleranalyse zu vereinfachen bzw. zu automatisieren, z.B. mithilfe von Machine Learning Algorithmen [14]. Klassische relationale Datenbanken stoßen bei einer ansteigenden Datenheterogenität an ihre Grenzen und können die vorhandene Menge an komplexen Relationen oft nicht beherrschen [15]. Die genannten Herausforderungen können in drei Punkten zusammengefasst werden:

- Durch Vielzahl an Messgeräten entsteht eine große semantische Heterogenität [7]
- Zur Auflösung der großen semantischen Heterogenität wird eine einheitliche formale und semantische Datenbeschreibung benötigt [12]
- Die heterogenen Daten mit ihren formalen und semantischen Beschreibungen müssen gemanagt werden, wobei die vorhandene Datenvielfalt und die komplexen Relationen abgebildet werden [15]

Im betrachteten Szenario der Fehleranalyse werden Daten in unterschiedlichen Datenquellen generiert, die verknüpft werden müssen. Zudem sind darüber hinaus häufig weitere Daten, z.B. aus der Produktion (Technische Datenblätter) oder aus dem Lebenszyklus des Bauteils (Umgebungsbedingungen) vorhanden, welche für die Fehleranalyse relevant sein können. Mithilfe von Ontologie-basierten Daten Integration (OBDI) Ansätzen können Daten in eine definierte Ontologie integriert werden. Ontologie-basierte Daten Integration ist eine weit verbreitete Technik, um die zuvor genannten Herausforderungen zu lösen und besteht aus drei Komponenten [16, 17]:

- Die Datenquelle, welche die Daten erzeugt
- Eine Ontologie zur Repräsentation des vorhandenen Wissens in Form einer Wissensbasis
- Ein Mapping zwischen diesen beiden Komponenten

Ontologie-basierte Daten Integration-Ansätze können in drei verschiedene Arten unterteilt werden. Dem Single Ontologie Ansatz, dem Multiple Ontologie Ansatz und dem hybriden Ontologie Ansatz. Beim Single Ontologie Ansatz wird eine globale Ontologie als Referenz-Modell aufgesetzt, welche die gesamte Domäne beschreibt. Im Gegensatz dazu wird beim Multiple Ontologie Ansatz jede Datenquelle semantisch von einer eigenen lokalen Ontologie beschrieben und zwischen den unterschiedlichen Ontologien Austauschformate bestimmt. Der hybride Ontologie Ansatz kombiniert diese beiden, in dem er lokale Ontologie für die einzelnen Datenquellen nutzt und diese dann in eine übergeordnete Ontologie integriert [18].

Es gibt viele vorhandene Ansätze und Projekte, bei denen globale Ontologien für eine Domäne entwickelt wurden. Bekanntes Beispiel dafür ist z.B. die Semantic Sensor Network (SSN)-Ontologie aus einem Projekt des World Wide Web Konsortiums (W3C) [19]. Diese Ontologie soll Sensorik und deren Beobachtungen beschreiben. Weiter gibt es die Smart Appliances Reference (SAREF)-Ontologie [20], welche durch das ETSI (European Telecommunications Standards Institute) standardisiert wurde. Diese Ontologie soll Daten verknüpfen und Kommunikation über verschiedene Protokolle für IoT Geräte in smarten Anwendungen ermöglichen. Der Fokus lag dabei auf Gebäude Anwendungen. Um diese Ontologie auf weitere Bereiche anzuwenden, wurden Erweiterungen z.B. für die Industrie und Produktions Domäne mit SAREF4INMA, entwickelt [21].

Häufig wird der hybride Ontologie Ansatz genutzt, da eine globale Ontologie nicht alle spezifischen Bereiche der betrachteten Domäne abdecken kann [22]. In [23] wurde eine Ontologie für die Produktion von Halbleiterbauelementen entwickelt. Die entwickelte Semantik wurde mit geographischen Produktionsdaten angereichert und bietet die Möglichkeit, Rückschlüsse für die Produktion aus den semantisch beschriebenen Daten zu ziehen. Der Siemens Industrial Knowledge Graph [24] soll das Problem von isoliert und heterogen vorliegenden Produktionsdaten (Datensilos) lösen. Dafür soll der Knowledge Graph eine integrierte Sicht über diese verteilten Daten auf einer erhöhten Ebene liefern. In [25] prüfte Ford das Potenzial von föderierten, verteilten Ontologien, um damit Schlussfolgerungen in der Produktion zu unterstützen. Insgesamt haben Ontologien mit Knowledge Graphen das Potenzial, die beschriebenen Herausforderungen zu lösen, die Daten automatisiert zu integrieren und in einer semantisch einheitlichen und formalen Beschreibung bereit zu stellen. Damit kann semantische Heterogenität aufgelöst werden, was eine eindeutige Bedeutung der Daten und damit maschinen-lesbarkeit für die weitere Verarbeitung sicherstellt.

3. Automatisierte Datenintegration für den Fehleranalyseprozess von Halbleiterbauelementen mithilfe von Ontologien und Graphen

Als Basis für eine intelligente Fehleranalyse von Halbleiterbauelementen ist eine automatisierte Datenintegration notwendig. Dabei soll vorhandenes Expertenwissen in der Modellierung genutzt werden und die Daten für die weitere Analyse vorbereitet und bereitgestellt werden. Wie im vorherigen Kapitel gezeigt, eignen sich Ontologien und Graphen zur Bewältigung der auftretenden Herausforderungen.

Eine globale einheitliche Beschreibung von allen Geräten ist nicht möglich, da die unterschiedlichen lokalen Messgeräte sehr heterogen sind. Eine grundlegende, übergeordnete Fehleranalyse-Struktur kann jedoch abgebildet werden. Aufgrund dessen wird ein hybrider Ontologie-Ansatz, wie in Bild 3 gezeigt, gewählt. Für jede auftretende Datenquelle (1...n) wird das Wissen eines Datenexperten (1...n) in Kombination mit dem Wissen eines Ontologieexperten genutzt, um eine lokale Lightweight-Ontologie dieser Datenquellen zu erzeugen. Durch diesen Schritt können die Daten dieser Datenquelle erfasst werden. Mithilfe der modellierten Lightweight-Ontologie werden die Daten vorbereitet und in die passende lokale Modellierung gebracht. Für eine übergeordnete Ontologie Fehleranalyse sollen möglichst einzelne Teile von bestehenden Ontologien (z.B. SAREF oder SSN), so genannte Ontologie Design Patterns (ODP), wiederverwendet werden [7]. Durch die Wiederverwendung können Entwicklungsaufwände reduziert. Zudem sind diese Ontologien bereits über einen langen Zeitraum entwickelt und eingesetzt worden und haben ihre Nutzbarkeit gezeigt. Zur Anpassung bzw. Erweiterung der wiederverwendeten Design Patterns ist Domänenwissen notwendig. Innerhalb der Ontologie Fehleranalyse werden schließlich die lokalen Lightweight-Ontologien automatisiert kombiniert und verknüpft, z.B. können über die einheitliche ID eines Bauelements Messungen aus unterschiedlichen Datenquellen einem gleichen Bauteil zugeordnet werden. Die Ontologie

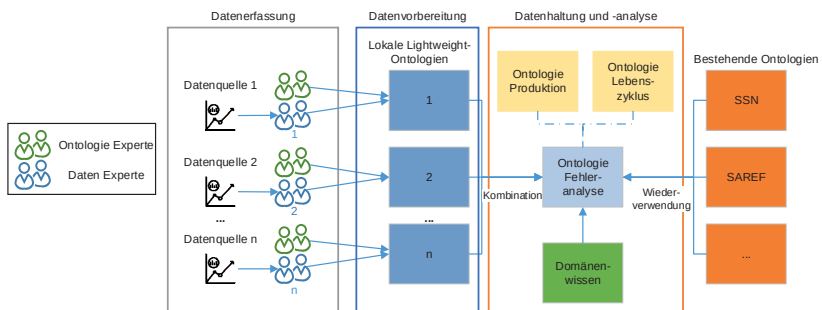


Bild 3: Konzeptioneller Ansatz für die automatisierte Datenintegration für den Fehleranalyseprozess von Halbleiterbauelementen

Fehleranalyse ermöglicht die einheitliche und zentrale Datenhaltung sowie die weitere Datenanalyse. Die Ontologie Fehleranalyse kann mit weiteren Ontologien (z.B. einer Ontologie Produktion oder Ontologie Lebenszyklus) verknüpft werden. Dadurch sollen Erkenntnisse aus der Fehleranalyse mit Informationen aus der Produktion oder dem Lebenszyklus verknüpft werden. Zukünftig sollen dadurch bisher unbekannte Verknüpfungen zwischen Produktionseinflüssen, Umgebungsbedingungen im Feld und einem aufgetretenen Fehler intelligent detektiert werden. Für eine Datenquelle (hier Time-Domain Reflektometrie) sieht der Daten- und Informationsfluss mit diesem Konzept wie in Bild 4 dargestellt aus.

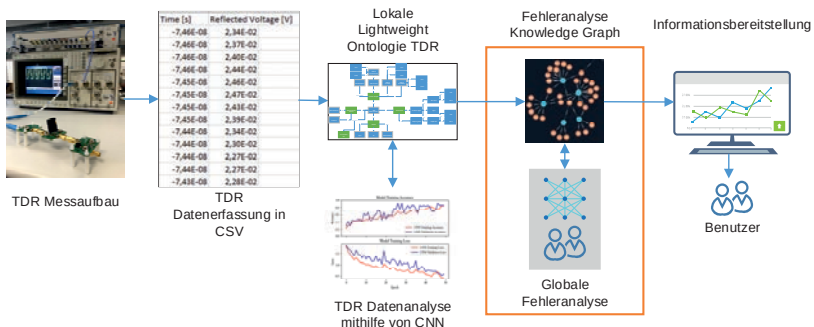


Bild 4: Daten- und Informationsfluss einer Datenquelle (hier Time-Domain Reflektometrie)

Zunächst werden die Daten einer Datenquelle erfasst und gegebenenfalls vor- oder aufbereitet vom Messgerät (z.B. in einer CSV-Datei). Die erfassten Daten und Datentypen variieren dabei stark von Gerät zu Gerät. Zusätzlich zu der Messung können auch beschreibende Daten der Messung (z.B. Parametereinstellungen) oder Umgebungsdaten (z.B. Temperatur) miterfasst werden. Die erfassten Daten fließen in die erstellte lokale Lightweight Ontologie ein, die für jede Datenquelle variieren kann und das Informationsmodell der Datenquelle darstellt. Auf dieser Stufe kann eine erste lokale Datenanalyse auf Basis der erfassten Daten durchgeführt werden. Bei der Fehleranalyse von Halbleiterbauelementen kann z.B. eine erste Schätzung des Fehlertyps oder -orts durchgeführt werden. Diese Analyse kann durch einen Experten oder automatisiert, z.B. durch ein Machine Learning Modell [26], durchgeführt werden. Die Ergebnisse der lokalen Datenanalyse werden an die lokale Ontologie zurück übermittelt und dort hinzugefügt. Diese lokale Ontologie wird in die Top-Level Ontologie integriert und mit weiteren vorhandenen Daten aus anderen lokalen Ontologien verbunden. Auf diesem Level kann eine „globale“ Datenanalyse auf Basis aller verfügbaren Daten durchgeführt werden. Diese

Datenanalyse kann ebenfalls durch einen Experten oder ein Programm automatisiert durchgeführt werden. Die Komponenten der Datenintegration, der Datenhaltung und der Datenanalyse auf diesem Level formen den Knowledge Graph. Der Knowledge Graph bietet die Möglichkeit, Analysen der vorhandenen Relationen durchzuführen (z.B. Detektion von fehlerhaften Daten oder Relationen, Prädiktion fehlender Relationen, etc.) oder mithilfe von Graph Neural Networks eine globale Fehleranalyse aus den vorhandenen Daten durchzuführen und damit neue, bisher unbekannte Zusammenhänge zu detektieren (z.B. Zusammenhänge zwischen Fehlern und Produktionsparametern bei der Kombination von Fehleranalyse- und Produktionsdaten). Der Knowledge Graph bietet weiter die Möglichkeit einer intuitiven Datenvisualisierung in Graph-Form sowie eines effizienten Datenzugriffs mithilfe von Graph Query Sprachen wie Cypher oder SPARQL.

4. Realisierung und Prototypische Umsetzung

Für die Realisierung wurden Daten des Time-Domain Reflektometrie Verfahrens genutzt. Dieses Verfahren bietet die Möglichkeit, einen Fehler in einem Bauelement mithilfe eines elektrischen Signals zu detektieren. Dabei wird ein elektrisches Signal inzidiert und das reflektierte Signal gemessen. Anhand des Reflektionssignal kann der Zustand des Bauelements detektiert werden. Eine beispielhafte Messung für einen SiC Transistor SCT2160KE der Firma Rohm ist in Bild 5 zu sehen.

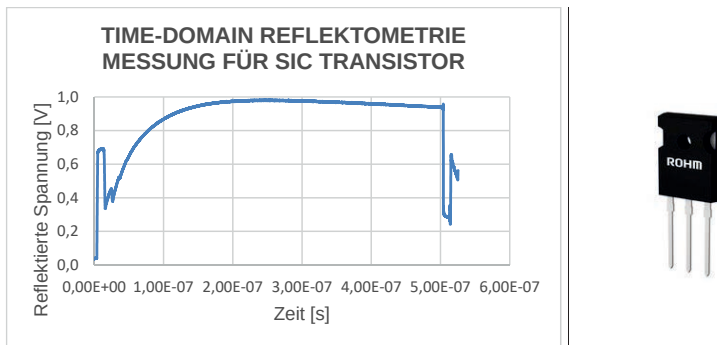


Bild 5: Beispielhafte Time-Domain Reflektometrie Messung (links) für SiC Transistor SCT2160KE der Firma Rohm (rechts)

Für diese Datenquelle wurde ein Informationsmodell, die lokale Lightweight-Ontologie aufgebaut. Diese ist in Bild 6 dargestellt. Diese und auch weitere lokale Ontologien können zukünftig iterativ erweitert werden.

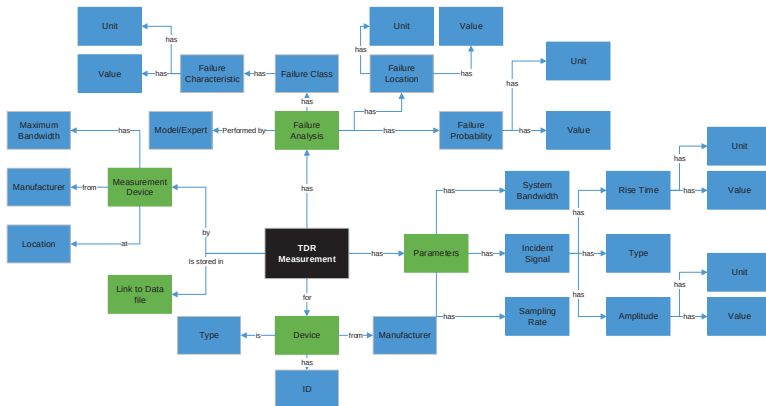


Bild 6: Lokale Lightweight-Ontologie für das Time Domain Reflektometrie-Verfahren

Der Ausgangspunkt der lokalen Ontologie bildet die *Time-Domain Reflektometrie (TDR) Measurement* (schwarz). Die grün dargestellten Sub-Klassen *Measurement Device*, *Link to Data File*, *Device*, *Parameters* und *Failure Analysis* sind für jede Datenquelle zu erwarten und werden von der Ontologie Fehleranalyse wiederverwendet. Die darunter aufgeführten Attribute variieren von Datenquelle zu Datenquelle. Die Eigenschaften von *Measurement Device*, *Link to Data File*, *Device*, *Parameters* werden direkt bei der Datenerfassung befüllt. Die Informationen für die Sub-Klasse *Failure Analysis* werden von der lokalen Datenanalyse geliefert. Im hier gezeigten Beispiel wurde eine hybride Modellierungsmethodik eingesetzt. Dabei wurde ein Convolutional Neural Network in Kombination mit physikalischen Modellen genutzt, um die Attribute der Fehleranalyse (Fehlerklasse, Fehlerort, Fehlerwahrscheinlichkeit und Fehlercharakteristik) automatisiert zu bestimmen und zu befüllen. In Bild 7 wird die befüllte lokale Ontologie für eine beispielhafte offene Verbindung gezeigt. Diese Instanz der lokalen Ontologie wurde in Neo4j realisiert. Die Daten für die lokale Ontologie werden in einer JSON-Datei bereitgestellt.

Die Vorteile der verwendeten Modellierung mithilfe der separaten lokalen Ontologien sind die **Erweiterbarkeit**, **Änderbarkeit** und die **partielle Wiederverwendbarkeit**. Es können neue Datenquellen (z.B. ein neues Messgerät) hinzugefügt werden, bestehende Modelle können angepasst werden und vorhandene Teil-Modellierungen (z.B. für die Sub-Klasse *Device*) können für neue Datenquellen wiederverwendet werden. Diese befüllte lokale Ontologie wird an die globale Ontologie weitergeleitet, in der diese Daten mit bestehenden Daten verknüpft werden. Auf diesem Level können weitere Datenanalysen durchgeführt werden. Dadurch kann

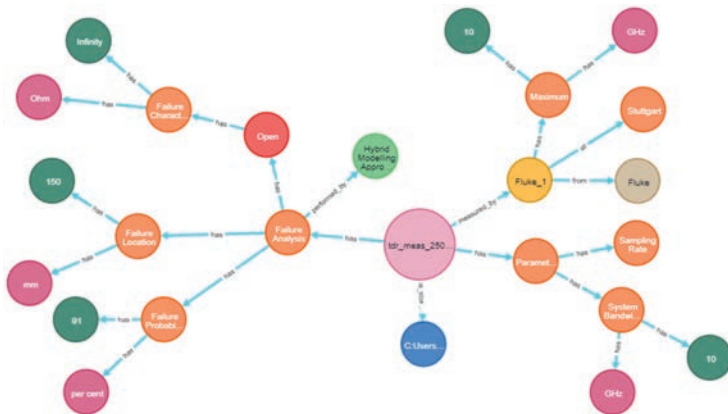


Bild 7: Befüllung der lokalen Lightweight-Ontologie auf Basis einer beispielhaften Time Domain Reflektometrie-Messung

eine globale Fehleranalyse auf Basis aller verfügbaren und verknüpften Daten durchgeführt werden. Die Daten sind für den Benutzer in der Fehleranalyse an einer zentralen Stelle verfügbar, wodurch die Fehleranalyse automatisiert und optimiert werden kann. Die Daten der einzelnen Datenquellen werden dabei automatisiert von der lokalen Datenerfassung mithilfe eines Mappings in den Knowledge Graphen integriert.

5. Zusammenfassung und Ausblick

In diesem Beitrag wurde ein Konzept für die automatisierte Datenintegration für den Fehleranalyseprozess von Halbleiterbauelementen mithilfe von Ontologien und Graphen vorgestellt. Das Konzept basiert auf einem hybriden Ontologie-Modellierungsansatz.

- Der Ansatz ermöglicht eine einfache Erweiterbarkeit, Änderbarkeit und partielle Wiederverwendbarkeit, um neue Datenquelle hinzuzufügen zu können.
- Einzelne lokale Datenquellen können mit weiteren Datenquellen einfach verbunden werden
- Der Ansatz ermöglicht sowohl eine lokale als auch eine globale Datenanalyse, in dem mehrere Datenquellen kombiniert für die Analyse genutzt werden können
- Es können über einen zentralen Zugriff alle verknüpften Daten abgefragt werden
- Weitere externe Datenquellen (z.B. Datenblätter aus dem Internet oder verfügbare Produktionsdaten) können mit dem aufgebauten Knowledge Graphen verknüpft werden, um weitere Informationen einzubinden.

Zukünftig wird weiter die automatisierte Ontologieerstellung und Wiederverwendung mithilfe von Ontology Design Pattern und Methoden wie Ontology Matching oder Ontology Alignment untersucht. Dadurch soll der Aufwand für die Entwicklung der Ontologien reduziert werden. Die globale Datenanalyse auf Basis des Knowledge Graph ist ebenfalls Teil zukünftiger Arbeit. Dafür werden aktuelle Ansätze wie Knowledge Graph Embeddings und Graph Neural Networks untersucht, mit deren Hilfe Daten aus unterschiedlichsten Datenquellen gemeinsam für die Fehleranalyse genutzt werden sollen, um das Ergebnis zu verbessern.

Acknowledgement

Diese Arbeit wurde vom deutschen Bundesministerium für Bildung und Forschung (BMBF) im Projekt FA4.0 gefördert. Dieses Projekt ist eine gemeinsame Initiative mit tschechischen, französischen und schwedischen Konsortien in den europäischen EUREKA-Clustern EURIPIDES2 und PENTA.

6. Literaturverzeichnis

- [1] J. Falck *et al.*, "Reliability of power electronic systems: An industry perspective," *IEEE Industrial Electronics Magazine*, vol. 12, no. 2, pp. 24–35, 2018.
- [2] Toshiba Electronic Devices & Storage Corporation, "Reliability Handbook," Toshiba, 2018. Accessed: Apr. 20 2021. [Online]. Available: <https://toshiba.semicon-storage.com/content/dam/toshiba-ss/shared/docs/design-support/reliability/reliability-handbook-tdsc-en.pdf>
- [3] R. Isermann, *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*: Springer Science & Business Media, 2005.
- [4] C. M. Furse *et al.*, "Fault diagnosis for electrical systems and power networks: A review," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 888–906, 2020.
- [5] F. Altmann *et al.*, "Failure analysis techniques for 3d packages," in *2018 IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, 2018, pp. 1–8.
- [6] C. Feilmayr and W. Wöß, "An analysis of ontologies and their success factors for application to business," *Data & Knowledge Engineering*, vol. 101, pp. 1–23, 2016.
- [7] C. Hildebrandt *et al.*, "Ontology Building for Cyber-Physical Systems: Application in the Manufacturing Domain," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 3, pp. 1266–1282, 2020.
- [8] M. Hepp, "Ontologies: State of the art, business potential, and grand challenges," *Ontology Management*, pp. 3–22, 2008.

- [9] L. Ehrlinger and W. Wöß, "Towards a Definition of Knowledge Graphs," *SEMANTiCS (Posters, Demos, SuCCESS)*, vol. 48, pp. 1–4, 2016.
- [10] A. Singhal, *Introducing the Knowledge Graph: things, not strings*. [Online]. Available: <https://blog.google/products/search/introducing-knowledge-graph-things-not/> (accessed: Apr. 20 2021).
- [11] M. Galkin *et al.*, "Enterprise Knowledge Graphs: A Semantic Approach for Knowledge Management in the Next Generation of Enterprise Information Systems," in *ICEIS (2)*, 2017, pp. 88–98.
- [12] V. Jirkovsky and M. Obitko, "Semantic Heterogeneity Reduction for Big Data in Industrial Automation," *ITAT*, 2014.
- [13] T. Mueller *et al.*, "Cyber-Physical Production Systems: enhancement with a self-organized reconfiguration management," *Procedia CIRP*, 2020.
- [14] L. Wang, "Heterogeneous data and big data analytics," *Automatic Control and Information Sciences*, vol. 3, no. 1, pp. 8–15, 2017.
- [15] B.-H. Yoon *et al.*, "Use of graph database for the integration of heterogeneous biological data," *Genomics & informatics*, vol. 15, no. 1, p. 19, 2017.
- [16] G. de Giacomo *et al.*, "Using ontologies for semantic data integration," in *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years*: Springer, 2018, pp. 187–202.
- [17] M. Cheatham and C. Pesquita, "Semantic data integration," in *Handbook of big data technologies*: Springer, 2017, pp. 263–305.
- [18] M. Sabou *et al.*, "Semantic web technologies for data integration in multi-disciplinary engineering," in *Multi-disciplinary engineering for cyber-physical production systems*: Springer, 2017, pp. 301–329.
- [19] M. Compton *et al.*, "The SSN ontology of the W3C semantic sensor network incubator group," *Journal of Web Semantics*, vol. 17, pp. 25–32, 2012.
- [20] L. Daniele *et al.*, "Created in close interaction with the industry: the smart appliances reference (SAREF) ontology," in *International Workshop Formal Ontologies Meet Industries*, 2015, pp. 100–112.
- [21] M. de Roode *et al.*, "SAREF4INMA: a SAREF extension for the industry and manufacturing domain," *Semantic Web*, Preprint, pp. 1–16, 2020.
- [22] N. Sahlab *et al.*, "Dynamic Context Modeling for Cyber-Physical Systems Applied to a Pill Dispenser," in *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2020, pp. 1435–1438.

- [23] S. Schabus and J. Scholz, "Semantically Annotated Manufacturing Data to support Decision Making in Industry 4.0: A Use-Case Driven Approach," in *Data Science-Analytics and Applications*: Springer, 2017, pp. 97–102.
- [24] A. Fishkin, "Industrial knowledge graph at siemens," in 2017.
- [25] D. Ostrowski *et al.*, "Integration of big data using semantic web technologies," in 2016 *IEEE Tenth International Conference on Semantic Computing (ICSC)*, 2016, pp. 382–385.
- [26] B. Maschler *et al.*, "Deep industrial transfer learning at runtime for image recognition," *at-Automatisierungstechnik*, vol. 69, no. 3, pp. 211–220, 2021.

Wie IIoT, Predictive Maintenance und proaktiver Support den Einsatz von Analysatoren und komplexer Sensorik ermöglichen

Dr. Markus Nick, Fabio Evola,
Endress+Hauser Group Services;
Lars Bondzio, Achim Gahr, André Lemke, Sebastian Schmidt,
Endress+Hauser Liquid Analysis

Kurzfassung

Aktuelle Megatrends treiben die Komplexität der Messgeräte in der Prozessindustrie und damit auch den Bedarf an entsprechender Expertise und Erfahrung. Damit ergibt sich die Resilienz eines Prozesses oder einer Anlage aus der Gesamtheit des sozio-technischen Systems [4]. Mangels Kompetenzen beim Betreiber geht dieses sozio-technische System über die Grenzen des betreibenden Unternehmens hinaus. Das Papier beschreibt Konzept und Pilotierung eines solchen soziotechnischen Systems zwecks Sicherstellung von Resilienz über Unternehmensgrenzen hinweg. Neben der neuartigen Kombination ergeben sich damit auch Chancen für neue Geschäftsmodelle.

1. Einleitung

Aktuelle Megatrends treiben die Komplexität der Messgeräte in der Prozessindustrie. So führt z.B. der Megatrend „Sustainability“ in der Umweltbranche zum Einsatz komplexerer Sensorik und Analysatoren, die ihren Ursprung in der Laboranalytik haben. Vor Ort fehlt insbesondere bei kleinen Anlagen die Expertise und Erfahrung, um solche Geräte im Prozess korrekt und effizient zur Realisierung des Mehrwerts zu betreiben. Damit ergibt sich die Resilienz eines Prozesses oder einer Anlage aus einer Vielzahl von Resilienzen des dahinter befindlichen sozio-technischen Systems [4]. Der Mangel an Kompetenzen beim Betreiber wird versucht durch die Einbindung von externen Ressourcen zu kompensieren. Dadurch erweitert sich das sozio-technische System über die Grenzen des betreibenden Unternehmens hinaus.

Im BMBF-Projekt CheckMATE [8] wurden kleinste erfahrungsbasierte Regelkreise auf Prozessschritzebene über eine Kombination von fallbasiertem Schließen und in Produktions- und Wartungsprozess integriertes Wissensmanagement realisiert. Ein Seeding der Fallbasis erfolgte über FMEA. Dies diente der Reduktion von Prozessdrifts mit Blick auf eine Quasi-Null-Fehler-Produktion in der Stückgutfertigung. Die Idee der engen Integration von

Wissensmanagement in Produktions- und Arbeitsprozesse ist eine der Grundlagen für unseren Ansatz. Das im BMBF-Projekt ProduFlexil [9] entwickelte Konzept und Prototyp fokussierten auf die Überprüfung von Änderungen an der Anlagensoftware. Neuartige Änderungen an Anlagenaufbau und Anlagensoftware wurden erst mit Simulation und/oder Freigabe durch Herstellerexperten zugelassen. Dies sollte u.a. im sogenannten Betreibermodell/XaaS das Risiko des Herstellers durch unzulässige Änderungen seitens des Anlagennutzers verringern. Diese Grundidee der Reduktion des Risikos durch Konnektivität und Simulation überträgt sich auf heutige IIoT in Kombination mit präventiver und prädiktiver Wartung.

Unser Ansatz stärkt die Resilienz über eine Kombination von IIoT, präventiver und prädiktiver Wartung, Remote Support und Wissensmanagement. IIoT und cloud-basierte Systeme ermöglichen die kontinuierliche Kollaboration. Remote Support und Wissensmanagement bringen hierbei Expertise über Unternehmensgrenzen hinweg effizient vor Ort. Präventive und prädiktive Wartung „gießen“ Expertise in Algorithmen, die Probleme vermeidet. Die Gesamtheit der Maßnahmen stellt dem Betreiber Assistenzsysteme und Expertise zur Verfügung, die zu einer kosteneffizienten Erhöhung der Verfügbarkeit der Anlage und korrekter Funktion der Messgeräte beiträgt. Dadurch erhöht sich schlussendlich die Resilienz des sozio-technischen Systems.

Damit einher geht auch eine Änderung des Geschäftsmodells. Der Messgerätehersteller kann damit eine stärkere Verantwortung übernehmen. In diesem Kontext wird aus dem unbezahlten technischen Support ein Service-Produkt, das nicht nur den Feldservice in die digitale Welt transportiert, sondern auch eine kontinuierliche Überwachung und proaktive Unterstützung realisiert. Auf dem Weg in Richtung XaaS/Betreibermodell ist dies ein wichtiger Schritt [6,7].

Abschnitt 2 stellt das Gesamtkonzept und seine Elemente vor. Abschnitt 3 berichtet, wie der Ansatz mit kolorimetrischen Analysatoren für Wasser und TDLAS-Gas-Analysatoren¹ mit Kunden von Endress+Hauser erfolgreich validiert und pilotiert wurden. Abschnitt 4 fasst die Ergebnisse zusammen und gibt einen Ausblick.

¹ TDLAS = tunable diode laser absorption spectroscopy

2. Soziotechnisches System “Connected Support” über Unternehmensgrenzen hinweg

2.1 Übersicht

Unser Ansatz löst die Problematik über eine Kombination verschiedener existierender und neuer Methoden und Technologien. Die Digitalisierung ist hierbei der „Enabler“. Das folgende Diagramm gibt einen Überblick über unseren Ansatz.

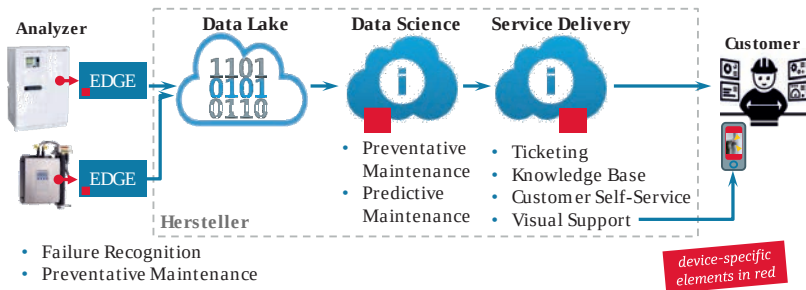


Bild 1: Soziotechnisches System für Resilienz über Unternehmensgrenzen hinweg.

IIoT-Konnektivität bringt Daten über einen zweiten, sicheren Kanal zu den Experten des Herstellers [2,3]. Damit kann der Hersteller den Betreiber proaktiv unterstützen und nicht durchgeführten einfachen Wartungsmaßnahmen effektiv nachzugehen.

Präventive/vorbeugende Wartung und prädiktive/vorhersagende Wartung (Predictive Maintenance) stärken die Resilienz durch die Vermeidung von Problemen. Die Algorithmik für präventive und prädiktive Wartung wird je nach Prozessnähe und Komplexität im Gerät oder in der Cloud platziert. Gerade die Ergebnisse von komplexeren Algorithmen wie KI bedürfen i.d.R. Erläuterungen durch Experten – auch um das Vertrauen in die Ergebnisse herzustellen. Beispiele dafür sind die Vorhersage des Reinigungsbedarfs eines Photometers, Health-Status-Übersichten und Messwertplausibilitätsreports. Details in Abschnitt 2.2.

Remote Support ermöglicht die Effizienz: (1) Vor-Ort-Serviceeinsätze werden reduziert, indem Diagnose oder sogar Problembehebungen fernassistiert durchgeführt werden. (2) Remote Support hilft auch den Technikern des Betreibers den Umgang mit neuen komplexen Geräten „on the job“ zu lernen statt durch die neue Technologie überfordert zu sein. Wichtig für die Effizienz ist „*Visual Support*“ [5], d.h. die Nutzung der Smartphone-Kamera eines Technikers

vor Ort, um dem „Remote“-Experten die Situation vor Ort zu zeigen und sich zeigen zu lassen, an welcher Stelle was zu tun ist (siehe Bild 2). (3) Remote Monitoring ist unabdingbar, um präventive und prädiktive Wartung in Kombination mit Support und Beratung seitens des Herstellers einzubringen.

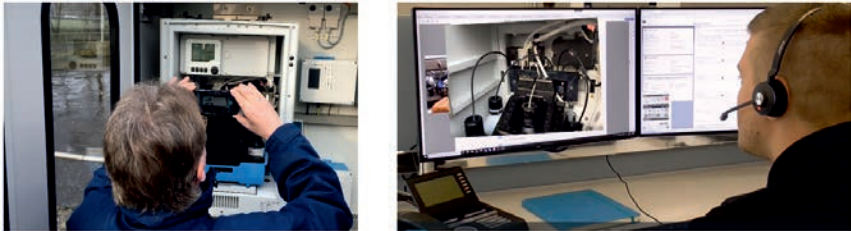


Bild 2: Visual Support gibt dem Support-Techniker Augen vor Ort.

Ein *Knowledge-Management-System* stellt dem Kunden Anleitungen und Videos für einfache Wartungsmaßnahmen im Self-Service zur Verfügung. Aus dem Supportprozess heraus wird das Wissen kontinuierlich erweitert und gepflegt [1]. Details in Abschnitt 2.3

Die Gesamtheit der Maßnahmen stellt dem Betreiber Assistenzsysteme und Expertise zur Verfügung, die zu einer kosteneffizienten Erhöhung der Verfügbarkeit der Anlage und korrekter Funktion der Messgeräte beiträgt. Dadurch erhöht sich die Resilienz des sozio-technischen Systems.

2.2 Präventive und prädiktive Wartung

Präventive und prädiktive Wartung stärken die Resilienz durch die Vermeidung von Problemen. Die zugehörige Algorithmik ist damit ohne unmittelbaren Einfluss auf die Produktion. Damit kann die Algorithmik auch entfernt vom Prozess installiert werden.

Der Platz für die Algorithmik wurde bei uns mit folgenden Kriterien festgelegt:

- *Einfachheit der Nutzung der Information:* Sofern der Betreiber die Information einfach verstehen und nutzen kann, wird die Algorithmik und Anzeige im Gerät platziert.
- *Wartbarkeit:* Updates von Firmware im Messgerät im Feld sind mit hohen Kosten verbunden. Updates an einem zentralen Platz in der Cloud sind hingegen einfach und mit relativ geringen Kosten verbunden.

- **Geschäftsmodell:** Ein differenziertes Lizenzmodell für z.B. Subscription-Modelle in Kombination mit Services ist in der Cloud einfacher zu realisieren.

Da die Ergebnisse von komplexeren Algorithmen wie i.d.R. Erläuterungen durch Experten bedürfen, ist präventive und prädiktive Wartung in unserem Ansatz inhärent mit Remote Monitoring verbunden.

Die Vorhersagen und Wartungsintervalle müssen Reaktionszeiträume für solche Aktivitäten lassen. Selbst bei klaren Aktivitäten wie Reinigung in Folge von Verschmutzung muss Zeit für Reaktion durch Betreiber und ggf. Unterstützung durch Hersteller vorgesehen werden.

2.3 Wissensmanagement als „Enabler“ für Informationsbereitstellung und systematischen Wissensaufbau

Wissensmanagement in Kombination einem Customer-Self-Service ermöglicht dem Hersteller eine Verlagerung der Supportarbeit von einfachen Themen, die Betreiber selbst lösen können, auf höherwertige Arbeiten, die insbesondere einen Mehrwert für den Betreiber darstellen. Weiterhin reduziert Wissensmanagement beim Ticketing die Zeit zur Lösungsfindung. Einer der wichtigsten Erfolgsfaktoren für Wissensmanagement ist dessen Integration in den Arbeitsprozess bzw. die Arbeitsweise.

Die Methodik Knowledge-Centered-Service (KCS) ermöglicht durch eine enge Integration von Wissensmanagement in den Arbeitsprozess die Wissenserfassung und -pflege als Nebenprodukt der Supportarbeit [1]. Die Methodologie existiert seit über 25 Jahren und wird von verschiedensten Unternehmen erfolgreich eingesetzt - z.B. SAP, Intel, Rockwell, Oracle, Ericsson, HP, DELL Technologies, Salesforce, Schneider Electric. Bei Endress+Hauser werden mittlerweile über 40% der Tickets im After-Sales-Support nach KCS bearbeitet. Die Knowledge-Base umfasst mehr als 15000 Artikel in Englisch. Dazu kommen Übersetzungen in verschiedenste Sprachen - für einige Sprachen auch maschinell übersetzt.

Für Analysatoren im Fokus der Evaluation wurden weiterhin systematisch Artikel erstellt, um Zusatzinformationen zur Betriebsanleitung anzubieten. Dazu gehören auch Videos, die zu nachweislich weniger Fehlern in der Durchführung führen. Im Serviceportal werden Artikel mit einem gewissen Reifegrad den Betreibern direkt zur Verfügung gestellt.

Die Diagnose in Gerät und/oder Cloud zusammen mit der Verbindung zum Wissensmanagement ermöglichen es, den Betreiber bei einem Vorfall/Wartungsbedarf automatisch mit aktuellen Zusatzinformationen zu versorgen. Weiterhin kann damit auch nachvollzogen werden, ob der Betreiber in der Lage war, das Problem zu beheben.

3. Erfahrungen aus Evaluation mit Pilotbetrieb

3.1 Ablauf der Evaluation

Ziel der Evaluation war die Analyse der Relevanz und des Business-Modells für Messgeräte an kritischen Stellen im Prozess. Das Modell sollte sowohl für wartungsintensive als auch für wartungsarme Geräte untersucht werden. Als wartungsintensive Messgeräte wurden kolorimetrische Analysatoren für Wasser/Abwasser, als wartungsarme Messgeräte TDLAS-Gas-Analysatoren gewählt. Der Ansatz wurde validiert mit Kunden von Endress+Hauser.

Die Evaluation lief in drei Stufen ab. Die Stufen orientierten sich an Risiko-Aspekten und Aufwand. In der ersten Stufe wurden Kundenbedarf und Customer Journey / Use Cases über Interviews mit Betreibern validiert. Weiterhin wurden Erfahrungen aus dem Betrieb von mehreren Niederlassungen von Endress+Hauser analysiert. Damit war eine erste Anpassung des Konzeptes möglich. In der zweiten Stufe wurde ein mehrmonatiger Pilotbetrieb gefahren. Der Pilotbetrieb diente der Validierung der Customer Journey und Customer Experience (CX) einerseits und der Technologie andererseits. Nach Abschluss des Piloten wurde abschließend die Kaufbereitschaft abgefragt.

3.2 Aufbau für Pilotbetrieb

Für Connectivity und eine sichere Cloud wurde das Endress+Hauser-eigene „Netilion“ verwendet [2]. Für Support und Service wurde die Service Cloud von Salesforce verwendet. Für Visual Support wurde SightCall wegen der vorhandenen Integration in Salesforce verwendet [5]. Die Algorithmik wurde für den Pilot prototypisch implementiert. Der Einsatz von verfügbaren Standardkomponenten beschleunigte die Entwicklung.

3.3 Kolorimetrische Analysatoren

Gerätetyp kolorimetrischer Analysator: Für die kolorimetrische Analyse ist Bereitstellung einer feststofffreien Probe notwendig. Dafür kommt eine applikationsspezifischen Probenaufbereitung zum Einsatz. Ein Teil des gewonnenen Permeats wird Probenaufbereitungssystem in die Misch-/ Reaktionskammer des Analysators gefördert. Das spezifische Farbreagenz wird in

einem definierten Mischungsverhältnis exakt zudosiert. Durch die chemische Reaktion entsteht die charakteristische Färbung der Probe. Ein multi-spektrales Photometer ermittelt die von der Probe bzw. der entstandenen Färbung verursachte Absorption bei definierten Wellenlängen. Die ausgewerteten Wellenlängen, als auch deren Verhältnisse zueinander, sind parameterspezifisch. Mit der Stärke der Absorption des Lichtes kann durch die Proportionalität direkt auf die Konzentration des Analyseparameters in der Probe rückgeschlossen werden.

Einsatzbereich und Branchenbedarf: Solche Geräte werden z.B. zur Online-Überwachung verschiedener Parameter mit Ammonium, Ortho-Phosphat oder chemischen Sauerstoffbedarf (CSB) bei der Wasser- und Abwasseraufbereitung eingesetzt. Für eine schnelle Reaktion auf tages- und jahreszeitliche Schwankungen der Abwassermatrix im Zulauf sind solche Geräte unabdingbar, da die herkömmliche Steuerung über eine Probenentnahme und Laboranalyse erst nach ca. 1-2 Tagen greift und damit bereits viel Abwasser nur suboptimal aufbereitet wurde und damit die Umwelt unnötig belastet oder unter Umständen gesetzlich definierte Einleitergrenzwerte nicht eingehalten werden können. In vielen Ländern wird solche komplexe Online-Messtechnik in der Ausbildung für die Umweltbranche nicht hinreichend adressiert. Kommunal betriebene Kläranlagen müssen oft mit wenig Personal auskommen und die Kosten der Abwasseraufbereitung niedrig zu halten. Damit ist es unmöglich, Spezialisten für Messtechnik im eigenen Personal zu haben oder aufzubauen.

Technologien für unseren Ansatz: Neben den allgemeinen Verfahren des Ansatzes aus Abschnitt 2 wurden die folgenden Verfahren eingesetzt:

- Die kolorimetrischen Analysatoren von Endress+Hauser verfügen über einfache Mechanismen im Gerät, um Fehlersituationen und Wartungsbedarf zu erkennen und die Wartung über Diagnosecodes anzufordern. So erkennt z.B. eine konduktive Füllstandserkennung am Probevorlagegefäß das Fehlen von genügend Prozessflüssigkeit, und eine auf einem Zähler basierende Software-Funktion warnt rechtzeitig bevor eine Betriebsflüssigkeit (Reagenz/Standard/Reiniger) verbraucht ist. Solche Diagnosemeldungen werden allerdings gerade bei neuen Installationen anfangs oft vom Kunden „übersehen/ignoriert“. Das führt dann zu falschen oder fehlenden Messungen mit entsprechenden Konsequenzen für den Prozess.
- Das Spiegeln des Wartungsbedarfs in die Cloud und zum Support ermöglicht es dem Hersteller, dem Betreiber beim Eintreten eines solchen Wartungsfalls über ein Knowledge-Management-System auch Anleitungen und Videos für einfache Wartungsmaßnahmen im Self-Service zur Verfügung zu stellen. Aus dem Supportprozess heraus

wird das Wissen gepflegt und aktualisiert [3]. Dem Betreiber werden diese Informationen dann über einen geeigneten Kanal gepusht, wenn der Fall eintritt. Weiterhin kann der Hersteller nicht durchgeführten einfachen Wartungsmaßnahmen effektiv nachzugehen und den Betreiber proaktiv zu unterstützen.

- Komplexere Algorithmen betreffen z.B. die Vorhersage der Verschmutzung des Photometers in einem Analysator und dem damit verbundenen Wartungsbedarf. Der eingesetzte Algorithmus nutzt die im Gerät vorhandenen Logdaten. Dazu wird bei jeder Messung die Rohabsorption des Photometers (vor Farbumschlag) mit der für den Prozess typischen verglichen. Eine erhöhte Absorption deutet dabei auf eine Verschmutzung in der Küvette. Zur Zeit der Pilotierung war die automatische Übertragung der Logdaten in die Cloud noch nicht umgesetzt. Der Algorithmus wurde daher im Supportcenter auf manuell eingeschickten Logdaten von Geräten evaluiert. Dazu wurde der Verlauf der Rohabsorptionen analysiert; bei einer sauberen Küvette sollte sich dabei eine Horizontale ergeben.

Erstes Betriebsjahr mit unserem Ansatz: Das erste Betriebsjahr ist gekennzeichnet von „Adoption“ und Wissenstransfer. Zur Inbetriebnahme vor Ort gehört auch eine Einweisung des Betreibers. Die einfacheren Wartungsmaßnahmen werden vom Betreiber durchgeführt – bei Bedarf mit Unterstützung durch Hersteller via Visual Support. (Bild 3)

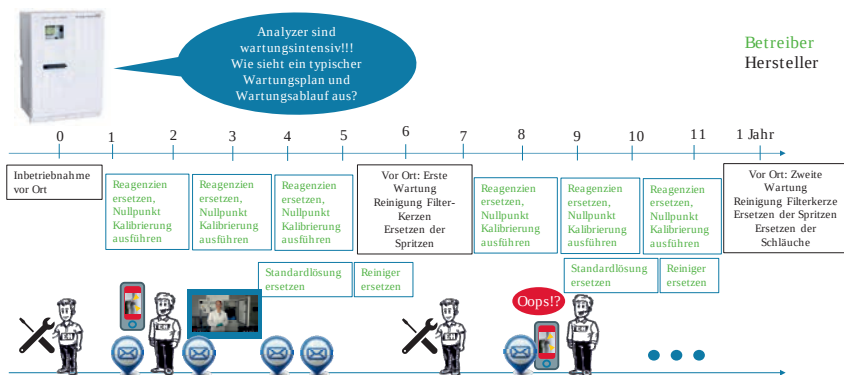


Bild 3: Beispielhafter Ablauf des ersten Betriebsjahr für kolorimetrische Analysatoren

Ergebnisse und Erfahrungen: Die verwendeten Technologien wurden von den drei Pilotteilnehmern gut akzeptiert. Selbst Kontaktaufnahme seitens Messgerätehersteller für einfache

Themen wie Erläuterung von Diagnosecodes wurde als proaktive Kundenerfahrung (CX) gewertet. Das zeigt, dass sich die Resilienz auch gefühlt für den Kunden erhöht hat. Das Geschäftsmodell insgesamt wurde via Kaufbereitschaft als akzeptabel bewertet.

3.4 TDLAS-Gasanalysatoren

Gerätetyp: Gasanalysatoren sind außerordentlich zuverlässig für die Messung von Spurengaskomponenten mithilfe der TDL-Technologie (Tunable Laser Diode). Die TDL-Absorptionsspektroskopie (TDLAS) ist eine hochauflösende Infrarottechnik, mit der bestimmte Gase präzise gemessen werden können, während Interferenzen vermieden werden, die bei herkömmlichen Infrarotanalysatoren häufig auftreten. Konkret können Gasanalysatoren Feuchtigkeit (H_2O), Kohlendioxid (CO_2), Schwefelwasserstoff (H_2S), Ammoniak (NH_3), Acetylen (C_2H_2), Sauerstoff (O_2) und vieles mehr messen.

Einsatzbereich und Branchenbedarf: TDLAS Gasanalysatoren werden eingesetzt, um die Prozesseffizienz, den Durchsatz und die Sicherheit für die Erdgasproduktion, den Transport, die Lagerung, den Vertrieb, die Gasverarbeitung, die Raffination und die Spezialgasherstellung zu verbessern. Die extraktiv arbeitende TDLAS-Technologie besticht durch genaue und robuste Messungen mit extrem schnellen Reaktionszeiten, was zu einer verbesserten Leistung, Zuverlässigkeit und Verfügbarkeit bei deutlich geringeren Gesamtkosten für Installation und Besitz im Vergleich zu alternativen Lösungen führt.



Dadurch hat sich TDLAS zur bevorzugten, optisch basierten Prozessanalysetechnologie sowohl in der Öl- und Gas- als auch in der chemischen Industrie entwickelt.

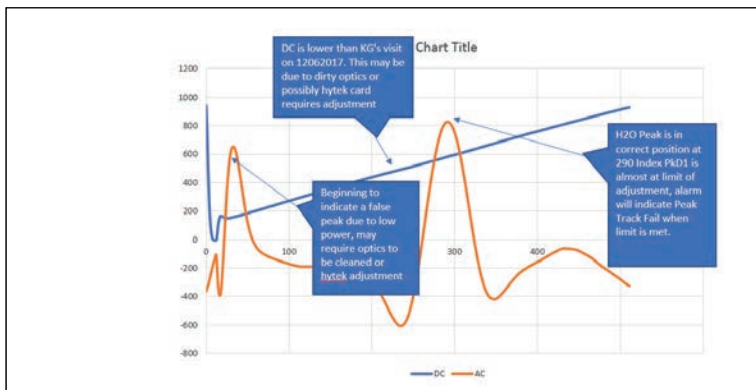
Gasanalysatoren sind in der Regel wartungsarm, das heißt, es bedarf im Regelbetrieb keines regelmäßigen, präventiven Ersatzes bestimmter Komponenten. Lediglich die Laserdiode sollte nach einer Betriebszeit von etwa 5 Jahren ausgetauscht werden.

In vielen Ländern wird solche komplexe Online-Messtechnik in der Ausbildung für die Öl, Gas- und Chemiebranche nicht hinreichend adressiert. Chemisches und physikalisches Spezialwissen sind zwingend erforderlich, um diese komplexe Technologie genauestens auf die Produktionsprozesse anzupassen. Es zeigt sich, dass Betreiber nur wenige oder gar keine Spezialisten für Messtechnik im eigenen Personal haben oder aufbauen.

Technologien für unseren Ansatz:

Neben den allgemeinen Verfahren des Ansatzes aus Abschnitt 2 wurden die folgenden Verfahren eingesetzt:

- Messwerte aus dem laufenden Betrieb sowie geräte-spezifische Informationen zu den Gasanalysatoren von Endress+Hauser können über eine proprietäre Softwareanwendung ausgelesen und analysiert werden. Aktuelle Messdaten werden dafür graphisch visualisiert. Dabei können sowohl Erkenntnisse über ein mögliches Fehlverhalten des Analysators als auch klare Aussagen über Prozess-Störungen mit entsprechendem Zeitstempel und möglicher Ursache getroffen werden.
- Die Messdaten wurden in 2 Szenarien analysiert:
 - o Regelmäßige, proaktive remote Überwachung der Analysatoren durch Endress+Hauser mit anschließendem „Health status“-Bericht an die Kunden
 - o Reaktive Analyse der Messdaten nach Anfrage der Kunden, im Falle, dass die vom Kunden festgelegten Grenzwerte für die Messungen über- oder unterschritten werden. Auch hier erfolge ein schriftlicher Bericht mit entsprechenden Handlungsanweisungen



Lebenszyklus mit unserem Ansatz: Das erste Betriebsjahr ist gekennzeichnet von „Adoption“ und Wissenstransfer. Zur Inbetriebnahme vor Ort gehört auch eine Einweisung des Betreibers. Dadurch, dass die Geräte jedoch wartungsarm und dabei sehr langlebig sind (Betriebszeit i.d.R. über 15 Jahre) hat sich allerdings gezeigt, dass die Betreiber im Laufe des Lebenszyklus

nur sehr wenige Berührungspunkte und damit wenig Wissen über die Technologie im Betriebspersonal erhalten können. Im Falle einer Störung fehlt die Expertise, die Situation richtig zu bewerten und oftmals wird die Verlässlichkeit der Messung des Analysators in Frage gestellt. Hier kam dann die Remote Unterstützung durch den Endress+Hauser Technischen Support ins Spiel. Durch die "Health status"-Berichte konnte zum einen die Verlässlichkeit der Messung selbst glaubhaft belegt werden, zum anderen konnte aber durch die graphische Analyse mögliche Ursachen für die Störung erläutert werden. Weitere Unterstützung durch den Hersteller zur Fehlerbehebung erfolgte „remote“ durch Anleitung des Betriebspersonals. Ad-hoc-Feldeinsätze durch Endress+Hauser-Spezialisten konnten dadurch vollständig vermieden werden.

Ergebnisse und Erfahrungen: Die verwendeten Technologien wurden von den drei Pilotteilnehmern gut akzeptiert. Die regelmäßigen Statusberichte und Handlungsempfehlungen wurden als positive Kundenerfahrung (CX) gewertet. Das zeigt, dass sich die Resilienz auch gefühlt für den Kunden erhöht hat. Das Geschäftsmodell insgesamt wurde via Kaufbereitschaft als akzeptabel bewertet.

4. Zusammenfassung und Ausblick

Zunehmende Komplexität der technischen Gesamtsysteme führt dazu, dass mehr und mehr Spezialwissen erforderlich ist, um den Mehrwert zusätzlicher Messsysteme zu realisieren. Kleinere und mittelgroße Standorte haben Zugang zu solchem Spezialwissen nur durch Mitarbeiter ihrer Zulieferer - d.h. im konkreten Fall der Messgerätehersteller.

Digitalisierung ermöglicht automatisierte Fernüberwachung und "digitales Vor-Ort gehen" ohne unproduktive Reisezeiten. Diese Kombination ermöglicht es Spezialisten auf neuartige Art- und Weise mit dem Betreiber zusammenarbeiten. Der Betreiber fokussiert auf den Prozess. Der Spezialist fokussiert darauf, dass die von ihm bereitgestellten Elemente zuverlässig funktionieren.

Es zeigt sich, dass Resilienz ist kein pures Technologie-Thema ist. Versäumnisse der menschlichen Komponente können zu suboptimalem Betrieb oder Störungen im Betrieb führen. Die Resilienz einer Anlage ergibt sich aus dem Zusammenspiel von Technologie und Mensch über Unternehmensgrenzen hinweg. Durch digitale Technologien können entsprechende Maßnahmen zur Erhöhung der Resilienz nunmehr in effizienter Weise umgesetzt werden.

Damit einher geht ein Shift im Geschäftsmodell. Die Bedeutung der Remote-Services bewegt sich damit auf das Niveau des Feldservices. Trends wie XaaS und Outsourcing lassen erwarten, dass sich der Shift im Geschäftsmodell bzgl. stärkerer Übernahme von Verantwortung durch die Messgerätehersteller fortsetzt.

Nachfolgend der positiven Evaluation mit Pilotbetrieb beabsichtigt Endress+Hauser die erprobte Lösung kurzfristig zur Marktreife zu entwickeln und global anzubieten.

Literatur

- [1] Consortium for Service Innovation: Knowledge Centered Service (KCS®). www.serviceinnovation.org/kcs/
- [2] Endress+Hauser Netilion <https://netilion.endress.com/>
- [3] NAMUR Open Architecture. <https://www.namur.net/en/focus-topics/namur-open-architecture/>
- [4] *Resilient Control Systems: Next Generation Design Research*; C.G. Rieger, D.I. Gertman, M.A. McQueen. 2nd IEEE Conference on Human System Interaction, Catania, Italy, May 2009.
- [5] SightCall <https://sightcall.com/>
- [6] B4B – How Technology and Big Data Are Reinventing the Customer-Supplier Relationship; J.B. Wood, Todd Hewlin, Thomas Lah. TSIA 2013.
- [7] Why Services is the Key to Digitally Transforming Equipment Manufacturers; Vele Galovski and Harald Kopp. TSIA 2019.
- [8] CheckMATE - Erfahrungsmanagement für Service-Roboter zur Realisierung von Self-Healing in Produktionsanlagen; Markus Nick, Sören Schneickert, Jürgen Grotepaß, and Ingo Heine. In German Journal on Artificial Intelligence (Zeitschrift "Künstliche Intelligenz"); Ausgabe 4, 2007; special issue on "Experience Management"; BöttcherIT Verlag, 2007.
- [9] Design and engineering processes in highly adaptive plants with ambient intelligence techniques; Miriam Schleipen, Michael Baumann, Michael Okon, Martin Neukäufer, Christian Fedrowitz, Martin Feike, Nataliya Popova, Markus Nick, Sören Schneickert, Martin Wessner.

In Proceedings of 42nd CIRP Conference on Manufacturing Systems, June 2009.

Zögerliche Digitalisierung der Geschäftsmodelle Ergebnisse einer repräsentativen Studie

Prof. Dr.-Ing. **Gerrit Sames**,
Technische Hochschule Mittelhessen, Gießen

Kurzfassung

Durch intensive Berichterstattung in vielen Medien ist das Thema Digitalisierung und Industrie 4.0 inzwischen in den deutschen Unternehmen angekommen. Oft wird darunter jedoch primär die Digitalisierung von Geschäftsprozessen verstanden. Noch viel wichtiger aber ist die Frage, in wieweit durch die Möglichkeiten der Digitalisierung bereits Veränderungen in die Geschäftsmodelle eingezogen sind. Die vorliegende Studie hat diese Fragestellung bei mittelständischen Unternehmen untersucht.

Abstract (optional)

Based on multiple publications digitization and industry 4.0 has finally arrived in German enterprises. Mostly it is interpreted as digitization of business processes. However, more important are the opportunities of digitization for business models. We have investigated at which degree chances of digitization for business models have found access in SME's.

1. Ausgangssituation

Beginnend mit der Hannover-Messe 2011 ist der Begriff Industrie 4.0 in die Öffentlichkeit gelangt und wird seitdem vor allem in den Medien lebhaft diskutiert. Dazu gesellte sich dann immer stärker die Begrifflichkeit Digitalisierung. Auffallend ist, dass in Deutschland meist eine sehr produktionszentrierte und prozessorientierte Sicht auf die Thematik in den Vordergrund gestellt wird. Digitalisierte Prozesse sind wichtige „Enabler“ für Geschäftsmodelle. Eine besondere Bedeutung für den wirtschaftlichen Erfolg eines Unternehmens kommt dem Geschäftsmodell zu. Unter Geschäftsmodell soll die Summe aller Aktivitäten eines Unternehmens verstanden werden, um für Kunden einen Mehrwert zu schaffen und damit Geld zu verdienen. Aufstrebende Technologien der Digitalisierung und von Industrie 4.0 sowie veränderte Kundenbedürfnisse sorgen dafür, dass sich neue datengetriebene Dienstleistungsangebote und Wettbewerber auf dem weltweiten Markt etablieren. Demnach sehen sich die deutschen Unternehmen mit einer kontinuierlichen Verschärfung des Wettbewerbsumfelds konfrontiert [1].

In Deutschland verantwortet insbesondere der Mittelstand den wirtschaftlichen Erfolg des Landes [2]. 99,5 % aller deutschen Unternehmen gehören zu der Kategorie kleine und mittlere Unternehmen (Stand 2017). Des Weiteren haben sie im Jahr 2017 mit einem Umsatzanteil von 35,0 % einen maßgeblichen Beitrag zum Gesamtumsatz in Deutschland geleistet [3]. Um weiterhin im weltweiten Wettbewerb bestehen zu können, ist es von entscheidender Bedeutung, dass der Mittelstand sein Geschäftsmodell an das digitale Zeitalter anpasst [4].

2. Umfang und Zielgruppe der Studie

Es wurden insgesamt 759 mittelständische Unternehmen per E-Mail zur Teilnahme an der Online-Befragung eingeladen. Insgesamt haben 107 Unternehmen an der Online-Befragung teilgenommen. Die größte Teilnehmeranzahl mit 37,4 % kommt aus dem Maschinen- und Anlagenbau, einer für Deutschland besonders wichtigen Branche. Bekannterweise spricht man Deutschland die Rolle des Ausrüsters der Welt mit Maschinen und Anlagen zu.

Deutschlands Industrie ist stark mittelständisch geprägt. So arbeiten 72% der Industriebeschäftigten in Unternehmen mit weniger als 1000 Mitarbeitern. In unserer Studie liegt der Anteil mit 87% sogar noch darüber (s. Bild 1) und ist ein klarer Beleg für eine stark mittelständisch geprägte Struktur.

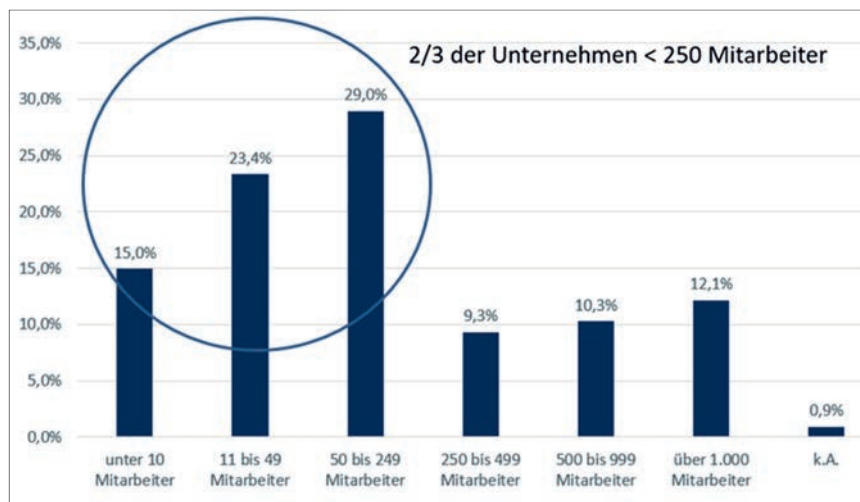


Bild 1: Verteilung Unternehmen nach Firmengröße

3. Services als Bestandteil des Geschäftsmodells

Mit Überraschung mußten wir feststellen, dass bei den Teilnehmern der Studie 17% gar keinen Umsatz mit Services machen, und 54% höchstens 20 % ihres Umsatzes mit Services erzielen. Der Anteil mit digitalen Services lag mit 36% noch deutlich darunter. Lediglich 7 % der befragten Unternehmen erwirtschaften mithilfe von Services über 40 % ihres Umsatzes. Digitale Services stellen bei einem Großteil der teilnehmenden Unternehmen (79 %) folglich noch immer keine bzw. keine wesentliche Einnahmequelle dar.

4. Digitalisierung im Vertrieb

Die Auswertungsergebnisse zeigen an, dass aktuell 68% der mittelständischen Unternehmen Online-Vertriebskanäle als Geschäftsmodell-Erweiterung noch nicht nutzen. Bild 2 hebt hervor, dass erst 16 % der Unternehmen ihre Produkte über einen eigenen Online-Shop vertreiben. Ähnlich ist das Angebot eines digitalen Showrooms für die eigenen Produkte, der nur in 6 % der Fälle für einen Großteil der Produkte zum Einsatz kommt.

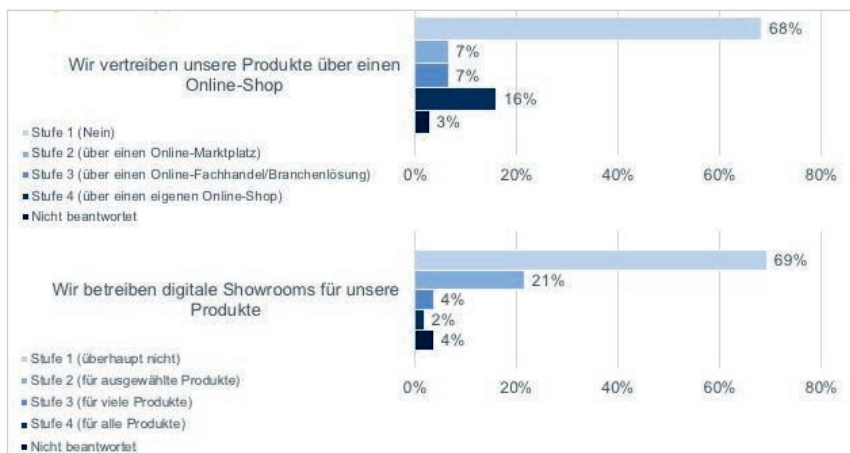


Bild 2: Online-Shop und digitaler Showroom

Auch Chatbots spielen im Mittelstand noch keine signifikante Rolle. 81 % der befragten Unternehmen nutzen noch keinerlei Chatbots. Das gilt sowohl für produktbezogene Anwendungsempfehlungen als auch für Problemlösungen zu den Produkten.

4. Digitale Services

Überraschend mußten wir feststellen, dass Condition Monitoring, Verfügbarkeitsgarantien und auch Predictive Maintenance als Geschäftsmodell-Erweiterung nur in einem kleinen prozentualen Anteil bei den Unternehmen eingezogen ist (s. Bild 3). Das steht in deutlicher Diskrepanz zu den vielen Publikationen zu der Thematik. Offenbar findet sich das als potentielle Einnahmequelle noch nicht in den mittelständischen Unternehmen wieder. Predictive Maintenance ist dadurch gekennzeichnet, dass anhand von Datenanalytik der jeweiligen Maschinen- und Anlagendaten ein optimaler Wartungszeitpunkt für den Kunden ermittelt werden kann.

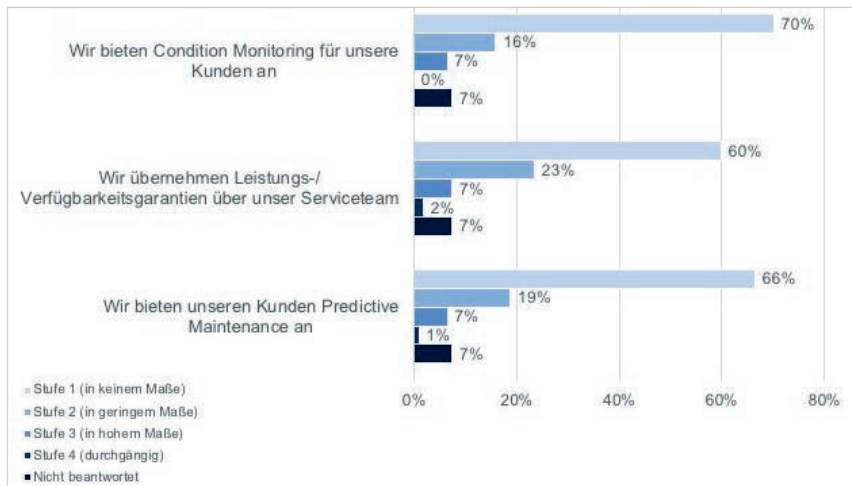


Bild 3: Condition Monitoring und Predictive Maintenance

Auch die Digitalisierung von Informationen zum Produkt läuft noch sehr verhalten. So ist die Bereitstellung eines digitalen Zwillings bei 72% der Teilnehmer noch kein Bestandteil des Geschäftsmodells.

5. Online-Plattformen und Daten

Wenn schon das Geschäftsmodell primär auf den Verkauf von Erzeugnissen an die Kunden abzielt, könnte man aber zusätzlich eine Plattform anbieten, auf der die Kunden ungenutzte Kapazitäten transparent machen. Eine Information über Maschinen-/Anlagenkapazitäten auf Online-Plattformen als Teil des Geschäftsmodells ist aber bei weit über 80 % der Unternehmen noch nicht im Fokus (s. Bild 4). Auch die Hersteller-eigenen Fertigungskapazitäten werden kaum über Online-Plattformen Dritten gegenüber angeboten.

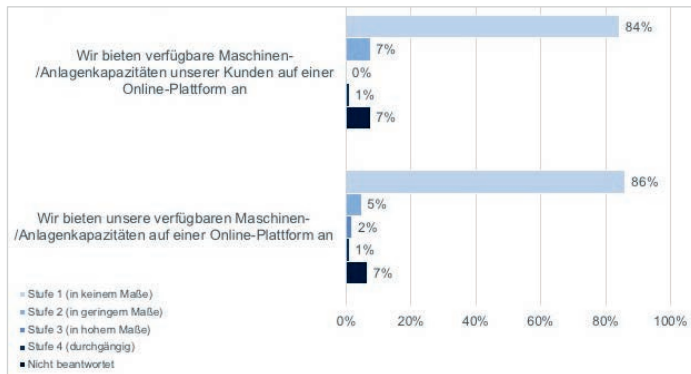


Bild 4: Fertigungskapazitäten auf Online-Plattformen

Deutliches Steigerungspotential ist auch bei der Unterstützung der Kunden durch Erfassung und Analyse von Kundenbetriebsdaten auf einer Online-Plattform vorhanden. Lediglich 15 % der Befragten bieten ihren Kunden diese Möglichkeit an. Eine Integration weiterer Partner des Kunden auf einer Online-Plattform wird nur von 8 % der Unternehmen angeboten.

Auch die weitere Bereitstellung von datenbasierten Informationen für die Kunden hat kaum Einzug in die Geschäftsmodelle gefunden. Lediglich 12 % der Befragten bieten Informationsprodukte auf Basis ihrer gesammelten Daten an. Nur 8 % der teilnehmenden Unternehmen stellen ihren Kunden Benchmarks zur Verfügung.

6. Monetarisierung

Auf Bild 5 kann man deutlich sehen, dass die Monetarisierung in einem Großteil der befragten Unternehmen bislang kaum auf digitalen Ansätzen beruht. Die Maschinen-/Anlagenverfügbarkeit dient nur bei 11 % der Befragten als Abrechnungsgrundlage. Darüber hinaus verwenden lediglich 8 % der teilnehmenden Unternehmen ein nutzungsabhängiges Abrechnungsmodell. Nur bei 13 % der Befragten orientiert sich die Fakturierungsbasis an der Erzielung von vereinbarten Outputs/Resultaten.

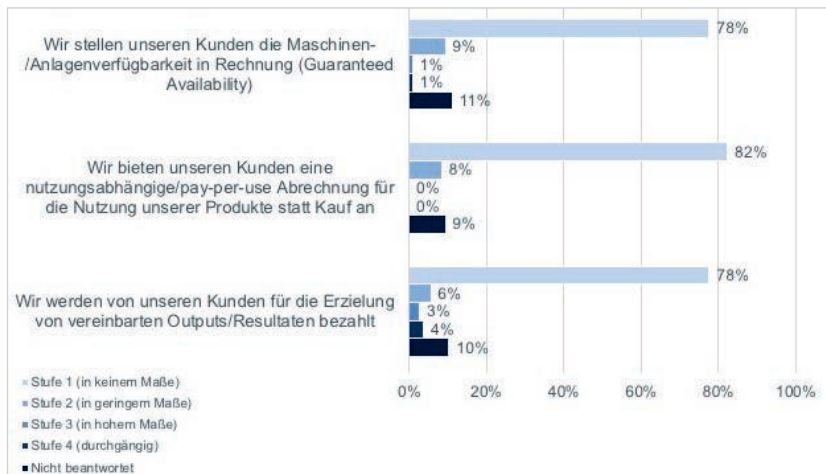


Bild 5: Fakturierungsbasis

Erst 15 % der befragten Unternehmen verlangen für einen Teil des Produktspektrums („in geringem Maße“) einen monatlichen Festpreis für eine vereinbarte Leistung statt des Produktverkaufs. Insofern ist es dann auch nur schlüssig, dass die Blockchain-Technologie für fälschungssichere Abrechnungszwecke bei pay-per-use oder pay-per output Abrechnungen noch keine Rolle spielt. So haben 88 % der befragten Unternehmen angegeben, dass sie derzeit die Blockchain-Technologie nicht für die Abwicklung von Zahlungsprozessen nutzen. Lediglich 4 % verwenden sie in einem geringen Maße für Zahlungsprozesse.

Zusammenfassend lassen die Auswertungsergebnisse der Kategorie „Monetarisierung“ den Schluss zu, dass die konventionelle Monetarisierung (also der Verkauf von physischen Produkten) weiterhin einen wesentlichen Bestandteil des Geschäftsmodells darstellen. Derzeit werden digitale Ansätze dazu nur von einzelnen Vorreitern eingesetzt.

7. Zusammenfassung und Empfehlungen

Um den (durchschnittlichen) Digitalisierungsgrad der 12 Kategorien berechnen zu können, wurde den erhobenen Daten in einem ersten Schritt ein entsprechender Zahlenwert zugeordnet. Die Befragungsdaten erhielten bei Fragestellungen mit vier möglichen Ausprägungsstufen einen Zahlenwert zwischen eins (Stufe 1) und vier (Stufe 4). Bei Fragestellungen mit zwei Ausprägungsstufen konnten die Befragungsdaten den Zahlenwert eins oder vier einnehmen. Anschließend wurde für jede Fragestellung ein Mittelwert berechnet. In einem abschließenden

Schritt wurden die berechneten Mittelwerte der Fragestellungen pro Kategorie summiert und durch die Anzahl der Fragestellungen dividiert, um den Digitalisierungsgrad der einzelnen Kategorien zu ermitteln. Ein Auswertungsergebnis in Höhe von eins repräsentiert den niedrigsten Digitalisierungsgrad. Der Wert vier stellt den maximalen Digitalisierungsgrad dar. Bild 6 zeigt den durchschnittlichen Digitalisierungsgrad der Geschäftsmodell-Erweiterungen in den 12 Kategorien auf.

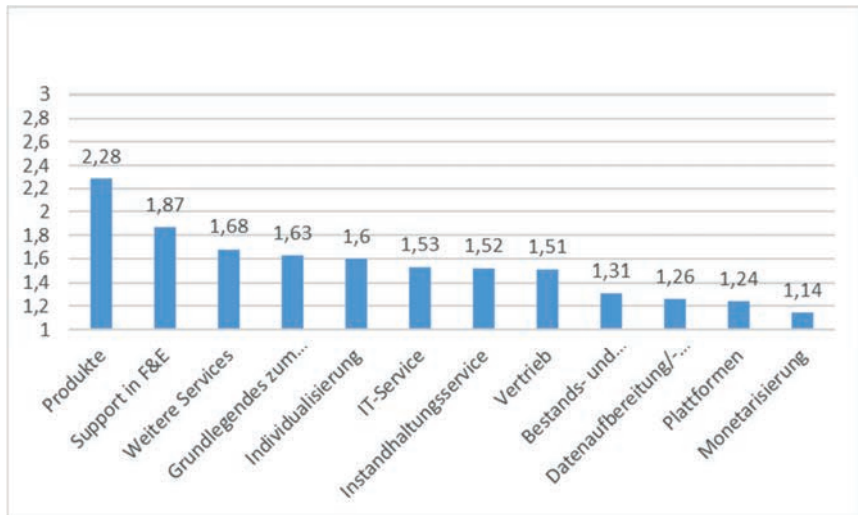


Bild 6: Digitalisierungsgrad von Geschäftsmodellen in den einzelnen Kategorien

Es wird deutlich, dass die Kategorie „Produkte“ mit einem Mittelwert in Höhe von 2,28 den höchsten durchschnittlichen Digitalisierungsgrad aufweist. Nachfolgend findet sich die Kategorie „Support für F&E“ mit einem durchschnittlichen Digitalisierungsgrad in Höhe von 1,87 wieder. Besonders hervorzuheben ist die beachtliche Differenz in Höhe von 0,41 zwischen diesen beiden Kategorien. Den niedrigsten durchschnittlichen Digitalisierungsgrad verzeichnet die Kategorie „Monetarisierung“ mit dem Wert 1,14. Auch das zeigt sehr deutlich auf, dass in den mittelständischen Unternehmen nach wie vor traditionelle Geschäftsmodelle mit dem Verkauf von physischen Produkten im Vordergrund der Geschäftstätigkeit stehen.

Interessant ist auch die Betrachtung des Digitalisierungsgrades nach Branchen (Bild 7).

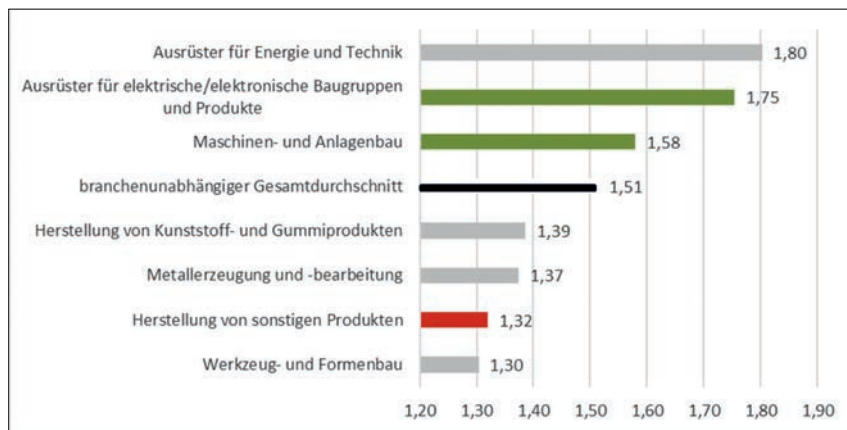


Bild 7: Digitalisierungsgrad von Geschäftsmodellen nach Branchen

Ausrüster für elektrische/elektronische Baugruppen und Produkte verzeichnen den zweithöchsten Digitalisierungsgrad in Höhe von 1,75. Nachfolgend ist der Maschinen- und Anlagenbau mit einem Digitalisierungsgrad von 1,58 vertreten. Der branchenunabhängige Gesamtdurchschnitt beträgt 1,51. Zusammenfassend lässt sich feststellen, dass die Digitalisierung erst in wenigen Branchen eine Bedeutung erlangt hat. Insbesondere die Hersteller von sonstigen Produkten, hinter denen sich eine Vielzahl unterschiedlicher Unternehmen verbirgt, haben noch bedeutenden Spielraum zur Anpassung ihrer Geschäftsmodelle durch Digitalisierung.

Nicht ganz unerwartet hängt der Digitalisierungsgrad sehr deutlich von der Firmengröße ab (s. Bild 8).

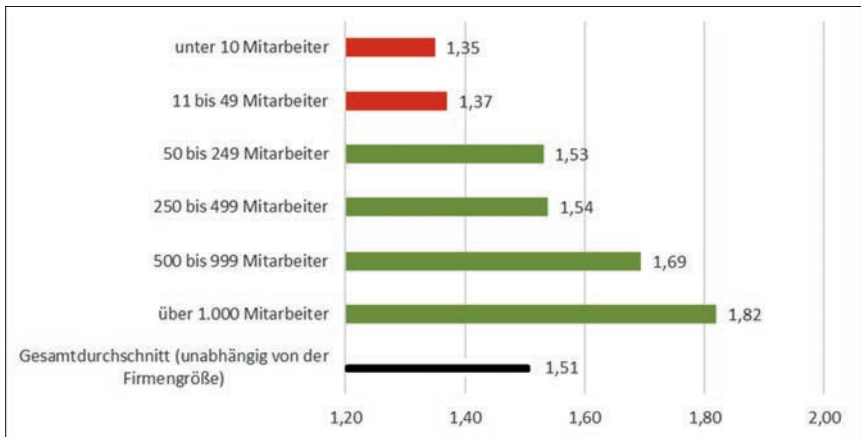


Bild 8: Digitalisierungsgrad nach Firmengrößen

Unternehmen, die mehr als 1.000 Mitarbeiter beschäftigen, weisen den höchsten Digitalisierungsgrad in Höhe von 1,82 auf. Im Vergleich zum maximalen Digitalisierungsgrad in Höhe von 4,00 handelt es sich hierbei aber immer noch um einen niedrigen Digitalisierungsgrad.

Nachfolgend sind die wesentlichen Erkenntnisse und Schlussfolgerungen zusammengefasst:

- Es lässt sich festhalten, dass die Geschäftsmodelle im Mittelstand nach wie vor noch produktlastig sind, der Umsatz also überwiegend durch den Verkauf physischer Produkte erzielt wird. Damit liegen noch große Chancen in der Erschließung von Services.
- Die Individualisierung von Produkten mithilfe von digitalen Tools steckt noch in den Anfängen. Online-Konfiguratoren wären hier eine hilfreiche Erweiterung.
- Online-Shops und digitale Showrooms sind bislang kaum in den Geschäftsmodellen vertreten. Hier kann man dringenden Handlungsbedarf attestieren.
- Weitergehende Unterstützung der Kunden durch digitales Ersatzteilmanagement und Bewirtschaftung von Beständen eigener Produkte beim Kunden ist im Mittelstand noch wenig ausgeprägt und bietet damit noch große Potentiale.

- Plattformen zum transparentmachen von unterausgelasteten Fertigungskapazitäten gibt es im Mittelstand ebenfalls bislang kaum. Auch das sind noch weitgehend unerschlossene Marktpotentiale.

Die vollständige Studie kann kostenfrei abgerufen werden [5].

- [1] Hanschke, I.: Digitalisierung und Industrie 4.0 – einfach und effektiv, Systematisch und lean die digitale Transformation meistern, München, 2018, S. 1.
- [2] Deloitte (Hrsg.): Industrie 4.0 im Mittelstand - Auf der Agenda der Top-Entscheider angekommen, abgerufen unter: https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Mittelstand/Deloitte-Industrie-4.0_%20im_%20Mittelstand_2019.pdf, 2019, S.6
- [3] https://www.ifm-bonn.org/fileadmin/data/redaktion/ueber_uns/ifm-flyer/Ifm-Flyer-2020.pdf
- [4] Sauer, R. / Dopfer, M. / Schmeiss, J. / Gassmann, O.: Das Geschäftsmodell: Gral der Digitalisierung, in: Gassmann, O. / Sutter, P. (Hrsg.): Digitale Transformation gestalten - Geschäftsmodelle, Erfolgsfaktoren, Checklisten, 2. Aufl., München, 2019, S. 19 – 32.
- [5] Sames, G. / Lapa, J.: Stand der Digitalisierung von Geschäftsmodellen zu Industrie 4.0 im Mittelstand; THM-Hochschulschriften Band 13; http://diqdok.bib.thm.de/volltexte/2020/5357/pdf/THM_Hochschulschriften_13_Endfassung.pdf, 6/2020, 55 S.

Smart Resilience Services for Industrial Production

S. Janzen, N. Öksüz,

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI),
Saarbrücken;

J. Sporkmann, M. Schlappa,

WHU - Otto Beisheim School of Management, Vallendar;

J. Gerhard,

Feintool System Parts Jena GmbH, Jena;

L. Ortjohann,

Machine Tools and Production Engineering, RWTH Aachen University,
Aachen;

P. A. Becker, deZem GmbH, Berlin

Kurzfassung

Im Projekt SPAICER sind wir davon überzeugt, dass nur resiliente Unternehmen zukunftsfähig sein können. Ziel des Projekts ist daher die Entwicklung eines datengetriebenen Ökosystems auf Basis lebenslanger und kollaborativer Smart Resilienz-Services (SRS) unter Verwendung führender KI-Technologien und Industrie 4.0-Standards. In diesem Beitrag stellen wir ein Vorgehensmodell zur Spezifikation von SRS vor, die für die industrielle Produktion relevant sind, um ein kontinuierliches Resilienz-Management im Unternehmen zu etablieren. Hierzu kombinieren wir Ansätze aus der Design Science Forschung, dem Requirements Engineering und dem Service Engineering.

Abstract (optional)

In the SPAICER project, we are convinced that only resilient companies can be fit for the future. Thus, the objective of the SPAICER project is the development of a data-driven ecosystem based on lifelong, collaborative and low-threshold Smart Resilience Services (SRS) using leading AI technologies and Industry 4.0 standards. In this paper, we present a method for specifying SRS relevant for industrial production for establishing a resilience business continuity. Therefore, we combine approaches from Design Science, Requirements Engineering and Service Engineering.

1. Motivation

Nowadays in our global and interconnected economy, disruptions in industrial production systems are one of the leading business risks. In general, disruptions and their implications on a production system can be manifold: damaged production facilities resulting from natural catastrophes such as hurricane Katrina, lockdown of economies during the COVID-19 pandemic, trade restrictions due to political conflicts such as the US-China conflict or supply shortages such as those resulting from the Ever-Given container vessel being stuck in the Suez Canal. Hence, managing production systems need to become more and more centered around questions of resilience to withstand the effects of such disruptions. In general, companies are characterized as “resilient” when they are capable of permanently adapting to internal and external changes and disruptions [1].

Tracking the overall performance of the production system, anticipating potential disruptive events and deploying effective measures in advance are among the most crucial capabilities in resilience management of global production systems. However, in a survey carried out by the Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), only 35% of the 60 surveyed company representatives stated that they have a systematic resilience management in their company, whereas the majority of 60% saw an urgent need for action under the circumstances that disruptions can occur at least occasionally.

The SPAICER project directly addresses the companies' urgent need for action with regard to resilience management: SPAICER aims for the development of Smart Resilience Services (SRS). By leveraging leading artificial technologies (AI) and Industrie 4.0 standards, we are confident that SRS can enable companies to anticipate disruptions earlier in time and can help them adapt their production planning accordingly at any time. For instance, in terms of resilience management, the great potential of AI has been investigated in both academia and practice [2,3,4,5].

Yet, despite the general promise of SRS, there is – to the best of our knowledge – currently no standardized approach describing how to develop an SRS. Therefore, SPAICER defined a new design method for the development of an SRS that can immediately be adopted by researchers and practitioners alike. Based on this blueprint we hope to further accelerate the proliferation of SRS in real-world applications.

In this paper, we discuss our 6-step design method for specifying SRS combining proven concepts from Design Science (e.g., Design Thinking), Requirements Engineering (e.g., Service Use Cases) and Service Engineering (e.g., Situation-Service-Fit). As we already applied the 6-step design method, we also reflect on our experience with our use cases and share a first technical proof of concept for an SRS.

2. Smart Resilience Services

Let us think of a supplier in the automotive industry who develops precision products for car and truck engines. As the supplier relies on just-in-time delivery of parts with low inventory levels on-side of its production facilities, the supplier might be particularly vulnerable to disruptions in the supply chain, such as supply shortages of raw materials due to trade barriers emerging from political conflicts or delayed deliveries resulting from an ill-suited transport system [6] - a rather simplified, but realistic scenario to illustrate the underlying concept of SRS. In this scenario, the urgent need for an AI-based SRS becomes clear. SRS are smart services that represent individually configured packages of products and services that improve resilience. In a nutshell, we refer to SRS as an adaptive IT application for a more proactive, data-driven approach to resilience management. To allow for more-informed decisions, SRS leverage large data sets from various sources - sources can be internal or external to a company - and perform "intelligent" analyses, such as identification of disruptions in advance. Introduced as a concept for processing data and providing services to various stakeholders, special emphasis is given to AI-based data analytics components and real-time decision-making by smart services [7,8,9]: For example, generic and specific AI modules that can be applied to data streams according to needs and be aggregated into SRS are required. Generic AI modules provide basic functionalities for SRS such as connecting internal or external data sources, pre-processing sensor data, computing similarity measures, encrypting data (e.g., homomorphic encryption), securing data exchange by means of eContracts etc. Specific AI modules meet domain-driven requirements and satisfy specific needs of SRS characteristics for certain production industries; e.g. (fault) pattern recognition, planning of optimized action sequences, (semantic) representation of data, assets, services, knowledge, etc. For our above-described scenario, a potential SRS could leverage multiple data (e.g., trend analysis for raw material prices, sentiment analysis for political articles) that are relevant to perform impact analyses on, for instance, whether and when supply shortage might occur. Finally, the SRS could provide recommendations for action such as expanding the supplier network, optimizing order volumes, distributing production to more locations, adjusting inventories and revising engine components (across models). Thus, the SRS helps the supplier to strengthen its resilience significantly.

3. Design Method for SRS Specification

"How should an SRS 'look like'? What functional space should be encompassed? What relevant problem in industrial production should be solved by the SRS?" When discussing these questions with the interdisciplinary team in SPAICER, different answers were given from ex-

parts of process and metalworking industry, platform providers, computer scientists, business people, users and data scientists [10,11]. This ambiguity is due to differences in how experts communicate and understand the same problem. One Challenge was to translate the implicit understanding of all project members and stakeholders regarding the intended system of SRS into explicit requirements understandable to the entire team. Design methods help in coping with this challenge by modeling SRS systems so that they can be analyzed from different perspectives as well as for different purposes [11]. As deficiencies in the analysis and requirements specification phase represent the main reason for failure in IT-related projects (cf. Glass' law, Boehm's first law), we searched for a method for bringing experts and various stakeholders "around the table" to develop a common understanding of what problem should be solved by SRS.

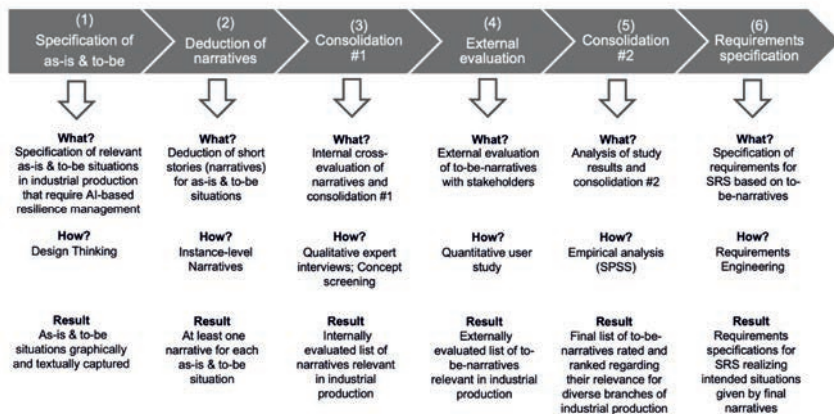


Fig. 1: Design method for specifying smart resilience services

Existing engineering methods from Design Science (e.g., User-Centered Design (UCD), Participatory Design, Design Thinking), Artificial Intelligence [12], Requirements Engineering (e.g., KAOS, i* model, scenarios) [13,14,15,16] and Software Engineering (e.g., Rational Unified Process, Prototyping, SCRUM) lacked an explicit analysis phase with structured procedure for step-by-step development of the desired functional space in the team and/or were not responsive to the central role of data and AI components within SRS. Therefore, we introduced a design method for specification of AI-based SRS according to the needs of the interdisciplinary project team (cf. Fig. 1). Within the method we combine approaches from

design science (Design Thinking, narrative thinking), information systems (qualitative / quantitative studies and focus groups), product development (concept screening) as well as from requirements engineering. Consisting of six phases, we started phase 1 with the specification of relevant as-is and to-be situations in industrial production that require future AI-based resilience management.

In our specific case, we started with three broad use cases for which SRS are to be developed that span across various parts of an organization from a micro to macro level. At micro level, planning horizons are local (e.g., on the machine), at meso level they are company-wide. Last, at macro level the whole value chain of the company is considered from a network-oriented perspective. In phase #1, we conducted a Design Thinking workshop working in four groups focusing on industrial resilience issues. Each group was characterized by a mix of experts of process or metalworking industry, platform providers, computer scientists, business people, users and data scientists. As a result, the team specified four as-is-situations representing the status quo and the issue to be solved as well as four to-be situations answering the question "If AI-based resilience management already existed, how would it be used?". All situations were captured graphically and textually. Within the next sections, the phases #2-#6 as well as outcomes will be elaborated.

4. Deduction of Narratives

Next, within each use case, as-is and to-be situations (phase #1) are defined and of short stories (narratives) described as part of a conceptual modelling method. Conceptual models are means by which a designer expresses his or her understanding of an envisioned information system such as our SRS. In our case, the goal of especially the to-be narratives is to answer the question "If the SPAICER platform already existed, how would it be used?". In the context of information systems design, narratives are a natural form for describing situations of intended information systems using textual descriptions [17]. They are short stories that describe which actors interact with the service or one another. Studies have shown that the more structured a conceptual language, the better the mental representations and, consequently, the learning and recall results [18]. They are significantly more effective than general-purpose conceptual modeling languages with respect to comprehend and recall of concepts and relationships [18]. Additionally, they are suitable for the external evaluation by experts from various sectors of the manufacturing industry in phase #4 due to their advantages specified above. Therefore, we used the concept of narratives for both, the as-is and to-be situations. For the deduction of narratives, we defined five criteria:

- (1) Focus on situation entities (roles, information, environment) and their interactions

- (2) Description on instance-level, not type-level
- (3) No technical or implementation-related aspects
- (4) Understandable for everyone
- (5) Short and sweet

Based on the four as-is and four to-be situations, each group that were formed during the design thinking workshop worked on the derivation of the narratives. First, the groups used brainstorming sessions to break the as-is and related to-be situations down into relevant situations focusing on very specific and single industrial processes. Then, each group generated instance-level narratives for each specified industrial process by following the criteria mentioned above. As a result of phase # 2, a list of narratives for each as-is & to-be situation has been derived (in total = 14 as-is and 14 to-be narratives).

Within the subsequent consolidation (phase #3), an internal cross-evaluation of narratives was conducted (cf. Fig. 1). Therefore, narratives for as-is and to-be situations of AI-based SRS were discussed with business partners and experts from process and metalworking industry. Objective of these qualitative interviews was to capture the relevance of SRS described in narratives with respect to the daily business of experts. Results of interviews showed application opportunities of SRS in further industrial domains as well as potential for ideation, e.g., forming hybrids of two narratives. For combining qualitative interviews with a quantitative approach for consolidation, we applied concept screening methods from product development for addressing multi-criteria decision problems, e.g., Pugh Controlled Convergence Method [19,20]. Therefore, a screening matrix "narratives of to-be situations vs. criteria" was set up. Criteria are independent in concept screening and represent a kind of target specification. In our case, we defined five criteria:

- (1) Compliance with SPAICER; i.e., narrative describes a to-be situation of AI-based resilience management
- (2) Relevance; i.e., narrative describes an issue whose solution is relevant and challenging in industrial production
- (3) Research gap; i.e., narrative describes a scientifically demanding topic whose solution goes beyond existing state of the art
- (4) Feasibility; i.e., narrative describes a usage situation of a service that can be prototypically implemented within the next year
- (5) Innovation; i.e., narrative indicates a technical solution of AI-based resilience management in industrial production going beyond and being not comparable with any existing system or out-of-the-box solution at market

In absence of a unique reference concept, evaluation of narratives was conducted by rating them with “+”, “-“ and “o” with respect to the whole set of narratives as well as given criteria. In case of “o” either the narrative’s value was quite similar to other narratives or the difference towards other narratives was controversial. Because additional information was required, the narratives were rated in a workshop with all stakeholders, i.e. the mix of experts of industrial production, platform providers, computer scientists, business people, users and data scientists. Special emphasis was given on ratings by business people and experts of industrial production in case of criteria (2), on ratings by R&D people for criteria (3), as well as on ratings by platform providers for criteria (4). The purpose of counting the points was not to identify a single winning narrative, but to reduce the number of narratives, i.e. eliminate weak concepts, form hybrids or gain a better understanding of SRS design issues. As a result, a consolidated and internally evaluated list of six final narratives was specified (cf. Table 1):

Table 1: Consolidated and internally evaluated list of six final narratives relevant in industrial production

Narrative 1: Machine parameter optimization to avoid defective prefabricated components A plant operator supervises the manufacturing process of prefabricated components on a plant. At the beginning of order processing, the plant reports and visualises recommendations for parameter optimization to the plant operator, such as optimal infeed and outfeed settings. The recommendations are based on analyses of the manufacturing parameters as well as sensor-based measurements on the raw material. The line operator releases the parameters for the line. After testing the first finished parts, which are completely in order, series operation can be started directly.
Narrative 2: Wear prediction for cost-effective tool maintenance and avoidance of downtimes A manufacturing process on a plant takes place continuously without disturbances. Relevant signal parameters of the plant are recorded in the background. The system reports and visualises: "Deviation 20% from normal system status, frequency range atypical! Repair of tool XYZ recommended!" The tool is removed, brought to an optimal condition and reinstalled. In the newly started series process, the proactive maintenance of the tool means that there are no out-of-order parts, the active elements of the tool can be reused many times through timely reworking and downtimes due to tool maintenance have been reduced.
Narrative 3: Knowledge transfer to avoid wrong decisions and additional costs An experienced employee supervises a manufacturing process at a plant. Accompanying the manufacturing process, the employee's expert knowledge is continuously recorded. The machine asks for feedback in the form of problem-solving patterns and process step evaluations and checks the plausibility of this information. When the experienced employee is absent and has to be replaced by an inexperienced colleague, a problem occurs on the line. The machine recommends a course of action: "Jammed part in the tool. Problem has already occurred frequently. Cleaning and additional use of non-oil fluid HLW3000 solved the problem on 13 June.2020." Through the transfer of knowledge, the problem can be solved quickly and optimally despite the lack of experience of the representative.
Narrative 4: Analysis of the consumer market to proactively adjust production planning. A company relies on analyses of the relevant consumer market in its strategic planning. A system

proactively provides regular detailed forecasts on the development of the consumer market (e.g. for the mobile phone and camera sectors). The forecasts are based on the evaluation of various data sources (e.g. export data, interviews, quarterly reports, blogs, etc.). This enables the company to make necessary internal changes and conversions/upgrades in production planning in good time (e.g. arrange for the purchase of additional machines at an early stage).

Narrative 5: Scenario simulation to prepare for external disruptive factors

A company is considering relocating one of its production sites. In addition to immediate investment costs, the aspect of "resilience" in particular is to be taken into account when deciding on a relocation. Therefore, the company analyses whether a relocation can contribute to an improvement or deterioration of the company's resilience. For this purpose, the risk management system simulates economic risks (e.g. financial crises), environmental and natural risks (e.g. earthquakes) as well as geopolitical risks (e.g. trade strike), social risks (e.g. poaching of skilled personnel), health risks (e.g. pandemics) and technological risks (e.g. cyber attacks). Based on the simulation, the company receives recommendations for action for setting up a new location.

Narrative 6: Signal analysis to secure and optimize supply of raw materials

For the production of a prefabricated components, the security of supply with raw materials plays a decisive role, as the machines cannot be temporarily switched off or changed over to another finished part. In order to be able to recognize potential risks such as delivery failures at an early stage, the buyers use a system that filters relevant sources of information regarding raw materials, suppliers, countries, etc. based on preferences, bundles messages and examines them for signals that could have an influence on the company's raw material supply. At the same time, the purchasing department can evaluate the relevance of displayed messages, so that the system is constantly improving. This saves time, as purchasing can quickly check and evaluate notifications from the system to secure the supply of raw materials.

5. External Evaluation

To conduct an external evaluation of the internally consolidated narratives by experts from various sectors of the manufacturing industry as the stakeholders of our planned SRS, we designed a quantitative study. The study was conducted in form of an online study. The participants had to read the six narratives describing AI-based resilience management and rate each narrative according to how well it would fit into the participant's everyday work routine. For the rating scale, an unforced bipolar seven-point Likert scale ranging from "1 = I completely disagree with the statement" to "7 = I completely agree with the statement" has been used. To get a first idea of the trend, we aimed to reach $n > 30$ [21]. In total, 33 subjects participated in the online study. The respondents came from wide range of industries with, with the top 3 industries being the manufacture of motor vehicles and parts (21%), mechanical engineering (18%) and leather manufacturing and production (12%). 65 % of the participants are working in enterprises located in Germany, whereas 13% are located in Europe and 22% worldwide. For this effort, we involved participants with deep knowledge of their respective industries to quickly get a good overview. This being said, the majority of the participants have more than 8 years of working experience (74%), are over 45 years of age (54%) and work in big enterprises with ≥ 250 employees (58%). The results show that all six narratives

seem to be relevant for the manufacturing industry (mean of every narrative >4.00). Moreover, narrative 2 has the highest relevance amongst the presented narratives (mean = 6.06, SD = 1.06) followed by narrative 1 (mean = 5.88, SD = 1.24). Therefore, it can be concluded that narrative 2 is perceived being the most relevant for the participants' everyday working routine. Narrative 4 "Analysis of the consumer market to proactively adjust production planning" has been rated as the least relevant narrative (mean = 4.85, SD = 1.56).

Table 2: Relevance Assessment of Narratives

	Narrative 1	Narrative 2	Narrative 3	Narrative 4	Narrative 5	Narrative 6
Mean	5.88	6.06	5.82	4.85	5.12	5.13
SD	1.24	1.06	1.07	1.56	1.24	1.50
Variance	1.55	1.12	1.15	2.44	1.55	2.27
Minimum	2	2	3	1	2	1
Maximum	7	7	7	7	7	7
p50	6,00	6,00	6,00	5,00	5,00	5,00

As a result of phase #4 (quantitative user study), we received an externally evaluated list of to-be-narratives relevant in the industrial production. After phase #5 (consolidation #2), we received a final list of to-be narratives rated and ranked regarding their relevance for diverse branches of industrial production. Thus, in the following section, we present the requirements specification for to-be narratives exemplarily for narrative 2 as being evaluated as the most relevant narrative amongst all presented narratives.

6. Requirements Specification

Based on first trend results of phase #5, we started the requirements specification based on the list of to-be-narratives in phase #6 (cf. Fig. 1). For realizing application scenarios of AI-based resilience management as described by the narratives, underlying information systems are required. Information systems are compounds of social systems, information spheres, and service systems that use information technology infrastructures for the realization of desired situations, as given by narratives [22,23]. In extension to information systems that exclusively process information objects, design models for information systems embedded in physical environments also require means for representing physical entities (e.g., IoT infrastructures, sensors). Adopting the Abstract Information System Model (AISM) representing a conceptual, i.e., logical description of information systems, we focused on four layers during requirements specification: information sphere, social system, service system and

physical object system [24] Information sphere covers a set of information objects used within the realm of the IS, e.g., raw data streams and aggregated data products (cf. GAIA-X data assets). The social system represents a set of roles available with rights, obligations, and prohibitions. In context of SPAICER, users of SRS can take multiple roles, e.g., data owner, provider or consumer of SRS etc. (cf. GAIA-X roles of participants [25]). Third component - the service system - covers the set of services available within situations in which the information system can be used, i.e., SRS that enables wear forecasting but also service modules realizing helper functions (cf. GAIA-X services and service instances). Last, the physical object system defines the set of physical entities available within situations in which the information system can be used. In SPAICER, this applies to IoT devices and sensors across production lines that produce raw data streams as well as entities offering interfaces to SRS (cf. GAIA-X nodes). Based on this four-piece pattern, narratives were analyzed regarding information objects, actors and roles, services and physical objects during requirements specification. Table 3 shows exemplarily the specification of requirements for narrative 2 based on the aforementioned pattern.

Table 3: Requirements specification of narrative 2 with respect to information objects, actors and roles, services and physical objects (cf. [24])

Information objects	(Which (raw) data / data streams are required / already available?)			
	• Sensory data in form of time series (temperature, sound etc.) of production run • Machine type (machine ID) • Specification of used raw material • Characteristics of materials (quality control etc.) • Machine parameters			
Roles	(Which aggregated data products are required / already available?)			
	• Aggregation of time series data (variance, mean values etc.) • Used tools (age in form of cycles, cycles since last maintenance, number of maintenances, etc.) • Interruptions (duration, reasons)			
	x	Data provider (Manufacturer of raw material, producer)	x	Data owner (Manufacturer of raw material, producer)
	x	Service consumer (machine/systems operators, quality assurance manager, production manager)		Service provider
		Service instance provider	x	Node provider (Manufacturer of raw material, producer)

Interface services	<i>(Which interface services with direct user interaction are required?, e.g., SRS)</i> - Wear forecasting with information about propagation of the fault into subsequent processes - Action Recommender (generation of recommended actions (with probability of success), e.g., based on error messages, problem reports in repair diary) - Feedback Integrator (recording and integration of user feedback (e.g., in case of detected anomalies)) - Visualizer (generation of graphical representations based on generated results, e.g., diagrams, traffic light representations, etc.) - Communicator (generation of textual or dialog-based representations of results or request options with regard to SRS offer)
Internal services	<i>(Which internal services without direct user interaction are required?, e.g., internal service modules)</i> - Data Preparation (collection, aggregation, and preparation of sensor data/raw data for use in SRS) - Annotator (semi-automatic description / annotation of data as well as roles and services where applicable) - Local and global anomaly detection - Explanation Generator (present explanations for recommendations for action, forecasts, etc. (Explainable AI)) - Performance Indicator (illustration of success rate of SRS; predicted results vs. reality) - Data Observer (monitoring availability of data streams / sensors and dealing with missing data (while checking remaining forecast quality))
Physical objects	<i>(Which IoT devices / sensors are required to produce data streams and/or realizing SRS interface?)</i> - Sensors for measuring machine parameters (temperature etc.) - Acoustic emission box (sound emission data) - IoT devices for realizing SRS interface (e.g., tablet, display on the machine)
Objective	<i>(How does the SRS contribute to resilience optimization?)</i> Predicting production line malfunctions/maintenance requirements for cost-effective maintenance and avoidance of downtimes

As a result of phase #6, we received requirements specifications for SRS realizing intended situations given by the final narratives. Serving as starting point for detailed technical specifications, based on determined information objects data-oriented requirements can be derived. Identified actors and roles can be applied for specifying non-functional requirements whereas interface and internal services represent the starting point for determining functional requirements.

7. Future Work

Application and results of our 6-step method showed that the approach can significantly help in developing innovative SRS relevant in various industries. For validating results of the trend analysis and for capturing critical technical risks identified at service level at an early stage, a first technical proof of concept for resilience optimization of production machines through AI-based wear prediction (cf. narrative 2) was implemented (see Fig. 2). For process industry a SRS was prototypically implemented that enables reduction of production failures and costs by avoiding wear and production downtime. The SRS helps (1) in monitoring wear of components, (2) to detect drifting of process states from "normal state" at an early stage, and (3) to be able to intervene. Various sensors are installed in the production system that continuously

transmit data. The service analyzes these data streams in real time locally for individual process sections using AI-based anomaly detection, provides anomaly metrics and aggregates them into a global view of the production line respectively the current production run. The SRS reports an anomaly in this section, indicates that no action is required, and asks for feedback from the operator (reinforcement learning). Thereby, the SRS thus enables a production run to be stopped in time before product quality deteriorates as well as targeted and local repairs or changes of components by detecting local anomalies in specific process sections.

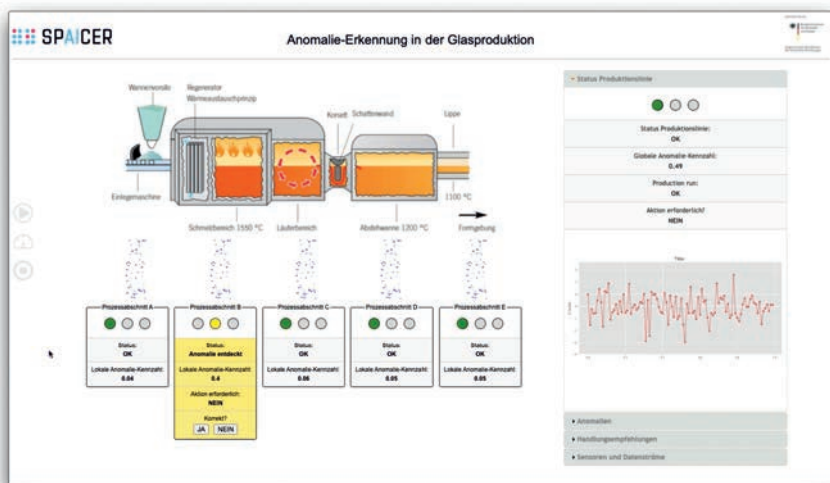


Fig. 2. Proof of concept of smart resilience service on wear protection described by narrative 2

Based on our first trend analysis, additionally planned workshops for further evaluation of our narratives as well as requirements specifications, we will: (a) develop PoCs for challenging technical issues generated from narratives identified as being the most relevant for production industry, and (b) implement SRS according to results of PoC developments. For implementing SRS on a SPAICER platform compliant with GAIA-X, reference implementations of AI core modules are planned where necessary modules will be incrementally integrated in SRS, tested, and optimized. Thereby, a successful deployment of relevant SRS for production industry on the envisioned SPAICER platform will be realized.

8. Literature

- [1] Zhang, W.J., & Van Luttervelt, C.A. (2011). Toward a resilient manufacturing system. *CIRP Annals*, 60(1), 469-472.
- [2] Zohuri, B., & Rahmani, F. M. (2019). Artificial Intelligence Driven Resiliency with Machine Learning and Deep Learning Components. *International Journal of Nanotechnology & Nanomedicine*, 4(2), 1-3.
- [3] Saravi, S., Kalawsky, R., Joannou, D., Rivas Casado, M., Fu, G., & Meng, F. (2019). Use of artificial intelligence to improve resilience and preparedness against adverse flood events. *Water*, 11(5), 973.
- [4] Petrenko, S. (2019). *Cyber Resilience*. Aalborg: River Publishers.
- [5] Petrelli, A., & Lau, S.-H. (2019, 12. Dezember). *Measuring Resilience in Artificial Intelligence and Machine Learning Systems*. Abgerufen am 03. Mai 2021, von <https://insights.sei.cmu.edu/blog/measuring-resilience-in-artificial-intelligence-and-machine-learning-systems/> [retrieved on: 20 April. 2021].
- [6] Mancini, L., De Camillis, C., & Pennington, D. (eds.). (2013). *Security of supply and scarcity of raw materials. Towards a methodological framework for sustainability assessment*. European Commission, Joint Research Centre, Institute for Environment and Sustainability, Publications Office of the European Union, Luxemburg.
- [7] Maaß, W., Shcherbatyi, I., Marquardt, S., Kritznier, A., & Moser, B. (2017). Real-Time Smart Farming Services. In: *International Conference on Agricultural Engineering (AgEng-2017)*, November 10-11, Hannover: VDI 2017.
- [8] Öksüz, N., & Maaß, W. (2020). *A SITUATION-SPECIFIC SMART RETAIL SERVICE BASED ON VITAL SIGNS*, International Conference on Information Systems (ICIS), 41th International Conference on Information Systems digital ICIS 2020.
- [9] Beverungen, D., Müller, O., Matzner, M., Mendling, J., & vom Brocke, J. (2019). Conceptualizing smart service systems. *Electronic Markets*, 29, 7–18.
- [10] Winograd, T., & Flores, F. (1986), *Understanding Computers and Cognition: A New Foundation for Design*, Intellect Books.

- [11] Clements, P., Bachmann, F., Bass, L., Garlan, D., Ivers, J., Little, R., ... & Stafford, J. (2010). *Documenting Software Architectures: Views and Beyond* (2. Auflage). Pearson Education.
- [12] Negnevitsky, M. (2011). *Artificial Intelligence: A Guide to Intelligent Systems* (3. Auflage). Addison Wesley.
- [13] Sutcliffe, A. (1998). Scenario-based requirements analysis. *Requirements engineering*, 3(1), 48-65.
- [14] Eric, S. Y. (2009). Social Modeling and i. In *Conceptual Modeling: Foundations and Applications* (99-121). Berlin Heidelberg: Springer.
- [15] Horkoff, J., & Yu, E. (2016). Interactive goal model analysis for early requirements engineering. *Requirements Engineering*, 21(1), 29-61.
- [16] Dardenne, A., van Lamsweerde A., & Fickas, S. (1993). Goal-directed requirements acquisition. *Sci Comput Program*, 20(1–2), 3-50.
- [17] Kuechler, W. L., & Vaishnavi, V. (2006). So, talk to me: The effect of explicit goals on the comprehension of business process narratives. *Mis Quarterly*, 30, 961-979.
- [18] Maaß, W., & Storey, V. C. (2015). Logical design patterns for information system development problems. In Johannesson, P., Lee, M.-L., Liddle, S.W., Opdahl, A.L., López, O.P., (Bearb.) *Conceptual Modeling - 34th International Conference, ER 2015, Stockholm, Sweden, October 19-22, 2015, Proceedings*. Volume 9381 of *Lecture Notes in Computer Science*, pages 134-147, Springer, 2015.
- [19] Pugh, S. (1996). *Creating Innovative Products Using Total Design*. Addison-Wesley.
- [20] Pugh, S. (1990). *Total Design*. Addison-Wesley.
- [21] Hogg, R. V., Tanis, E. A., & Zimmerman, D. L. (1977). *Probability and statistical inference* (Vol. 993). New York: Macmillan.
- [22] Lamb, R., & Kling, R. (2003). Reconceptualizing Users as Social Actors in Information Systems Research. *MIS Quarterly*, 27(2), 197-235.

- [23] Orlikowski, W. J., & Barley, S. R. (2001). Technology and Institutions: What can Research on Information Technology and Research on Organizations Learn from each others? *MIS Quarterly*, 25(2), 145-165.
- [24] Maaß, W., & Varshney, U. (2012). Design and Evaluation of Ubiquitous Information Systems and Use in Healthcare, *Decision Support Systems*, 54(1), 1-24.
- [25] Bundesministerium für Wirtschaft und Energie. (2020). *GAIA-X Technical Architecture*. Federal Ministry for Economic Affairs and Energy (BMWi).

Flexible On-Premises Datenanalyse für KMU

Dr. G. Steinhagen, Dr. M. Banf, fabforce GmbH & Co. KG, Netphen

Kurzfassung

Im Zeitalter von Industrie 4.0, dem Aufkommen des Internets der Dinge (IoT) und neuartiger Lernalgorithmen aus dem Bereich des Machine Learnings, erleben wir eine Revolution in Bezug auf die Möglichkeiten in der automatisierten Überwachung von Produktionsprozessen, der Optimierung von Prozessabläufen, bis hin zur assistierten Produktentwicklung.

Auch kleine und mittlere Unternehmen (KMU) können von diesen Entwicklungen profitieren, allerdings fangen die Hindernisse hier häufig schon mit der Erfassung und der Verarbeitung der Daten an. Für die Nutzung von Datenbanktechnologien und Datenanalyse sind in den meisten Fällen keine internen Ressourcen vorhanden. Auf der anderen Seite liegt häufig ein sehr heterogener Maschinenpark vor, der Maschinen aus verschiedenen Generationen vereint.

Dieses Paper stellt eine flexible Architektur vor, die das sichere Sammeln, Verarbeiten und Analysieren von Daten On-Premises ermöglicht. Durch die Flexibilität wird die Anbindung der verschiedensten Maschinen schnell umsetzbar. Außerdem werden verschiedene Analysen gezeigt, die mit begrenzten Ressourcen einer On-Premises Serverlösung stabil funktionieren und nicht die Rechenkapazität einer Cloud benötigen. Dabei handelt es sich um unüberwachte Analysemethoden, Edge-Anwendungen aus der Umformüberwachung und Schwingungsanalysen, sowie innovative Prozessvisualisierungen, die Prozessspezialisten detaillierte Einblicke in die Produktion liefern.

1. Einleitung

Das Thema Künstliche Intelligenz hat viele Hoffnungen auf mehr Effizienz in der Produktion geweckt. Gleichzeitig steigt aber auch der Wettbewerbsdruck, sich mit diesem Thema und der Digitalisierung allgemein zu beschäftigen, um nicht den Anschluss zu verpassen. Insbesondere bei kleinen Unternehmen kommt dabei die Frage auf, wie diese den Einstieg in das Thema finden, ohne einen Locked-In Effekt in die falsche Technologie zu riskieren. Außerdem sind organisatorische Fragen, wie das Hosting und die Wartung der digitalen Infrastruktur, zu klären.

Zunächst stellt sich die Frage, wo und in welcher Form die notwendige Hard- und Softwarearchitektur verwaltet werden soll. Die beiden großen Kategorien sind hier Cloud basierte oder lokale OnPremises Lösungen. Während bei Cloud basierten Ansätzen jeweils ein Drittanbieter die Serverlandschaft zur Verfügung stellt und verwaltet, bleibt bei OnPremises Lösungen die Hardwarearchitektur - und somit gegebenenfalls sensible Daten - innerhalb des eigenen Unternehmens. Dies bietet nach wie vor, trotz der teilweise vorteilhaften Anreize von Cloud Lösungen im Hinblick auf Hardware- und Verwaltungskosten, einen nicht zu vernachlässigenden Vorteil im Bereich Datensicherheit.

Mit Hinblick auf die Software-Architektur bieten auf Microservices basierende Ansätze die Möglichkeit, Softwarepakete flexibel zu kombinieren und diese auf vielfältigen Systemen zu installieren. Gleichzeitig sind diese Systeme deutlich einfacher erweiterbar und damit zukunfts-sicherer als eine monolithische Software eines Anbieters. Dies erleichtert den Einstieg in die Digitalisierung.

Schwerpunkte in der produktionsbezogenen Datenanalyse liegen zunächst im Bereich der permanenten Zustandsüberwachung (condition monitoring). Hier können Verfahren aus der Statistik und des Machine Learnings verwendet werden, um z.B. Defekte in Werkstücken zu erkennen (quality inspection) oder die verbleibende Lebensdauer von Werkzeugen und Maschinenbauteilen abzuschätzen (predictive maintenance).

Innerhalb des Machine Learnings selbst lassen sich ebenfalls zwei große Methodenbereiche unterscheiden, das überwachte (supervised) und unüberwachte (unsupervised) Lernen. Beim überwachten Lernen werden Trainingsdaten zur Verfügung gestellt, die im Vorfeld bewertet und kategorisiert (annotiert) wurden. Diese Annotationen ermöglichen es, den Lernalgorithmus und die Rohdaten mit Hinblick auf diese Kategorisierungen hin zu analysieren. Im produktiven Einsatz versucht ein solches System dann neue Daten gemäß ihrer Ähnlichkeit zu den Trainingsdaten zu einer der gegebenen Kategorien zuzuordnen. Auf diese Weise können neben Klassifikations- auch Vorhersagemodelle angelernet werden.

Anwendungen in der Produktion sind beispielsweise Erkennung von Oberflächendefekten [1, 2], die Materialzustandsbestimmung von Getrieben, Kugellagern [3], oder Fräsmaschinen [4]. In Abschnitt 5 wird eine Anwendung zur Klassifikation von Fräsprozessen basierend auf annotierten Schwingungsdatensätzen vorgestellt.

Unüberwachte Algorithmen arbeiten direkt auf den Rohdaten ohne zusätzliche Annotationen. Sie bieten sich im Besonderen an, um einzelne Ausreißer oder Anomalien zu erkennen, wenn keine oder zu wenige Datensätze zum Training eines überwachten Ansatzes zur Verfügung stehen. Dies ist in der Produktion von Vorteil, wenn das manuelle Annotieren zeit- und kostenintensiv ist bzw. Fehlertypen nicht konkret im Vorfeld definiert werden können oder nur sehr

vereinzelte auftreten. Des Weiteren erlauben unüberwachte Ansätze effektive Verfahren zur Reduktion der eigenständigen Clusterbildung und der Visualisierung von Daten. Anwendungen in der Produktion sind zum Beispiel Auffälligkeitsanalyse von Rotationsmaschinen [5] oder innerhalb von Metallumformprozessen [6]. In [7] wurde ein Beispiel auf Basis mehrerer Druck- und Ventilensorsignale gezeigt, die für einen Hydroforming Prozess eingesetzt wurden.

Innerhalb dieser beiden Kategorien hängt die Auswahl des geeigneten Verfahrens dann zusätzlich von den gegebenen Datenquellen - z.B. Vibrations-, Temperatur-, oder Drucksensoren - als auch Datentypen - z.B. Sensordaten, dokumentierte Maschinenhistorie, Wartungspläne etc. ab.

2. Systemarchitektur

Um eine möglichst flexible Software bereitstellen zu können, wurde das System als eine Microservices-Architektur umgesetzt. Zur Implementierung der einzelnen Microservices wird eine Container-Technologie basierend auf dem Docker Framework verwendet. Microservices können sowohl auf lokalen Systemen eingesetzt, als auch in großen Serverarchitekturen verwendet werden. Die verschiedenen technologischen Funktionen sind so in Subsysteme aufspaltbar, die über einheitliche Schnittstellen kommunizieren.

Ein weiterer Vorteil ist, dass verwendete Technologien schneller erweitert oder ausgetauscht werden können. Dies ist insbesondere bei schnellen Entwicklungszyklen sehr wichtig, auf die sonst nur sehr träge reagiert werden könnte.

Die Funktionen die notwendig sind, um Daten effizient analysieren zu können sind das Bereitstellen von Maschinenschnittstellen, das Speichern und Verwalten der Daten sowie die Analyse der Daten. Zudem müssen die Daten für den Nutzer visualisiert werden. Darüber hinaus ist es sinnvoll, die Services möglichst flexibel zu halten und die Parameter über eine zentrale Konfiguration einzustellen. Somit lassen sich Analysen, Schnittstellen und Datenbanken präzise auf individuelle Maschinentypen und Fragestellungen abstimmen.

Wichtig ist auch, dass sich auf diese Weise generische Analysen, die über Konfigurationen individualisiert werden, genauso einbinden lassen, wie individuelle Analysen, die auf proprietären Kundendaten basieren und somit Kern-Know-How dieser Kunden darstellen.

Die Anbindung an weitere Softwarekomponenten einer bestehenden IT-Landschaft ist ebenfalls an verschiedenen Stellen im Informationsfluss möglich. Ein Aufbau des Datenflusses ist in Bild 1 dargestellt.

Auf diese Weise können auch schnell vielfältige Maschinen- und IT-Schnittstellen eingebunden werden, ohne große Änderungen an der restlichen Software vornehmen zu müssen.

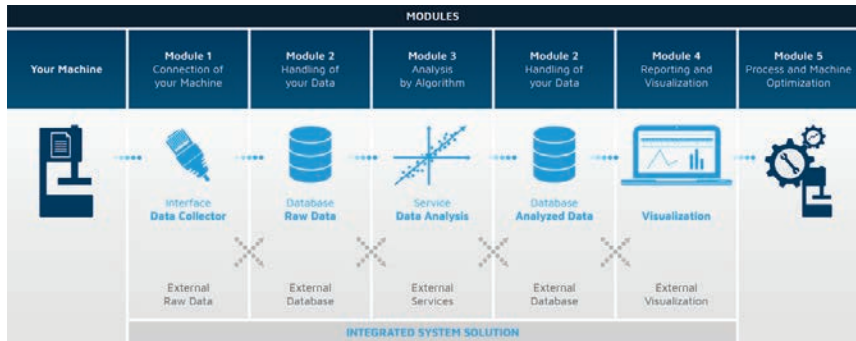


Bild 1: Datenfluss des Systems

3. Datenverarbeitung

Die Art der Daten, die verarbeitet werden müssen, lassen sich sowohl nach der Datenstruktur, als auch nach dem Analysegrad unterscheiden. Die in der Automatisierung wohl relevantesten Daten sind Zeitreihendaten; darüber hinaus sind matrixartige Daten, wie Bilder von Kameras oder analysierte 2D oder 3D Daten relevant. Außerdem sind Metadaten zu den genannten Daten wichtig, da über Annotationen ein überwachtes Lernen möglich wird. Beim Analysegrad kann zwischen Rohdaten und bereits analysierten Daten unterschieden werden. Während bei der Datenstruktur insbesondere die Auswahl des richtigen Datenbankformats wichtig ist, beeinflusst der Analysegrad eher die notwendige Speicherdauer und die Art der Visualisierung. Es ist nicht sinnvoll große Mengen von Rohdaten, die bereits analysiert sind, langfristig vorzuhalten. Wenn die Analysen bereits Kernaussagen über den Prozess bereitstellen, kann die Datenmenge so deutlich reduziert werden. Bild 2 zeigt den Aufbau unserer Datenverarbeitung, welche die oben genannten Anforderungen erfüllt. Die Microservices-architektur ermöglicht es auch hier die vielfältigen Anforderungen in sinnvolle Teilaufgaben und damit verbundene Services aufzuteilen. Um die Daten zu koordinieren, wird dabei ein Integrationsservice verwendet, der auf die Datenbanken zugreift. Auf diese Weise lassen sich auch auf einem relativ kleinen System vielfältige Daten verwalten.

Um die Verwaltung auf die individuellen Einsatzfälle anzupassen, wurde ebenfalls eine Schnittstelle zur zentralen Konfiguration geschaffen. So lässt sich zum Beispiel die Datenvorhaltungsdauer gezielt einstellen. Zusätzlich ist es aber auch von Vorteil die Möglichkeiten zur Datenaufbereitung moderner Datenbanken zu nutzen. So können beispielsweise Rohdaten vor

der Übertragung direkt von der Zeitreihendatenbank entsprechend der Anforderungen gesammelt werden. Dies verringert den Analyseaufwand, sowie den Datenverkehr. Außerdem können direkt virtuelle Variablen erzeugt und für die weitere Analyse genutzt werden.

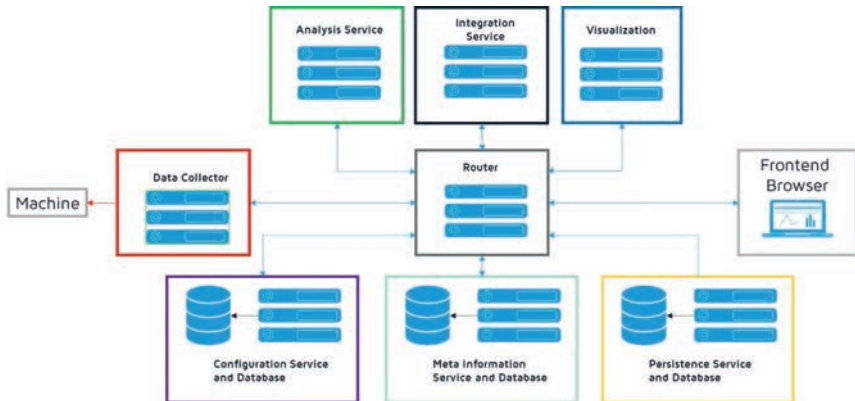


Bild 2: Aufbau der Multi-container basierten Microservices-Systemarchitektur

4. On-Premises Datenanalyse

Um die gesammelten Daten möglichst schnell nutzen zu können bieten sich unüberwachte Lernalgorithmen an, da hier lange Anlernphasen der Algorithmen entfallen. Die Strukturierung nach Chargen, Zyklen, Produktarten, Maschinenkomponenten und Fehlerzusammenhängen ermöglicht einen gezielteren Einblick in die Daten. Es entstehen Möglichkeiten analog zum Online Analytical Processing (OLAP) wie es im Bereich Business Intelligence bekannt ist. Im Folgenden werden verschiedene Beispiele vorgestellt, die eine schnelle Nutzung der Daten ermöglichen.

Perzentilen basierte Analyse

Produktionsprozesse unterliegen in der Regel Schwankungen, welche anhand der verschiedenen Sensorwerte einer Maschine erkennbar sind. Meist wird ein Prozess mit wenigen Probestücken eingerichtet. Eine weitere Analyse in der normalen Produktion findet dann erst im Fehlerfall statt. Auch hier werden, sofern überhaupt vorhanden, die Daten des fehlerhaften Prozessdurchlaufs analysiert. Einige Probleme und Prozessinstabilitäten lassen sich jedoch bereits früher entdecken, wenn die Streuung der Werte betrachtet wird. Eine Möglichkeit hierfür bietet die Analyse der Perzentilverläufe verschiedener Sensordaten [1]. Bild 3 zeigt am

Beispiel eines Ventilstellwerts dessen Perzentile für eine Produktionscharge. Es sind Signaltbereiche zu sehen, die eine sehr breite Streuung zeigen, und Bereiche, die ein sehr homogenes Verhalten aufweisen.

Ausreißer können mit diesem Vorgehen ebenfalls gut und in Echtzeit analysiert werden, da sich der Vergleich der aktuellen Produktionsdaten mit den bereits berechneten Perzentilen einer laufenden Charge bereits nach wenigen Durchläufen durchführen lässt. Der Vorteil ist, dass weder ein Hochleistungsrechner noch aufwendiges, manuelles Annotieren von Daten notwendig sind.

Um ein genaueres Abbild vom Prozess zu bekommen, kann anstatt einzelner Perzentile auch eine kontinuierliche Perzentilverteilung pro Sensor abgebildet werden. Hierdurch zeigen sich noch genauere Einblicke, in welchem Bereich die Werte konzentriert sind.



Bild 3: Perzentildarstellungen der Sensorsignale einer Produktionscharge zur Visualisierung der Signalstreuung

Analyse mit Hilfe nicht-parametrischer Verteilungen

Perzentile stellen eine gute Möglichkeit dar, unimodal verteilte Daten darzustellen, also basierend auf der Annahme eines einzigen, dominanten bzw. idealen Prozessverlaufes. Bei komplexeren Verteilungsmustern eines Prozessverlaufes, in denen beispielsweise Nebenmaxima enthalten sind, kann es durch die Perzentildarstellung zu Fehldarstellungen bzw. -interpretationen kommen. Zur Analyse dieser Fälle bieten sich hier nicht-parametrische Verteilungsfunktionen zur Darstellung und Analyse an. Durch diesen zusätzliche Detailgrad können wertvolle Informationen gewonnen werden. Liegt zum Beispiel ein bistabiler Produktionsprozess vor, der die Gesamtqualität beeinflusst? Handelt es sich um eine spezifische Fehlerhäufung, die näher analysiert werden kann, um eine Vermeidungsstrategie zu entwickeln? Auf dieser Grundlage können zum Beispiel weitere Clusteranalysen für Fehler erstellt werden oder über

PARETO Diagramme aus dem Six Sigma Werkzeugkasten die Häufigkeit und Beeinträchtigung durch die auftretenden Fehler analysiert werden.

Bild 4 zeigt ein Beispiel mit einem zunächst gleichmäßig verteilten Signal, welches sich dann am Ende in zwei verschiedene, stabile Muster aufspaltet, die sich im Zeitbereich unterscheiden. Die zugrundeliegenden Zustände können damit näher ermittelt werden.

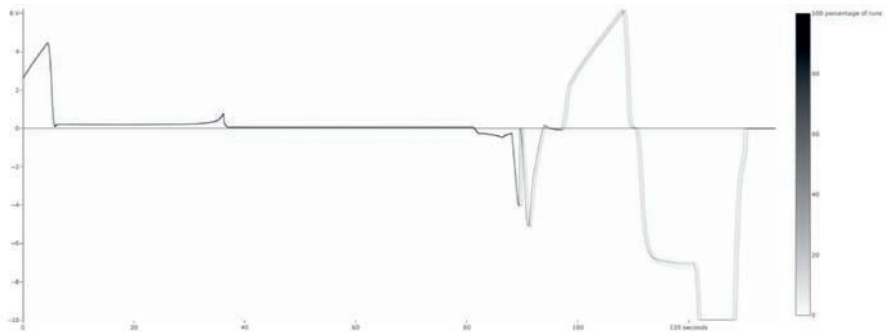


Bild 4: Nichtparametrische Verteilung eines Ventilsignals

Qualitätsbewertung einzelner Produktionsdurchläufe

Zur Qualitätsbewertung, sowie deren übersichtlicher Darstellung, einzelner Produktionsdurchläufe z.B. einer Produktionscharge, ist es zunächst hilfreich eine Kennzahl für einen einzelnen Produktionsdurchlauf zu definieren. Hierbei ist die Ähnlichkeit bzw. Abweichung des Produktionsdurchlaufes zum idealen bzw. "mittleren" Produktionsdurchlauf ein sinnvoller Ausgangspunkt. Zunächst kann - ausgehend von der Grundannahme, dass eine Charge deutlich mehr fehlerfreie als fehlerhafte Produktionsdurchläufe beinhaltet - aus sämtlichen Produktionsdurchläufen der Charge ein "idealer" Durchlauf errechnet werden. Dieser kann im Weiteren verwendet werden, um die Abweichung jedes Einzeldurchlaufes zu diesem "idealen" Durchlauf zu ermitteln und in kompakter Form als skalaren Wert pro Durchlauf darzustellen.

Neben der Bewertung und Darstellung der Produktionsdurchlaufbewertungen einzelner Sensoren können so ebenfalls sehr einfach Bewertungshierarchien aufgebaut und visualisiert werden. So ist es möglich die Durchlaufbewertungen einzelner Sensoren zu physischen oder funktionalen Sensorgruppen oder auch auf Maschinenebene zusammenzufassen und gemeinsam auszuwerten. So können Sensorsignale, welche zu bestimmten Komponenten einer Maschine gehören oder die in bestimmten Fehlerfällen Auffälligkeiten zeigen, zusammen ausgewertet werden. Damit ist es möglich auch bei einer größeren Anzahl von Einzelsensoren

eine schnelle und effiziente Fehleranalyse zu gewährleisten. Ein weiterer Vorteil der kompakten visuellen Darstellung ist die Möglichkeit, möglichst schnell Abweichungsmuster, wie zum Beispiel Trends oder phasenweise Abweichungen zu erkennen, so dass frühzeitig reagiert werden kann.

Im folgenden Abschnitt wird ein Beispiel erläutert, wie sich hieraus ein effizienter Drill-Down Ansatz zur Fehlersuche ableiten lässt.

Effiziente Abweichungsanalyse als Drill-Down Ansatz

Die oben genannten einzelnen Verfahren lassen sich zu einem effizienten Drill-Down Ansatz in der Fehleranalyse verknüpfen. Mithilfe eines passenden Frontends lassen sich die Verfahren geschickt kombinieren. Bild 5 skizziert dieses Vorgehen für die Analyse mehrerer Prozessdurchläufe einer hydraulischen Presse. Auf der obersten Ebene lassen sich durch die Gruppen der Maschine die Komponenten oder Funktionsgruppen sehr kompakt darstellen. Es ist erkennbar, dass die ersten vier Werkstückedurchläufe pro Sensorgruppe (Stößel, Kraft, Systemdruck) eine deutliche Abweichung im Vergleich zum Rest der Produktionscharge aufweisen (Abb. 5a).

Diese lassen sich auch in den einzelnen Kraftmessungen an den Pressensäulen wiederfinden, die in der Gruppe „Kraft“ zusammengefasst sind (Abb. 5b). Das Frontend bietet die Möglichkeit, mit einem übergreifenden zeitlichen Zoom schnell auf die interessanten Bereiche zu fokussieren. So können schnell die Durchläufe mit den entsprechenden Daten aufgerufen werden. Hier kann die abweichende Gruppe aufgelöst nach Einzelsensoren analysiert werden, um zu überprüfen, für welchen Sensor die relevanten Abweichungen auftreten (Abb. 5c).

Letztendlich können die Rohdaten im Vergleich zu den Perzentilen aufgerufen und für den Abweichungsbereich analysiert werden. Somit ist schnell erkennbar an welcher Stelle im Prozess die Sensorwerte ausbrechen und Prozess- und Maschinenexperten bekommen ein genaues Bild der Störung, ohne manuell nach diesen Abweichungen suchen zu müssen.

Im Beispiel zeigt sich anhand der Abweichung der ausgewählten Werkstückdurchlaufsignale (rot) im Vergleich zu den korrespondierenden Perzentilen eine deutliche Veränderung der Kraftverläufe, was auf ein mechanisches Blockieren oder Abweichungen im Ausgangswerkstück hindeuten könnte. Der Ansatz kann auch insoweit angepasst werden, dass Sensoren, die Störungen besonders deutlich zeigen oder die Qualität gut repräsentieren, ebenfalls kompakt überwacht werden. So zeigt sich zum Beispiel im Hydroforming, dass insbesondere der Positionswert des Druckübersetzers ein guter Indikator für die Qualität und Störungen im Prozess ist.

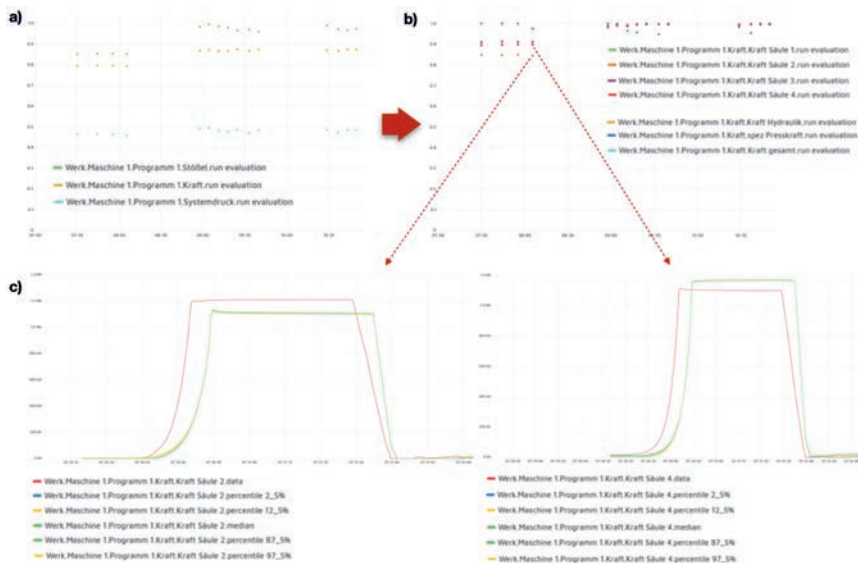


Bild 5: Drill-Down Vorgehen für die Abweichungsanalyse mehrerer Prozessdurchläufe einer hydraulischen Presse

5. Edge Computing zur direkten Datenaufbereitung

Statt einer Auswertung im Rahmen des gesamten Systems kann es in verschiedenen Fällen sinnvoller sein, die Daten direkt in einer Edge Computing Anwendung maschinennah auszuwerten. Dies kann beispielsweise der Fall sein, wenn die Daten über eine Livevisualisierung direkt an der Maschine für den Maschinenbediener genutzt werden oder die Menge der Rohdaten eine Vorverarbeitung zur Datenreduktion notwendig macht. Im Folgenden werden aus beiden Bereichen je ein Beispiel vorgestellt, die sich mit dem oben bereits vorgestellten Auswertungskonzepten kombinieren lassen.

Visualisierung für das Schrittbiegen von Rohren

In der Produktion von großen Konstruktions- und Pipelinerohren ist einer der Produktionsschritte das Schrittbiegen der Rohre. Aus einem geraden Blech wird in einem Dreipunktbiegeverfahren schrittweise ein Rohr geformt. Für die Messung in diesem Prozess

gibt es zwei Herausforderungen: Es ist jeweils im Werkzeug nur ein Teilausschnitt des ganzen Rohres bzw. Bleches zu sehen und jeder Biegeschritt ist einer Rückfederung unterworfen. Aktuell wird der Solldurchmesser des Rohrs mit dem gemessenen Durchmesser beim Einlernen der Biegepositionen anhand einer Schablone manuell vom Bediener der Anlage abgeglichen. Je nach Rohrhersteller werden dann auch die angelerneten Positionen des Biegevorgangs teilweise wieder mit der Schablone überprüft, da in Materialchargen Schwankungen der Materialparameter vorliegen können.

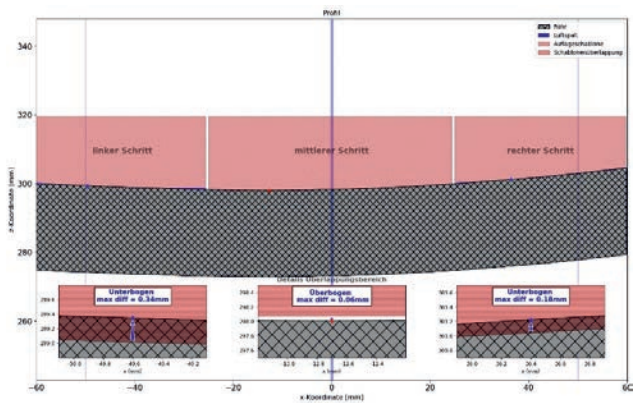


Bild 6: Visualisierung mit digitaler Schablone für das Schrittbiegen von Rohren

Um den Bediener ein einfach zu verwendendes Assistenzsystem zu bieten, wird ein Lasersensor verwendet, der die aktuelle Außenkontur des zu biegenden Bleches im Werkzeug misst. Daraus wird der relevante Innenradius berechnet und für den Bediener als Matching mit einer digitalen Schablone je Teilabschnitt dargestellt. In Bild 6 ist zu sehen, dass zudem noch eine maximale Abweichung für die Teilabschnitte gegeben ist. Diese Information kann mit den oben vorgestellten Analysen wieder gekoppelt werden.

Darüber hinaus wird für das Rohr die lokale Krümmung berechnet. Diese kann bei Bedarf ebenfalls für den Bediener dargestellt werden. Die Krümmung ist sehr hilfreich, um zum Beispiel die Einflusszonen des Biegeschrittes zu bestimmen und somit ggf. die Schrittweite zu optimieren oder die Rückfederung besser abschätzen zu können. Dadurch kann eine schnellere Kompensation erreicht werden [8]. Der Verlauf der Krümmung oder deren Maxima sind ebenfalls eine gute Grundlage für nicht überwachte Lernmethoden, um zum Beispiel Rückschlüsse über die Materialqualität oder Informationen für Folgeprozesse zu generieren.

Schwingungsdatenanalyse

Bei der Analyse von Schwingungsdaten, zum Beispiel aufgenommen zur Analyse von Fräsprozessen, treten je nach Frequenzspektrum sehr hohe Datenmengen auf. Beschleunigungssensoren liefern Messwerte beispielsweise im Bereich von 10kHz. Die Übertragung solcher Rohdatenmengen an einen Server ist nicht sinnvoll.

In ersten Untersuchungen der Datenanalyse eines Fräsprozesses hat sich herausgestellt, dass relevante Aspekte wie der Verschleiß der Wendeschneidplatten am Fräskopf, die Spannung von Transmissionsriemen oder auch die Maschinenstabilität erkennbar sind.

Die Detektion ist ebenfalls nach einer Reduktion der aufgezeichneten Vibrationsrohdaten mittels vorausgewerteter statistischer Merkmale, wie zum Beispiel dem quadratischen Mittel (Root Mean Square) oder dem Scheitelfaktor (Crest Factor) des Rohsignals zuverlässig. Diese reduzierte Datenmenge kann dann ebenfalls an den Server weitergeleitet werden, um dort Analysen vorzunehmen.

In den Voruntersuchungen hat sich gezeigt, dass Entscheidungsbaumverfahren (Decision Trees) eine gute Möglichkeit bieten, die Daten überwacht zu analysieren. Der auf dem Server angelernte Entscheidungsbaum kann auch wieder zurück in den lokalen Prozess gespielt werden, um dort anhand der Livedaten relevante Ereignisse im Fräsprozess zu klassifizieren. Darüber hinaus ist die Analyse von zeitkontinuierlichen Verschleißvorhersagemodellen auf Basis der aufgenommenen und annotierten Schwingungsdaten in Planung.

6. Zusammenfassung und Ausblick

Es wurde eine prinzipielle Softwarearchitektur vorgestellt, die es ermöglicht, verschiedene Analysen durchzuführen und diese auch zu erweitern. Mit verschiedenen Beispielen wurde gezeigt, wie sich diese sehr schnell für verschiedene Anwendungsfälle einsetzen lassen. Durch seine Flexibilität ermöglicht der Ansatz insbesondere kleinen Unternehmen einen einfachen Einstieg in die Themen Digitalisierung und Künstliche Intelligenz.

Durch die Einführung eines digitalen Logbuchs wird es einem Kunden zukünftig ermöglicht, wichtige Informationen zum Maschinenstatus schnell zu dokumentieren und in Echtzeit für alle relevanten Personen bereitzustellen. Das integrierte konsistente Labeling der Daten kann dann als Grundlage für kundenspezifische, überwachte Lernalgorithmen verwendet werden. Die daraus entwickelten Algorithmen können schnell in die bestehende Architektur integriert und die Produktivität weiter erhöht werden.

Literatur

- [1] M. Rudolph, B. Wandt and B. Rosenhahn. "Same same but different: Semi-Supervised Defect Detection with Normalizing Flows." WACV (2021)
- [2] S. Qi, J. Yang, and Z. Zhong. "A Review on Industrial Surface Defect Detection Based on Deep Learning Technology." MLMI, NY, USA, (2020): 24–30
- [3] G.E. Kondhalkar, G. Diwakar. "Crest Factor Measurement by Experimental Vibration Analysis for Preventive Maintenance of Bearing." ICRRM, Springer (2019)
- [4] E. Traini, G. Bruno, G. D'Antonio, F. Lombardi. "Machine Learning Framework for Predictive Maintenance in Milling." IFAC 52 (2019): 177-182
- [5] C. Lu, Z. Wang, W. Qin, J. Ma. "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification." Signal Processing 130 (2017): 377-388
- [6] B. Lindemann, F. Fesenmayr, N. Jazdi, M. Weyrich. "Anomaly detection in discrete manufacturing using self-learning approaches." Procedia CIRP 79 (2019): 313-318
- [7] G. Steinhagen, A. Hoffmann. "Process surveillance in hydroforming based on machine learning algorithms." Procedia Manufacturing 27 (2019): 57-64.
- [8] G. Steinhagen, F. Kapp. „Neuer Ansatz für das automatische Schrittbiegen von Großrohren.“ 31. ASK Umformtechnik (2018)

Towards an Artificial Intelligence based Approach for Manufacturing Resilience

N. Öksüz, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Saarbrücken;

S. Bouschery, Institute for Technology and Innovation Management (TIM), RWTH Aachen University;

M. Schlappa, J. Sporkmann, WHU - Otto Beisheim School of Management, Vallendar;

M. Unterberg, Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen

Kurzfassung

Die aktuelle COVID19-Pandemie hat bereits massive Effekte auf unsere globalisierten Gesellschafts- und Wirtschaftssysteme. Um zu verstehen, wie sich die negativen Auswirkungen von Störereignissen auf komplexe Systeme der produzierenden Industrie messen lassen und wie langfristig erfolgreiche Anpassungsstrategien aussehen können, nutzt das Projekt SPAICER das Konzept der Resilienz. Doch es stellt sich die Frage, wie die Unternehmensresilienz gemessen werden kann, um darauf basierend Handlungsempfehlungen zu geben. Dieser Beitrag stellt einen Ansatz einer KI-basierten Messung der Resilienz in Produktionsunternehmen vor. Darüber hinaus werden auf den Ansatz basierend Chancen und Hürden einer KI-basierten Resilienzmessung erläutert.

Abstract (optional)

The current COVID19 pandemic is already having massive effects on our globalized social and economic systems. In order to understand how the negative effects of disruptive events on complex systems in the manufacturing industry can be measured and what successful long-term adaptation strategies can look like, the SPAICER project uses the concept of resilience. But the question arises as to how corporate resilience can be measured in order to provide recommendations for action. This paper presents an approach for an objective, AI-based measurement of resilience in manufacturing companies. In addition, opportunities and hurdles of an AI-based resilience measurement based on the approach are explained.

1. Introduction

Industrial production systems are prone to disruptions such as natural disasters or political unrest. While some of these events can be predicted, others are simply unpredictable or

even unknowable before the fact. Most recently, besides the COVID-19 pandemic, the supply chain incident with the Ever-Given container vessel being stuck in the Suez Canal highlighted the vulnerability of our global and subtly balanced economies. Once again, we have learned the hard way that resilience is key to successfully managing production systems. Resilience describes – written in simple words – the level to which companies can “handle disruptions” or “keep the lights on”, meaning preparing for, resisting against, reacting to, and adapting after disruptions.

Yet, only few companies actively manage their resilience. In a survey carried out by the Deutsches Forschungszentrum für Künstliche Intelligenz GmbH in 2019, only 35% of the 60 surveyed company representatives stated that they have an active resilience management in their company, whilst the majority (60%) saw an urgent need for action, considering that disruptions can occur at least occasionally. In this context, we are reminded of economist Peter Drucker, who rightly said: “What gets measured gets managed” [1]. One reason for not actively managing resilience might be that some companies lack concepts to gauge their current level of resilience. Existing concepts do not seem to be adequate as they come with several shortfalls.

The BMWi-funded SPAICER project – a consortium of researchers and practitioners – takes up the discussion on resilience management and, among other things, aims to advance existing concepts with a new resilience index that is at the forefront of technology. In a nutshell, our resilience index measures a company's resilience based on how and to what extent a company's performance alters when it is affected by potential disruptions. A company's performance is derived from established key performance indicators (KPI) such as EBIT or on-time delivery. Leading AI technologies lay the foundation of our resilience index which allows us to address the shortfalls of existing concepts and pursue a holistic, data-driven approach. Our resilience index enables companies to more objectively measure their level of resilience and take a more proactive approach to resilience management. Such a new data-driven approach will become particularly valuable as companies operate under increasingly volatile, uncertain, complex, and ambiguous conditions.

The remainder of this paper is organized as follows. Section 2 reviews literature on resilience and approaches to measuring it, Section 3 introduces our methodology, Section 4 presents a discussion and Section 5 concludes the paper.

2. Literature Review

Past research in different domains like psychology, supply chain management, and engineering has provided a plethora of definitions and concepts for the term resilience. Originally

stemming from the Latin term *resilire* (to bounce or spring back), Merriam-Webster provides two definitions for the term resilience: From a physical point of view, resilience denotes “the capability of a strained body to recover its size and shape after deformation [...]”[2]. In the psychological domain, resilience describes “an ability to recover from or adjust easily to misfortune or change” [2]. Vugrin et al. expand the definition of resilience to systems and define resilience as the “ability [of a system] to reduce efficiently both the *magnitude* and *duration* of *deviation* from targeted system performance levels” when facing one or more disruptive events [3].

Hosseini et al. cluster past approaches to resilience of the *organizational, social, economic, and engineering* domain [4]. To improve the resilience of a system, e. g. a manufacturing system, the status quo has to be assessed [5]. Approaches to assessing resilience can be differentiated into *qualitative* and *quantitative* approaches. Because our suggested approach towards determining resilience falls under the latter category, the following focuses on existing quantitative frameworks.

On a supply chain level, Pettit et al. distill the resilience of a SC to its *vulnerability* and its *capability* to withstand and recover from disruption [6]. The authors propose the Supply Chain Resilience Assessment and Management (SCRAM) questionnaire containing questions to measure 14 capabilities like *flexibility in sourcing, market position, and financial strength* with 71 sub-factors like *alternate distribution channels* or *multi-sourcing* [6]. Seven *vulnerabilities* like *external pressures* and *connectivity*, further distinguished into 40-sub factors like *product purity* or *restricted materials* are gathered within a 5-point Likert scale [6]. Smith et al. emphasize the original meaning of resilience as *bouncing back* and propose the Brief Resilience Scale (BRS) that measures the ability of a person to recover from stress with a questionnaire [7]. On an organizational level, Shirali et al. distinguish between the six resilience indicators *top management commitment, trust culture, learning culture, awareness and opacity, preparedness, and flexibility* [8]. For a specific case in the process industry, these indicators have been measured with a questionnaire and analyzed using principal component analysis (PCA) [8]. Since these approaches rely on an assessment by individuals, they might be vulnerable to subjectivity and thus potential biases.

An approach to the resilience of community infrastructure towards earthquakes is proposed by Bruneau et al. [9]. The authors define the *loss of resilience* as the total deviation of the quality of a building over time from the initial quality between the time of occurrence of the earthquake and the time of full recovery [9]. Chang et al. propose an approach for the assessment of the resilience of community infrastructure towards earthquakes by measuring the *loss of performance* and the *length of recovery* [10]. A threshold for *maximum acceptable*

loss of performance and *maximum acceptable recovery time* is defined manually [10]. The probability of exceeding the thresholds yields a metric for resilience and is measured through a series of disruption simulations [10]. Although the aforementioned authors utilize their approaches for an assessment of infrastructure towards earthquakes, the proposed methodologies might also be considered for different domains like organizational or economic resilience [10]. However, applied to the organizational domain, these approaches are rather static and do not systematically account for steadily changing conditions and circumstances of a company.

A recent approach by Hosseini et al. utilizes Bayesian networks to account for uncertainty in supplier selection based on their *vulnerability* and *recoverability* [11]. For the case of an automotive OEM, supplier resilience is quantified and assessed regarding disruption propagation based on an “implicit consideration of disruption probabilities” [11]. An approach to SC resilience utilizing digital technology was proposed by Ivanov et al. [12]. The authors state that “today and looking at the near future, the SC will be as good as the digital technology behind it” and highlight the potential of digital technologies (e. g. real-time analysis of sensory data) on supply chain management and subsequently the effect of supply chain management on controlling the ripple effect in SCs [12]. By continuous analytics of relevant data, disruptions could be anticipated (*proactive* stage) or at least identified in real-time (*reactive* stage) [12]. Utilizing simulations with the knowledge of the type of disruptions, both reactive adaptations to the SC execution dynamics and proactive SC optimization could be carried out [12]. None of the aforementioned approaches provides a holistic view on resilience (e.g., neglecting internal dependencies and external data sources like weather, news, traffic, etc.) that would ultimately enable a granular resilience management.

We identified three shortfalls of past approaches to the assessment of resilience: Past approaches (1) *largely rely on assessments of individuals and may be vulnerable to subjectivity and biases*, they are (2) *rather static and do not systematically account for steadily changing conditions of a company* and (3) *don't provide a holistic view on resilience needed to enable a granular resilience management*.

3. Methodology

Given the shortfalls described above, we suggest a new approach to measuring a company's resilience that is based on an iterative, AI-driven simulation: in a nutshell, resilience is measured by looking at a bundle of KPIs, simulating the impact of potential disruptions on KPIs and learning from the simulation to adapt the different parts of the analysis (KPIs, metrics, model, and scenarios) iteratively (see Fig. 1). The underlying idea of this simulation approach

is that the performance (as measured by KPIs) of resilient companies should show relatively little changes when faced with disruptions because resilient companies should generally fare better in times of disruption than non-resilient ones. And given that we use this approach iteratively and continuously, we don't see the resilience index itself as a one-off project but rather as a continuous part of a company's daily business with dedicated resources.

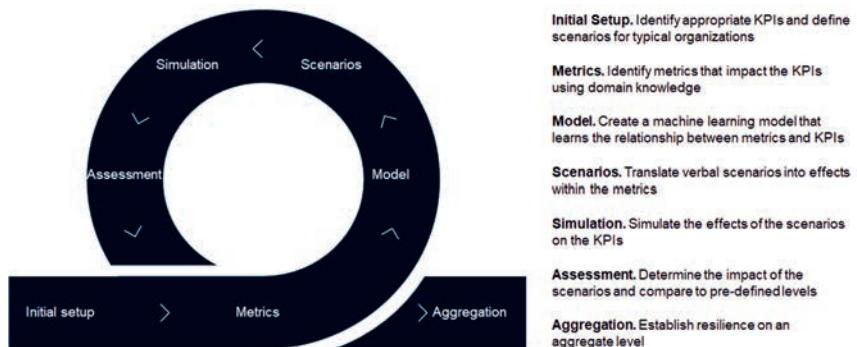


Fig. 1: Resilience Index Process

Step 1: Initial setup

The initial set-up of our approach is twofold: on the one hand, the relevant KPIs to measure a company's performance across its different departments are identified; and on the other hand, scenarios of potential disruptions are defined.

KPIs are metrics that allow for measuring the performance of a company and for the identification of performance gaps between current and desired performance [13]. Based on KPIs, managers conclude on measures to safeguard or improve current performance levels. In this regard, KPIs need to be aligned with a company's critical business objectives [13]. Based on literature [14,15,16,17,18,19,20,21,22,23] and interviews with industry partners in the project SPAICER, we developed a comprehensive list of KPIs (across different departments of a company): for instance, for (a) production: production cost per unit, machine availability, mean time between failures, downtime, reject rate; (b) supply chain management: on-time delivery of goods, lead time, timeliness; (c) human resources: absence days (illness, vacations), time-to-fill (vacancies), overtime hours, fluctuation rates.

In order to allow comparability, these shocks are initially pre-defined in a number of scenarios which represent 'typical disruptions' to companies. Literature suggests that shocks can have different origins: sources of risk that are *internal to the company* (e.g., processes, con-

tol), *external to the company but internal to the supply chain network* (e.g., demand, supply) and *external to the network* (e.g., environmental, socio-political, economic or technological) [24,25]. In the initial setup, typical disruptions are defined within these categories, such as natural catastrophes or an unexpected demand shortage. Each disruption scenario is described in plain language and with a sufficient level of detail to facilitate the understanding of the people interacting with the resilience index within the company. For instance, one such scenario might describe a long-lasting (2 weeks) snowfall at the company's (main) production site. On top of that, categories for the resilience evaluation after the simulation are pre-defined for each scenario to facilitate an assessment of the resilience level. To increase adaptation, we suggest a simple scale (low, medium, and high resilience) that defines for each scenario what impact we would expect to see in terms of magnitude as well as duration. These classifications can be based on expert assessments and in the long run, once many companies have used this resilience index, a benchmark can be made across all companies from a certain location and/or industry.

There are indicators measured internally or externally that might help to understand changes in KPIs [26]. If certain shocks occur, they will change the internal and external metrics (e. g. snowfall as an external parameter and absence of workers as an internal metric can negatively influence the KPI *on-time delivery of goods*). Thus, taking a closer look at these metrics as input parameters to learn dependencies and structures as well as finding the most influencing metrics for relevant KPIs might help to measure resilience in a company.

Step 2: Metrics

Metrics are internally and externally collected indicators that can help to understand changes in the KPIs, i.e. they have an influence on a company's KPIs [26]. Let's take *on-time delivery of goods* as an example for a relevant KPI in the area of supply chain management. In this case, there are several metrics that serve as internal input parameters having an impact on *on-time delivery of goods* per month (e.g.: set-up time, machining time, absence days of workers). Furthermore, there are several external input parameters that might have an impact on *on-time delivery of goods* (e.g.: natural catastrophe at a production location, shocks such as the recent COVID-19 pandemic). Domain experts have the knowledge about which metrics might be helpful for understanding specific KPIs [27]. Changes in these internal and external metrics caused by certain shocks might have an impact on the KPIs of a company (e. g. low stock level and delayed transportation of raw material due to COVID-19 regulations might lead to delay of raw materials which leads to a reduction of on-time delivery of goods). The more resilient a company is against internal and external disruptions or shocks that are visible in form of changes in the metrics, the less sensitive a KPI might react to these chang-

es. Our approach aims to use methods of machine learning to find patterns regarding dependencies between the metrics and KPIs. This way, we intend to extract the metrics with the highest influence on relevant KPIs in a specific company.

Step 3: Model

Machine learning (ML) as a subset of artificial intelligence is able to find patterns in data and use these to make predictions. More specifically, ML enables systems to automatically learn and improve from the given data without being explicitly programmed [28]. For our AI-based resilience index approach, we take the influencing metrics as input variables and the observed KPI as output variable and train the ML algorithm to come up with the best performing model predicting the output variable, namely KPI, for the next period and, thus, the best representation of the relationship between input and output variables (which is important for the simulation in step 4).

Let us take the example of *on-time delivery of goods* for the supply chain management department again. Due to the fact that we take *on-time delivery of goods* as our relevant KPI, we define it as the output variable in our machine learning model.

The *on-time delivery of goods* can be represented as a percentage giving information about how high the ratio between items that have been delivered on time and the total number of ordered items in a specific month is [29]. Our goal is to find the metrics that have the highest impact on *on-time delivery of goods*. Thus, with the help of domain expert knowledge about possibly high impact metrics and the availability of metrics in the company, we could consider the following metrics as input features for the machine learning model:

Internal metrics: (1) set-up time, (2) machining time, (3) waiting time, (4) stock, (5) product mix, (6) production volume, (7) volume of orders, (8) absence days of workers

External metrics: (1) natural catastrophe at production location, (2) rainfall volume at production location, (3) information about traffic jams

By training various ML models using the defined metrics as input variables and the on-time delivery of goods (in percentage) as output metrics as well as using methods such as feature engineering, we come up with the best performing model predicting the output (e.g. whether on-time delivery will increase or decrease in the next two months) by comparing the performances of the models. As black-box ML models are increasingly being employed to make important predictions in various contexts, the demand for transparency is increasing from the various stakeholders in AI [30]. A risk associated with inferring decisions from black-box models is making decisions that are not legitimate, justifiable, or that do not allow obtaining detailed explanations of their behavior [31]. Thus, explanations supporting the output of a model are crucial, e. g. in our example, where we require more information from the model

than the prediction for the KPI alone, i. e. transparency regarding the specific impact of metrics on relevant KPIs. Thus, concepts of Explainable AI (XAI) are needed in our approach, where the results can be comprehended by humans [32].

Step 4: Scenarios

The model as described above represents the company in a steady-state where everything is running relatively smoothly and as planned. This being said, the simulation in the subsequent step intends to assess the reaction of the representation of the company to various shocks. These shocks are already defined as scenarios in plain language terms in the initial setup phase. To make them usable for the machine learning model, they have to be transformed from verbal descriptions into machine-readable changes of the input parameters: a scenario with increased snowfall, for instance, means that the input parameter representing the amount of snow is increased by a certain percentage or value representing the difference between ground level and snow level in centimeters. So, each scenario has a varying effect on some of the input parameters that are related to KPIs (and eventually resilience) of different parts of the company.

In order to make the simulations comparable across companies, scenarios have to be constructed equally while, to a certain extent, the actual impact on the input parameters (e.g., minus 5% or minus 10% in inventory) might differ across companies since it is highly dependent on the actual organization or industry.

Step 5: Simulation

Equipped with a set of pre-defined scenarios and their translation in changes in certain input parameters, the scenarios can be simulated and the effects of these assessed. Again, following the literature and the definition of resilience, the resilience of a company can be analyzed across two dimensions, the magnitude of the disruption (measured as the deviation of a certain KPI from a baseline level) and the duration of the disruption (measured as the time to recover to that baseline level) (see, e.g., [33,34]). For instance, the scenario described above (2-week, unexpected snowfall) might lead to a reduction of on-time-delivery of goods by 10% compared to the baseline level and it might take 4 weeks to recover.

Step 6: Assessment

As described above, the effect of a disruption should be lower on resilient companies than on non-resilient ones. The effect itself can be measured as the magnitude and the duration of the disruption. Thus, the lower the magnitude and the shorter the duration, the higher the

resilience since the company does a better job at resisting the disruption and, if there is any impact, recovering to normality. After the impact on both is simulated, it can be evaluated using the evaluation categories that were already defined in the initial setup phase. For instance, we might expect a very resilient company to have a reduction in sales of maximally minus 2% and a time to recovery of maximally two weeks in the scenario described above.

Step 7: Aggregation

The six steps above describe the process to determine the resilience of a certain part of the organization measured by a specific KPI. Since we suggest using more than one KPI for each part of the company, the individual assessments of each of the KPIs need to be aggregated into one assessment of the resilience of the part of the organization. This way, individuals on the different levels of the organization can get an understanding of the resilience of the company and areas in which resilience might be lagging behind.

One idea for aggregating the KPIs for every part of the company could be to weigh each KPI equally and to adjust it in case there might be high dependencies between several KPIs. For example, let us assume that we defined the following six important KPIs for production: production cost per unit, machine availability, mean time between failures, downtime, reject rate. We could weigh each of these KPIs with $1/6$. If we assume high dependencies between the KPIs *mean time between failures* and *downtime*, we could weigh each of these two KPIs with $1/10$ and the remaining four KPIs with $1/5$.

4. Discussion

Taking an AI-based approach to measuring resilience is advantageous in many regards: First, our approach enables companies to objectively identify areas for improvement within their resilience strategy without being limited to subjective measures and experiences of individuals whose assessment might be prone to several biases – a fact we have witnessed in our work with industry partners. A holistic, data-based approach can go beyond what subjective, questionnaire-based indices reveal and, in the end, allows for a far more targeted and granular resilience analysis. This enables companies to not only understand their current level of resilience but also to make more informed decisions regarding the implementation of resilience measures as it opens up the opportunity to move beyond the tried-and-true measures and simulate the effectiveness of measures not previously explored.

This case-specific implementation of our resilience index allows for greater scalability and adaptability of the index that can more realistically capture company-specific characteristics like the organizational structure, product and service offering, or the companies' environment. Therefore, we intentionally do not position our resilience index as a plug-and-play solu-

tion but rather an evolutionary process. Its iterative nature does not only enable the continuous improvement of a company's resilience, but also of the index itself. Our model is not static and uses the continuous data inflow to bring model and reality closer together with each iteration step. While this approach might require some more initial set-up efforts, we are convinced that the benefits of having a tailored resilience index significantly outweigh the effort needed for implementation.

Furthermore, our approach is well-suited to perform quantitative cost-benefit analyses about the right set of resilience measures. Different from existing resilience indices, our AI-based resilience index, for instance, accounts for the steadily changing conditions of a company. It is designed to consider and process a large amount of data from both internal or external data sources in the resilience assessment. Hence, once the AI-based resilience index has been implemented, it provides a convenient way to continuously assess a company's resilience capabilities, while also eliminating human-capital intensive aspects of previous approaches at the same time. We believe that such a continuous assessment can help to foster a resilience mindset and build a resilience culture by constantly highlighting areas of success and opportunities for improvement. Not only do we view our approach as a valuable tool for improving a companies' resilience but also as a reliable and objective way to signal resilience towards external stakeholders such as potential partners or investors.

Resilience indices based on simulations, in particular, entail the risk that not all possible events and disruptions are exhaustively reflected in the defined set of simulated scenarios. The definition of scenarios is to a certain extent limited to what is knowable or at most imaginable by the designer – hence, so-called unknown unknown events might remain unconsidered in the assessment. Yet, while not fully eliminating the problem of so-called unknown unknowns, modeling and simulating considerable amounts of possible scenarios enables companies to explore disruptions that they would not be able to analyze with backward-looking approaches. Therefore, our approach helps to strengthen resilience capabilities and provides a way of considering future developments when analyzing a company's resilience. Consequently, practitioners can not only make more informed decisions but also take relevant counter-measure earlier in the process, providing a clear competitive advantage.

Apart from that, we noticed some practical limitations leading to the fact that the resilience index cannot immediately be implemented at every company. From a technical standpoint, the proliferation of our AI-based resilience index is mostly impeded by what we refer to as limited AI-readiness. Despite the promising advantages, AI adoption is currently falling short [35]. One major factor causing limited AI-readiness resides in a company's data ecosystem: for instance, steadily increasing data-privacy regulations and in discrete data silos from dis-

parate applications [36]. Furthermore, studies have shown that 82% of German companies believe that their data quality must be optimized for AI solutions and that only 2% of German companies are completely ready for AI [37,38]. However, building also on numerous in-depth discussions between researchers and practitioners of the SPAICER consortium, we are reasonably confident that the limited AI-readiness is temporary only and will be overcome soon. We recommend that companies further deploy a holistic 'data strategy' including investments for IT infrastructure harmonization and structural approaches to a more consistent data ecosystem.

Beyond AI-readiness, companies lack departments or personnel dedicated to resilience management. While dedicated risk management departments or positions are quite common [39]), our exemplary internet-based research reveals that equivalent roles for resilience are typically missing. A fact which is rather counterintuitive when bearing in mind that, in a Capgemini survey, more than 60% of 1,000 surveyed supply chain executives from different organizations state that "increasing supply chain resilience is a key priority for their supply chain organization, post COVID-19" [40]. The lack of adequate organizational structures puts the resilience index at risk of remaining unconsidered in practice. Consequently, we recommend defining a standardized approach through which the AI-based resilience index can be best embedded into managerial practice; including a standardized solution for managerial processes, structures, and roles. For instance, organizational research could help establish clear roles and responsibilities (e.g., resilience manager) and provide guidelines towards how to use the resilience index for managerial decision making (e.g., frequency of usage, ownership, etc.).

Finally, we also observed that there is no common understanding – neither in academia nor in practice – about what is meant by resilience and ultimately how it can be consistently measured (see also Section 2). Yet, building on the previously introduced aphorism of Peter Drucker, we are of the opinion that only what is understood gets measured and what gets measured gets managed. For the sake of active resilience management, we thus claim for an industry standard for resilience measurement and are reasonably confident, that our resilience index as a holistic interpretation lays the foundation for such.

5. Conclusion

Through our AI-based resilience index, we contribute to the further development of the field of resilience management. Moving away from questionnaire-based indices that might suffer from subjectivity, we provide a resilience measure that is at the forefront of technology and provides opportunities for further theoretical and practical developments.

In our research, we sat out to address shortfalls of existing resilience indices and provide companies with a more objective and data-driven approach to measuring resilience as we believe that improving companies' resilience plays a major role in ensuring continued economic success. The goal was to enrich the discussion on using quantitative measures to measure resilience and foster active resilience management. Through our close collaboration with industry partners in the SPAICER project, we identified shortfalls of existing approaches and tried to leverage these insights to create a solution that supports practitioners in their quest to becoming more resilient. Our paper describes a new AI-based resilience index approach and how it can be used to establish a resilience-enhancing loop. Fundamentally, it uses machine learning methods to establish connections between previously identified KPIs and input metrics and provides a quantitative measurement of resilience based on simulations of potential disruptions. While we firmly believe that our approach solves many of the identified shortfalls of previous models, we also identified several areas that need further research and improvement. Not all relate to the index itself but rather the lack of organizational preconditions, and we are optimistic that these can be overcome in the future. Informal conversations with practitioners confirm our optimism since data is generally widely available (yet not always accessible) and the recent disruptions have once again highlighted the need for a better understanding of a company's resilience before catastrophes happen. As soon as more companies become AI-ready, there will be many opportunities to validate and continuously improve our approach. Future research should continue the discussion on AI-based resilience indices and work towards improving companies' resilience.

References

- [1] Prusak, L. (2010, 07. Oktober). *What can't be measured*. Retrieved on 10 April 2021, from <https://hbr.org/2010/10/what-cant-be-measured>, [retrieved on: 12 April 2021]
- [2] Merriam-Webster Online Dictionary. Abgerufen am 04. Mai 2021, von <https://www.merriam-webster.com/dictionary/resilience>
- [3] Vugrin, E. D., Warren, D. E., Ehlen, M. A., & Camphouse, R. C. (2010). A Framework for Assessing the Resilience of Infrastructure and Economic Systems. In Gopalakrishnan, K., & Peeta, S., *Sustainable and Resilient Critical Infrastructure Systems* (S. 77–116). Berlin, Heidelberg: Springer.
- [4] Hosseini, S., Barker, K., & Ramirez-Marquez, J. E. (2016). A review of definitions and measures of system resilience. In Soares, C. G (Bearb.), *Reliability Engineering & System Safety* (Vol. 145(c), S. 47–61). Elsevier.
- [5] Francis, R. A., & Bekera, B. B. (2014). A metric and frameworks for resilience analysis of engineered and infrastructure systems. *Reliability Engineering & System Safety*, 121(c), 90–103.
- [6] Pettit, T. J., Croxton, K. L., & Fiksel, J. (2013). Ensuring Supply Chain Resilience: Development and Implementation of an Assessment Tool. *J Bus Logist*, 34(1), 46–76.
- [7] Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: assessing the ability to bounce back. *Int. J. Behav. Med.*, 15(3), 194–200.
- [8] Shirali, G. A., Mohammadfam, I., & Ebrahimipour, V. (2013). A new method for quantitative assessment of resilience engineering by PCA and NT approach: A case study in a process industry. *Reliability Engineering & System Safety*, 119, 88–94.
- [9] Bruneau, M., & Reinhorn, A. (2006). Overview of the resilience concept. In Proceedings of the 8th US national conference on earthquake engineering (Vol. 2040, pp. 18–22).
- [10] Shinozuka, M., Chang, S. E., Cheng, T. C., Feng, M., O'rourke, T. D., Saadeghvaziri, M. A., ... & Shi, P. (2004). Resilience of integrated power and water systems (pp. 65–86). Multidisciplinary Center for Earthquake Engineering Research.
- [11] Hosseini, S., Ivanov, D., & Dolgui, A. (2019). Review of quantitative methods for supply chain resilience analysis. *Transportation Research Part E: Logistics and Transportation Review*, 125, 285–307.
- [12] Ivanov, D., Dolgui, A., & Sokolov, B. (2019). The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. *International Journal of Production Research*, 57(3), 829–846.

- [13] Weber, A., & Thomas, R. (2005). Key performance indicators. Measuring and Managing the Maintenance Function, *Ivara Corporation*, Burlington.
- [14] Badawy, M., Abd El-Aziz, A. A., Idress, A. M., Hefny, H., & Hossam, S. (2016). A survey on exploring key performance indicators. *Future Computing and Informatics Journal*, 1(1-2), 47-52.
- [15] International Organization for Standardization. Automation systems and integration. Key performance indicators (KPIs) for manufacturing operations management, 22400-2, 2014.
- [16] Sahu, A. K., Datta, S., & Mahapatra, S. S. (2017). Evaluation of performance index in resilient supply chain: a fuzzy-based approach. *Benchmarking: An International Journal*.
- [17] Tukamuhabwa, B. R., Stevenson, M., Busby, J., & Zorzini, M. (2015). Supply chain resilience: definition, review and theoretical foundations for further study. *International Journal of Production Research*, 53(18), 5592-5623.
- [18] Pettit, T. J., Croxton, K. L., & Fiksel, J. (2013). Ensuring supply chain resilience: development and implementation of an assessment tool. *Journal of business logistics*, 34(1), 46-76.
- [19] Soni, U., Jain, V., & Kumar, S. (2014). Measuring supply chain resilience using a deterministic modeling approach. *Computers & Industrial Engineering*, 74, 11-25.
- [20] Sheffi, Y. (2001). Supply chain management under the threat of international terrorism. *The International Journal of logistics management*, 12(2), 1-11.
- [21] Jüttner, U., & Maklan, S. (2011). Supply chain resilience in the global financial crisis: an empirical study. *Supply Chain Management*, 16(4), 246-259.
- [22] Kang, N., Zhao, C., Li, J., & Horst, J. A. (2016). A Hierarchical structure of key performance indicators for operation management and continuous improvement in production systems. *International Journal of Production Research*, 54(21), 6333-6350.
- [23] Gabčanová, I. (2012). Human resources key performance indicators. *Journal of competitiveness*. 4(1), 117-128.
- [24] Christopher, M., & Peck, H. (2004). Building the Resilient Supply Chain. *The International Journal of Logistics Management*, 15(2), 1-14.
- [25] Mason-Jones, R., & Towill, D. (1998). "Shrinking the Supply Chain Uncertainty Cycle", *Control*, S.17-22.

- [26] Wetzstein, B., Leitner, P., Rosenberg, F., Brandic, I., Dustdar, S., & Leymann, F. (2009, September). *Monitoring and analyzing influential factors of business process performance*. In 2009 IEEE International Enterprise Distributed Object Computing Conference (pp. 141-150). IEEE.
- [27] Medium. Why domain knowledge is important in Data Science. Retrieved on 09 March 2021, from <https://medium.com/@anand0427/why-domain-knowledge-is-important-in-data-science-anand0427-3002c659c0a5>
- [28] Lison, P. (2015). An introduction to machine learning. *Language Technology Group (LTG)*, 1(35).
- [29] Shannon, J. (2019). *How to Measure the On-Time Delivery KPI*. Retrieved on 14 April 2021, from <https://www.visualsouth.com/blog/on-time-delivery-kpi>.
- [30] Preece, A., Harborne, D., Braines, D., Tomsett, R., & Chakraborty, S. (2018). Stakeholders in Explainable AI, Presented at AAAI FSS-18: Artificial Intelligence in Government and Public Sector, Arlington, Virginia, USA.
- [31] Gunning, D. (2017). Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA)*, *nd Web*, 2(2).
- [32] Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? *Copmuter Science, Mathematics*, arXiv preprint arXiv:1712.09923.
- [33] Falasca, M., Zobel, C. W., & Cook, D. (2008). A decision support framework to assess supply chain resilience. In *Proceedings of the 5th international ISCRAM conference*, S. 596–605.
- [34] Ponomarov, S. (2012). *Antecedents and consequences of supply chain resilience: A dynamic capabilities perspective* (Dissertation). University Tennessee, Tennessee.
- [35] Fountaine, T., McCarthy, B., & Saleh, T. (2019). *Building the AI-Powered Organization*. Retrieved on 04. Mai 2021, from <https://hbr.org/2019/07/building-the-ai-powered-organization>
- [36] Gottlieb, J. Ho, T., Kanagasabai, K., Rowshankish, K., & Weinberg, A. McKinsey (Ed.) (2019). *How to maximize the returns from your data*. Retrieved on 04 May 2021, from <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/digital-blog/how-to-maximize-the-returns-from-your-data>.
- [37] PWC, Künstliche Intelligenz in Unternehmen, 2018. Retrieved on 28 April 2021, from <https://www.pwc.de/de/digitale-transformation/kuenstliche-intelligenz/studie-kuenstliche-intelligenz-in-unternehmen.pdf>.

- [38] Deloitte, Thriving in the era of pervasive AI, 2020. Retrieved on 13 April 2021, from <https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/about-deloitte/deloitte-cn-dtt-thriving-in-the-era-of-persuasive-ai-en-200819.pdf>.
- [39] Ruggeri, C., Vanuga, C., Calagna, K., Vitters, C., & Fray, M. Deloitte (Ed.). (2019). *Reimagine risk: Thrive in your evolving ecosystem: Deloitte's 2019 survey of risk management*. Retrieved on 04 May 2021, from <https://www2.deloitte.com/us/en/insights/topics/risk-management/cro-risk-management-survey-results.html>.
- [40] Capgemini Research Institute (Ed.). (2020). *Fast Forward: Rethinking supply chain resilience for a post-COVID-19 world*. Retrieved on 04 May 2021, von <https://www.capgemini.com/research/supply-chain-resilience/>.

Nutzung der BaSys 4.2 Middleware – Lösung in der Umgebung von Industriesteuerungen

S. Schäfer, D. Schöttke, T. Kämpfe, K. Ralinovski,
Hochschule für Technik und Wirtschaft Berlin;
H. Klempnow,
KleRo GmbH Roboterautomation

Kurzfassung

Der stetig steigenden Produkt- und Variantenvielfalt begegnen erfolgreiche Hersteller mit wandelbaren Produktionsumgebungen. [1] Unternehmen gelten als wandlungsfähig, wenn sie die Strukturen der Organisation und der Ressourcen permanent an veränderte Bedingungen und Situationen der Auftragslage kurz-, mittel- und langfristig anpassen können. [2] In diesem Beitrag bezieht sich die Wandelbarkeit auf die technische Fähigkeit eines Systems, auf neue Anforderungen so reagieren zu können, dass nur geringfügige Programmänderungen notwendig sind. Diese technische Fähigkeit lässt sich ebenfalls mit dem Begriff der rekonfigurierbaren Produktionssysteme wie er bereits in [3] verwendet wird, beschreiben. Die Verwendung von domänenspezifischen Fachsprachen der DIN EN 61131-3 führt jedoch zu starren Verkettungen der Steuerung mit ihrer Sensorik und Aktorik sowie mit dem Prozess, so dass spätere Änderungen der Anlagenkonfiguration einen neuen Engineering-Prozess verbunden mit neuen Test- und Inbetriebnahme-Phasen notwendig machen. Rekonfigurierbare Produktionssysteme, verfolgen demgegenüber das Paradigma, dass Änderungen am Anlagendesign nur zu geringen Eingriffen in die Programmierung und dem Inbetriebnahme-Prozess führen. Dieser Beitrag beschäftigt sich damit, konventionelle Produktionssysteme durch die Verwendung der BaSys 4.2-Middleware in ein rekonfigurierbares Anlagendesign zu überführen.

1. Grundlagen – Einführung

Basys 4.2-Middleware

Für die hier skizzierte Infrastruktur wird auf die verfügbare Middleware BaSys 4.0 zurückgegriffen, die eine Infrastruktur für die Erstellung und Nutzung von I4.0 Komponenten bietet. BaSys 4 ist eine virtuelle Middleware für wandelbare Produktionsanlagen, welche die notwen-

digen Dienste zur Verfügung stellt. Dazu definiert BaSys 4 eine Referenzarchitektur für Produktionssysteme, die die Transformation zu Industrie 4.0 ermöglicht. Um dies zu erreichen, stellt BaSys 4 einen digitalen Zwilling, d.h. mit einer standardisierten Struktur.

Die Open-Source-Middleware Eclipse BaSyx ist eine Referenzimplementierung der BaSys 4-Konzepte. Sie bietet gebrauchsfertige Komponenten und erweiterbare Software Development Kits. [4] In diesem Beitrag wird auf die C#-Version der Fa. Bosch zurückgegriffen, welche als Entwicklungsplattform für die hier vorgestellten Infrastrukturkomponenten dient.

Die Verwaltungsschale (eng. Asset Administration Shell (AAS))

Um rekonfigurierbare Produktionssysteme zu betreiben, wird ein digitaler Zwilling benötigt. Die Plattform Industrie 4.0 beschreibt die Verwaltungsschale (AAS) als Umsetzung des digitalen Zwillings für die Industrie 4.0. Dabei ist die Verwaltungsschale das digitale Abbild eines Gegenstands. Der Gegenstand, also das physische Asset und die Verwaltungsschale zusammen bilden die sogenannte I4.0-Komponente. Im Rahmen dieses Beitrags werden sowohl physische als auch virtuelle Objekte (z. B. Dokumente, Software) als Assets betrachtet, welche Gegenstände der Verwaltungsschale sein können. Verwaltungsschalen stellen die Schnittstellen für eine I4.0 konforme Kommunikation zur Verfügung. [5]

Typen von Verwaltungsschalen

Derzeit werden drei Typen von Asset Administration Shells unterschieden:

Typ 1 Verwaltungsschalen bestehen aus serialisierten Dateien, z. B. XML- oder JSON-Dateien. Serialisierte Verwaltungsschalen enthalten statische Informationen und können als Dateien verteilt werden. Eclipse BaSyx-Komponenten können Verwaltungsschalen vom Typ 1 senden und empfangen. Das Datenmodell von Verwaltungsschalen des Typs 1 ist durch das AAS-Metamodell definiert.

Typ 2 Verwaltungsschalen existieren als Laufzeitinstanzen. Sie werden auf Servern gehostet und können sowohl statische, d. h. sich weniger häufig ändernde Informationen enthalten, aber auch mit anderen Komponenten interagieren. Auf diese Weise bieten Typ 2 Verwaltungsschalen ein Frontend z. B. zu Gerätediensten, Live-Daten von Sensoren, Produkten oder Echtzeit-Verfügbarkeit und Lieferzeiten für Ersatzteile. Die Typ 2 Verwaltungsschalen stellt Eigenschaften und Operationen zur Verfügung und ist in der Lage, veränderte Bedingungen mit Ereignissen zu signalisieren. Das Datenmodell von Typ 2 Verwaltungsschalen wird ebenfalls durch das AAS-Metamodell definiert. Darüber hinaus definiert das AAS eine generische Laufzeitschnittstelle, die den Zugriff auf Eigenschaften, Operationen und Ereignisse ermöglicht. Typ 2

Verwaltungsschalen können daher einheitliche Schnittstellen zu hetero-genen Entitäten realisieren.

Typ 3 Verwaltungsschalen erweitern die Verwaltungsschalen vom Typ 2. Sie implementieren zusätzlich ein aktives Verhalten, d. h. sie können selbständig in die Kommunikation und Verhandlung eintreten. Die VDI/VDE 2139 definiert eine Sprache für Typ 3 Verwaltungsschalen. [4]

Teilmodelle

Verwaltungsschalen können aus einem oder mehreren Teilmodellen bestehen. Jedes Teilmodell (Submodel (SM)) enthält eine strukturierte Menge von Merkmalen. Teilmodelle und Merkmale können Typ- oder Instanz-bezogen sein, in denen u.a. Informationen und Funktionalitäten eines Gegenstandes beschrieben werden. Merkmale (Properties) sind im Kontext Industrie 4.0 das übergreifende Konstrukt, um statische Merkmale, Parameter, Methoden, Fähigkeiten, Operationen, Zustände und viele andere Dinge auszudrücken. [6]

Kommunikation

Im Kontext der Industrie 4.0 wird eine Kommunikation zwischen den Verwaltungsschalen benötigt. Hinsichtlich der Anforderung an die Kommunikation wird in drei Szenarien unterschieden. Das erste Szenario beschreibt die Anforderungen an die „Echtzeitdaten-kommunikation in flexiblen Produktionssystemen“. Das zweite Szenario stellt auf die „Kommunikation für Engineering und Asset-Management“ und das dritte Szenario fokussiert auf die „Produktdaten-kommunikation über die Lebenszeit“. [7] Die in diesem Beitrag beschriebene Kommunikation basiert auf dem Feldbusprotokoll OPC UA. Für die weitergehende Kommunikation wird die im BASYX-Framework bereits integrierte REST- Schnittstelle verwendet. Für den Zugriff auf die benötigten Ressourcen wird ein einheitlicher Ressourcenbezeichner, englisch Uniform Resource Identifier, abgekürzt URI benötigt.

2. Anwendungsszenarien

Die Einführung einer Verwaltungsschale als Digitaler Zwilling in der Industrie 4.0-Anlage erlaubt eine Vielzahl neuer Anwendungsfälle entlang des gesamten Lebenszyklus der Anlage. Dies kann die Unterstützung im Engineering bei Anlagenmodifikationen, die Gewinnung weiterer Erkenntnisse im operativen Betrieb, vorausschauende und zustandsorientierte Wartungsszenarien oder auch Entsorgungskonzepte sein. Die Anwendungsszenarien müssen nicht auf die Maschine oder Anlage selbst beschränkt sein, sondern können auch auf ganze

Fabriken oder Branchen skaliert werden. In [8] und [9] finden sich mehrere Beispiele für Anwendungsszenarien, die entweder mit Hilfe einer Verwaltungsschale als digitalem Zwilling unterstützt oder erst durch Verwaltungsschalen ermöglicht werden. So können Verwaltungsschalen beispielsweise die Prozessführung/-steuerung und das Life-Cycle-Management einer Produktionsanlage unterstützen.

3. Anwendungsfall – Bestandsumgebungen

Bestandsanlagen

Industriesteuerungen, die in den Fachsprachen nach DIN EN IEC 61131-3 programmiert werden, sind gemäß DIN EN 62264-1 in der funktionalen Hierarchie auf der Ebene 2 für die Überwachung und Steuerung der physischen Prozesse angesiedelt. [10] Durch ihre physikalischen Schnittstellen sind sie häufig fest mit den Betriebsmitteln auf der Ebene 1 verbunden. Auch der Einsatz von Feldbussen, welche adress- oder speicherbasiert arbeiten, ändert an diesem Paradigma nichts. So gehen Änderungen an der Anlagenkonfiguration auf den Ebenen 0 bis 2 mit umfangreichen Engineering-Tätigkeiten einher, welche bei vernetzten Produktionssystemen über die einzelne Anlage hinausgehen können. Die Einführung der Verwaltungsschale als digitalem Zwilling kann die Engineering-Abläufe erheblich vereinfachen. Die Verwaltungsschale entspricht einer neuen Hardwareabstraktionsschicht und hebt die bisherige starre Kopplung in der Anlage zugunsten einer flexiblen Verbindung auf.

Bestandsaufnahme

Um eine bestehende Anlage zu migrieren, ist zunächst eine Bestandsaufnahme der Hard- und Software und der Dokumentation erforderlich. Möglicherweise haben sich im Lebenszyklus der Anlage bereits Modifikationen in den Anforderungen und der Gestaltung der Systemumgebung ergeben, wodurch eine Anpassung der Systemkomponenten notwendig wird. Weitere Schwierigkeiten können sich dadurch ergeben, dass es während der Implementierungsphase zu Abweichungen von der Anforderungsphase kam und diese Abweichungen nicht immer konsistent abgeglichen wurden. Diese Effekte treten vor allem bei fehlender Versionierung und Dokumentation der Projektunterlagen auf. Gegebenenfalls ist je nach Stand und Alter der Anlage eine neue Engineering-Phase beginnend bei der Anforderungsermittlung notwendig. Die Vorgehensweise für Anforderungsermittlungen wird unter anderem in [11] erläutert. In [12] und [13] finden sich Vorgehensmodelle und Methoden für die strukturierte Modellierung von Programmen, welche in den Fachsprachen nach DIN EN IEC 61131-3 erstellt wurden. In den genannten Beiträgen wird u.a. von der Erstkonfiguration von Anlagen und den notwendigen Strukturen ausgegangen. Die Migration von Bestandsanlagen, also sogenannten

Brownfield-Anlagen, in ihrer Gesamtkonstellation oder in Teilbereichen hin zu I4.0-Anlagen stellt im Gegensatz zu der Neuerrichtung einer Greenfield-Anlage auf der „grünen Wiese“ als wandelbares Produktionssystem eine anspruchsvollere Herausforderung dar. Die Erstellung neuer Programm-Organisationseinheiten (Programming Organization Units, POU) und auch die Erweiterungen und Anpassungen bereits funktionierender POU's kann zu ungewollten Seiteneffekten führen, so dass hier, basierend auf den neuen Anforderungen an die I4.0-Anlage, umfangreiche Testszenarien zum Tragen kommen können. Allerdings kann die Verwendung bereits getesteter Programmbausteine auch zu erheblichen Vereinfachungen für die Migration führen. Da hier automatisierte Teststrategien für die Anlagenmigration bislang noch fehlen, ist hier eine jeweilige Einzelfallprüfung notwendig. Dieser Aufwandsaspekt ist bei der Überführung von Bestandsanlagen in moderne I4.0-Strukturen nicht zu unterschätzen.

Da in vielen Fällen im Prozess des Engineerings eine asynchrone Aufbereitung der Anlagenbestandteile erfolgt, stehen für die Testphase nicht immer die realen Anlagenumgebungen zur Verfügung. Auch für diese Fälle der virtuellen Inbetriebnahme stehen digitale Zwillinge, die das Anlagenverhalten in vielen Aspekten der Automatisierung abbilden können, zur Verfügung. Beispiele für virtuelle Inbetriebnahmen finden sich in [14] und [15]. Modellierungsumgebungen wie in [16] und [17] beschrieben, stellen bereits umfangreiche Bibliotheken und das funktionale und zeitliche Verhalten von Anlagen zu simulieren.

Eine alternative Möglichkeit es Testens von Automatisierungsfunktionen ist die Verwendung eines digitalen Zwillings als Verwaltungsschale wie er in [5] beschrieben ist.

Wenn grundlegende Funktionen auf der Steuerungsseite sinnvoll gekapselt werden, kann die Steuerung und Synchronisation der Funktionen anschließend über einen Repräsentanten der Steuerung auf der Ebene der Verwaltungsschale erfolgen. Hierdurch ergeben sich flexible und leicht anpassbare Strukturen innerhalb einer Domäne.

4. Beschreibung der I4.0 – Infrastruktur

Sofern die Engineering-Dokumente aus der Bestandsaufnahme vollständig und widerspruchsfrei vorliegen, können die Anforderungen nun in einen maschinenlesbaren Funktionskatalog unabhängig von der jeweiligen Hardware oder vorhandener Software übertragen werden. Für die I4.0-Anlage wird die frei zugängliche BaSys 4 Referenzarchitektur für Produktionssysteme verwendet. Die Open-Source Middleware Eclipse BaSyx ist eine Referenzimplementierung der Konzepte von BaSys 4. Eine Einführung in die Middleware liefert die dreiteilige Serie in der Fachzeitschrift iX unter [18].

Erzeugung von Verwaltungsschalen

Für I4.0-Anlagen werden Verwaltungsschalen als digitale Zwillinge benötigt. Diese lassen sich prinzipiell mit jedem Texteditor erstellen. Als Werkzeug für die Erzeugung und Beschreibung von Typ 1-Verwaltungsschalen steht auch der frei verfügbare AASX-Package-Explorer zur Verfügung. Der AASX Package Explorer ist ein Software-Tool, mit dem Unternehmen Verwaltungsschalen in den Formaten XML und JSON erzeugen und editieren können. [19]

Beispiele für die Vorgehensweise der Erzeugung von Verwaltungsschalen der Typen 1 bis 3 finden sich in [20] und in [21]. Die mit dem AASX-Package Explorer beschriebenen Typ 1-Verwaltungsschalen können u.a. in ein maschinenlesbares JSON-Format exportiert werden. Diese JSON-Datei lässt sich mittels Analyse, Filterung, und Erweiterung in eine PLCopen-konforme XML-Datei überführen. [22]. Die überführte PLCopen-konforme XML-Datei lässt sich wiederum als Template in diversen Projektierungsumgebungen, wie beispielsweise Codesys V3.x importieren und nutzen.

Startkomponenten

Eine funktionierende I4.0-Anlage benötigt eine Minimalumgebung. Die in diesem Beitrag vorgeschlagene Konfiguration besteht aus den Startkomponenten für die möglichst generische Erzeugung der benötigten Datenstrukturen, Schnittstellen und Typ 2-Verwaltungsschalen.

In diesem Beispiel besteht die Startumgebung aus einer Registry, als einem Verzeichnis, in dem sich die Typ 2-Verwaltungsschalen mit ihren Submodellen registrieren können, dem AAS-Generator, welcher Typ 1-Verwaltungsschalen in Typ 2-Verwaltungsschalen überführt. Der AAS-Generator stellt zudem die Schnittstellen für den Typ 2-Verwaltungsschale und der mit ihr gekoppelten Industriesteuerung zur Verfügung und überprüft u.a. die Erreichbarkeit der Industriesteuerung. Der AAS-Generator verfügt über eine Datenbank, in welcher bereits Typ 1-Verwaltungsschalen für die spätere Verwendung abgelegt wurden.

Die Ablage dieser Verwaltungsschalen erfolgt mittels eines Storage-Services, der in festgelegten Intervallen einen Dateiordner regelmäßig auf neue JSON-Dateien überprüft. Sollten neue Dateien vorhanden sein, prüft der Storage-Service anschließend die Dateien auf das Vorhandensein bestimmter Merkmale wie z.B. AAS oder SM und lädt diese dann in die Datenbank. Aufgrund der objektorientierten Struktur der Verwaltungsschale haben sich die Autoren für eine MongoDB entschieden.

Die Industriesteuerung auch Speicherprogrammierbare Steuerung (SPS, engl. programmable logic controller, PLC) genannt, ist aufgrund ihrer Verbreitung in der industriellen Automation ein wesentlicher Bestandteil auch künftiger I4.0-Umgebungen.

Damit die Industriesteuerung sowohl auf Anforderungen anderer I4.0-Komponenten als auch auf Signale und Ereignisse des unterlagerten Prozesses reagieren kann, wurde in der Steuerung ein Zustandsautomat analog der PackML-Spezifikation Funktionsbausteines programmiert. Der PLC-interne Zustandsautomat hat darüber hinaus die Aufgabe, die Typ 2-Verwaltungsschale über jeden Zustandswechsel zu informieren. Da in der Typ 2-Verwaltungsschale der gleiche PackML-konforme Zustandsautomat ausgeführt wird, findet im Produktivbetrieb die permanente Synchronisation der SPS mit Ihrer verknüpften Typ 2-Verwaltungsschale statt.

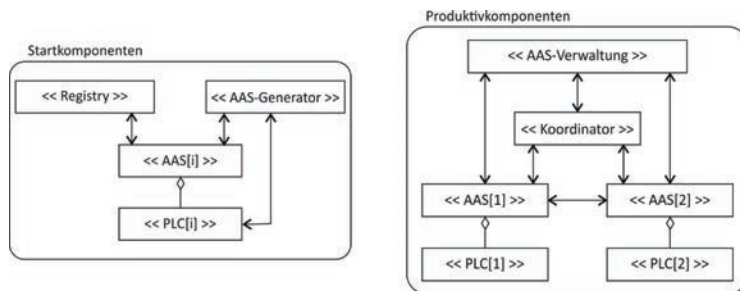


Bild 1: Komponenten der Start- und Produktivphase (Auszug)

Produktivkomponenten

Für den Produktiveinsatz werden darüber hinaus weitere Produktivkomponenten benötigt. Die Koordinierung von Arbeitsaufträgen erfolgt hier durch den Koordinator. In der Eclipse BaSys Referenzimplementierung wird dieser auch als Kontrollkomponente bezeichnet. Im einfachsten Fall führt der Koordinator ein einfaches Mapping von Eigenschaften innerhalb der Submodelle von Verwaltungsschalen durch. Es lassen sich aber auch Koordinatoren als Zustandsautomaten oder als BPMN-Modelle implementieren. Eine zusätzliche Auftragsverwaltung beispielsweise in Form eines MES-Systems könnte auch den Koordinator oder auch Typ 2- oder Typ 3-Verwaltungsschalen direkt ansprechen.

5. Generierung & Nutzung von Verwaltungsschalen

Generierung von Typ 2-Verwaltungsschalen

Im Folgenden soll auf den Generierungsprozess der Typ 2-Verwaltungsschalen näher eingegangen werden. Das in Bild 2 dargestellte Sequenz-Diagramm erläutert den Generierungsprozess für die Erzeugung einer Verwaltungsschale und der benötigten Kommunikationsschnittstellen. Um Verwaltungsschalen vom Typ 1 in Typ 2-Verwaltungsschalen zu überführen, wurde im Rahmen dieses Forschungsvorhabens ein AAS-Generator entwickelt. Der AAS-Generator ist von seiner Struktur selbst eine Typ 3-Verwaltungsschale.

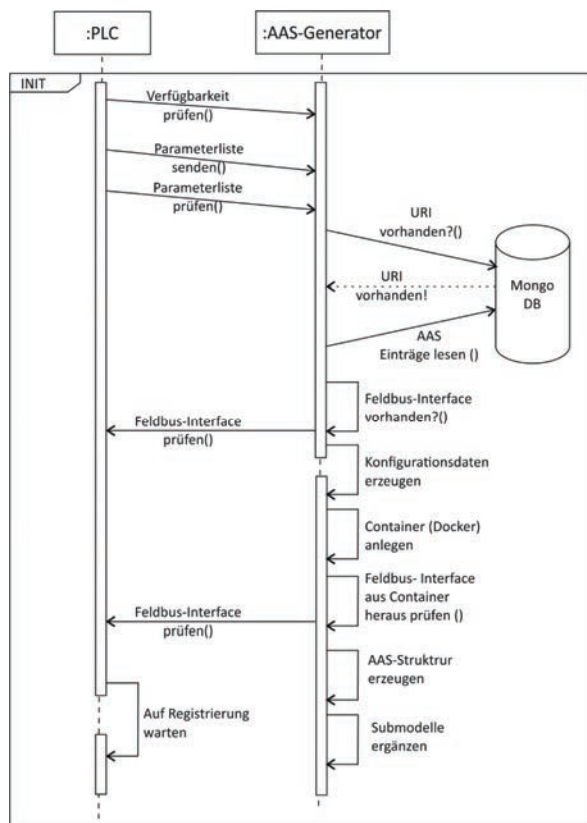


Bild 2: Sequenz des Generierungsprozesses (Auszug)

Die SPS besitzt eine REST-Schnittstelle und benötigt für die Kommunikation eine URI und die Kenntnis der Submodelle und deren Properties. Der AAS-Generator verfügt über ein Submodell, welches die Eigenschaft „Heartbeat“ als Zeichenkette mit dem Inhalt „alive“ beinhaltet. Die

SPS prüft die Verfügbarkeit des Generators, indem sie eine GET-Anfrage versendet. Der AAS-Generator liefert, sofern er verfügbar ist, die Zeichenkette „alive“ zurück

Im Falle einer fehlerhaften Kommunikation erhält man einen Timeout. Wenn der Generator erreichbar ist, sendet die SPS einen POST-Befehl. Dieser POST-Befehl enthält eine weitere URI, welche unter anderem die Verwaltungsschalen-ID, sowie weitere zusätzliche Submodell-IDs enthält.

Außerdem werden an den AAS-Generator zusätzliche Informationen für die spätere Feldbuskommunikation zwischen der SPS und ihrer Verwaltungsschale gesendet. Dies sind beispielsweise künftigen IP-Adressen, benötigte Port-Adressen und das zu verwendende Kommunikationsprotokoll. Der AAS-Generator liest die korrespondierende Typ 1-Verwaltungs-schale aus der zuvor beschriebenen MongoDB-Datenbank aus. Wenn eine Verwaltungs-schale mit einer übereinstimmenden ID vorhanden ist, liest der AAS-Generator die Einträge für die Verwaltungsschale und die Submodelle aus der MongoDB und erzeugt ein Laufzeitobjekt im Speicher.

Nun prüft der AAS-Generator die Verfügbarkeit der Feldbuskommunikation. Beispielsweise erzeugt der AAS-Generator im Falle einer OPC UA-Kommunikation als OPC-Client eine Session und stellt eine Verbindung zur SPS, die als OPC Server fungiert, her.

Der AAS-Generator erzeugt nun eine Konfigurationsdatei als JSON-Datei, die alle Information der Typ1-Verwaltungsschale und die Submodelle nebst Feldbus-Schnittstellen enthält. Die Konfigurationsdatei enthält darüber hinaus ebenfalls die Informationen, welche für die Kommunikation zwischen der Typ 2-Verwaltungsschale und der Registry benötigt werden.

Dann legt der AAS-Generator legt einen Docker-Container an. Dieser Docker-Container wird aus einem Container-Image gestartet, das als einfache Datei leicht von einem System auf ein anderes übertragbar ist und somit die Anforderungen an verteilte Umgebungen erfüllt. Innerhalb des Docker-Containers werden nun von einem mit dem Container gestarteten AAS-Prozess Laufzeit-Objekte erzeugt. Diese verschiedenen Laufzeit-Objekte werden dann zu einer Typ2-VWS zusammengeführt.

Registrierung der Verwaltungsschalen (hier Typ 2)

Das folgende Sequenz-Diagramm stellt die Kommunikation zwischen der aktiven Verwaltungsschale, der Registry und der jeweiligen SPS dar.

Ein Prozess Namens „AAS“, der mit dem Docker-Container gestartet wurde, liest die zuvor im JSON-Format erstellte Konfigurationsdatei ein. Der AAS-Prozess prüft aus dem Container heraus erneut die Feldbus-Kommunikation. Dieser erneute Kommunikationstest ist notwendig,

damit sichergestellt ist, dass der Docker-Container, auch wenn er verlagert wurde, die Feldbuskommunikation zur SPS aufbauen kann.

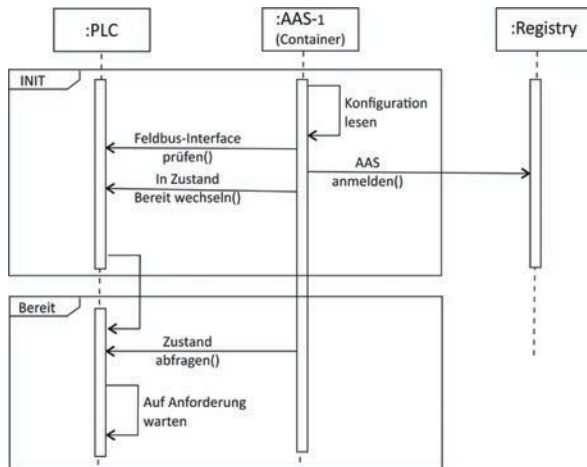


Bild 3: Registrierung des AAS (Auszug)

Sollte die Feldbus-Kommunikation erfolgreich verlaufen sein, meldet sich die Verwaltungsschale an der Registry an und übergibt der Registry die Beschreibungen der Verwaltungsschale, deren Submodule und deren Kommunikations-Endpunkte, so dass die Verwaltungsschale nun als aktive Ressource zur Verfügung steht.

Nach erfolgter Anmeldung an der Registry sendet die Verwaltungsschale der SPS den Befehl für den Zustandswechsel von „STOP“ in den Zustand „IDLE“.

Um eine ressourceneffizienter Feldbuskommunikation zu nutzen, wurde für die Kommunikation zwischen SPS und ihrer Verwaltungsschale der erweiterte Kommunikationsmechanismus OPC UA Publish-Subscribe (Pub/Sub) etabliert.

Bei diesem erweiterten Kommunikationsprotokoll sendet die SPS als Publisher ihre Daten verbindungslos ins Netzwerk und Verwaltungsschale kann als Subscriber diese Daten nutzen, ohne dass eine ressourcenverbrauchende Verbindung aufgebaut werden muss. Wenn sich nun auf Seiten der SPS eine Wertänderung (Change of value) ergibt, wird die Verwaltungsschale hierüber informiert. Die SPS wechselt nun in den Zustand betriebsbereit („IDLE“) und wartet auf Anforderungen aus dem Prozess oder weiterer I4.0-Komponenten

Produktivbetrieb - Nutzung von Verwaltungsschalen

Der Produktivbetrieb wird nun anhand des folgenden Sequenzdiagramms näher erläutert. Als weitere Komponente für den operativen Betrieb wird ein Koordinator benötigt. Um die Auftragsausführung flexibler zu gestalten, soll der Koordinator künftig über weitergehende Fähigkeiten verfügen. Je nach Auftragsanforderung soll dann der Koordinator Registry nach Verwaltungsschalen mit den für den Auftrag benötigten Eigenschaften und Merkmalen durchsuchen und diese Ressourcen dann buchen. Nach erfolgter Ressourcenbuchung kann die I4.0-Anlage nun in den operativen Betrieb wechseln.

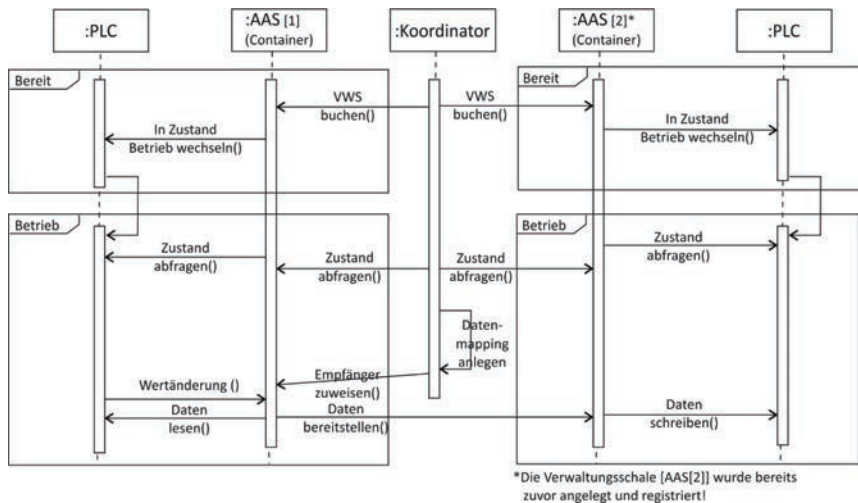


Bild 4: Anforderungs- und Produktivphase (Auszug)

Hierfür sendet die Verwaltungsschale dann die Anfrage für den Zustandswechsel von IDLE auf EXECUTE an die SPS. Wenn dieser Wechsel innerhalb der SPS erfolgt ist, wird die Verwaltungsschale über die OPC Schnittstelle informiert und synchronisiert ihren Zustandsautomaten entsprechend.

Nachdem die Verwaltungsschalen gebucht wurden und sich im EXECUTE-Modus befinden, erfolgt das Datenmapping, also das Mapping zwischen Eigenschaften und Merkmalen innerhalb von Submodellen der gebuchten Verwaltungsschalen. Jede Verwaltungsschale hat ein subscription Submodel. Innerhalb dieses Submodels gibt es den Funktionsaufruf „subscribe“, welcher die drei Parametern URI, ID des benötigten Properties, welches zugeordnet werden

soll, und ID des Ziel-Properties enthält. Der Koordinator ruft die „Subscribe“-Funktion innerhalb des Submodels auf und schreibt die jeweiligen Properties in die verkettete Liste des jeweiligen Quell-Properties. Dies erfolgt in der anderen Verwaltungsschale analog, so dass die beiden Verwaltungsschalen über ihre jeweiligen gespiegelten Einträge in den verketteten Listen miteinander gekoppelt sind. Nach erfolgtem Daten-Mapping agieren die Verwaltungsschalen als digitale Zwillinge ihrer zugeordneten Steuerungen.

6. Zusammenfassung und Ausblick

Die hier skizzierte Infrastruktur einer Industrie 4.0 Anlage zeigt, welche Voraussetzungen in zunächst erfüllt werden müssen, damit eine Bestandsanlage in eine Industrie 4.0-Anlage überführt werden kann.

Der Weg zu einer rekonfigurierbare Industrie 4.0-Anlage führt über die Erzeugung von Verwaltungsschalen, welche als digitale Zwillinge ihrer industriellen Partner-Steuerungen fungieren, zur Implementierung moderner und offener Schnittstellen für eine ressourcenfreundliche Kommunikation und der Einführung zusätzlicher Verwaltungs- und Managementkomponenten für die Initialisierungsphase und den operativen Betrieb.

Mit Hilfe des in diesem Beitrag gezeigten Transformationsmodells für Verwaltungsschalen lassen sich nun Templates generieren, welche den Engineering-Aufwand bei der Projektierung wesentlich vereinfachen können.

Im Beitrag wird u.a. das Zusammenspiel der ergänzenden Verwaltungsschalen im Kontext der BaSyx - Middleware mit Industriesteuerungen dargestellt. Das in C# entwickelte Eclipse BaSyx Software Development Kit der Fa. BOSCH liefert hierfür eine stabile und professionell dokumentierte Infrastruktur, auf welcher kommende Projekte und Anwendungen aufsetzen können. Die neuen Typ 2-Verwaltungsschalen als digitale Zwillinge ihrer Partner-Steuerungen eröffnen dem Anlagenbetreiber eine Vielzahl neuer Möglichkeiten, um seine Anlage einfacher an veränderte Anforderungen anzupassen und darüber hinaus auch weitere Erkenntnisse für eine effizientere oder stabilere Produktion zu gewinnen. Die hier vorgestellte Infrastruktur soll anhand weitere Praxisbeispiele künftig weiter ausgerollt werden.

Diese Forschungsarbeit wurde im Forschungsvorhaben „OpenBasys 4.0“ durch das Bundesministerium für Bildung und Forschung BMBF unter der Nummer (01IS1900A) gefördert und ist ein BaSys 4 Satellitenprojekt für anwendungsorientierte Projekte.

Die Autoren bedanken sich für die Unterstützung des BaSys 4-Teams.

Literatur

- [1] Wolf, W.: Entwicklung eines agentenbasierten Steuerungssystems zur Materialflussorganisation im wandelbaren Produktionsumfeld, 2021
- [2] Westkämper, E. u. Zahn, E.: Wandlungsfähige Produktionsunternehmen. Berlin, Heidelberg: Springer Berlin Heidelberg 2009
- [3] Stehle, T. u. Heisel, U.: Konfiguration und Rekonfiguration von Produktionssystemen. In: Spath, D. u. Westkämper, E. (Hrsg.): Handbuch Unternehmensorganisation: Strategien, Planung, Umsetzung. Springer Reference Technik. Berlin and Heidelberg: Springer Vieweg 2016, S. 1–35
- [4] BaSyx - Eclipsepedia. 2021. <https://wiki.eclipse.org/BaSyx>, abgerufen am: 27.02.2021
- [5] Plattform Industrie 4.0: Die Verwaltungsschale im Detail - von der Idee zum implementierbaren Konzept: [AKTUALISIERTE VERSION]. 2021. <https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/verwaltungsschale-im-detail-pr%C3%A4sentation.html>, abgerufen am: 08.05.2021
- [6] Plattform Industrie 4.0: Verwaltungsschale in der Praxis: Wie definiere ich Teilmodelle, beispielhafte Teilmodelle und Interaktion zwischen Verwaltungsschalen (Version 1.0). 2021. <https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/2020-verwaltungsschale-in-der-praxis.html>, abgerufen am: 08.05.2021
- [7] ZVEI Fachverband Automation: Kommunikation im Industrie-4.0-Umfeld: Welchen Herausforderungen hat sich die industrielle Kommunikation im Kontext von Digitalisierung und Industrie 4.0 zu stellen? Fachverband: Whitepaper - Teil 4. 2021. https://www.zvei.org/fileadmin/user_upload/Presse_und_Medien/Publikationen/2018/April/Kommunikation_im_Industrie-4.0-Umfeld/Kommunikation_im_Industrie-4.0-Umfeld_Download-Neu.pdf, abgerufen am: 08.05.2021
- [8] Plattform Industrie 4.0: Fortschrittsbericht 2021. online 2021. <https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/2021-fortschrittsbericht.html>, abgerufen am: 09.05.2021
- [9] Plattform Industrie 4.0: Fortschreibung der Anwendungsszenarien der Plattform Industrie 4.0. 2016. <https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/fortschreibung-anwendungsszenarien.html>, abgerufen am: 09.05.2021
- [10] Beuth Verlag: DIN EN 62264-1 2014-07: Integration von Unternehmensführungs- und Leitsystemen, DIN EN 62264-1 2014-07. Berlin: Beuth Verlag GmbH 2014.
- [11] Rupp, C.: Requirements-Engineering und -Management: Aus der Praxis von klassisch bis agil. München: Hanser 2014

- [12] Sieber, C., et al.: Strukturierte Modellierung von Validierungsregeln. atp magazin 60 (2018) 10, S. 50–59
- [13] Schäfer, S., et al.: Flexible Erweiterung von Arbeitsräumen. atp-edition Automatisierungstechnische Praxis, Band 58 Nr. 07-08 (2016), S. 58–67
- [14] Chan, R. u. Kraus, M.: Virtuelle Inbetriebnahme in der Prozessindustrie. http://ojs.di-verlag.de/index.php/atp_edition/article/view/2251, abgerufen am: 10.05.2021
- [15] Wolf, G., et al.: Integrierte virtuelle Inbetriebnahme: Modellanforderungen, Workflows und Anwendungsbeispiele. <https://www.vulkan-shop.de/integrierte-virtuelle-inbetriebnahme-1153>, abgerufen am: 10.05.2021
- [16] Siemens Deutschland: SIMIT Simulation. 2021. <https://new.siemens.com/de/de-produkte/automatisierung/industrie-software/simit.html>, abgerufen am: 10.05.2021
- [17] WINMOD: Montage- und Verkettungssysteme: Virtuelle Inbetriebnahme mit dem Digitalen Zwilling von WinMOD. 2021. https://www.winmod.de/de/loesungen/fabrik-automation/mechatronische-systeme/?gclid=Cj0KCQjws-OEBhCkARIsAPHokIZEMHdeAd-MQgrkb0UeLIZ4Mz2INBXDFIKwViKAZPmS6mRG60CvouwYaAuxqEALw_wcB, abgerufen am: 10.05.2021
- [18] Kuhn, T., Schnicke, F. u. Ziesche, C.: Tutorial, Teil 1: BaSys und Eclipse BaSyx – Middleware für die wandelbare Produktion. Heise (2019)
- [19] Plattform Industrie 4.0: Direkter Einstieg für Entwickler: Plattform Industrie 4.0 veröffentlicht aktualisierte Verwaltungsschalen-Spezifikation mit OPC UA-, AML- und RDF-Mapping. https://www.plattform-i40.de/PI40/Redaktion/DE/Pressemitteilungen/2019/2019-11-25-VWS_SPS.html, abgerufen am: 11.05.2021
- [20] Bader, S., et al.: Details of the Asset Administration Shell : Part 1: The exchange of information between partners in the value chain of Industrie 4.0, Federal Ministry for Economic Affairs and Energy (BMWi) 2020
- [21] Plattform Industrie 4.0: Praxisbeispiel 2: Verwaltungsschale selbst gemacht: der AASX Package Explorer. 2021. <https://www.plattform-i40.de/PI40/Redaktion/DE/Newsletter/2019/Ausgabe21/2019-21-Praxisbeispiel2.html>, abgerufen am: 28.02.2021
- [22] PLCOpen.org: Downloads | PLCopen. 2021. https://plcopen.org/downloads?field_category_target_id=164, abgerufen am: 10.05.2021

Refaktorisierung von Steuerungssoftware cyber-physischer Produktionssysteme – Potentiale und Nutzen

Eva-Maria Neumann, Univ.-Prof. Dr. **Birgit Vogel-Heuser**,
Juliane Fischer, Lehrstuhl AIS, Technische Universität München;
Dr. **Philipp Puntel Schmidt**, Dr. **Martin Witte**, Siemens AG, Nürnberg;
Nikolai Langer, Brückner Maschinenbau GmbH & Co. KG, Siegsdorf;
Lucia Becerril, ID-Consult GmbH, München;
Dr. **Matthias Bengel**, teamtechnik Maschinen und Anlagen GmbH,
Freiberg am Neckar

Kurzfassung

Der Funktionsumfang von Steuerungssoftware in automatisierten Produktionsanlagen steigt kontinuierlich an, was einen erheblichen Zuwachs der Softwarekomplexität nach sich zieht. Um eine hohe Softwarequalität und Wartbarkeit zu gewährleisten und dadurch langfristig wettbewerbsfähig zu bleiben, sind gezielte Refaktorisierungsmaßnahmen erforderlich, um die Struktur der historisch gewachsenen Software zu optimieren, ohne dabei ihre Funktion zu ändern. Obwohl Codeanalysemaßnahmen ein mächtiges Werkzeug zur Identifikation von Refaktorisierungspotentialen sind, werden diese aktuell kaum in der industriellen Praxis genutzt, z.B. aufgrund des initial hohen Aufwands bei der Einführung. In diesem Beitrag werden drei Use Cases vorgestellt, wie Codeanalyse durch geeignete Toolunterstützung in verschiedenen Phasen des Lebenszyklus von verschiedenen Stakeholdern genutzt werden können, um systematisch Optimierungspotentiale in der Software zu identifizieren und dadurch die Qualität zu verbessern. Die Use Cases wurden mit Experten aus dem Maschinen- und Anlagenbau in Deutschland evaluiert und die Nützlichkeit der Codeanalysemechanismen bestätigt.

1. Einführung und Motivation für die Refaktorisierung von Steuerungssoftware

Systematische Wiederverwendung und hohe Qualität von Steuerungssoftware sind für Unternehmen im Maschinen- und Anlagenbau wesentliche Voraussetzungen, um langfristig wettbewerbsfähig zu sein. Aufgrund des langen Lebenszyklus automatisierter Produktionssysteme (aPS) von bis zu mehreren Jahrzehnten ändern sich die Anforderungen an diese Systeme während ihrer Betriebsphase. Die Änderungen werden zu großen Teilen bevorzugt durch Anpassungen in der Software realisiert, da diese im Vergleich zu mechanischen und elektrischen

Komponenten kurzfristig leichter angepasst werden kann. Aufgrund des hohen Zeit- und Kostendrucks in der Entwicklung sowie mangelnder Infrastruktur wird Software jedoch oftmals durch unstrukturierte Funktionserweiterungen weiterentwickelt, was langfristig zu einem unberechenbaren Wachstum der Software mit hoher Variantenvielfalt führt. Mithilfe einer gezielten Refaktorisierung, also der Optimierung der Softwarestruktur ohne Änderung des Verhaltens [1], kann dem entgegengewirkt werden. In der Informatik gibt es bereits eine Vielzahl verschiedener Ansätze aus der Codeanalyse, um modulare Software von hoher Qualität zu erstellen oder Verbesserungspotentiale aufzudecken. Diese sind jedoch auf Steuerungssoftware in aPS nur bedingt übertragbar, z.B. aufgrund von Einflüssen aus anderen Disziplinen (z.B. Elektrotechnik und Mechanik) oder der Forderung nach Anpassbarkeit während und nach der Inbetriebnahme durch Techniker ohne Softwarehintergrund. Obwohl es auch für das Software Engineering im Maschinen- und Anlagenbau immer mehr Forschungsansätze gibt, um Softwarequalität, Wiederverwendung, und dadurch die Effizienz insgesamt zu verbessern, werden solche Ansätze in der Industrie bisher wenig genutzt, da beispielsweise der initiale Aufwand für die Einführung der Ansätze als zu groß eingeschätzt wird. Zudem ist für Softwareentwickler in der industriellen Praxis oft unklar, welchen Nutzen derartige Ansätze und Mechanismen im Alltag bieten können.

Um aufzuzeigen, wie tool-gestützte statische Codeanalyse Softwareevolution und -refaktorisierung in verschiedenen Situationen unterstützen kann, werden in diesem Beitrag drei Use Cases zur Anwendung von Codeanalysemechanismen durch verschiedene Stakeholder in verschiedenen Lebenszyklusphasen eines aPS vorgestellt – von der Entwicklung der Software in-house durch die Integration von Zuliefercode über die strukturelle Refaktorisierung historischer Systeme bis hin zur Integration der Software auf der Baustelle des Kunden. Es wird diskutiert, welche Methoden der Codeanalyse sich zur Adressierung der Use Cases eignen und wie diese durch ein prototypisches Analysetool automatisiert umgesetzt werden können. Die Use Cases werden mit Experten aus drei Unternehmen des Maschinen- und Anlagenbaus evaluiert.

2. Stand der Technik in der Analyse und Refaktorisierung von Steuerungssoftware

aPS sind komplexe mechatronische Systeme mit einer Lebensdauer von typischerweise mehreren Jahrzehnten. Die Steuerungssoftware läuft dabei meist auf Speicherprogrammierbaren Steuerungen (SPS) und wird gemäß des Standards IEC 61131-3 bzw. SIMATIC STEP 7 mit textuellen sowie graphischen Sprachen programmiert. Zur Kapselung und Wiederverwendbarkeit einzelner Softwareteile wird die Software in modulare Teile untergliedert, sog. Programmorganisationseinheiten (POE). Der Informationsaustausch zwischen POEs kann entweder

direkt erfolgen durch die Übergabe von Parametern während eines Bausteinaufrufs oder indirekt [2] durch Lese- und Schreibzugriffe auf Werte in einer globalen Variablenliste bzw. in globalen Datenbausteinen (DBs).

Die lange Lebensdauer sowie die damit einhergehenden Funktionserweiterungen, die häufig primär in Form von Softwareanpassungen umgesetzt werden, führen langfristig häufig durch mangelnde Dokumentation und Wartung der dabei entstehenden Variantenvielfalt zu einem unkontrollierten Wachstum der Software. Abhilfe schafft hier gezielte Refaktorisierung [1]. Dabei kann statische Codeanalyse bei der Identifikation von Optimierungspotentialen unterstützen, indem die Software hinsichtlich bestimmter Eigenschaften geprüft wird, ohne den Code dabei auszuführen. Für die Refaktorisierung spielt insbesondere die Bewertung von Modularität [3] eine wesentliche Rolle, z.B. durch die Berechnung von Schnittstellenmetriken [2] oder Softwarekomplexität. Fischer et al. [4] identifizierten fünf Arten von Komplexität in Steuerungssoftware sowie geeignete Metriken, um diese quantitativ zu bewerten. Dazu wurden sowohl neue Metriken entwickelt, z.B. zur Bewertung der Datenstrukturkomplexität, als auch bewährte Metriken aus der Hochsprachenprogrammierung auf Steuerungssoftware übertragen, darunter die Komplexitätsmetriken nach Halstead [5] zur Bewertung von Länge, Schwierigkeit und verwendetem Vokabular, die zyklomatische Komplexität nach McCabe [6] zur Bewertung des Kontrollflusses sowie der FanIn-FanOut-Metrik nach Henry und Kafura [7] zur Erfassung der ein- und ausgehenden Datenströme in ein Modul.

Tabelle 1: Komplexitätskategorien und Metriken zur Quantifizierung nach Fischer et al. [4]

Komplexitätskategorie	Metrik
Umfang	Halstead Länge
Softwarewissenschaft	Halstead Vokabular
	Halstead Schwierigkeit
Kontrollfluss	McCabe's zyklomatische Komplexität
Informationsfluss	Koppelung (intermodulare Komplexität), FanIn-FanOut
Datenstruktur	Datenstrukturkomplexität

Die Interpretierbarkeit von Codeanalyseergebnissen und damit die Anwendbarkeit in der industriellen Praxis erfordert eine geeignete Aufbereitung und Visualisierung der Erkenntnisse. Die Darstellung des Datenaustauschs z.B. kann helfen, Schwachstellen in der Modularisierung zu identifizieren. Mithilfe einer Design-Struktur-Matrix (DSM) [8] kann beispielsweise in Form einer Matrixdarstellung mit Softwareteilen als Zeilen bzw. Spalten angezeigt werden, ob und wieviel Information ausgetauscht werden. Durch einen Cluster-Algorithmus können zusammengehörende Einheiten basierend auf dem Datenaustausch gruppiert werden. Eine weitere

Möglichkeit zur Visualisierung des Datenaustauschs sind Sankey-Diagramme [9], die den Informationsaustausch zwischen Einheiten in Form eines Flussdiagramms abbilden, wobei die Anzahl der ausgetauschten Informationen mit der Breite der Verbindungen korreliert. Im Rahmen des Beitrags wird u.a. untersucht, wie DSM und Sankey-Diagramme zur Visualisierung von Codeanalyseergebnissen für verschiedene Use Cases genutzt werden können.

3. Use Cases zur Anwendung von Codeanalyse in der industriellen Praxis

Toolgestützte statische Codeanalyse ist ein effektives Mittel, um Schwachstellen in historisch gewachsener Steuerungssoftware oder während der Softwareentwicklung systematisch aufzudecken. Obwohl immer mehr Plattformhersteller entsprechende Analysewerkzeuge zur Verfügung stellen, wird das volle Potential statischer Codeanalyse in der industriellen Praxis aktuell kaum ausgeschöpft. Im Folgenden werden deshalb drei Use Cases vorgestellt, in denen Codeanalysemechanismen die Softwareentwicklung in der Praxis unterstützen können.

3.1 Use Case 1 – Qualitätssicherung bei Integration von zugeliefertem Code

In vielen Unternehmen im Maschinen- und Anlagenbau wird Steuerungssoftware nicht vollständig innerhalb des eigenen Unternehmens geschrieben, sondern extern von Zulieferunternehmen zugekauft. Obwohl es dazu i.d.R. strenge vertragliche Bestimmungen gibt, um Umfang, Funktionalität und Qualität der gelieferten Software zu gewährleisten, ist eine objektive Bewertung der Einhaltung dieser Vorgaben in der Praxis schwer umsetzbar. Darüber hinaus führt die Integration von Zuliefercode insbesondere dann zu Herausforderungen, wenn die extern geschriebene Software für die internen Mitarbeiter schwer verständlich ist, da die Bausteine beispielsweise in Umfang und Komplexität stark von der unternehmensinternen Software abweicht. Um sicherzustellen, dass zugelieferter Code einerseits funktionsfähig ist und andererseits grundlegende Anforderungen hinsichtlich festgelegter Qualitätskriterien erfüllt, können Mittel der statischen Codeanalyse eingesetzt werden. Im Rahmen dieses Beitrags wird evaluiert, ob mit Hilfe von *Komplexitätsmetriken* die Identifikation von Qualitätsabweichungen zugelieferter Software unterstützt werden kann.

3.2 Use Case 2 – Strukturelle Refaktorisierung

Insbesondere im Bereich historisch gewachsener Legacy Software wird oftmals das Ziel verfolgt, Softwarequalität und -wiederverwendbarkeit zu erhöhen. Ein Mittel, um dieses Ziel zu erreichen, ist die strukturelle Refaktorisierung, um durch eine modulare, wiederverwendbare Struktur mehr Flexibilität, erleichtertes Variantenmanagement, und dadurch langfristig geringere Kosten und bessere Wettbewerbsbedingungen zu erzielen.

Ziel einer strukturellen Refaktorisierung ist die Optimierung der Softwarestruktur, ohne dabei das Verhalten zu ändern. Dies umfasst häufig z.B. die Neustrukturierung der Funktionalitätsverteilung, was sich auf die Abhängigkeiten zwischen Teilen des Automatisierungssystems auswirkt. Codeanalyse kann diese Prozesse unterstützen, indem Abhängigkeiten unter Berücksichtigung der Automatisierungshardware explizit gemacht und wo möglich reduziert werden, um Teillösungen zu modularisieren und dadurch wiederverwendbar zu machen. Durch die *Visualisierung vorhandener Abhängigkeiten* kann z.B. die vorhandene historisch gewachsene Software mit der im Rahmen der Refaktorisierung intendierten Struktur abgeglichen werden, um dadurch ungewollte Abweichungen zu identifizieren. In diesem Beitrag wird deshalb konkret untersucht, wie die Visualisierung von Abhängigkeiten in Form einer DSM oder als Sankey-Diagramm genutzt werden kann, um die strukturelle Refaktorisierung zu unterstützen.

3.3 Use Case 3 – Integration von Code auf der Baustelle

Im Maschinen- und Anlagenbau sind Anpassungen während der Inbetriebnahme an Prozessabläufen, Ansteuerungen und Kommunikation mit anderen Automatisierungsgeräten, SCADA oder MES gängige Praxis. Aufgrund des hohen Zeitdrucks während der Inbetriebnahme werden diese Änderungen in der Regel selten dokumentiert. Es gibt oftmals keine Kommunikation zwischen den Inbetriebnehmern auf der Baustelle und den Softwareentwicklern im Unternehmen, welche Stellen der Software geändert wurden. Somit ist keine Reflexion darüber möglich, wie sich diese Änderungen langfristig auf die Wartbarkeit der Software auswirken können oder ob z.B. einzelne Bausteine grundsätzliche Anpassungen durch die Softwareentwicklung erfordern, da sie während der Inbetriebnahme zu Schwierigkeiten führen und deshalb vor Ort geändert werden. Codeanalyse kann hier helfen, die Auswirkungen einer Veränderung im Steuerungscode greifbar zu machen, in dem durch die *Visualisierung vorhandener Abhängigkeiten* (z.B. DSM oder Sankey-Diagramme) der Datenaustausch zwischen Bausteinen bzw. Softwaremodulen sichtbar gemacht und somit die Auswirkung von Änderungen abgeschätzt und ggf. reduziert werden kann. Dies ermöglicht zudem abzuwägen, ob ein Eingriff im laufenden Betrieb oder erst während eines Anlagenstillstands gemacht werden kann.

4. Prototypische Implementierung der Codeanalysemechanismen

Die nachfolgenden Abschnitte beschreiben die prototypische Implementierung konkreter Codeanalysemechanismen, um diese für die in Abschnitt 3 genannten Use Cases in der industriellen Praxis nutzbar zu machen. Dazu wurde ein Analysetool entwickelt, um im Siemens TIA Portal entwickelte Steuerungssoftwareprojekte automatisiert zu quantifizieren. Als Basis dazu dienen die Quellcodes, die über die Public API (Openness Interface) des TIA Portals als

XML-Format exportiert werden (vgl. Bild 1). Anschließend werden diese in das Codeanalysetool eingelesen, um die in Abschnitt 3 genannten Codeanalysemechanismen (*Komplexitätsmetriken* und *Abhängigkeitsvisualisierung*) automatisiert durchzuführen. Die Ergebnisse werden anschließend in Form geeigneter Darstellungen (z.B. Sankey Diagramme, DSM) für den Anwender aufbereitet.

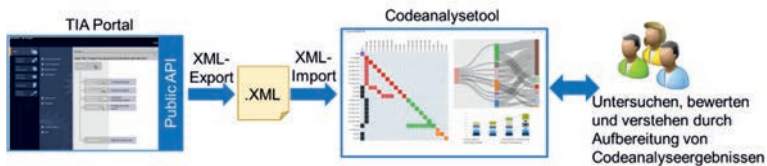


Bild 1: Werkzeugkette zur Umsetzung der prototypischen Implementierung der Codeanalysemechanismen

4.1. Automatisierte Berechnung von Komplexitätsmetriken

Die Analyse industrieller Steuerungssoftware ist eine komplexe Herausforderung, da es keine eindeutige Definition von guter oder schlechter Software gibt. Stattdessen ist das Empfinden von Softwarequalität und -komplexität subjektiv je nach Sicht auf die Software: Während dem Inbetriebnehmer vor Ort ein prozeduraler Aufbau der Software wichtig ist, da die Spiegelung des technischen Prozesses in der Software die Inbetriebnahme erleichtert, priorisiert ein Bibliotheksentwickler ggf. einen modularen Ansatz mit kleinen POE-Schnittstellen nach dem Kapselungsprinzip, um die Software langfristig wartbar zu halten. Dies führt dazu, dass Software nicht pauschal als komplex oder nicht komplex bewertet werden kann, sondern dass für eine differenzierte Aussage verschiedene Kategorien verglichen werden müssen. Für eine differenzierte Komplexitätsbewertung wurden deshalb die durch Fischer et al. [4] evaluierten Metriken zur Erfassung verschiedener Komplexitätskategorien prototypisch implementiert und in Form von Balkendiagrammen visualisiert, sodass für jeden Baustein des Projekts die Ergebnisse in den verschiedenen Komplexitätskategorien bewertet werden können.

4.2 Visualisierung von Abhängigkeiten in Steuerungssoftware

Für die Visualisierung von Abhängigkeiten in der Steuerungssoftware werden zwei Sichten fokussiert: die DSM zur Darstellung von Aufrufbeziehungen zwischen Bausteinen und Sankey-Diagramme zur Nachverfolgung des Datenflusses zwischen POEs über DBs.

In der DSM (vgl. Bild 2) sind die im Projekt enthaltenen Bausteine und ihre Beziehung zueinander ersichtlich. Ein ausgefülltes Feld gibt an, dass eine Aufrufbeziehung von einem Bau-

stein auf der vertikalen Achse mit einem Baustein in der horizontalen Achse erfolgt. Im verwendeten Beispiel erfolgt der Bausteinaufruf rein sequenziell, d.h. ohne Rückkopplungen. Entsprechend ist in der DSM nur das untere Dreieck der Matrix (feed forward) belegt [8]. Zudem kann durch eine entsprechende Farbgebung angegeben werden, welche Bausteine derselben Funktionsgruppe angehören (Zuordnung in Siemens TIA z.B. über Ordnerstruktur möglich), um Querbeziehungen sichtbar zu machen. Mit Hilfe eines Clustering-Algorithmus können zudem Bausteine mit einer hohen Anzahl an Aufrufbeziehungen gruppiert werden.

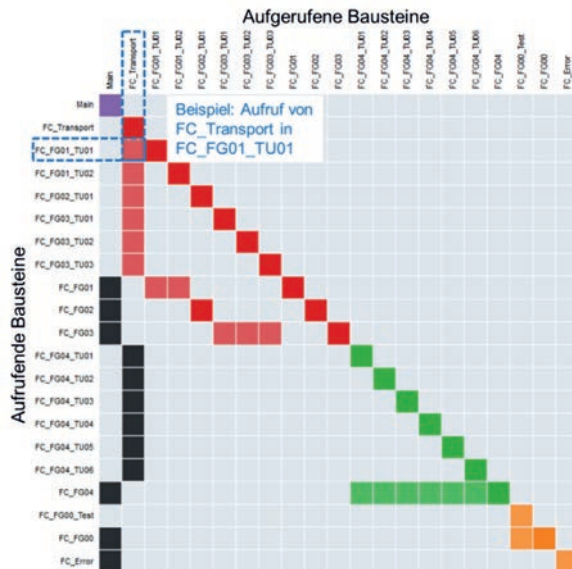


Bild 2: Design-Struktur-Matrix (DSM) für Aufrufbeziehungen zwischen Bausteinen (Farbe kennzeichnet Bausteinaufrufe innerhalb einer Funktionsgruppe, schwarz entspricht Queraufrufen zwischen Funktionsgruppen)

Mit Hilfe eines Sankey-Diagramms kann visualisiert werden, über welche globalen DBs innerhalb eines Steuerungssoftwareprojekts Bausteine Informationen austauschen, also welche Bausteine eine Variable im DB schreiben und welche diese lesen. Dazu werden die Bausteine und die dazwischen geschalteten DBs jeweils als Balken dargestellt, wobei die Länge der Balken mit der Anzahl der gelesenen bzw. geschriebenen Variablen korreliert. In Form von Mouse-Over-Events erhält der Nutzer zusätzliche Informationen darüber, auf welche Variable sich ein bestimmter Teil der grauen Verbindungsbalken bezieht. Die Darstellung als Sankey Diagramm kann auch für einen einzelnen DB erfolgen. So sind einzelne Datenfelder innerhalb eines DB und Ihre Verwendung innerhalb eines Projekts identifizierbar (vgl. Bild 3).

aussagekräftigen Vergleich zugelieferter Bausteine ähnlicher Funktionalität unterstützen. Fällt ein Baustein durch einen auffallend hohen oder niedrigen Wert einer bestimmten Metrik auf (vgl. Bild 4, rechts, rot markierter Bereich), kann dies ein Indiz dafür sein, dass die detailliertere Prüfung der Bausteinimplementierung sinnvoll ist. Die Identifikation von Ausreißern (vgl. Bild 4, links) hilft zudem bei der Abschätzung, ob der gelieferte Code hinsichtlich seines Umfangs den Vorgaben entsprechen kann. Code von zweifelhafter Qualität kann so frühzeitig erkannt und Gegenmaßnahmen eingeleitet werden. Darüber hinaus ermöglichen Metriken zur Bestimmung der Verständlichkeit eine Abschätzung darüber, ob der zugelieferte Code vergleichbar ist zu der intern entwickelten Software oder ob der Zulieferer nachbessern bzw. die hohen Komplexitätswerte argumentieren sollte.

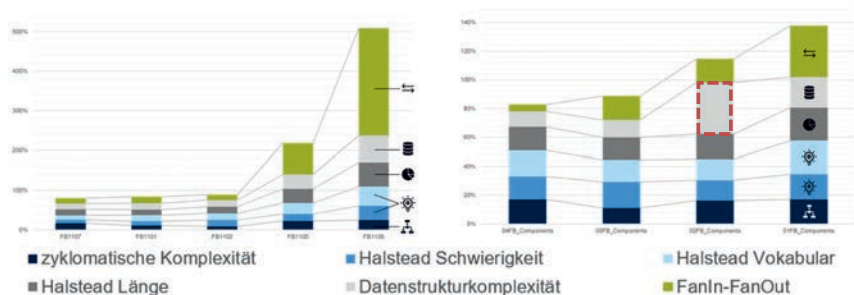


Bild 4: Anwendung von Komplexitätsmetriken (links: Identifikation von Ausreißern, rechts: Vergleich von Bausteinen ähnlicher Funktionalität)

5.2 Evaluation Use Case 2 – Strukturelle Refaktorisierung

Die implementierten Codeanalysemechanismen zur Visualisierung von Abhängigkeiten innerhalb eines Projekts wurden mit Experten aus zwei Unternehmen aus dem Bereich Medizintechnik, Automotive und Solar bzw. Folienstreckanlagen evaluiert. Diese bestätigten den praktischen Nutzen der Abhängigkeitsvisualisierung für die nachfolgend beschriebenen Aspekte bei der strukturellen Refaktorisierung von Steuerungssoftware.

Bietet der Plattformhersteller ein Upgrade seiner Programmierungsumgebung an (z.B. im Fall von Siemens durch die Einführung des TIA Portals als Nachfolgesystem des STEP 7 SIMATIC Managers), lohnt es sich für Anwender in der Regel aufgrund des erweiterten Funktionsumfangs ein Umstieg auf das neue System, um langfristig wettbewerbsfähig zu bleiben. Im Fall des Siemens TIA Portals z.B. wurde als ein Upgrade im Vergleich zum SIMATIC Manager die

Möglichkeit eingeführt, Steuerungssoftware in Ordner zu strukturieren. Ein Anwender mit historisch gewachsener Steuerungssoftware im Altsystem kann somit den Umstieg auf das neue System nutzen, um mit Hilfe von Ordnern die Software in Module zu strukturieren, also zu POE-Gruppierungen, die aufgrund einer gemeinsam implementierten Funktionalität oder der Ansteuerung einer zusammenhängenden Hardwareeinheit zusammengehören. Durch eine DSM kann anschließend abgeglichen werden, ob die vorgesehene Modulstruktur durch Sichtbarmachen der vorhandenen intermodularen Abhängigkeiten sinnvoll ist oder ob die Modulschnittstellen aufgrund zu großer oder unübersichtlicher Datenflüsse ungünstig gewählt sind. Wie oben angedeutet wird die Möglichkeit, im Siemens TIA Portal POEs eines Softwareprojekts in Form von Ordnern zu strukturieren, in der Praxis häufig genutzt, um die Software in zusammengehörende Funktionsgruppen zu unterteilen. Durch die Darstellung von Abhängigkeiten in Form einer DSM in Kombination mit der Ordnerstruktur kann somit sichtbar gemacht werden, an welchen Stellen Abhängigkeiten über Ordner- bzw. Modulgrenzen hinweg vorliegen. Häufig nutzen Unternehmen definierte Regeln, an welchen Stellen modulübergreifender Datenaustausch zulässig ist (z.B. für den Zugriff auf zentrale Stellen, z.B. das HMI) und wie genau dieser umzusetzen ist. Liegen Verstöße gegen diese Richtlinien vor, kann dies durch eine Darstellung in einer DSM analog zu Bild 2 sichtbar gemacht werden.

Bei der Wiederverwendung von Steuerungssoftware können je nach Ausgangssituation und vorhandenen Randbedingungen unterschiedliche Strategien zur Wiederverwendung genutzt werden. Bei einem template-basierten Ansatz beispielsweise werden ausgereifte Module aus der Bibliotheksentwicklung in die Applikationssoftware kopiert und an eindeutig vorgesehenen Stellen für den jeweiligen Anwendungsfall anpasst. Wird gegen die Programmierrichtlinien verstoßen und dadurch ungewollte Abhängigkeiten an der falschen, nicht dafür vorgesehenen Stelle eingefügt, erschwert dies die Wiederverwendung des Moduls in anderem Kontext sowie die Verständlichkeit, da die Struktur nicht dem Programmierstandard entspricht. Solche ungewollten Abhängigkeiten könnten sowohl in einer DSM als auch in Form eines Sankey-Diagramms sichtbar gemacht werden, wenn die Informationsflüsse an den Schnittstellen eines Moduls von der vorgesehen Struktur abweicht, da erforderliche Verknüpfungen des Moduls fehlen oder zu falschen Stellen führen.

Darüber hinaus kann eine geeignete Darstellung ungewollter Abhängigkeiten dazu dienen, erforderliche Maßnahmen zur Refaktorisierung verständlich aufzuzeigen und deren Aufwand nachvollziehbar abzuschätzen. Die abstrahierte Darstellung der Verstöße hilft nicht nur bei der Kommunikation zwischen Softwareentwicklern, sondern auch Abteilungsleiter können so schnell die Qualität der Software nachvollziehen.

5.3 Evaluation Use Case 3 – Integration von Code auf der Baustelle

Auch für die Unterstützung des Use Cases, Änderungen während der Baustelle zu dokumentieren und deren Quereffekte abzuschätzen kann die Implementierung der DSM und Sankey-Diagramme genutzt werden. Der Use Case wurde erneut mit den in 5.2 genannten Experten aus den beiden genannten Unternehmen evaluiert.

Werden Module während der Inbetriebnahme kopiert, kann dies zu ungewollten Abhängigkeiten führen, die sowohl durch die DSM als auch durch Sankey-Diagramme sichtbar gemacht werden können. Ein häufiger Fehler beispielsweise ist das Vergessen von Namensänderungen, sodass das kopierte Modul nicht korrekt in seine Umgebung eingebunden ist. Durch Anwendung der Codeanalysemechanismen zur Visualisierung von Abhängigkeiten wird eine solche Abweichung direkt sichtbar und der Fehler kann behoben werden.

Darüber hinaus kann Codeanalyse dabei unterstützen, Kommunikationsengpässe zwischen der Softwareentwicklung im Haus und der Inbetriebnahme im Feld zu überbrücken. Beispielsweise kommt es häufig vor, dass Techniker bei der Inbetriebnahme wiederkehrend auf dasselbe Problem stoßen, da z.B. die Funktionalität eines Bausteins nicht der Erwartung entspricht und seine Schnittstellen angepasst werden müssen. Wird diese Änderung nicht dokumentiert und der Entwicklungsabteilung zurückgemeldet, besteht keine Möglichkeit, solche wiederkehrenden Schwierigkeiten durch eine Anpassung des Bausteins oder verbesserte Dokumentation zu unterbinden. Auch in diesem Fall können Codeanalysemechanismen unterstützen, indem z.B. Sankey Diagramme oder DSM vor und nach der Inbetriebnahme für mehrere Beispiele verglichen werden, um graphisch zu visualisieren, welche Stellen der Software am meisten verändert wurden und um wiederkehrende Muster zu finden. Darüber hinaus kann durch einen Vergleich von Komplexitätsmetrikwerten vor und nach der Änderung auf der Baustelle abgeschätzt werden, wie umfangreich ein bestimmter Baustein geändert wurde.

5. Zusammenfassung und Ausblick

Im Rahmen dieses Beitrags wurde evaluiert, wie Codeanalyse verschiedene Use Cases in der industriellen Steuerungssoftwareentwicklung für verschiedene Stakeholder bei der Optimierung der Softwarequalität unterstützen kann. Die Use Cases adressieren verschiedene Phasen des Lebenszyklus von der Softwareerstellung durch Integration von Zuliefercode bis hin zur Inbetriebnahme auf der Baustelle. Die Relevanz der Anwendungsfälle sowie der Mehrwert von Codeanalyse, um Software zu bewerten und Richtlinienverstöße zu identifizieren, wurden mit Experten aus Unternehmen des Maschinen- und Anlagenbaus evaluiert und bestätigt. Dabei unterstützt eine abstrahierte Darstellung der Ergebnisse z.B. in Form von Sankey-Diagrammen oder als DSM sowohl bei der Kommunikation zwischen Softwareentwicklern als auch

beim übersichtlichen Nachvollziehen von Änderungsauswirkungen im Code durch Abteilungsleiter.

In zukünftigen Arbeiten ist geplant, den Mehrwert von Codeanalysen weiteren Anwendungsfällen im Softwareentwicklungsworkflow zu untersuchen, z.B. bei der methodischen Gestaltung von Steuerungssoftware für verbesserte Konfigurierbarkeit, Wiederverwendbarkeit, und projektübergreifendes Variantenmanagement. Zudem ist geplant, die Methoden in der Codeanalyse für SPS-Software zu erweitern, z.B. in Form weiterer Metriken zur Bewertung des Datenflusses oder durch Indikatoren zur Bewertung des indirekten Datenaustauschs über globale DBs.

Referenzen

- [1] M. Fowler, *Refactoring: Improving the design of existing code*. Reading Mass. u.a.: Addison-Wesley, 1999.
- [2] J. Fischer, B. Vogel-Heuser, C. Huber, M. Felger, and M. Bengel, "Reuse Assessment of IEC 61131-3 Control Software Modules Using Metrics – An Industrial Case Study," in *19th IEEE International Conference on Industrial Informatics (INDIN)*, Mallorca, Spain, Jul. 2021, [eingereicht].
- [3] B. Vogel-Heuser, T. Bauernhansl, and M. ten Hompel, Eds., *Handbuch Industrie 4.0: Produktion, Automatisierung und Logistik*. Wiesbaden: Springer Fachmedien Wiesbaden, 2015.
- [4] J. Fischer, B. Vogel-Heuser, H. Schneider, N. Langer, M. Felger, and M. Bengel, "Measuring the Overall Complexity of Graphical and Textual IEC 61131-3 Control Software in Five Classes – An Industrial Case Study," in *IEEE 17th International Conference on Automation Science and Engineering (CASE)*, Lyon, France, 2021, [eingereicht].
- [5] M. H. Halstead, *Elements of software science*, 3rd ed. New York: North Holland, 1979.
- [6] T. J. McCabe, "A Complexity Measure," *IEEE Trans. Software Eng.*, SE-2, no. 4, pp. 308–320, 1976.
- [7] S. Henry and D. Kafura, "Software Structure Metrics Based on Information Flow," *IEEE Trans. Software Eng.*, SE-7, no. 5, pp. 510–518, 1981.
- [8] S. D. Eppinger and T. R. Browning, *Design structure matrix methods and applications*. Cambridge, Mass: MIT Press, 2012. [Online]. Available: <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=458479>
- [9] H. R. Sankey, "The Thermal Efficiency of Steam Engines. (Including Appendixes)," *Minutes of the Proceedings of the Institution of Civil Engineers*, vol. 125, no. 1896, pp. 182–212, 1896.

Interoperable Interaktionen semantisch heterogener Industrie 4.0 Komponenten auf Basis einer Capability Ontologie für die Technische Gebäudeausrüstung

N. Maisch, M. Both, Prof. Dr. J. Müller,
Technische Hochschule Köln, Köln

Kurzfassung

Industrie 4.0-Verwaltungsschalen bieten die Möglichkeit, ihre Informationen durch Referenzen auf externe Ontologien semantisch zu beschreiben. Neben Properties können in Verwaltungsschalen auch die Fähigkeiten eines Assets in Form von Capabilities definiert werden, die für die Übernahme gemeinsamer Funktionen notwendig sind. Im Bereich der Technischen Gebäudeausrüstung gibt es keine Ontologie für Capabilities, auf die Verwaltungsschalen von Assets der Gebäudetechnik referenzieren können, um beispielsweise ihre Funktionalität innerhalb eines Energieerzeugungssystems semantisch auszeichnen zu können. In diesem Paper wird die Entwicklung einer solchen Ontologie und die Methode zur Ableitung der zentralen Begriffe beschrieben. Die Ontologie stellt zudem einen Rahmen vor, um die übergeordneten Funktionalitäten aus granularen Asset-Funktionen zusammenzusetzen. Dadurch soll erreicht werden, dass aus Referenzen zu den jeweiligen Funktionen einzelner Assets auf die Funktionalität ihres Gesamtsystems impliziert werden kann.

1. Einleitung

Durch die Industrie 4.0 (I4.0)-Verwaltungsschale werden Merkmale von Assets auf standardisierte Weise dargestellt. Dabei werden die Informationen eines Assets innerhalb einer Verwaltungsschale gesammelt, gemäß des Metamodells strukturiert und maschinenlesbar abgebildet. Zentrale Bausteine der Verwaltungsschale sind *Teilmodelle*, welche Eigenschaften und Funktionalitäten der Assets und deren Inhalte abbilden. Teilmodelle wiederum enthalten *Teilmodellelemente*, welche die tatsächlichen Informationen der Komponenten enthalten [1].

Trotz der standardisierten Darstellung ist kein semantisch homogenes Umfeld garantiert, da die Teilmodellelemente hersteller- oder domänenspezifisch beschrieben werden können. Bild 1 zeigt das Einbinden semantisch heterogen beschriebener Assets in eine Anwendung, das

wegen der verschiedenen semantischen Beschreibungen mit großem händischem Engineering verbunden ist.

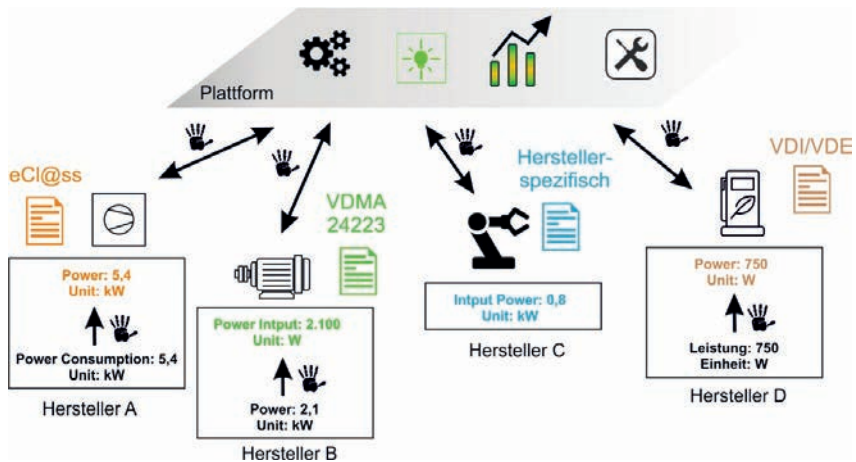


Bild 1: Händisches Engineering zum Einbinden der Leistung semantisch heterogen beschriebener Assets

Ein Ansatz zur Herstellung eines semantisch homogenen Raums ist es, einen einheitlichen Standard zur Beschreibung von Informationen zu schaffen, den alle Hersteller zur Beschreibung ihrer Assets verwenden, z.B. ECLASS [2]. Ein weiterer Ansatz referenziert Teilmodellelemente auf online verfügbare und allgemein zugängliche Ontologien. Werden die Teilmodellelemente einer Verwaltungsschale zu einem Eintrag einer Ontologie referenziert, können interagierende Assets oder Anwendungen dessen Beschreibung und die Beziehungen des Eintrags zu anderen Einträgen ermitteln [1, 3]. Mittels Linked Data werden Verknüpfungen zwischen verschiedenen hersteller- und branchenspezifischen Ontologien hergestellt [4], wodurch der Informationsaustausch zwischen semantisch heterogen beschriebenen Verwaltungsschalen ermöglicht wird (Bild 2).

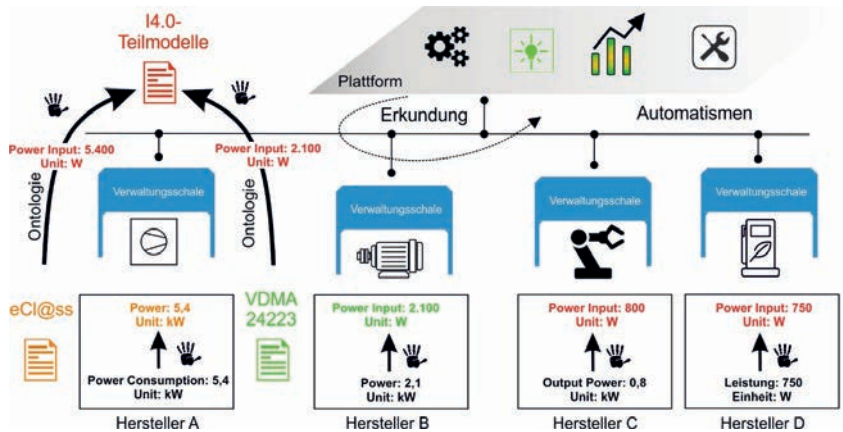


Bild 2: Interoperabilität durch die Abbildung semantisch heterogener Standards und Merkmalbeschreibungen auf Ontologien

Das Teilmodellelement *Capability* beschreibt die Fähigkeiten eines Assets in der Verwaltungsschale [1]. Durch Referenzen auf Ontologien werden die Capabilities semantisch ausgezeichnet, sodass Anwendungen oder andere Verwaltungsschalen diese verstehen und einordnen können [3]. Dies ist die Grundlage für die Übernahme gemeinsamer Funktionen wie z.B. die Wärmeversorgung eines Gebäudes bildet.

Im Bereich der Technischen Gebäudeausrüstung (TGA) fehlt eine Ontologie, die grundsätzliche Funktionalitäten zur Beschreibung von Versorgungsvorgängen innerhalb von Gebäuden abbildet. In dieser Arbeit wird das Vorgehen beim Erstellen einer solchen Ontologie erklärt. Ziel ist es, dass durch die Ontologie verschiedene Versorgungssysteme innerhalb eines Gebäudes durch den Austausch ihrer Versorgungsfunktionen interoperabel interagieren können.

2. Funktionalitäten im Betrieb der TGA

Eine Klassifizierung der übergeordneten Ziele im Betrieb von Gebäuden liefert [5]. Die Nutzungsphase von Gebäuden wird dabei in neun Teilgebiete unterteilt (Bild 3). Neben verwaltenden Tätigkeiten wie dem *Kosten-* oder *Flächenmanagement* ist insbesondere die Tätigkeit *Objekte ver- und entsorgen* relevant für den Bereich der TGA und der Gebäudeautomation. *Objekte versorgen* wird als die Bereitstellung von Energie und Medien in zweckdienlicher Form beschrieben [5].

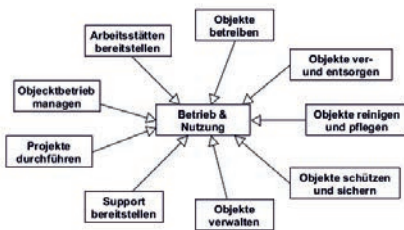


Bild 3: Aufgaben des Lebenszyklus *Betrieb & Nutzung* nach [5]

Deckungsgleich zu [5] unterteilt [6] das Monitoring von Versorgungsaufgaben in die Erfassung von Energie- und Medienverbräuchen. [7] unterteilt die Prozesse zum Erfüllen der Energieversorgungsaufgaben innerhalb von Gebäuden zudem in ihre Grundbestandteile: *Erzeugung*, *Verteilung* und *Übergabe*. Die Versorgungsaufgabe mit Medien kann analog dazu unterteilt werden (Bild 4).

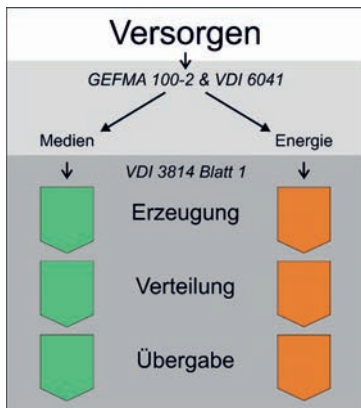


Bild 4: Ableitung der Grundfunktionen zur Versorgung von Gebäudeteilen mit Energie und Medien nach [5 bis 7]

Neben den übergeordneten Versorgungsfunktionalitäten innerhalb eines Gebäudes gibt es granulare Funktionen einzelner Assets. [8] liefert einen Pool von grundlegenden Asset-Funktionen und beschreibt eine dreistufige Klassifizierung und Kodierung, mit der funktionale Eigenschaften definiert und voneinander abgegrenzt werden. Sowohl die Aufgaben zur Erfüllung von Versorgungsaufgaben innerhalb von Gebäuden als auch die Asset-Funktionen

nach [8] können im Zusammenhang mit I4.0 als Capabilities beschrieben und damit Assets zugeordnet werden [1].

3. Capabilities in der Verwaltungsschale

Das Metamodell der I4.0-Verwaltungsschale stellt die Grundlage für den Informationsaustausch zwischen Assets dar, da sie einen Rahmen zur Darstellung von Teilmodellelementen und dadurch Asset-Informationen liefert [1]. Für „Plug and Produce“-Szenarios und flexible Wertschöpfungsnetzwerke, in denen Assets selbstständig gemeinsame Funktionen übernehmen, müssen neben beschreibenden Teilmodellelementen (Properties) die Fähigkeiten von Assets (Capabilities) in Verwaltungsschalen abgebildet werden [3]. Dadurch kann ein Austausch über mögliche Aktionen stattfinden und ein Matching mit erfordernten Fähigkeiten stattfinden. Nach einem Check der Fähigkeiten eines Assets (Capability Check) erfolgt für einen Interaktionsprozess auf Basis der spezifischen Bedingungen und Properties ein Check der Machbarkeit (Feasibility Check) [3].

Das Konzept basiert auf dem „Process, Product and Resource“ (PPR)-Modell, nach dem in bestimmten technischen Bereichen (z.B. fertigende Industrie) Produktionslinien auf Basis von Fertigungsfähigkeiten einzelner Anlagen modular geplant und automatisiert werden, um auf veränderte Bedingungen reagieren zu können und bei Bedarf Module der Fertigung austauschen zu können [3]. Dafür ist es notwendig, den Fertigungsvorgang als Zusammenspiel von Ressource, Prozess, Produkt und Fähigkeiten zu modellieren. Dabei wird ein bestimmter Fertigungsprozess, durch den ein Produkt entsteht, mithilfe einer Fähigkeit von einer Ressource erfüllt. Hinreichende Bedingung für die Realisierung des Prozesses durch die Ressource ist der Abgleich der Funktionen der Ressource mit den für den Prozess notwendigen Funktionen [9 bis 11].

Die notwendigen Funktionen eines Prozesses können durch Ontologien strukturiert formalisiert und orchestriert werden. Dadurch können auch Ressourcen für kompliziertere Prozesse, die mehr als eine Fähigkeit erfordern, ermittelt werden [12]. Die Ontologie C4I [13] liefert eine Methode zur Zuweisung von Capabilities an Ressourcen. Zusätzlich bietet sie die Möglichkeit, zwischen *associatedWithCapability* für eine allgemeine Zusammensetzung von Capabilities und der Beziehung *hasCapability* für die notwendige Ausführung eines Prozesses durch eine einzige Ressource zu unterscheiden.

4. Ontologien als semantische Basis von Verwaltungsschalen

Damit Assets gemeinsame Funktionen übernehmen können, ist es notwendig, dass die Definitionen der Capabilities semantisch aufeinander abgeglichen werden. Deshalb sollten die Definitionen außerhalb der Verwaltungsschalen liegen, was durch Referenzen auf externe Ontologien erreicht werden kann [3]. Eine Ontologie ist eine Menge von Axiomen, die Beziehungen zwischen Entitäten explizit und maschinenlesbar beschrieben. Durch die Logik der Formalisierungssprachen für Ontologien können aus dem explizit beschriebenen Wissen weitere Beziehungen und damit zusätzlich weiteres Wissen impliziert werden [14]. Mit einer Referenz auf bestimmte Einträge innerhalb einer Ontologie können Properties und Capabilities von Assets also semantisch beschrieben werden. Die Referenz auf einen bestimmten Eintrag einer Ontologie kann mithilfe einer Semantic-ID in Form einer eindeutigen URL erreicht werden [1]. Durch formalisiertes Wissen in der Ontologie können dann durch semantische Tripel definierte Beziehungen zwischen den Entitäten festgestellt werden. Durch die W3C-Standards für Sprachen zum Formalisieren von Tripeln „RDF“ [15] und „RDFS“ [16] können Beziehungen wie beispielsweise „Heizen durch Gaskessel rdfs:subClassOf Heizen“ beschrieben werden. Die „Web Ontology Language (OWL)“ [17] erweitert die Möglichkeiten zum Festlegen von Axiomen um weitere Beziehungsbeschreibungen. Damit sind auch kompliziertere Aussagen möglich, wie beispielsweise, dass Heizen und Kühlen stets disjunkt sind.

5. Ontologien in der TGA

Auch in der Domäne der Gebäudetechnik können Teilmodellelemente durch Ontologien semantisch ausgezeichnet werden. Bestehende Ontologien haben dabei unterschiedliche thematische Schwerpunkte und Herangehensweisen bei der Aufteilung der Gebäudetechnik.

Die Building Topology Ontology (BOT) ist eine Ontologie zur topologischen Beschreibung grundlegender Bestandteile eines Gebäudes [18 bis 20]. Sie teilt Gebäude in Zonen, Elemente (physische Dinge wie Wände, Türen oder Sensoren) und Interfaces (Übergänge zwischen Zonen und Elementen) ein. Die Klassen werden nicht weiter unterteilt, da für genauere Beschreibungen der Elemente auf domänenspezifischere Ontologien verwiesen wird.

Durch die Ontologie BRICK [21] können neben dem Gebäude als solches auch Assets der Gebäudetechnik und deren Metadaten und Beziehungen virtuell abgebildet werden. Bei Untersuchungen konnten zwischen 96 und 99% aller Datenpunkte eines Gebäudes durch Einträge der Ontologie abgedeckt werden.

Die Ontologie DogOnt [22, 23] ermöglicht es, die Regelungen und Steuerungen von SmartHome-Geräten abzubilden, mit dem Ziel Systeme verschiedener Hersteller semantisch verbinden zu können. Beispielsweise können An-Aus-Signale referenziert werden.

Mit der Smart Applications REference Ontology (SAREF) [24] können Assets innerhalb von Gebäuden beschrieben werden. Es werden die Funktionen zum Regeln eines Geräts beschrieben, jedoch nicht die grundlegenden Funktionalitäten von TGA-Komponenten wie beispielsweise Heizen.

Mit ifcOWL [25, 26] kann ein BIM-Modell (Building Information Modeling) eines Gebäudes in Form von semantischen Tripeln abgebildet werden. Eine Definition von Grundfunktionen der Assets innerhalb eines TGA-Systems ist nicht Teil des BIM-Modells.

Mit den Ontologien SSN (Semantic Sensor Network Ontology) und SOSA (Sensor, Observation, Sample und Actuator) [27, 28] werden Begriffe der Sensorik und Aktorik abstrahiert. Konkrete Fähigkeiten bezüglich der Gebäudetechnik werden nicht genannt.

Bestehende Ontologien in der Gebäudetechnik bilden bestehende Strukturen und Funktionen innerhalb von Gebäuden ab. Mit keiner der genannten Ontologien können Grundfunktionalitäten der TGA von Verwaltungsschalen referenziert werden, um interoperabel gemeinsame Aufgaben zu übernehmen. Im Folgenden wird die Entwicklung einer Ontologie beschrieben, die Versorgungsfunktionalitäten strukturiert beschreibt und durch die von granularen Asset-Funktionen auf die übergeordneten Funktionalitäten geschlossen werden kann.

6. Ableiten der Grundfunktionen und Aufbau der Ontologie

Bei der Entwicklung der Ontologie wurden alle Capabilities von der Klasse *Capability* nach [13] abgeleitet, um die Basis für Interoperabilität der Ontologie mit anderen Capability-Ontologien zu schaffen. Als grundlegende Capability wurden das Betreiben des Gebäudes („Operating_Building“) und dessen Unterklassen nach [5] definiert. Da sich der Umfang der Ontologie im Bereich der Gebäudetechnik, speziell im Versorgen von Gebäuden befindet, liegt das Hauptaugenmerk der Ontologie im Bereich *Objekte ver- und entsorgen*. Aus dieser Klasse wurden auf Basis von [7] die grundlegenden Capabilities „Generating“, „Distributing“ und „Transferring“ abgeleitet. Da aber durch kein Asset mit nur einer dieser drei Grundcapabilities eine komplette Versorgungsaufgabe übernommen werden kann, wurde die Klasse „Supplying“ als übergeordnete Klasse definiert, die aus den drei Grundfunktionen aggregiert werden kann (Bild 5). „Distributing“ kann dabei für den Fall ausgelassen werden, wenn eine

Versorgungseinheit im Sinne der Beziehung *hasCapability* nach [13] sowohl erzeugend als auch übertragend tätig ist, beispielsweise ein Kaminofen.

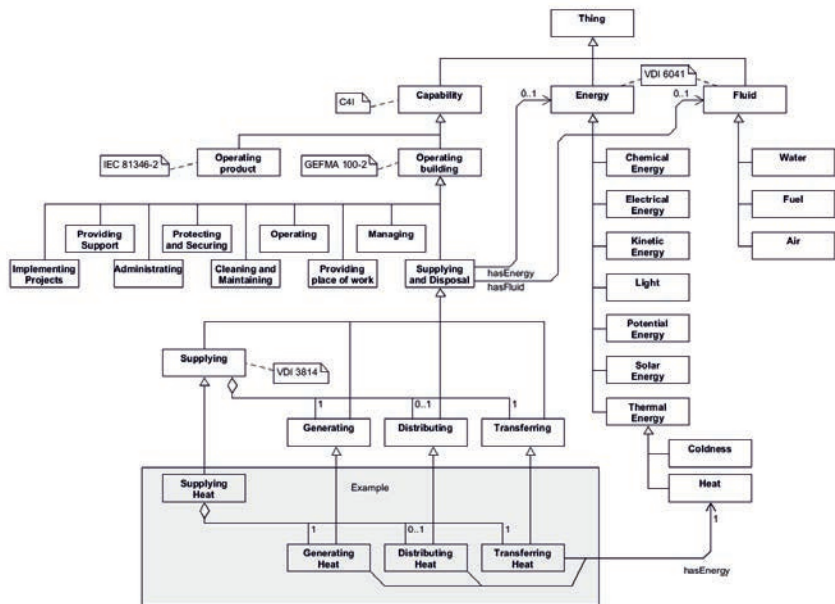


Bild 5: Struktureller Aufbau der Ontologie für Capabilities der Gebäudetechnik und beispielhafte Darstellung der Aggregation von Wärmeversorgung

Auf Basis von [6] werden die drei Grundcapabilities zudem mit einem Attribut in ihren Tätigkeitsbereich zur Versorgungsart aufgeteilt. Beispielsweise wird „Transferring“ in „Transferring_Energy“ und „Transferring_Fluid“ aufgeteilt. Dafür wurden die Object Properties „hasEnergy“ und „hasFluid“ eingeführt, die eine Capability mit einer der verschiedenen Klassen von möglichen Energien und Fluiden ausstattet.

Durch eine Versorgung mit Luft („Supplying_Air“) kann zudem zusätzlich eine Versorgung mit Wärme oder Kälte stattfinden, weshalb sich beispielsweise die Versorgung mit Wärme („Supplying_Heat“) zusätzlich aus der Versorgung mit Luft und des Übertragens von Wärme zusammensetzen kann (Bild 6).

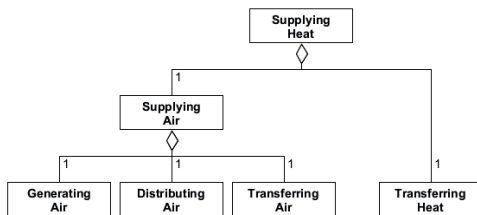


Bild 6: Aggregation der Wärmeversorgung durch eine RLT-Anlage

Durch die vorgestellte Ontologiestruktur können Versorgungsaufgaben innerhalb von Gebäuden modular und nach der Art der Versorgung zusammengesetzt werden. Hintergrund ist die Möglichkeit für Versorgungssysteme in Gebäuden, Capabilities referenzieren können, um Versorgungsaufgaben bereitzustellen und somit flexibel auf einen bestimmten Bedarf des Gebäudes reagieren zu können.

7. Matching der Versorgungsfunktionalitäten mit Asset-Grundfunktionen

Durch einzelne Assets (z.B. Ventilatoren) können keine Versorgungsaufgaben erfüllt werden. Deshalb muss sichergestellt werden, dass aus verschiedenen Capabilities der Geräteebene implizit auf die definierten übergeordneten Versorgungsaufgaben geschlossen werden kann. Deshalb wird zusätzlich zur Ontologie für Capabilities der Gebäudetechnik eine Ontologie für Capabilities auf der Asset-Ebene erstellt, die [8] nachempfunden ist. Durch das Herstellen von Äquivalenzen von Unterklassen der Basisfunktionen „Generating“, „Distributing“ und „Transferring“ mit Capabilities aus der Ontologie nach [8] können beide Ontologien verknüpft werden, sodass aus der Summe granularer Asset-Capabilities auf übergeordnete Funktionalitäten der Gebäudetechnik impliziert werden kann. In Bild 7 sind beispielhaft die Bauteile einer Vollklimaanlage mit allen vier thermodynamischen Luftbehandlungsfunktionen auf funktionaler Ebene einzeln beschrieben. Bei den für die grundlegende Versorgungsfunktionalität relevanten Funktionen sind die äquivalenten Funktionen aus der Ontologie aufgeführt.

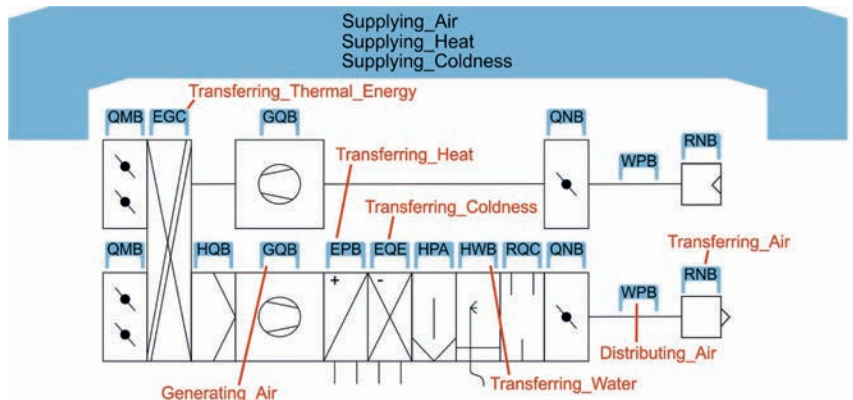


Bild 7: Asset-Grundfunktionen der Bauteile einer RLT-Anlage nach [8] und deren äquivalente Capabilities

Kann ein Bereich innerhalb eines Gebäudes von mehr als einer Anlage mit Wärme versorgt werden, so kann dies durch einen Capability Check ermittelt werden. Durch einen Feasibility Check kann unter Berücksichtigung aller Asset-spezifischen Merkmale dann der geeignetere Teilnehmer ermittelt werden (Bild 8).

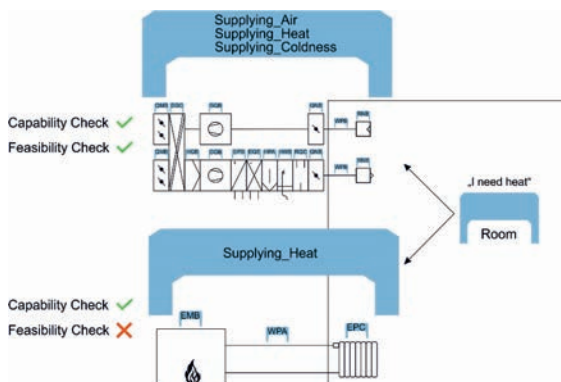


Bild 8: Beispielhafter Capability und Feasibility Check der Wärmeversorgung eines Raums

8. Fazit und Ausblick

Mit der Ontologie von Capabilities der Gebäudetechnik wurde eine Grundlage für die Übernahme gemeinsamer Funktionen verschiedener Assets innerhalb von Gebäuden gelegt. Durch eine Definition der Capabilities einzelner Assets kann durch die Ontologie impliziert werden, welche Funktionalität ein Teilnehmer innerhalb eines Versorgungssystems hat. Gebäude können durch flexibel aktivierbare Energieversorgungssysteme bedarfsgerechter und effizienter betrieben werden, weil diese auf Basis aktueller Betriebs- und Umweltdaten angesteuert werden können.

Der Einsatz von Ontologien und das Verlinken von verschiedenen domänenspezifischen Ontologien ist eine Möglichkeit, Interoperabilität in semantisch heterogenen Bereichen zu bewerkstelligen. Dennoch ist diese Vorgehensweise mit viel händischem Aufwand verbunden: Ontologien müssen aufgebaut, verlinkt und implementiert werden. Ergebnisse von Untersuchungen im Bereich der Künstlichen Intelligenz (KI) zeigen, dass mittels Natural Language Processing (NLP) eine automatisierte Abbildung verschiedener Ontologien und hersteller- und domänenspezifischer Wissensstandards aufeinander möglich ist [29]. Für die Entwicklung solcher Anwendungen kann die entwickelte Ontologie als Referenzwissensbasis dienen, um Machine Learning-Modelle zu trainieren.

Literaturangaben

- [1] Bundesministerium für Wirtschaft und Energie: Details Of the Administration Shell - Part 1. Part 1 - The exchange of information between partners in the value chain of Industrie 4.0 (2020)
- [2] ECLASS e.V.: ECLASS: ECLASS - Standard für Stammdaten und Semantik für die Digitalisierung. <https://www.eclass.eu/index.html>, abgerufen am: 17.11.2020
- [3] Bundesministerium für Wirtschaft und Energie: Describing Capabilities of Industrie 4.0 Components (2020)
- [4] McCrae, J. P.: The Linked Open Data Cloud. <https://lod-cloud.net/>
- [5] 2004. GEFMA 100-2: Facility Management. Leistungsspektrum
- [6] VDI-Richtlinie 6041. Facility-Management - Technisches Monitoring von Gebäuden und gebäudetechnischen Anlagen
- [7] VDI-Richtlinie 3814 Blatt 1. Gebäudeautomation (GA) - Grundlagen
- [8] IEC: IEC 81346-2:2019-06, Industrial systems, installations and equipment and industrial products - Structuring principles and reference designations. Part 2: Classification of objects and codes for classes (2019)

- [9] Backhaus, J. u. Reinhart, G.: Digital description of products, processes and resources for task-oriented programming of assembly systems. *Journal of Intelligent Manufacturing* 28 (2017) 8, S. 1787–1800
- [10] Pfrommer, J., Stogl, D., Aleksandrov, K., Schubert, V. u. Hein, B.: Modelling and orchestration of service-based manufacturing systems via skills. *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*. IEEE 2014 - 2014
- [11] Malakuti, S., Bock, J., Weser, M., Venet, P., Zimmermann, P., Wiegand, M., Grothoff, J., Wagner, C. u. Bayha, A.: Challenges in Skill-based Engineering of Industrial Automation Systems *. 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE 2018 - 2018, S. 67–74
- [12] Perzylo, A., Grothoff, J., Lucio, L., Weser, M., Malakuti, S., Venet, P., Aravatinos, V. u. Deppe, T.: Capability-based semantic interoperability of manufacturing resources: A BaSys 4.0 perspective. *IFAC-PapersOnLine* 52 (2019) 13, S. 1590–1596
- [13] Weser, M., Bock, J., Schmitt, S., Perzylo, A. u. Evers, K.: *Ontology-based Metamodell for Capability Descriptions* (2020)
- [14] Hitzler, P., Krötzsch, Markus, Rudolph, Sebastian u. Sure, Y.: *Semantic Web*. Berlin, Heidelberg: Springer Berlin Heidelberg 2008
- [15] World Wide Web Consortium: The RDF Concepts Vocabulary (RDF), 1999. <https://www.w3.org/1999/02/22-rdf-syntax-ns#>
- [16] World Wide Web Consortium: The RDF Schema vocabulary (RDFS), 2000. <https://www.w3.org/2000/01/rdf-schema#>
- [17] World Wide Web Consortium: The OWL 2 Schema vocabulary (OWL 2), 2002. <https://www.w3.org/2002/07/owl#>
- [18] Rasmussen, M. H., Pauwels, P., Hviid, C. A. u. Karlshøj, J.: Proposing a Central AEC Ontology That Allows for Domain Specific Extensions. *LC3 2017: Volume I - Proceedings of th Joint Conference on Computing in Construction* (2017), S. 237–244
- [19] Automated Ontology Matching in the Architecture, Engineering and Construction Domain - A Case Study: Schneider, Georg. *The 7th Linked Data in Architecture and Construction Workshop* (2019)
- [20] Rasmussen, M. H., Pauwels, P., Lefrançois, M. u. Schneider, G. F.: *The Building Topology Ontology (BOT)*, 2018. <https://w3c-lbd-cg.github.io/bot/bot.ttl>, abgerufen am: 25.04.2021
- [21] Fierro, G. T.: *Design of an Effective Ontology and Query Processor Enabling Portable Building Applications*. Technical Report No. UCB/EECS 2019-106 (2019)

- [22] Bonino, D. u. Corno, F.: DogOnt - Ontology Modeling for Intelligent Domotic Environments. The Semantic Web - ISWC 2008 (2008), S. 790–803
- [23] Bonino, D. u. Russis, L. de: DogOnt. An Ontology for Intelligent Environments, 2013. <http://iot-ontologies.github.io/dogont/index.html#>
- [24] Daniele, L., Hartog, F. den u. Roes, J.: Created in Close Interaction with the Industry: The Smart Appliances REference (SAREF) Ontology 225, S. 100–112
- [25] Pauwels, P. u. Terkaj, W.: EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. Automation in Construction 63 (2016), S. 100–133
- [26] ifcOWL, 2021. <https://technical.buildingsmart.org/standards/ifc/ifc-formats/ifcowl/>, abgerufen am: 25.04.2021
- [27] Haller, A., Janowicz, K., Cox, S. J. D., Lefrançois, M., Taylor, K., Le Phuoc, D., Liebermann, J., García-Castro, R., Atkinson, R. u. Stadler, C.: The SOSA/SSN Ontology: A Joint W3C and OGC Standard Specifying the Semantics of Sensors, Observations, Actuation, and Sampling. Semantic Web X (2018)
- [28] Haller, A., Janowicz, K., Cox, S. J. D., Le Phuoc, D., Taylor, K. u. Lefrançois, M.: Semantic Sensor Network Ontology, 2020. <https://w3c.github.io/sdw/ssn/>, abgerufen am: 25.04.2021
- [29] Both, M. u. Müller, J.: Deep Learning in Industrie 4.0 Umgebungen als Wegbereiter für automatisierte Abbildungen von Ontologien. Automation 2020 (2020), S. 675–686

Auswirkung und Folgen Technischer Schuld in mechatronischen Systemen und daraus resultierende Anforderungen

M.Sc. **F. Bi**, Univ.-Prof. **B. Vogel-Heuser**,
Technische Universität München AIS, München;
Dr. **H. Dettmering**, Prozesswerk, München

Kurzfassung

Die zunehmende Notwendigkeit zur flexiblen Entwicklung und Produktion führen zu technischen Kompromissen in Unternehmen und einer erhöhten Anhäufung von Altlasten mit direkter Auswirkung auf die Qualität des Systems. Das bewusste Eingehen technischer Kompromisse, die auf kurze Sicht Vorteile erbringen können, jedoch auf lange Sicht zu erheblichen Einbußen der Systemqualität und somit auch zu finanziellen Verlusten führen, ist unter dem Begriff „Technische Schuld“ bekannt. Oftmals werden diese Entscheidungen getroffen, weil ihre Tragweite, ihre Auswirkungen und Maßnahmen zur ihrer Behebung nicht eingeschätzt oder unterschätzt werden.

Die Technische Schuld liegt in mechatronischen Produkten jedoch nicht nur im Bereich der Software, sondern auch im Bereich der Mechanik und Elektrotechnik vor. Diese Schuld in der Mechatronik ist jedoch bisher wenig erforscht. Ungleich in Softwaresystemen, ergeben sich in mechatronischen Systemen aufgrund der zusätzlichen Phasen des Lebenszyklus (z.B. Produktion, Inbetriebnahme und Wartung vor Ort), der Vernetzung der verschiedenen Disziplinen sowie verkürzter und asynchroner Innovationszyklen integrierter Produkte, unvorhergesehene, schwer beherrschbare Typen / Arten von technischer Schuld mit weiteren Eigenschaften und Dimensionen.

Beispielsweise wird in einem Use Case mit mechatronischem Fokus anhand einer Presse gezeigt, welche Technische Schuld die Copy-Paste-Modify-Strategie nach sich ziehen kann. Die Mechanik-, Elektronik- und Softwareentwicklung greifen jeweils auf unterschiedliche Versionen ehemaliger Pressen zurück und führen diese für eine Neuentwicklung zusammen. Aufgrund unzureichender Testabdeckung der HW- und SW-Module, Silodenken und nicht-konkreten Systemschnittstellen, führt die Copy-Paste-Modify-Strategie (Ursachen) durch unreife Varianten und Versionen (Technische Schuld-Typ und -Subtyp) zu ungetesteten Sub-Systemen und einem langwierigen Trouble-Shooting Prozess in der Inbetriebnahme (Symptom). Die Lieferfrist wird massiv überschritten, da die Fehlersuche und -behebung aufgrund enormer Komplexität

und Redundanzen einen planbaren Rahmen übersteigt (Konsequenzen). Die so durch Copy-Paste-Modify entstandene Schuld ist bereits früh in der Entstehung zu erkennen, z.B. durch eine Reifegradbewertung über die Testabdeckung der HW-/SW-Komponenten.

In diesem Beitrag analysieren wir mit drei weiteren Use Cases typische Symptome der Technischen Schuld, um diese frühzeitig mittels Indikatoren (z.B. schwache Systemleistung, hohe Ersatzteilanzahl) zu erkennen und die Tragweite der Konsequenzen besser zu überblicken. Drei konkrete Anforderungen an das TD-Management System werden in diesem Beitrag hergeleitet.

1. Einleitung

Insbesondere in Zeiten von Krisen kann ein nicht sorgfältiges Engineering existenzielle Folgen für das Unternehmen bedeuten. Das Risiko nimmt insbesondere zu, wenn vorausgeplante Meilensteine oder Kostenvorgaben nicht erreicht werden, die Wartbarkeit und Erweiterbarkeit stark limitiert werden oder Fehler und Redundanzen gar zum Systemstillstand führen.

Mittels industrieller Use Cases aus Unternehmen des Maschinen- und Anlagebaus wird die angefallene Schuld aus mechatronischer, prozessspezifischer und globaler Sicht vorgestellt. Jeder Use Case beinhaltet diverse Technische Schuld-Teilaspekte, und verdeutlicht, wie sich die Technische Schuld zwischen Disziplinen, über Lebenszyklusphasen und gar über internationale Gewerke ausbreiten kann und die Systemqualität beeinflusst. Das Management der Technischen Schuld bietet ein großes Potential, komplexe Prozesse und Entscheidungen transparenter zu machen, sowie unternehmensübergreifende Schwächen zu identifizieren.

2. Stand der Wissenschaft

Technische Schuld (engl. Technical Debt (TD)), ein Begriff aus Softwaresystemen, bezeichnet die verursachte Schuld aus technischen Entscheidungen, die kurzfristigen Nutzen erbringen, jedoch dem System langfristig Schaden zufügen [1]. Nach der Klassifikation von Avgeriou et al. [1] setzt sich TD aus verschiedenen TD-Elementen (TD Item) zusammen (Abb. 1). TD-Elemente können voneinander abhängig sein, werden durch TD-Ursachen (TD Cause) ausgelöst, weisen TD-Symptome (TD Symptom) auf und führen zu TD-Konsequenzen (TD Consequences). Der Begriff TD-Vorfall bezeichnet das Bündel voneinander anhängiger TD-Elemente, die auf dieselben TD-Ursachen zurückzuführen sind und zu denselben TD-Konsequenzen führen. Um TD systematisch beherrschen zu können, werden für Softwaresysteme TD-Management (TDM)-Aktivitäten vorgeschlagen, welche aus der TD-Identifikation (Identification), TD-Messung (Mea-

surement), TD-Priorisierung (Prioritization), TD-Prävention (Prevention), TD-Überwachung (Monitoring), TD-Tilgung (Repayment), TD-Darstellung/Dokumentation (Representation/ Documentation) und TD-Kommunikation (Communication) bestehen [2].

TD in mechatronischen Systemen wurde erstmals in Form von Architectural TD in Produktionssysteme identifiziert [3]. Als besondere Herausforderungen von TD in der interdisziplinären Systemarchitektur solcher mechatronischer Produktionssysteme wurden der hohe Termindruck, die Montage und Übergabe an den Kunden vor Ort sowie unvorhergesehene Umwelteinflüsse über die Gesamtdauer des Betriebs von bis zu 50 Jahren identifiziert [3–5]. Mit dem V-Modell der Mechatronik als Vergleichsgrundlage betrachtet die Positionsstudie [6] TD disziplinspezifisch und entlang des Lebenszyklus von industriellen Produktionssysteme. TD tritt im Engineering aller drei Mechatronik-Disziplinen sowie in einigen der zusätzlichen Phasen des Lebenszyklus mechatronischer Systeme auf. Fertigung, Montage, Inbetriebnahme und Übergabe an den Kunden sowie Service sind diese zusätzlichen Lebenszyklusphasen in denen TD identifiziert wird [6]. Ebenso stellt die Folgestudie [7] exemplarisch die TD-Typen Requirements, Architectural, Design, Test, Infrastructure und Documentation TD nach Li et al. [2] fest. Als Ursachen für TD in Produktionssystemen werden exemplarisch Inkonsistenzen, Informationsasymmetrie zwischen mechatronischen Disziplinen, Missdeutung von technischen- und Kundenanforderungen durch unzureichende Spezifikationen sowie fehlendes Budget für Wartung identifiziert [6]. Eine Untersuchung mit 50 industriellen Experten aus zehn Branchen [8] zeigt insbesondere die abweichende Verteilung der TD-Typen zwischen mechatronischen Systemen und Softwaresystemen. In den ersten drei Phasen des Lebenszyklus werden die meisten TD-Typen identifiziert (Requirements TD, Architectural TD und Design TD), wobei Infrastructure TD unabhängig zum Lebenszyklus als häufigster TD-Typ genannt wird [8].

3. Technical Debt Use Cases

In diesem Beitrag analysieren wir anhand dreier Use Cases typische Symptome der Technischen Schuld, um diese frühzeitig mittels Indikatoren (z.B. schwache Systemleistung, hohe Ersatzteilanzahl) zu erkennen und die Tragweite der Konsequenzen besser zu überblicken.

TD-Vorfall I: Der Leiter der Abteilung Elektro- und Softwaretechnik eines mittelständischen Automobilzulieferers schilderte einen TD-Vorfall, der die Qualität des Systems neun Jahre lang belastete. Das Unternehmen beliefert Fahrzeughersteller mit einem integrierten Teilprozess für deren Serienproduktion. Basierend auf dem bestehenden Konkurrenzprodukt der vollintegrierten Roboterzelle wurde in den Anforderungen davon ausgegangen, dass jede Roboterzelle nicht

mehr als vier Roboter beinhalten und koordinieren kann. Obwohl vor der Entwicklung eine Marktanalyse und Recherche über den aktuellen Stand der Technik hätte durchgeführt werden müssen (*TD-Ursache: Fehlende Marktanalyse und Recherche*), wurde das Konkurrenzprodukt ohne weitere Recherche kopiert (*TD-Typ: Requirements TD*). Der gesamte mechanische Aufbau, die Steuerung und die Software wurden starr konzipiert und auf vier Roboter beschränkt (*TD-Ursache: Entscheidung ohne spezifisches Wissen*). Weniger als ein Jahr nach der Markteinführung forderten Kunden Zellen mit mehr als vier Robotern (*TD-Symptom: Kundenwunsch*). In den folgenden acht Jahren wurde jedes Mal, wenn mehr als vier Roboter angefordert wurden, das bestehende Produkt mehrfach verkauft. Neben diesem Mehraufwand mussten die Zellenschnittstellen für die Kommunikation jedes Mal aufwendig mit kundenspezifischer Software erweitert werden (*TD-Symptom: Softwareumbau*). Die Software hätte mit einer variablen Anzahl von Robotern in einer Zelle eingerichtet werden müssen (*TD-Typ: Architectural TD*). Dies hätte eine höhere Anfangsinvestition in die Programmierung erfordert, hätte sich aber in kurzer Zeit amortisiert. Um Kosten zu sparen, hielt das Unternehmen nach Entdeckung der Schuld neun Jahre lang an der alten Version der Software fest (*TD-Ursache: Kosten sparen*), bis ein Mitarbeiter von sich aus die Neuprogrammierung der Software veranlasst hatte (*TD-Typ: Infrastructure TD*). Neun Jahre nach der Identifikation der technischen Schuld programmierte ein Software-Ingenieur die gesamte Software mit einem neuen Architektur-Setup erneut (*TD-Folge: Neues Architektur-Setup; Neuprogrammierung der gesamten Software*).

TD-Vorfall II: Ein Prozessingenieur in einem großen Automobilunternehmen beschrieb einen TD-Vorfall, bei dem ein mechatronisches Werkzeug gekauft wurde, obwohl es das zu produzierende Endprodukt zerstören könnte. Im Automobilbereich werden Magnete installiert, um die Bleche für den Umformprozess in der Produktionslinie von Autokarosserien zu separieren. Es gibt Magnete, die elektrisch gesteuert werden können, andere hingegen sind Permanentmagnete. Die steuerbaren Magnete stellen die optimalere technische Lösung dar, da die permanenten eine Überbiegung der Bleche verursachen und diese somit beschädigen können. Die Einkaufsabteilung hat das Vorrecht, die Technologie auszuwählen (*TD-Typ: Infrastructure TD*) und wählt die wirtschaftlichste Variante, da sie keine Folgekosten begründen muss (*TD-Ursache: Kosten sparen; unterschiedliche KPIs in den Abteilungen*). Diese Überbiegung führt auch zu Schäden an anderen Geräten, z. B. leiden die Anwendungen zur Messung der Leiterplattendicke unter stärkerem Verschleiß und müssen daher häufiger gewartet werden (*TD-Typ: Defect TD; TD-Symptom: Kaputtes Teil/Fehler/Defekt*). Obwohl der Lieferant der Magnete die Funktionsfähigkeit der Dauermagnete garantiert, wurde die hohe Leistung der automatisierten Pro-

duktionslinie vernachlässigt (*TD-Ursache: Entscheidung ohne spezifisches Wissen*). Die Mechanikabteilung muss eine Lösung finden, um die beschädigten Teile und Geräte zu warten und zu reparieren. Der Sensor muss immer wieder exakt auf die gleiche Position zurückgesetzt und auf Null gestellt werden. Dieser Mehraufwand kostet zusätzliche Produktionszeit, weshalb eine zusätzliche Prüfstation für dieses Gerät gebaut wird (*TD-Symptom: Zusätzliche Gerätschaften; hohe variable Kosten*). Um Kosten zu sparen, wurden die erforderlichen Prüfungen und Wartungen vom Einkauf gestrichen (*TD-Typ: Maintenance TD; TD-Ursache: Kosten sparen*). Bei einem Produktionsstopp wird die Mechanikabteilung verantwortlich gemacht, nicht die Entscheidung der Einkaufsabteilung (*TD-Symptom: Gefühl der suboptimalen Lösung/bekanntes Risiko*). Der Lieferant der Magnete wird nicht die technisch bessere Variante mit den höheren Anschaffungskosten vorschlagen, da sie ihm geringere Folgekosten verursacht.

TD-Vorfall III: Der Abteilungsleiter des IoT-Consultings eines internationalen Industrieautomatisierungsunternehmens beschreibt seinen TD-Vorfall bei der Anwendung agiler Entwicklungsmethoden in der Mechatronik. Hier übergibt der Chief Product Owner (CPO) mehrere Acceptance Criteria (AC) an den Product Owner (PO). Der CPO muss abschätzen, wie viel ein PO und sein Team innerhalb eines Sprints umsetzen können. Der CPO neigt dazu, die technisch saubere Lösung zu wählen und nur die Anzahl der User Stories mit AC weiterzugeben, die in einem Durchgang erfüllt werden können. Leider wird das gleiche AC an mehrere Teams vergeben. Obwohl der CPO eine ganzheitliche Sicht hat, sollte ein PO nicht durch Aufgaben anderer Teams belastet werden, da er sich sonst nicht auf seine eigenen Aufgaben konzentrieren kann (*TD-Ursache: Zusätzliche Informationen können Entwickler belasten*). Aufgrund der begrenzten Zeit (*TD-Ursache: Zeitmangel/Deadline bis zur Markteinführung/Lieferzeit zum Kunden*), der Anzahl der zu koordinierenden Projekte und Personen bietet der CPO/PO kein ganzheitliches Bild und Informationen werden nicht rechtzeitig an die richtigen Personen weitergegeben (*TD-Symptom: Schlechte Performance*). Außerdem sind die Schnittstellen für die Teilgespräche und -arbeiten im Vorfeld nicht geklärt, so dass ein reibungsloser Zusammenführungsprozess behindert erscheint (*TD-Typ: Infrastructure TD*). Daher kommt es häufig vor, dass Arbeiten doppelt ausgeführt werden. Der CPO hält Anforderungen zurück, so dass spätere Erweiterungen schwieriger werden können (*TD-Typ: Requirements TD; TD-Symptom: Erweiterungen sind schwieriger zu implementieren*).

Herleitung erster Herausforderungen im und Anforderungen an das mechatronische TD-Management System

TD-Vorfälle sind fast in jedem technischen System vorhanden, an dem Menschen beteiligt sind. Anstatt TD generell aus dem System zu verbannen, ist es wahrscheinlicher, die Koexistenz von TD und einem gesunden System zu unterstützen. Bestehende TD-Vorfälle im System müssen identifiziert und beobachtet werden, während zukünftige TD-Folgen in ihrer Eintrittswahrscheinlichkeit und ihrem Ausmaß bewusst bewertet werden müssen, bevor eine Entscheidung getroffen wird. Wesentliche Herausforderungen der TD-Forschung im Software-Engineering sowie in der Mechatronik waren bisher die folgenden und müssen bei der Gestaltung der Datenerfassung und -organisation berücksichtigt werden:

- (1) **Schwierigkeiten bei der Quantifizierung qualitativer Informationen, vor allem der negativen Auswirkungen von TD.** Bisher wurde versucht, den Mehraufwand an zeitlichem und monetärem Mehraufwand durch Expertenschätzungen zu quantifizieren. Die Qualität der Expertenschätzungen wird besonders vage, wenn man bedenkt, dass die Geschäftsprozesse und mechatronischen Zusammenhänge breit und komplex sind. Daher erhalten die Experten typischerweise nur personelle Eindrücke und partielle Implikationen.
- (2) **TD-Vorfälle und deren Ausprägung in der Mechatronik hängen stark von organisatorischen und umweltbezogenen Faktoren ab.** Die Infrastruktur und Prozesse/Vorgehensweise innerhalb eines Unternehmens kann innerhalb von Projekten und Disziplinen variieren. Ereignisse hängen im interdisziplinären Lebenszyklus zusammen. Verschiedene (auch unbekannte) Faktoren sind unsicher. Daher benötigen wir eine entwicklungsfähige und leicht zugängliche Datenstruktur. Eine Analyse von mehrdimensionalen Faktoren muss es erlauben, Zusammenhänge zu erkennen. Dies impliziert, dass alle Faktoren auf vergleichbare Weise dargestellt werden müssen.
- (3) **TD-Konsequenzen eines TD-Items könnten die TD-Ursachen eines anderen TD-Items sein.** In früheren Studien wurden die TD-Ursachen und TD-Folgen getrennt analysiert. Es ist jedoch zu beobachten, dass TD-Folgen des einen Falles TD-Ereignisse eines anderen Falles beeinflussen. Wenn eine Organisation eine begrenzte Anzahl von personellen Ressourcen hat, während eine Aufgabe/Abteilung "ungeplante Überstunden für komplexe Fehlersuche" (TD-Konsequenz) benötigt. Dies impliziert, dass diese TD-Folge einen "Mangel an technischer Expertise" (TD-Ursache) in einer anderen Aufgabe/Abteilung verursachen kann. Daher können TD-Folgen TD-Ursachen beeinflussen und müssen daher in einer Ursache-Wirkungs-Beziehung analysiert werden.

4. Zusammenfassung und Ausblick

TD-Vorfälle und deren Ausprägung in der Mechatronik hängen stark von organisatorischen und umweltbezogenen Faktoren ab. Die Infrastruktur und Vorgehensweise innerhalb eines Unternehmens kann innerhalb von Projekten und Disziplinen variieren. TD-Konsequenzen eines TD-Items könnten die TD-Ursachen eines anderen TD-Items sein. In früheren Studien wurden die TD-Ursachen und TD-Folgen getrennt analysiert. Es ist jedoch zu beobachten, dass TD-Folgen des einen Falles TD-Ereignisse eines anderen Falles beeinflussen. Diese Studie deckt exemplarisch den „ansteckenden Charakter“ von TD auf und verifiziert, dass TD in einer Disziplin anfallen kann, aber eine anderen Disziplin belastet und dort behoben werden muss. Die Tatsache seines „ansteckenden Charakters“ unterstreicht die Notwendigkeit, TD zu verhindern, ebenso wie die Rückzahlung der bestehenden TD-Vorfälle.

Die Analyse des Zusammenhangs von TD-Items und ihren Untermerkmalen innerhalb eines TD-Vorfalles bietet einen möglichen Ansatz für zukünftige TD-Identifikation und Vorhersage. Die drei konkreten Herausforderungen sollten in der Anforderungsgestaltung eines mechatronischen TD-Management Systems berücksichtigt werden.

References

- [1] P. Avgeriou, N. A. Ernst, R. L. Nord, and P. Kruchten, "Technical Debt Broadening Perspectives Report on the Seventh Workshop on Managing Technical Debt (MTD 2015)," *SIGSOFT Softw. Eng. Notes*, vol. 41, no. 2, pp. 38–41, 2016, doi: 10.1145/2894784.2894800.
- [2] Z. Li, P. Avgeriou, and P. Liang, "A systematic mapping study on technical debt and its management," *Journal of Systems and Software*, vol. 101, pp. 193–220, 2015, doi: 10.1016/j.jss.2014.12.027.
- [3] B. Vogel-Heuser, A. Fay, I. Schaefer, and M. Tichy, "Evolution of software in automated production systems: Challenges and research directions," *Journal of Systems and Software*, vol. 110, pp. 54–84, 2015, doi: 10.1016/j.jss.2015.08.026.
- [4] B. Vogel-Heuser and D. Hess, "Guest Editorial Industry 4.0–Prerequisites and Visions," *IEEE Trans. Automat. Sci. Eng.*, vol. 13, no. 2, pp. 411–413, 2016, doi: 10.1109/TASE.2016.2523639.
- [5] A. Fay, B. Vogel-Heuser, T. Frank, K. Eckert, T. Hadlich, and C. Diedrich, "Enhancing a model-based engineering approach for distributed manufacturing automation systems with characteristics and design patterns," *Journal of Systems and Software*, vol. 101, pp. 221–235, 2015, doi: 10.1016/j.jss.2014.12.028.
- [6] B. Vogel-Heuser and S. Rösch, "Applicability of Technical Debt as a Concept to Understand Obstacles for Evolution of Automated Production Systems," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, Kowloon Tong, Hong Kong, 2015, pp. 127–132.
- [7] B. Vogel-Heuser, S. Rösch, A. Martini, and M. Tichy, "Technical debt in Automated Production Systems," in *2015 IEEE 7th International Workshop on Managing Technical Debt (MTD)*, Bremen, Germany, 2015, pp. 49–52.
- [8] B. Vogel-Heuser and F. Bi, "Interdisciplinary effects of technical debt in companies with mechatronic products — a qualitative study," *Journal of Systems and Software*, vol. 171, p. 110809, 2021, doi: 10.1016/j.jss.2020.110809.

Hygienische Qualitätsparameter für die Prozessautomation Vom Labor zum Prozess

Dr. **H. Tippe**, Endress+Hauser Group Services, Weil am Rhein;

Dr. **A. Gahr**, Endress+Hauser Liquid Analysis, Gerlingen;

Dr. **R. Möller**, Analytik Jena, Jena

Kurzfassung

Hygienische Aspekte in Hinblick auf Produktsicherheit und Umweltschutz gewinnen in der industriellen Produktion zunehmend an Bedeutung. Ein wichtiger, wenn auch indirekter Treiber hierfür ist die sich zuspitzende Wasserknappheit und die damit verbundene Notwendigkeit eines nachhaltigen Umganges durch Landwirtschaft, Industrie und Kommune. Die Schließung von Wasserkreisläufen zur Reduzierung des spezifischen Wasserverbrauchs sind erste Schritte, erhöhen jedoch das Risiko auch mikrobieller Kontaminationen mit negativen Folgen für den Anlagenbetrieb und die Produktsicherheit. Das Monitoring relevanter wasserhygienischer Parameter ist bis heute allerdings nur nachträglich, als offline Labor-Information für die Prozessführung nutzbar.

Dies führt zu der Frage, ob und wie hygienische Parameter aus dem Labor mit der Prozesswelt verknüpft werden können. Der Begriff „Lab2Field“ bezeichnet hier zukünftige Lösungen für die inline-Messung hygienischer Wasserparameter in einer, der jeweiligen Applikation angemessenen Informationsqualität.

Am Beispiel der PCR (Polymerase-Kettenreaktion, englisch: Polymerase Chain Reaction) werden anhand zweier Applikationen (Abwasser und Lebensmittelindustrie) die wesentlichen Herausforderungen eines Lab2Field- Ansatzes für hygienische Parameter dargestellt.

1. Handlungsraum für den Technologievergleich

Etliche Entwicklungen in Wissenschaft und Forschung arbeiten an Lab2Field-Ansätzen, um Labortechnologien für die Prozesswelt verfügbar, d. h. mobil und bestenfalls onlinefähig zu machen. Dabei ist für jeden Akteur die individuelle und auf die Messaufgabe ausgerichtete Konkretisierung des vorhandenen Handlungsraumes bedeutsam (Bild 1). Dieser Handlungsraum wird durch folgende Kriterien definiert:

1. Technologien, die sich zur Messung mikrobieller, d. h. hygienisch relevanter Parameter bewährt haben oder zukünftiges Anwendungspotential besitzen.

2. Identifizierte Zielapplikationen, wo die Erfassung der hygienischen Qualität heute oder in absehbarer Zukunft gefordert ist bzw. sein wird. Hierbei können die Gründe unterschiedlich sein, jedoch werden diese in engem Verhältnis zu Mehrwertthemen stehen wie Kosten, Produktsicherheit und/oder -qualität.
3. Informationsgüte, die durch Messtechnologien bereitgestellt bzw. in einer Anwendung gefordert wird. Es kann ein weiter Bereich betrachtet werden, beginnend bei der sehr selektiv-quantitativen Aussage bis hin zur Korrelation „traditioneller“ Parameter zur Risikoabschätzung.

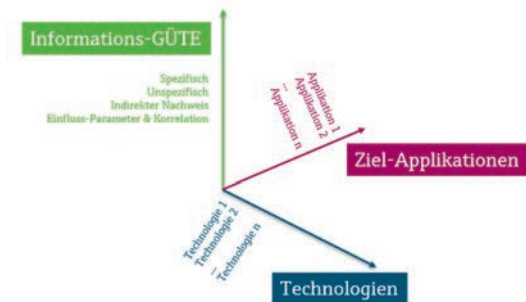


Bild 1: Skizze des sich ergebenden Handlungsraumes beim Technologievergleich.

Dieser Handlungsraum verbindet verfügbare Technologien bzw. zukünftige Technologiepotentiale mit klaren Anwendungsmöglichkeiten für Betreiber. Er bietet Raum sowohl für kurzfristig realisierbare Lösungen als auch für perspektivische Lab2Field-Zukunftsbilder. Voraussetzung ist die konkrete Identifizierung jener Zielapplikationen, für die sich heute oder in Zukunft aus der Messung eines hygienischen Qualitätsparameters ein signifikanter Mehrwert ergibt. Ein sinnvolles Clustern der verschiedenen Technologien zur Messung hygienischer Qualitätsparameter ist anhand der Informationsgüte möglich und sinnvoll. Tabelle 1 zeigt einen Vorschlag. Die Kultivierungsmethode im Labor ist separat aufgeführt, stellt sie heute den etablierten Standard zum Nachweis mikrobiologisch-hygienischer Parameter dar. Sie wird aber zur weiteren Betrachtung hinsichtlich eines Lab2Field-Ansatzes nicht weiter betrachtet.

Tabelle 1: Relevante Technologien für hygienische Qualitätsparameter; nach Informationsgüte geclustert

	Standard- methode	Indikator- parameter	Indirekte Parameter	Unspezifische Parameter	Spezifische Parameter
	Bakterien, Pathogene; nur lebend	pH/ORP, O ₂ , Cl, Tu, Lf, SAK, ...	Biofilm (Dicke)	Anzahl lebender und toter Bakterien	Spezies (Bakte- rien, Viren); lebend & tot
<i>Mess- methode</i>	Kultivierung auf selektiven Nährböden, Paddle Tester, Bart Tester	Chemisch- physikalische Methoden, Optische Methoden	Ultraschall, Elektro- chemisch, Wärme- transfer ...	Durchfluss- zytometrie, ATP, Enzyme, Optische Aus- zählung, (Anti- körper) ...	ELISA (Enzyme-linked- immuno-sorbent Assay), qPCR (quantita- tive Polymerase Chain Reaction), Genom- Sequencing, FISH (Fluores- zenz in-situ Hybridi- sierung), MALDI-TOF (Matrix- Assistierte Laser- Desorption- Ionisierung – Time of Flight), RAMAN, Hyperspektrale Bildererkennung
<i>Informa- tionsgüte</i>	Spez. Nachweis lebender Mikro- organismen	Parameter- veränderung zur Risiko- abschätzung, ggf. unter Nutzung von Korrelations- Algorithmen	Biofilm- Nachweis zur Risikoabschät- zung; CIP-Reinigung (Effizienz)	Quantitativer, direkter Nachweis der bakteriellen Belastung	Quantitative und qualitative Analyse der bakteriologi- schen Belastung
<i>Verfügbarkeit</i>	3 - 14 Tage	Sek - Min	Sek - Min	< 2 Stunden	< 8 Stunden

Welche Technologie verspricht nun in welcher Applikation den größten Nutzen und Mehrwert für den Betreiber? Da jede Technologie spezifische Einsatzmöglichkeiten aber auch Anwendungsgrenzen besitzt, kann diese Frage nur anhand konkreter Anforderungen der Zielapplikationen beantwortet werden. Das erfordert eine genaue Betrachtung der beiden Achsen *Technologien* und *Zielapplikation* (Bild 1) und die Formulierung geeigneter Differenzierungskriterien. Die *Informationsgüte*, die entweder applikativ gefordert oder technologisch bereitgestellt wird, stellt dabei den verbindenden Aspekt dar. Sie wird technologisch bestimmt durch die Art und Verfügbarkeit der Information und den Leistungskriterien der Messung (Nachweisgrenze, Präzision, Genauigkeit, etc.). Der Nutzen einer zur Verfügung gestellten Informationsgüte ist bedeutsam zur Bewertung technologischer Innovationsfelder und kann ausschließlich im Kontext konkreter Zielapplikationen beantwortet werden. Bild 2 zeigt ohne Anspruch auf Vollständigkeit eine Auswahl maßgeblicher Kriterien.

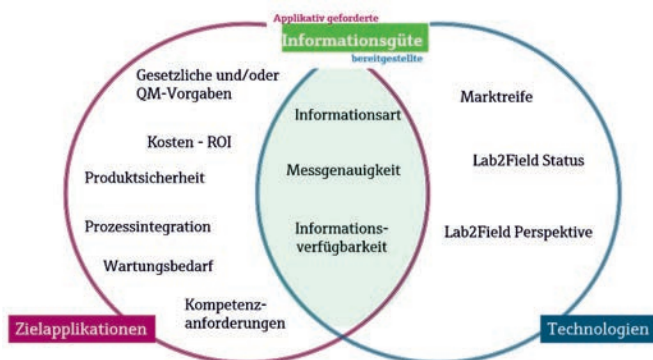


Bild 2: Kriterienauswahl zur Beurteilung von Technologien und Applikationen.

2. Lab2Field-Ansatz für ein PCR-basiertes Pandemiemonitoring im Abwasser

Die Messung mikrobiologisch relevanter Qualitätsparameter in der Wasserver- und Abwasserentsorgung ist als Labormethode state-of-the-art, nicht zuletzt, da auch gesetzliche Anforderungen hier seit langem greifen. Aber nicht nur der Gesetzgeber, sondern v. a. neue Ansätze des Risiko- und Qualitätsmanagements durch die Anlagenbetreiber führen zu weitergehenden Anforderungen im Hinblick auf die mikrobiologische Wasserqualität. Erste Ansätze des online-Monitorings hygienisch relevanter Parameter begannen vor Jahren insbesondere im Bereich der Trinkwassernetze. Mit Hilfe von Standard-Qualitätsparametern (pH, Sauerstoff, Trübung, freies Chlor u. a.), Algorithmen und Cloudlösungen können Verän-

derungen der Wasserqualität online verfolgt und hinsichtlich der hygienischen Aspekte kritisch bewertet werden.

Die Corona-Pandemie führte zu einer neuen, sehr drängenden Fragestellung: Ist es möglich, eine SARS-CoV-2-Analytik im Zulaufstrom der Kläranlagen zu realisieren und darüber das Pandemiegeschehen indikativ, aber relativ schnell zu bewerten? Um das zu realisieren, musste die qPCR, als Goldstandard für den SARS-CoV-2 Nachweis inzwischen allgemein bekannt, dem Kläranlagenprozess und dessen Medium Abwasser zugeführt werden. Eine Lab2Field-Herausforderung, an deren Beispiel die typischen Fragestellungen eines solchen Prozesses deutlich wurden und am Ende auch die Sinnfrage einer 100%igen Prozessintegration als online-Messgröße exemplarisch diskutiert werden kann.

Die **qPCR (quantitative Echtzeit Polymerase-Ketten Reaktion)** ist eine Methode zur thermisch getriggerten Vervielfältigung (Amplifizierung) von Nukleinsäuren eines ausgewählten DNA- oder RNA-Abschnittes (z. B. von SARS-CoV-2) inklusive deren Quantifizierung. Die Quantifizierung wird mit Hilfe von Fluoreszenz-Messungen durchgeführt; das steigende Fluoreszenz-Signal kann während der Amplifizierungszyklen beobachtet werden. qPCR ist heute eine reine Labormethode und beinhaltet die drei grundlegenden Prozessschritte Probenahme, Probenvorbereitung (Extraktion und Anreicherung) und qPCR Analytik (Vervielfältigung und Messung).

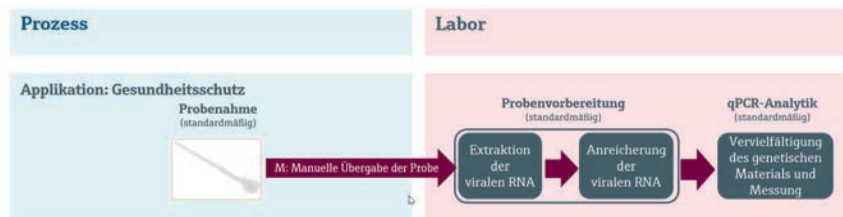


Bild 3: Übersicht der Prozessschritte einer qPCR als Labor-Standardmethode u.a. für den SARS-CoV-2 Nachweis.

Die qPCR ist also eine höchst selektive, analytische Methode, die sogar als Multiparametersystem gleich mehrere Zielorganismen (Bakterien, Viren, etc.) parallel quantifizieren kann. Das Zeitfenster zwischen Probenahme und Ergebnis beträgt 2 bis 6 Stunden, zuzüglich Probentransport. Originäres Anwendungsbiet der qPCR heute sind medizinische Fragestellungen. Primäre Anwender dieser Technologie sind Krankenhäuser, Großlabore und universitäre Forschungsstätten; alle mit hohem bis sehr hohem täglichen Probendurchsatz.

Angesichts der Komplexität der Messmethode und Verschiedenheit zwischen der originären Anwendung (Gesundheitswesen) und der Zielapplikation (Kläranlage – ungereinigtes Abwasser) war schnell klar, dass der Lab2Field-Ansatz schrittweise erfolgen musste. Ausgehend von einer leistungsstarken vorhandenen Laboranalytik erlangte die Frage nach der repräsentativen Probenahme und -vorbereitung oberste Priorität. Denn nur dann sind Messergebnisse aussagekräftig und für das Pandemie-Monitoring von Nutzen.

Abwasser ist ein schwieriges Medium, insbesondere im Zulaufbereich der Kläranlage, wo noch kein Reinigungsschritt erfolgte. Menge, Zusammensetzung und Schmutzfrachten variieren stark. Automatische Probenehmer ermöglichen eine repräsentative Probenahme, wobei i. d. R. hohe Probevolumina vom mindestens mehreren Hundert Millilitern genommen werden. Für die qPCR-Analyse ist die originäre Abwasserprobe bereits aus Gründen des Probevolumens nicht geeignet. Zudem stellt die geringe Konzentration der für die PCR-Analyse verwertbaren Virus-RNA im Rohabwasser eine Herausforderung dar. Der erste Schritt musste also der Methodenentwicklung für die Probenvorbereitung gewidmet werden, um eine reproduzierbare Aufkonzentrierung der originären Abwasserprobe zu sichern und damit einen reproduzierbaren qPCR-Nachweis von SARS-CoV-2 in Abwasser überhaupt erst zu ermöglichen.

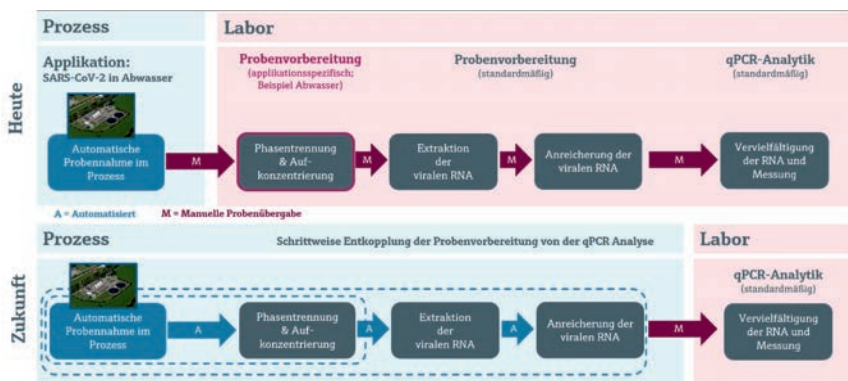


Bild 4: Lab2Field-Ansatz für die qPCR Analyse in Abwasser. Die Annäherung an die Prozesswelt erfolgt durch schrittweise Verlagerung der Probenvorbereitung in den Prozess und führt zur Entkopplung von der eigentlichen qPCR-Messung im Labor.

Dank weltweit vernetzter Forschungsarbeiten konnte grundsätzlich der Nachweis von SARS-CoV-2-Fragmenten in Abwasser zügig demonstriert werden.¹² Dabei stand der erfolgreiche Virusnachweis mittels qPCR als Hilfestellung zur Pandemiebekämpfung im wissenschaftlichen Fokus, aber nicht die Etablierung eines durchgängigen, prozessnahen Workflows von Probenahme bis zur Messung, wie es ein Field2Lab-Ansatz verlangt. Somit kamen eine Vielzahl unterschiedlicher Labormethoden zur Virusanreicherung aus Abwasserproben zum Einsatz, welche aber im Sinne einer Prozessintegration (Bild 4) nicht weiterentwickelt bzw. grundsätzlich kaum in Frage kamen.

Die Anforderungen an eine Probenvorbereitung aus Abwasserproben für eine prozessnahe qPCR-Analyse lassen sich wie folgt beschreiben:

- Repräsentative Probennahme aus dem Prozess (im Abwasserbereich mit Hilfe automatischer Probenehmer gegeben).
- Reproduzierbare, möglichst verlustfreie Überführung der hochverdünnt vorliegenden Ziel-DNA bzw. -RNA aus einer Abwasserprobe in eine, für die qPCR-Methode geeignete Vorlage.
- Hohes Automatisierungspotential und damit einfache, sichere Handhabung.

Versuche an einer realen Applikation im Zulauf einer Kläranlage (Bundesland NRW) zeigen, dass mittels elektronegativer Membranfilter (Porenweite 0,45 µm) eine geeignete Probenvorbereitung unbehandelten Abwassers für die spätere qPCR realisiert werden kann. Unter Ausnutzung der elektrostatischen Wechselwirkung zwischen Filtermaterial und Virusfragmenten (RNA und gebundene Proteine) werden letztere an die Filtermatrix gebunden und reproduzierbar aufkonzentriert. Die für den Nachweis relevanten RNA-Fragmente liegen stabilisiert im Filtermaterial vor - eine wichtige Voraussetzung für die sichere Handhabung und den Transport. Damit ist die erste Herausforderung der Prozessintegration von qPCR-Analytik an einen realen Abwasserprozess gelöst.

¹ Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands; Gertjan Medema, Leo Heijnen, Goffe Elsinga, Ronald Italiaander, and Anke Brouwer; Environmental Science & Technology Letters 2020 7 (7), 511-516; DOI: 10.1021/acs.estlett.0c00357

² Primary concentration – The critical step in implementing the wastewater based epidemiology for the COVID-19 pandemic: A mini-review; Dingnan Lu, Zhuangrong Huang, Jiayue Luo, Xiaoqi Zhang, Sha Sha; Science of The Total Environment; 2020 (747); <https://doi.org/10.1016/j.scitotenv.2020.141245>.

Es deutet sich darüber hinaus ein weit umfassenderer Entkopplungsprozess zwischen Probenvorbereitung und Messung an (Bild 4). Um einen möglichst hohen Automatisierungsgrad in der gesamten Probenaufbereitungskette mit nur geringfügigen manuellen Eingriffen zu erzielen, wurden für den SARS-CoV-2-Anwendungsfall auf der Kläranlage auch die nächsten Schritte *Extraktion und Virusanreicherung* optimiert und vereinfacht. Auch wenn hier eine volle Automation noch aussteht, stand damit aus Betreibersicht Aufwand und Nutzen in einem akzeptablen Verhältnis, während gleichzeitig die Verlässlichkeit und Routinefähigkeit des SARS-CoV-2 Nachweises garantiert werden konnte.

Das absehbare Szenario einer prozessnahen qPCR-Analytik ist also eng verbunden mit dem Automatisierungsgrad der Probenvorbereitungsschritte und deren Verlagerung in den Prozess bei gleichzeitiger Entkopplung von der eigentlichen qPCR-Messung. Letztere bleibt zunächst noch Aufgabe des (Betriebs-)Labors. Damit wird noch keine vollumfängliche Prozessintegration, d. h. keine online-fähige Messung der spezifischen Bakterien- oder Virenlast erreicht. Jedoch sollte eine vereinfachte, schnelle sowie kostengünstigere qPCR-Analytik dank repräsentativer, automatisierter Probenvorbereitung grundsätzlich das Potential zu einer Standard-Betriebsmethode besitzen.

3. Erfahrungen mit einer mobilen PCR für die Qualitätskontrolle in der Lebensmittelproduktion

Die mikrobielle Überwachung von Prozessen und Produkten hinsichtlich kritischer Mikroorganismen, z. B. von Salmonellen oder Listerien, sind in der Lebensmittelindustrie zwingend gefordert. Standardmethode zum Nachweis von Mikroorganismen ist auch hier das Kultivierungsverfahren im Labor; manchmal bereits ergänzt durch die modernere qPCR-Analyse. Für den Betreiber bedeuten beide Methoden ein erheblicher Zeitverzug zwischen Probenahme und Messergebnis. Dieser kann Stunden bis mehrere Tage betragen, da i. d. R. die Analytik an externe Speziallabore übertragen werden. Dabei liegt es nahe, dass in einer Branche mit so vielseitigen hygienischen Fragestellungen und Herausforderungen der Bedarf eines prozessnahen Monitorings hygienischer Parameter gegeben ist. So auch in einem chinesischen Groß-Schlachtbetrieb, in dem die mikrobielle Qualität des Rohfleisches in der Anlage direkt vom Betriebspersonal getestet werden sollte.

Die Anforderungen lagen hier im Vergleich zur o. g. Abwasserapplikation woanders. Die Probenahme (Abstrich am Fleisch) war dem der ursprünglichen PCR-Anwendung, dem Gesundheitswesen, sehr ähnlich und bedurfte keiner weiteren Betrachtung. Innovationsbedarf lag vielmehr in der Gestaltung einer durchgehend mobilen und automatisierten PCR-

Analyseneinheit, die in Analogie zur sog. Point-of-Care Lösungen der medizinischen Diagnostik eine PCR-Analyse direkt vor Ort durch normales Betriebspersonal erlaubt.



Bild 5: Das kompakte MobiLab² vereint alle notwendigen Geräte und Analysenschritte in einem transportablen System, um mobile PCR-Analysen von verschiedensten Proben durchzuführen.

Realisiert wurde ein tragbares System (MobiLab², Bild 5), das alle erforderlichen Geräte, Reagenzien und Arbeitstools beinhaltet und eine einfache Handhabung von Probe und Chemie durch Betriebspersonal erlaubt. Erreicht wurde dies durch eine Vielzahl kleiner Verbesserungen im Sinne einer weitergehenden Standardisierung. Beispielsweise durch vorpräparierte „ready-to-use“ Assays, die eine leichte Handhabung sowie eine gut verständliche visuelle Auswertung bieten. Um örtlich flexibel zu sein, ist das MobiLab² System zudem batteriebetrieben.

Für den Einsatz in der chinesischen Großschlachtereier war der Test erfolgreich, jedoch zeigten sich auch einige Schwachstellen, die einem größeren Einsatz in der Zielbranche Lebensmittelindustrie bisher im Wege stehen. Dazu gehören ein noch ungünstiges Kosten-Nutzen Verhältnis aufgrund der Limitierung, nur ein Bakterium oder Pathogen pro Messung nachweisen zu können. Eine Kritik, die mit heute inzwischen erhältlichen Multi-Parameter-Assays begegnet werden könnte. Als kritischer sind die Anforderungen an Nachweisgrenzen anzusehen, die für viele Applikationen in der Lebensmittelindustrie gegeben sind. Da das mobile PCR-System ohne vorherige Anreicherung arbeitet, stößt es verfahrensbedingt an Grenzen, die für eine Standard-PCR-Analyse i. d. R. noch nicht gelten.

Auf den anfangs eingeführten Handlungsraum zurückkommend (Bild 1), dient das vorgestellte MobiLab² als ein Beispiel, bei dem die von der Technologie bereitgestellte Informationsgüte (Bild 2) für die Zielapplikation nicht oder noch nicht ausreichte, trotz der Vorteile wie hohe Flexibilität, direkter Einsatz im Prozess, einfache Handhabung und vor allem beschleunig-

nigte Messergebnisse. Das bedeutet nicht, dass diese Schlussfolgerung für andere Zielapplikationen im selben Maße gelten muss oder zukünftige technologische Verbesserungen die bestehenden Limitierungen nicht aufheben könnten. Für Ersteres bedarf es aber eine Neubewertung der applikativen Anforderungen; für Zweiteres die technologische Weiterentwicklung des tragbaren PCR-Systems.

4. Fazit

Wir erleben im Grunde seit vielen Jahrzehnten einen kontinuierlichen „Lab2Field“-Prozess für Qualitätsparameter: Angefangen bei einfachen chemisch-physikalischen Parametern, fortgeführt mit optischen und kolorimetrischen Methoden bis hin zur RAMAN oder anderen komplexen Messmethoden, die heute der Prozessautomation zur Verfügung stehen. Daher ist es nicht erstaunlich, dass auch im Bereich der mikrobiologischen Qualitätssicherung Bewegung erkennbar ist und neue Technologien aus dem Labor in die Prozesswelt drängen. Natürlich gilt die Kultivierungsmethode immer noch als Standard für den mikrobiellen Nachweis, nicht zuletzt im regulatorischen Kontext. Doch wird diese Labormethode den Anforderungen einer modernen prozessnahen Eigenüberwachung der Anlagenbetreiber immer weniger gerecht. Denn neben der Einhaltung gesetzgeberischer Vorgaben zielt Prozesskontrolle und -automation auf die kontinuierliche Verbesserung der Anlageneffizienz und -verfügbarkeit (Energie, Produktivität, Wassernutzung u. v. m.) sowie der Produktqualität ab. Diese Motivation der Anlagenbetreiber wirkt in zweierlei Hinsicht: Als Treiber für regulatorische Anforderungen und gleichzeitig als Motivator für Innovationen. Das gilt auch für Lab2Field-Ansätze für hygienische Qualitätsparameter.

Vorgestellt wird der SARS-CoV-2-Nachweis in Abwasser durch eine prozessnahe qPCR-Analytik zur Beobachtung des Pandemiegeschehens. Dieses Beispiel demonstriert eindrucksvoll, wie eine neue applikative Anforderung eine komplexe Labormethode (hier: qPCR) zu gänzlich neuen Anwendungsfeldern und zu konkreten Schritten im Sinne des Lab2Field-Ansatzes führen kann. Die sog. prozessnahe qPCR-Analytik basiert zunächst auf eine Weiterentwicklung der Probenvorbereitung, um die Anbindung an den Kläranlagenprozess und eine repräsentative Probe überhaupt zu gewährleisten. Des Weiteren führt eine partielle Automatisierung der für die qPCR notwendigen Probenvorbereitungsschritte zur verbesserten Handhabbarkeit und Akzeptanz beim Betriebspersonal. Die komplette Verlagerung der Probenvorbereitung in den Prozess ist absehbar, was einer Entkopplung von Probenvorbereitung und qPCR-Messung gleichkommt. Damit wird perspektivisch eine anspruchsvolle, bisher den Großlaboren und Spezialisten vorenthaltende Labormethode auch für Betriebslabore greifbar.

Dagegen gelang es im zweiten Beispiel nicht umfänglich, die PCR in einem neuen kompakten Konzept als mobile Lösung für die Warenkontrolle in einem Schlachtbetrieb durchzusetzen. Die Vereinfachung der Methodik zugunsten der kompakten Bauform führte zur Verschlechterung der Informationsgüte des Gesamtsystems mit der Folge, dass sie den in der Lebensmittelindustrie geforderten Ansprüchen insbesondere in Hinblick auf den unteren Messbereich nicht mehr ausreichend genügte.

Natürlich ist ein Lab2Field-Ansatz für hygienische Parameter nicht auf die PCR-Methode beschränkt. Das zeigt der Handlungsraum, der durch die drei Achsen Technologie, Zielapplikation und Informationsgüte aufgespannt wird. Welche Technologien sich für die Messung hygienischer Qualitätsparameter auch in der Prozessautomation durchsetzen werden, ist noch offen. Hier ist das Kräfteverhältnis zwischen den zwei grundsätzlichen Zielstellungen der Prozessautomation entscheidend: Die Steuerung oder Regelung verfahrenstechnischer Prozesse zwecks Optimierung versus Monitoring der (Produkt-)Qualität. Ist das Monitoring der treibende Faktor, dann dominieren Anforderungen wie Selektivität und Genauigkeit und weniger die Geschwindigkeit der Messung. So darf beim Beispiel der SARS-CoV-2-Pandemieüberwachung auf Kläranlagen kritisch hinterfragt werden, ob der Lab2Field-Ansatz überhaupt bis zum Ende, d. h. bis zum online-fähigen qPCR-Messsystem getrieben werden muss! Gegebenenfalls liegt der maximale Nutzen bereits bei der automatisierten Prozessintegration von Probenvorbereitung und einer vereinfachten, kostengünstigen qPCR-Analytik im Betriebslabor.

Wird jedoch nach einem hygienischen Qualitätsparameter gefragt, weil eine weitergehende Prozessoptimierung nur so erreichbar ist, dann ist die Frage der online-Fähigkeit einer Messmethode prioritär zu bewerten. Unter Umständen liegt der optimale Mehrwert einer Lab2Field-Lösung dann eher im Kompromiss zwischen einer weniger spezifischen Messmethode, dafür aber schnellen-online Messung direkt im Prozess. Welche Technologien hier den besten Mehrwert bieten, ist aus Sicht des Autorenteams noch offen.

Literatur

- [1] Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands; Gertjan Medema, Leo Heijnen, Goffe Elsinga, Ronald Italiaander, and Anke Brouwer; Environmental Science & Technology Letters 2020 7 (7), 511-516; DOI: 10.1021/acs.estlett.0c00357
- [2] Primary concentration – The critical step in implementing the wastewater based epidemiology for the COVID-19 pandemic: A mini-review; Dingnan Lu, Zhuangrong Huang, Jiayue Luo, Xiaoqi Zhang, Sha Sha; Science of The Total Environment; 2020 (747); <https://doi.org/10.1016/j.scitotenv.2020.141245>.

OPC-UA-Applikationserkennung in Software-defined Networks für Industrie 4.0

Dr.-Ing. **D. Schulze-Zipper**,
Institut für Automation und Kommunikation e.V. (ifak), Magdeburg;
Dr.-Ing. **W. Mahnke**,
ascolab GmbH, Kalchreuth;
M.Eng. **D. Gerlach**, M.Eng. **J. Neubauer**,
SBSK GmbH & Co. KG, Schönebeck

Kurzfassung

Das Konzept Industrie 4.0 sieht eine steigende Flexibilität im Produktionsprozess vor. Dazu können, beginnend bei den Planungsaufgaben bis hin zur produktionstechnischen Ausführung, Aufgaben maschinell ausgehandelt und kooperativ durchgeführt werden. Immanent für Industrie 4.0 ist die Fähigkeit der Teilnehmer, dynamisch auf Anforderungen aller Ebenen zu reagieren und selbstständig geeignete Kommunikationspartner zu finden. Anforderungen können dabei hinsichtlich der Sicherheit oder der Echtzeiteigenschaften formuliert werden. In diesem Themenumfeld ist das Projekt MONAT (Modellbasierte und bedarfsgerechte Netzwerk-konfiguration für Netzwerke der Automatisierung und Telekommunikation, Förderkennzeichen: 16KIS0779) angesiedelt. Neben der Informationsmodellierung [1] und damit der für Industrie 4.0 benötigten semantischen Selbstbeschreibung bietet Open Plattform Communication Unified Architecture (OPC-UA) [2] als standardisierte Kommunikationsplattform den sicheren, zuverlässigen und herstellernunabhängigen Datentransport und garantierte Interkonnektivität über alle Ebenen der industriellen Produktion.

Der Einsatz eines logisch zentralisierten Netzwerkmanagements sowie die dynamische Programmierbarkeit der Infrastruktur bieten das Potenzial, den Anforderungen der Industrie hinsichtlich Hochverfügbarkeit und verzögerungsarmer Kommunikation zu begegnen und skalierbare Lösungen zu realisieren.

Die Integration einer Anwendungserkennung als Software-defined-Networking (SDN)-Applikation ist ein Ansatz, das existierende Netzwerkmanagement in seinen Aufgaben hinsichtlich Monitoring und Konfiguration zu unterstützen und zu ergänzen.

Das Projekt setzt sich mit sicherheitskritischen Fragestellungen aus den Bereichen OPC-UA und SDN auseinander.

OPC-UA bietet zwei grundsätzliche Mechanismen zur Kommunikation. Neben der etablierten Client-Server-Architektur gibt es inzwischen auch ein Publish-Subscribe-Modell. Dieses kann sowohl innerhalb eines AT-Netzwerks als auch zur Kommunikation in IT-Netzwerken verwendet werden. Im Beitrag werden die entwickelten OPC-UA-Prototypen [3] für beide Kommunikationsarten vorgestellt, inklusive der für die IT-Sicherheit benötigten zentralen Komponenten.

Weiterhin wird in diesem Beitrag die Entwicklung sowie die Etablierung einer OPC-UA-Anwendungserkennung [4] in ein SDN vorgestellt. Die binäre Klassifikation von Kommunikationsbeziehungen ist in diesem Zusammenhang ohne Berücksichtigung von Portnummern und Nutzdateninspektion realisiert und ermöglicht auch die Einordnung von verschlüsseltem Datenverkehr. Als Identifikationsmerkmale dienen dem Machine-Learning-gestützten Klassifikator statistische Paket- und Verbindungseigenschaften. Im Rahmen der betrachteten Anwendungsfälle wird eine Erkennungsgenauigkeit von über 95 % bezogen auf beide OPC-UA-Kommunikationsmodelle gezeigt.

Zudem soll das Netzwerkmanagement in einer SDN-Architektur aus sicherheitskritischer Sicht bewertet werden [5]. In diesem Zusammenhang werden speziell Failover-Mechanismen beschrieben. Für die Herangehensweise wird ein Prototyp zur Paketübertragung, bestehend aus mehreren Switchen und Hosts, aufgebaut. Die Steuerebene bei dem SDN-Netzwerk bildet der Open Network Operating System (ONOS) Controller. Mit der Definition verschiedener Testfälle wird gezeigt, dass bei Geräteausfällen ein Umrouting der Pakete funktioniert.

1. Anforderungen in der industriellen Umgebung

Die stetige Flexibilisierung von Produktionsprozessen steht im Fokus der industriellen Entwicklung. Entstehend durch steigende Warenangebote, Kleinserienfertigungen und immer kürzer werdenden Lieferzeiten, wird eine gesicherte, dynamische Verteilung von Daten über Kommunikationsnetze nötig. Zeitkritische Produktionsabläufe und eine bessere Ressourceneffizienz wird möglich. Neuartige Formen dienstintegrierter, breitbandiger und latenzarmer Netzwerke entstehen, welche zudem dynamisch auf veränderliche Anforderungen reagieren. Dabei treten Wechselbeziehungen zwischen unternehmensübergreifenden Diensten und lokalen Prozessen vermehrt in den Fokus. Bedarfsgerechte Lastverteilungen, automatisierte Verwaltungsaufgaben und herstellerübergreifende Kommunikationsprotokolle stellen dabei die Kernthemen dieser Entwicklung dar.

Drahtlose Kommunikationssysteme mit sehr geringen Latenzen helfen bei der Etablierung neuer Netzwerkkonzepte. Die Konvergenz heterogener Kommunikationsnetzwerke im industriellen Umfeld lässt neue Business-Cases entstehen. Die Kombination von Produktionsprozessen mit Cloud- und Edge-Cloud-Ressourcen erzeugt eine steigende Produktindividualität und kürzere Rüstzeiten. Die Harmonisierung verschiedener Kommunikationssysteme erfordert den Einsatz zeitkritischer und hochverfügbarer Netzwerke.

Ein denkbare Szenario stellt die Verwendung von Low-Power-Wireless-Netzwerkprotokollen für die Errichtung von Sensornetzwerken dar. Weiterhin birgt 5G großes Potential für die Vernetzung von Werkhallen, Campusnetzwerken sowie für die Weitverkehrsübertragungen. SDN bietet für die Etablierung heterogener Netzwerkkonzepte ganzheitliche Möglichkeiten für Security- und Network-Management sowie für das Traffic-Engineering.

Die Kombination von OPC-UA und SDN ermöglicht ein zentrales Management für Dienste und Anwendungen der Automatisierungstechnik. Das Ressourcen- und Konfigurationsmanagement kann Anpassungen automatisiert für die Komponenten im Netzwerk verteilen.

2. Vorstellung der Mechanismen zu OPC-UA

OPC-UA ist ein etablierter Standard zum Informationsaustausch in industriellen Umgebungen. Dabei werden zum einen Mechanismen zur Informationsmodellierung vorgegeben, die eine semantische Selbstbeschreibung der Daten und damit einen Informationsaustausch ermöglichen. Zum anderen zeigt die breite Anwendung von OPC-UA in unterschiedlichen Domänen eine Vielzahl der auf Basis von OPC-UA erstellten standardisierten Informationsmodelle, beispielsweise für Roboter, Werkzeugmaschinen, Bildverarbeitungssysteme, Pumpen, RFID-Lesegeräte, Waagen, Spritzgussmaschinen, kommerzielle Küchengeräte, SPSEN und Analysegeräte.

Zum anderen wird eine standardisierte Kommunikationsplattform bereitgestellt, die eine sichere, zuverlässige und herstellerunabhängige Kommunikation bietet. Neben verschiedenen Protokolloptionen und Kodierungen der Daten stellt OPC-UA dabei zwei unterschiedliche Kommunikationsmechanismen zur Verfügung: die Client-Server-Architektur sowie das Publish-Subscribe-Modell. Ersteres ist seit der ersten Version von OPC-UA definiert, wohingegen letzteres erst 2018 zu OPC-UA hinzugefügt wurde, um spezielle Anwendungsfälle besser zu unterstützen. Der grundlegende Unterschied der beiden Ansätze ist in Bild 1 dargestellt.

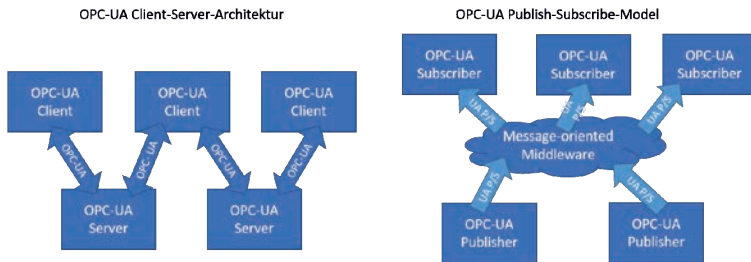


Bild 1: Verschiedene Kommunikationsmechanismen in OPC-UA

Bei der Client-Server-Architektur besteht immer eine individuelle Kommunikationsbeziehung zwischen Client und Server. Ein Server kann von mehreren Clients gleichzeitig angesprochen werden, und ein Client kann gleichzeitig mit mehreren Servern kommunizieren. Der Datenaustausch findet aber immer individuell per Kommunikationsbeziehung statt. Beim Publisher-Subscriber-Modell hingegen veröffentlicht ein Publisher Nachrichten, die von verschiedenen Subscribern empfangen werden können. Hier findet also nicht mehr gezwungenermaßen eine Eins-zu-Eins-Kommunikationsbeziehung statt, sondern ein Publisher kann zeitgleich an viele Subscriber dieselbe Nachricht schicken, und ein Subscriber kann natürlich auch Nachrichten von verschiedenen Publishern empfangen. Je nach Protokoll-Option kann hierbei noch ein Broker eingebunden werden, der die Nachrichten für die Subscriber bereitstellt (MQTT und AMQP), und damit auch eine zeitliche Entkoppelung von Publisher und Subscriber ermöglicht. Die Nachrichten können aber auch direkt ausgetauscht werden, entweder über UDP (als Multicast für mehrere Subscriber) oder über TSN (Time Sensitive Networking) für einen deterministischen Echtzeitdatenaustausch.

Im Publish-Subscribe-Modell findet lediglich eine unidirektionale Kommunikation statt, d.h. der Publisher schickt ausschließlich Daten, und der Subscriber empfängt ausschließlich Daten. Selbstverständlich können die Kommunikationspartner auch jeweils die andere Rolle einnehmen, sodass eine OPC-UA-Applikation als Publisher Daten verschickt und als Subscriber Daten empfängt. Im Gegensatz dazu findet bei der Client-Server-Architektur inhärent eine bidirektionale Kommunikation statt, d.h. der Client empfängt nicht nur Daten vom Server, sondern kann beispielsweise Werte schreiben oder Methoden des Servers aufrufen. Das reine Publish-Subscribe-Modell bietet damit lediglich eine Untermenge der Funktionalität der Client-Server-Architektur.

Häufig werden OPC-UA-Anwendungen deshalb sowohl Server-Funktionalität bereitstellen, als auch die Publisher- und / oder Subscriber-Rolle einnehmen. Tatsächlich bietet OPC-UA über das Informationsmodell die Möglichkeit, bei einem Server sowohl die Publisher- als auch die Subscriber-Rolle zu konfigurieren, wie in Bild 2 dargestellt.

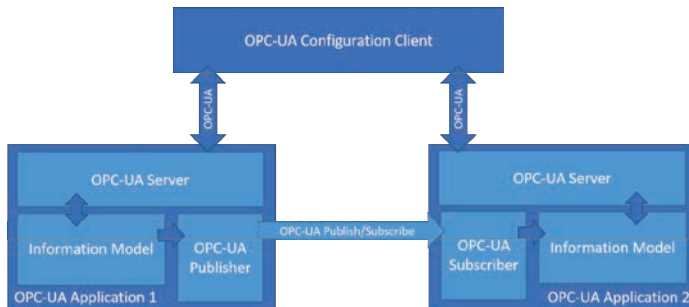


Bild 2: Beispielanwendung die OPC-UA Client/Server und Pub/Sub kombiniert

Die Anwendung 1 wird, ebenso wie die Anwendung 2, über die Client-Server-Architektur von einem Konfigurationswerkzeug konfiguriert, sodass Anwendung 1 Daten aus dem Informationsmodell als Publisher veröffentlicht, wohingegen Anwendung 2 als Subscriber konfiguriert wurde, der die Daten in Empfang nimmt und in sein Informationsmodell einbringt. Beispielsweise könnten das zwei SPSen sein, bei dem die eine ihre Ausgabe-Werte veröffentlicht, und die andere diese als Eingabe-Werte in Empfang nimmt.

Die in Bild 2 dargestellte Architektur wurde in Prototypen [3] implementiert, in denen Laufzeitdaten generiert werden, und diese dienten damit für beide Kommunikationsarten als Anwendungsbeispiel.

3. Vorstellung der OPC-UA-Anwendungserkennung

Um den heterogenen Anforderungen der beiden OPC-UA-Kommunikationsmechanismen an die Übertragung zu begegnen, ist die Sicherstellung der Quality of Service (QoS) elementar. Die Integration einer Anwendungserkennung in ein SDN ist ein vielversprechender Ansatz, um die Gewährleistung der Dienstgüte dynamischer und effektiver zu gestalten.

Aufgrund der steigenden Anzahl von Anwendungen, welche dynamische Ports zur Kommunikation nutzen, sowie der Verbreitung verschlüsselter Datenübertragung und Tunnelmechanismen, gilt die Port- und Protokollanalyse zum Zweck der Identifikation von Datenverkehr als

nicht mehr ausreichend. Die Untersuchung von Nutzdaten, Deep Packet Inspection (DPI), ist ein rechenintensives Unterfangen, welches die Verschlüsselung durch SSL/TLS Interception umgehen kann. Dieses Vorgehen entspricht einem Man-in-the-Middle-Angriff (MITM-Angriff), verletzt die Privatsphäre der Kommunikationsteilnehmer und bricht die Zusage einer Ende-zu-Ende-Verschlüsselung [6].

Statistische Modelle auf Grundlage maschineller Lernverfahren versprechen einen skalierbaren Ansatz, um Anwendungen anhand von Paket- und Verbindungseigenschaften zu identifizieren. Im Rahmen des Projekts MONAT wurde als Teil des Netzwerkmanagements eine prototypische Anwendungserkennung für OPC-UA-Kommunikationsbeziehungen entwickelt, welche den statistischen Ansatz für Nutzdatenverkehr implementiert.

Für die Erzeugung von Paketmitschnitten wurde ein Testaufbau realisiert, welcher verschiedene Anwendungsszenarien im Rahmen der Kommunikation auf Basis von OPC-UA Client/Server und OPC-UA Pub/Sub simuliert. Für die Realisierung der binären Klassifikation war es darüber hinaus notwendig, Paketmitschnitte der negativen Klasse (nicht-OPC-UA-Datenverkehr) zu generieren und einfließen zu lassen.

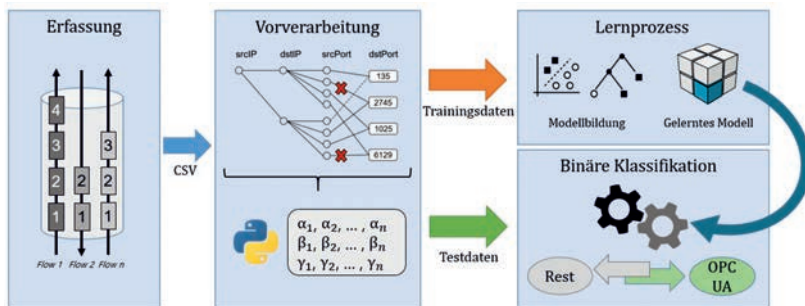


Bild 3: Architektur der OPC-UA-Anwendungserkennung

Bild 3 zeigt den Aufbau der Anwendungserkennung, die im ersten Schritt auf die Erzeugung von Paketmitschnitten beider Klassen angewiesen ist. Dieser wird entweder direkt als CSV-Datei erzeugt oder aus einem beliebigen Format umgewandelt. Im Rahmen der Vorverarbeitung werden für die erfassten Verbindungen auf Basis ihrer Zeitstempel- und Paketeigenschaften Kennzahlen ermittelt, welche den Kommunikationsverlauf abbilden. Zur Erzeugung von

Trainingsdaten ist es im nächsten Schritt notwendig einen Teil der Verbindungen einmalig händisch zu klassifizieren, um die Modellbildung im Rahmen des überwachten Lernprozesses zu ermöglichen. Ausgehend von dem erzeugten Klassifikationsmodell werden neue Verbindungen automatisch klassifiziert. Die Erkennung erfolgt dabei ohne den Bezug auf Portnummern oder die Analyse der Nutzdaten, sondern basiert ausschließlich auf statistischen Paket- und Verbindungseigenschaften. Paketmitschnitte verschiedener Kommunikationsszenarien wurden dazu genutzt, um Kenngrößen zu ermitteln, welche eine Klassifizierung der Verbindungen ermöglichen. Für die (binäre) Klassifikation des Netzwerkverkehrs wurden verschiedene überwachte maschinelle Lernverfahren (ML)¹ trainiert, verglichen und bewertet. Auf Basis dieser Ergebnisse wurde untersucht, welche Merkmale im Rahmen der Klassifizierung geeignet sind und wie die Klassifikationsgüte mit der Paketanzahl korreliert.

So zeigten die angelernten Klassifikatoren signifikante Unterschiede hinsichtlich der benötigten Paketanzahl zum Erzielen vergleichbarer Klassifizierungsergebnisse. Die acht unterschiedlichen Klassifikationsverfahren wurden hinsichtlich ihrer Klassifizierungsgüte angewendet auf die vorliegende Problemstellung miteinander verglichen. Zur Beurteilung der Klassifikationsgüte wurde der *Matthews correlation coefficient (MCC)* herangezogen, da dieser die Sensitivität und die Spezifität eines Klassifikators gleichermaßen berücksichtigt und somit auch für ungleichverteilte Datensätze aussagekräftige Ergebnisse liefert.

¹ Insgesamt wurden acht ML-Verfahren untersucht: Entscheidungsbäume (CART), (naiver) Bayes-Klassifikator, k-nächste Nachbarn (kNN), lineare Diskriminanzanalyse (LDA), logistische Regression (LR), Support Vector Machine (SVM), sowie die Ensemble-Modelle Bootstrap-Aggregation (Bagging) und Gradient Boosting

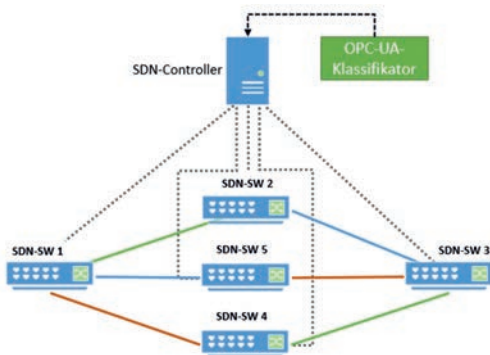


Bild 2: Integrationsvariante der Anwendungserkennung in eine SDN-Umgebung

Während der OPC-UA-Klassifikator im aktuellen Aufbau in der Data Plane positioniert ist, um Netzwerkverkehr zu erfassen und zu klassifizieren, wird perspektivisch die in Bild 4 dargestellte Implementierung als SDN-Applikation angestrebt, um den Prozess dynamischer und granularer zu gestalten und entsprechende Anpassungen im Netz zentral durch den SDN-Controller an die Data Plane zu kommunizieren.

Wird eine OPC-UA-Kommunikationsbeziehung identifiziert und festgestellt, dass die Übertragung aufgrund einer Störung oder Überlastung des genutzten Pfades im Netz nicht die geforderte Dienstgüte gewährleistet, besteht ein möglicher Lösungsansatz in der Umleitung der zugehörigen Pakete. Im folgenden Abschnitt wird für dieses Szenario ein Routing-Mechanismus beschrieben, welcher bei Erkennung von Paketverlust auf einem Übertragungspfad einen störfreien Alternativpfad ermittelt, um den Datenverkehr auf diesen zu verlagern.

4. Umrouting bei Paketverlusten beschrieben in einer SDN-Architektur

Das Engineeringkonzept basiert auf einer mathematischen Modellierung, welche das zeitliche Übertragungsverhalten der Pakete widerspiegelt. Der Ansatz wurde in [7] veröffentlicht. Gewählt wird hier eine ereignisdiskrete Methode mit der max-plus-Algebra, welche auf zeitbehafteten Pertinenzen basiert. Bereits in dem Anwendungsgebiet der drahtlosen Kommunikation wurde diese Methodik für die Sicherstellung eines Koexistenzmanagements angewendet [8, 9]. Das Ziel soll es sein, einen Routingmechanismus zu beschreiben. Zunächst ist es wichtig,

den Übertragungsweg hervorzuheben. In diesem Abschnitt soll eine störfreie Modellierung aufgezeigt werden. Die nachstehende Tabelle gibt eine Übersicht zur Analogie zwischen den Elementen aus dem Kommunikationsnetzwerk und der Petrinetznotation.

Tabelle 1: Analogie der Elemente: Kommunikation-Petrinetz

Kommunikationsnetzwerk	Bild im Petrinetz
Gerät	Platz
Zeitverhalten der Paketübertragung	Zeitbehaftete Transition
Paket	Marke (koleriert)
Eigenschaft einer Nachricht	Eigenschaft einer Marke

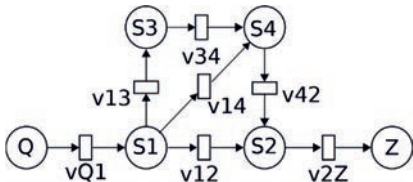


Bild 5: PROFINET-Netzwerk in Petrinetznotation

Bild 5 zeigt den Aufbau in Petrinetznotation auf. Die Switches werden durch die Plätze S1, S2, S3 und S4 beschrieben. Weiterhin sind auch die Plätze Q und Z als Quelle, hier als Beispiel der Controller, und das Ziel, hier als Beispiel der Actor, ausgewiesen. Die zeitbehafteten Transitionen werden durch vQ1, v12, v2Z, v13, v14, v42, v3C, v34 und v3C dargestellt. Der Pfad $Q \rightarrow S1 \rightarrow S2 \rightarrow Z$ ist der nominale Pfad, das heißt die Übertragung ist störfrei.

Die Systembeschreibung wird in der Zustandsraumnotation dargestellt mit

$$\begin{aligned}x(k + 1) &= f(x(k), u(k)) \\ y(k) &= h(x(k), u(k))\end{aligned}$$

Der Zustands- und Ausgabevektor wird mit $x(k)$ und $y(k)$ beschrieben, wobei k das Ereignis einer Paketübertragung widerspiegelt. Die Eingangsgröße wird mit $u(k)$ und die Zustands- und

Ausgabefunktion werden mit $f(\cdot)$ und $h(\cdot)$ dargestellt. Für das störfreie Übertragungsverhalten ergibt sich die nachstehende Notation

$$x(k) = \begin{pmatrix} \varepsilon \\ vQ1v12 \end{pmatrix} u(k)$$

$$y(k) = (\varepsilon \quad v2Z)x(k).$$

Das Verhalten im fehlerfreien Fall kann später für einen Reglerentwurf zur Erkennung und Berechnung eines neuen Pfades genutzt werden. In der nachstehenden Bild 6 ist der Aufbau schematisch beschrieben, der für die Routingmechanismen herangezogen wurde.

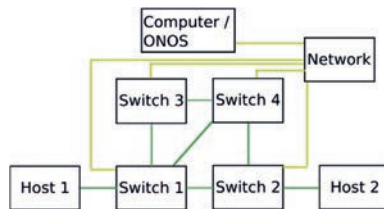


Bild 6: Schematischer Hardware in the Loop -Aufbau

Die Hardwarekomponenten in der Data Plane werden durch Host 1 und Host 2, sowie der einzelnen Switches 1, 2, 3 und 4 beschrieben. Diese Hardware ist durch einen separaten Netzwerkschwitch „Network“ mit dem ONOS Controller und in der Web GUI verbunden. Für die Hosts 1 und 2 werden Raspberry Pis verwendet, auf denen die Nachrichtenpakete erzeugt bzw. gesendet werden. Dies kann auch bidirektional erfolgen. Für die Switches 1, 2, 3 und 4 werden ZedBoards als Entwicklungsboards verwendet. Der Aufbau ist dargestellt in Bild 7.

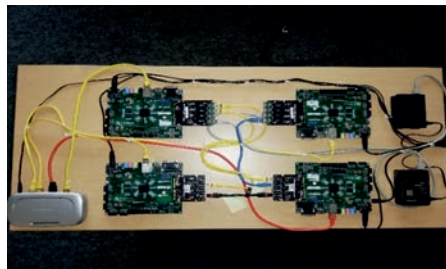


Bild 7: Hardware in the Loop -Aufbau für das SDN-Netzwerk

5. Zusammenfassung

Dieser Beitrag zeigt im Wesentlichen die Ergebnisse des Forschungsprojektes MONAT - Modellbasierte und bedarfsgerechte Netzwerkkonfiguration für Netzwerke der Automatisierung und Telekommunikation, Förderkennzeichen: 16KIS0779, auf. Im ersten Schritt wurden dazu die verschiedenen Mechanismen der OPC-UA-Anwendung aufgezeigt. Zum einen werden Mechanismen zur Informationsmodellierung vorgegeben, die eine semantische Selbstbeschreibung der Daten und damit einen Informationsaustausch ermöglichen. Zum anderen wird eine standardisierte Kommunikationsplattform bereitgestellt, die eine sichere, zuverlässige und herstellerunabhängige Kommunikation bietet. Weiterhin wird eine OPC-UA-Anwendungserkennung vorgestellt, welche auf Machine Learning basiert. Hierbei kamen verschiedene Verfahren zum Einsatz und wurden erprobt. Bezogen auf Kommunikation im Rahmen von OPC-UA Client/Server und der vorläufigen Betrachtung von OPC-UA Pub/Sub erzielten die Verfahren k-nächste Nachbarn (kNN) und Support Vector Machine (SVM) die besten Ergebnisse. Weiterhin wurde an einem Routingmechanismus gearbeitet, welche mit Hilfe der Petrinetznotation beschrieben werden kann. Der Ansatz beschreibt zunächst den störfreien Fall, der im späteren Verlauf auf einer Regelung erweitert werden kann.

Referenzen

- [1] W. Mahnke: Informationsmodellierung mit OPC Unified Architecture. In atp magazin, 03 2020, Vulkan-Verlag, 2020
- [2] W. Mahnke, S.-H. Leitner, M. Damm: OPC Unified Architecture, Springer-Verlag, 2009
- [3] Unified Architecture: C++ Based OPC UA Client/Server/PubSub SDK Documentation, <http://documentation.unified-automation.com/uasdkcpp/1.7.1/html/index.html>, 2020
- [4] J. Neubauer: Entwicklung einer Methode zur Erkennung von OPC-UA-Kommunikationsbeziehungen. Masterarbeit, Magdeburg, Hochschule Magdeburg-Stendal 2019
- [5] J. Ohms, M. Böhm, D. Wermser, M. Meier, D. Schulze-Zipper, M. Riedl: Flexible Kommunikation von TSN für industrielle Kommunikationsnetze. In AUTOMATION 2019, VDI Verlag GmbH, 2019
- [6] Bakhshi , Taimur ; Ghita , Bogdan: On Internet Traffic Classification: A Two-Phased Machine Learning Approach. 2016
- [7] J. Ohms, M. Böhm, D. Wermser, M. Meier, D. Schulze, M. Riedl: Testen von Konfigurationen in industriellen Netzen mit zentraler Koordinierungsinstanz. KommA 2019 – 10. Jahreskolloquium „Kommunikation in der Automation“, Magdeburg
- [8] D. Schulze, L. Rauchhaupt, M. Kraetzig und U. Jumar: Coexistence Plant Model for an Automated Coexistence Management. In: 20th IFAC World Congress (WC'17), IFAC-PapersOnLine, Band 50, Seiten 355-362, Toulouse, Frankreich, 9. - 14. Juli 2017.
- [9] Schulze, D. und H. Zipper: A Decentralised Control Algorithm for an Automated Coexistence Management. In: 57th IEEE Conference on Decision and Control (CDC'18), Seiten 4187–4193, Miami Beach, USA, 17. - 19. Dezember 2018.

Increased reliability and safety in Coriolis flowmeter operation by monitoring its physical response to process conditions

Using the Heartbeat Sensor Integrity (HBSI) parameter to monitor sensor wear and predicting maintenance needs

Dr. **Daniel Persson**,

Endress+Hauser Group Services, Reinach, Switzerland;

Dr. **Johan Pohl**,

Endress+Hauser Flow, Reinach, Switzerland;

Dr. **Alfred Rieder**, Dr. **Michael Wiesmann**,

Endress+Hauser Flow, Freising

Abstract

The Heartbeat Sensor Integrity (HBSI) parameter is available on all Endress+Hauser Coriolis flowmeters featuring Heartbeat Technology. This paper describes how the HBSI parameter is derived to monitor the Coriolis sensor's integrity and how it can be used to provide increased reliability and safety in operations by enabling early detection of any corrosive or abrasive wear mechanisms in the process. Three use cases are presented to demonstrate the capabilities of the HBSI parameter and its potential to increase operational safety and efficiency as well as reducing operational expenses by providing insights on the application and device performance, thus enabling the user to predict maintenance needs and plan maintenance activities accordingly.

1. Introduction

During the last decades, Coriolis flowmeters have become a preferred measurement technology choice for numerous industries and applications since they measure mass flow with an unmatched accuracy and repeatability. In addition, they also reliably measure a wide variety of other process parameters such as volume flow, density, temperature, concentration and viscosity [1, 2]. All process parameters are derived by evaluating specific Coriolis sensor signals which are generated from the physical response of the sensor to the prevailing process conditions. The main sensor components involved are the measuring tubes, the exciter and the pickup sensors (see Fig. 1).

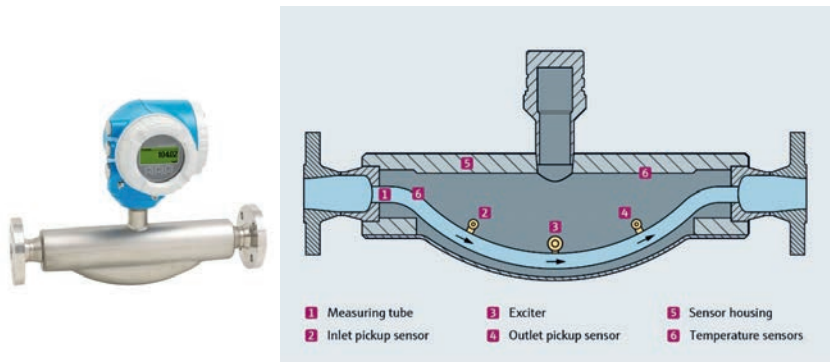


Fig. 1: Left: Promass F 300 Coriolis flowmeter consisting of a transmitter (blue) and a sensor (metallic). Right: Schematic cross section of the sensor and its key components. Parts in contact with the process fluid are marked in light blue.

The sensor housing is also essential for establishing a reliable measurement, as it provides a stable and dedicated environment for the sensor components to optimally perform their measuring task. In addition, the responses of the integrated temperature sensors allow reliable determination of the thermal load distribution in the Coriolis sensor. This enables an accurate compensation of the measurement parameters in case of any temperature effects. The signals generated by the sensor components are sent to the transmitter where they are processed to derive all measurement parameters, which subsequently are passed on to the process control system via the desired communication protocol.

If a Coriolis flowmeter is operated within its intended range of use and its wetted parts are chemically compatible with the fluid, then its measuring performance is expected to be very stable over time, as process conditions will have a minimal influence on the sensor components. However, for particularly challenging process conditions some impact may prevail. For example, an unforeseen presence of abrasive particles in the fluid or an increase in flow rate, where abrasive particles are already present, could trigger abrasion of the sensor's measuring tubes. Other potential challenges include unforeseen changes of the fluid composition or its concentration, an unexpected process temperature increase, or the use of a more aggressive cleaning agent. In such cases, the sensor's wetted parts may no longer be compatible with the process and corrosion could occur, thereby jeopardizing both the performance and reliability of the device. If preventive measures are not taken in a timely manner, even the mechanical integrity of the device could be at risk.

This paper illustrates how various physical responses of the Coriolis sensor have been combined to a dedicated parameter which reliably identifies deterioration in sensor integrity. This parameter is called Heartbeat Sensor Integrity (HBSI). Practical use cases demonstrate how this technology can be applied to increase operational safety and reliability under challenging abrasive and corrosive process conditions. The HBSI parameter provides the user a direct “window” into the process, enabling detection of harmful process conditions. By monitoring the HBSI parameter over time, insights are created allowing the user to plan and optimize maintenance actions efficiently, resulting in cost savings and increased productivity.

2. Heartbeat Technology

The HBSI parameter is available on the majority of Endress+Hauser's Coriolis flowmeters as a part of the Heartbeat Technology offer. Also available for other flowmeters as well as for level, pressure, temperature and analytical measuring devices, Heartbeat Technology represents an integral part of Endress+Hauser's design concept for process instrumentation diagnostics, aiming to improve the plant availability while reducing operative expenses by providing the user with an optimal balance of excellent measurement functionality, reliability, safety and ease of use. The foundation of all devices featuring Heartbeat Technology is the best-in-class diagnostic coverage which minimizes undetected device failures and consequently provides a high level of reliability. Built-in verification and monitoring functionalities can additionally be selected via a configurable option. These allow the user to assess device functionality, improve process performance and provide insights, enabling predictive maintenance in a reliable and comfortable manner. Consequently, all functionality related to the HBSI parameter requires either that the corresponding Heartbeat Technology option is selected when ordering the device, or that it is activated later in the life cycle by contacting Endress+Hauser.

3. The Heartbeat Sensor Integrity (HBSI) parameter

The HBSI parameter reliably identifies deterioration in integrity of the sensor components (see Fig. 1), where the condition of the measuring tubes is a key contributing factor. If the process fluid exposes them to a wear mechanism like corrosion or abrasion, then their elastic properties will change due to the corresponding reduction in wall thickness. This is also likely to impact the measurement accuracy for density and mass flow by way of overreading due to the reduced measuring tube stiffness caused by the wear mechanism. If not acted upon in a timely manner, a severe wear mechanism could over time lead to device failure and potentially process fluid escaping from the measuring tube. In such cases, the process fluid will divert into the sensor housing, where it will be contained. By evaluating the HBSI parameter, the user

can confirm if a wear mechanism is present in the process or not, resulting in either a high confidence in the sensor's integrity and measurement performance or allowing for maintenance actions to be planned accordingly.

A Coriolis flowmeter measures mass flow by exciting the main oscillation mode of its measuring tubes as well as the contained process fluid at the resonance frequency f_0 (see Fig. 2) and thereafter evaluating the phase difference between the signals of the inlet and outlet pickup sensors to generate the mass flow value. To determine the HBSI parameter, the tube oscillation is excited at a higher frequency f_{HBSI} (see Fig. 2) with a constant current, and the resulting oscillation amplitude is evaluated to generate the HBSI value [3]. The frequency f_{HBSI} is kept fixed relative to the frequency of the main mode f_0 . Consequently, if f_0 increases or decreases due to a change in fluid properties, so will f_{HBSI} .

There are several advantages in using a single relative frequency for the determination of sensor integrity. Firstly, both f_0 and f_{HBSI} will shift when the properties of the process fluid changes, making the HBSI parameter independent of the prevailing process conditions as it is always derived at a higher frequency compared to the main mode f_0 . Secondly, the influence from the process fluid's viscosity on the measuring tube oscillation is negligible at frequencies distant from the main resonance mode f_0 , making the HBSI parameter insensitive for any change in viscosity of the process fluid. Thirdly, the use of a single frequency avoids the introduction of numerical noise that may be introduced by fitting algorithms required in a multiple frequency approach. Finally, determination of sensor integrity can be performed at frequent time intervals in a resource constraint system.

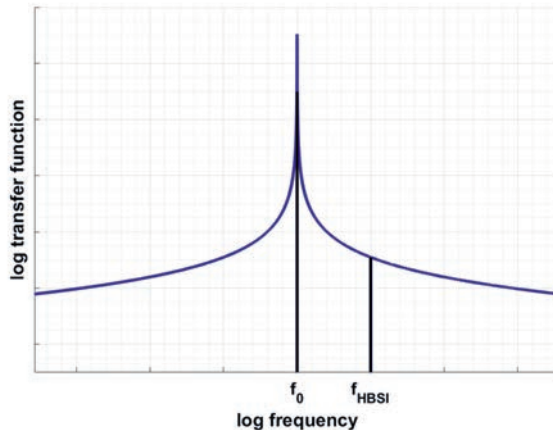


Fig. 2: Logarithmic sensor transfer function of a Coriolis flowmeter featuring Heartbeat Technology. The main oscillation mode of the measuring tubes and the contained process fluid is occurring at the resonance frequency f_0 . For determining the HBSI parameter, an additional higher frequency f_{HBSI} is excited at a fixed frequency ratio relative to f_0 .

A new Endress+Hauser Coriolis flowmeter has a HBSI value of 0% at reference conditions. For most applications, the device will not be exposed to any wear mechanism and the HBSI value will therefore remain close to zero and the measurement accuracy will meet specifications. Consequently, the user can expect reliable long-term performance from the Coriolis flowmeter if operated within its given specifications. However, if there is a wear mechanism present, the sensor integrity will deteriorate over time and the HBSI value will increase accordingly. The most common wear mechanisms are corrosion or abrasion caused by an aggressive process fluid.

Even though the effects of various process conditions on the HBSI value are excluded by design or compensated, a slight residual influence on the HBSI value may remain. When the device is operating close to reference conditions, the residual effect on the HBSI value is negligible, but if process conditions are unstable, an influence on the HBSI value cannot be excluded. Consequently, to securely identify the presence of a wear mechanism in the process, the HBSI value should be observed over time. This will enable the user to determine if the HBSI is just affected by varying process conditions, or if the HBSI value is gradually increasing due to the presence of a wear mechanism.

The HBSI parameter is a reliable indicator of sensor wear, but it will not provide a generic relationship between its value and the expected impact on measurement accuracy which fits all Coriolis flowmeter models and applications, as various wear patterns will impact the elastic properties of the measuring tubes differently. Consequently, also the oscillations at f_0 (used to derive mass flow and density) and f_{HBSI} (used to derive the HBSI value) will react differently.

However, if the process conditions and corresponding wear mechanism remain constant, a relationship between the HBSI value and the measurement performance can be established, if the effort is made to calibrate the meter at various HBSI levels. Such built-up experience can also be used to predict the behavior of a replacement Coriolis flowmeter of the same type if the application conditions are kept constant. Fig. 3 shows the correlation between density and mass flow measurement accuracy and the HBSI value for three different wear conditions.

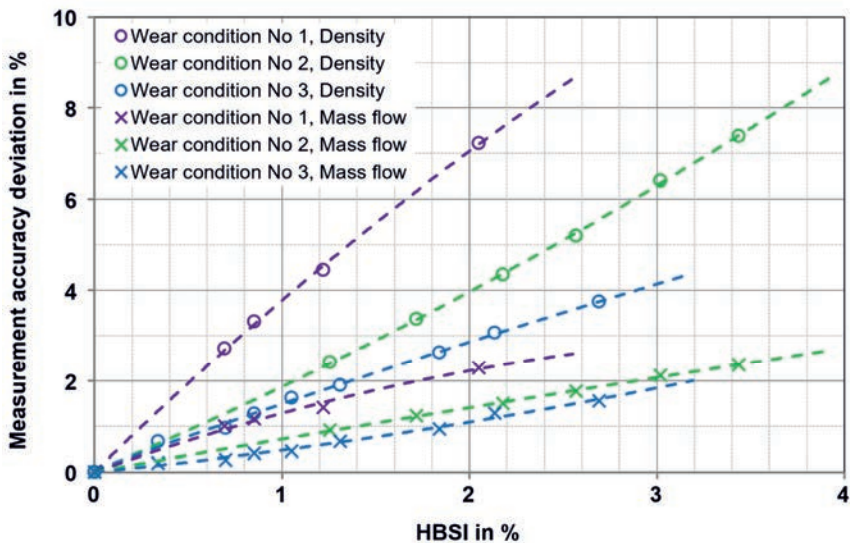


Fig. 3: Correlation of density (o) and mass flow (x) measurement accuracy deviation and the HBSI value for three different wear tests on Coriolis flowmeters in a laboratory environment.

When a Heartbeat Verification is performed, the current HBSI value will be included in the Heartbeat Verification report. Thus, a series of Heartbeat Verifications over time will provide information about the development of the HBSI value.

The user can also schedule the HBSI parameter to be automatically generated in certain time intervals based on application needs, e.g. once per day. For the single straight-tube Promass I, which generates the HBSI signal from the always active torsional frequency of the measuring tube [4], the HBSI value can even be measured continuously. The HBSI parameter can be monitored for both methods by configuring it to an analog output or reading it via a digital protocol.

No process interruption is required to determine the HBSI value as it is generated by the device itself. The primary measurement variables also remain continuously available. Thus, obvious efficiency gains and risk reduction are achieved compared to removing the device for inspection or calibration.

When the HBSI value exceeds 4% during a Heartbeat Verification, the HBSI section of the verification report as well as the overall verification result will be marked as failed. This indicates, with a very high level of confidence, that the meter is no longer fit for purpose and is recommended to be replaced. The user can also choose to perform certain maintenance actions at specific HBSI values, e.g. based on any previous application experience or requirements. The remaining time until a certain HBSI level is expected to be reached can be predicted if the HBSI value is monitored over time and the application conditions remain constant. However, the more common use case is that the HBSI value is monitored to confirm the absence of a wear mechanism in the process. In that case the HBSI value will remain at a stable low level.

The following chapter illustrates three typical use cases. These demonstrate how the HBSI parameter can provide insights on the application and device performance in real applications.

4. Use cases

In the first use case (Chapter 4.1), the user wanted to make sure that the previously experienced wear mechanism was no longer present in his application. In the next use case (Chapter 4.2), the HBSI parameter was monitored in an application with significant wear to predict maintenance needs. In the final use case (Chapter 4.3), the HBSI parameter enabled detection

of an unexpected wear mechanism in the application, allowing the user to apply corrective actions to avoid the wear mechanism. The effect of this optimization was also confirmed by the continued monitoring of the HBSI parameter.

4.1 Proving the absence of a wear mechanism

A user had an application containing a potentially abrasive fine slurry. The initial Coriolis flowmeter used showed clear signs of abrasion and failed after only six months in service. The abrasion was evident both from a drift in density measurement, calibration results and visual inspection. After consultation, the initial meter which featured a bent dual-tube design was replaced with a single straight-tube Coriolis flowmeter as the latter measuring tube geometry often provides good resistance to abrasion. However, based on the rapid deterioration of the initial meter, the user was concerned that abrasion might still take place. As this would cause him another process interruption and a potential safety issue, the flowmeter was equipped with the Heartbeat Technology option, enabling the HBSI parameter to be continuously monitored to confirm that no wear mechanism was present. At the time of writing, the meter had been used for more than four years and is still in operation. During this time, the HBSI parameter showed a flat-line behavior (see Fig. 4) and the user has regained full confidence in the measuring point.

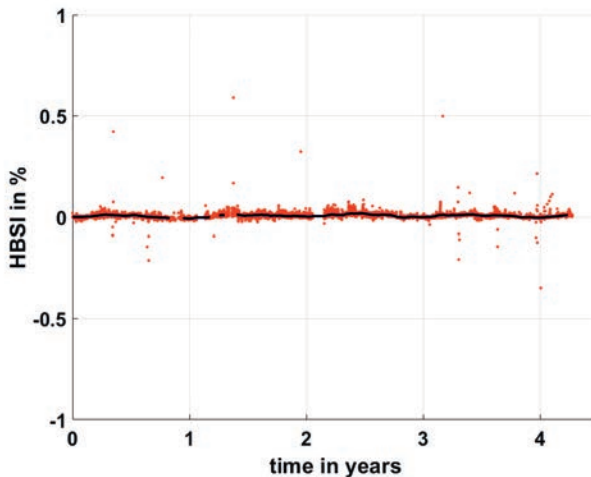


Fig. 4: The HBSI parameter shows a flat-line behavior for more than four years in a potentially abrasive application, thereby indicating that no abrasion has taken place.

Additional confirmation of the sensor integrity and measurement accuracy was obtained by analyzing the reduced data set from the in-line cleaning process. The process line is in regular intervals flushed with water for cleaning. As any significant wear mechanism would negatively impact the density measurement, the measured density of the cleaning water provides a good indication if the meter is still measuring accurately. Fig. 5 shows both the density and the temperature of the cleaning water as measured by the Coriolis flowmeter. While the density showed a certain seasonal dependence due to temperature variations, no long-term drift in the density data was observed.

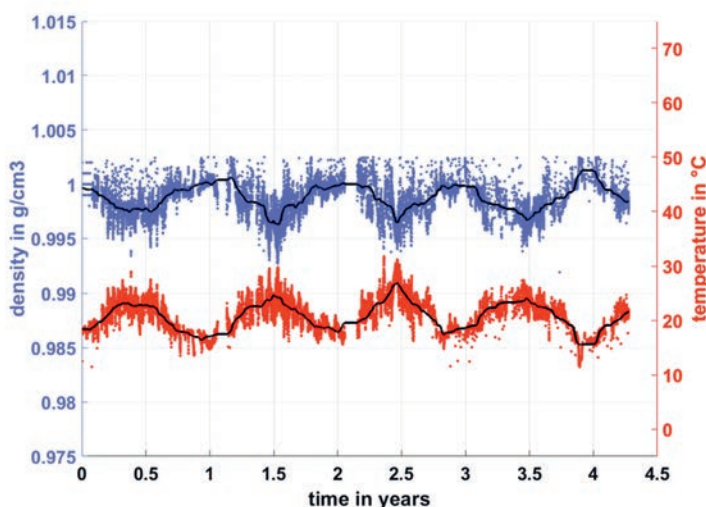


Fig. 5: Measured density (top) and temperature (bottom) of the water used for the cleaning process. The measured density average (black line) shows a clear seasonal dependence, but no trend which would indicate the presence of abrasion.

The data confirms that the meter is still measuring density according to its specification, and since both the density and the mass flow measurement accuracy are related to the HBSI parameter, it also provides a strong indication that the mass flow is measured accurately. The increased confidence made it possible for the user to optimize his calibration intervals, which previously had to be scheduled on a frequent basis, as no additional insights into the sensor condition were available. Thus, calibration efforts and connected costs were reduced, and the process availability increased.

4.2 Predicting maintenance needs in presence of a wear mechanism

A user had an application with a highly abrasive coarse slurry which was flowing with high velocity through the flowmeter (up to 8 m/s). Even though the user knew the application was challenging and that abrasion was likely to occur regardless of which flowmeter was selected, he also needed to measure both flow rate and density as accurately as possible. In addition, he expressed a wish to be able to predict the point in time when the meter would need replacement. Considering all the above, a Promass Q Coriolis flowmeter was selected and installed as seen in the left picture of Fig. 6. This sensor both provides the best mass flow and density measurement performance currently on the market and features the Heartbeat Technology option, enabling use of the HBSI parameter.



Fig. 6: Left: The installed Promass Q Coriolis flowmeter. Right: Optical investigation of the interior of one of the flowmeter's measuring tubes after 11 months in operation shows clear signs of abrasion.

As the plot in Fig. 7 reveals, an increase in the HBSI value was observed shortly after operations started, indicating the presence of abrasion. After tracking the HBSI for the first half year a clear linear trend was established. After 11 months the flowmeter was removed for investigation and the abrasion identified by the HBSI parameter was confirmed by the optical investigations (see Fig. 6, right picture). However, the abrasive process had not removed so much material that the overall sensor integrity would be jeopardized, and the measurement performance was also within the acceptable range for the user. Consequently, the flowmeter did not

need to be replaced and was re-installed. The user wished to perform another evaluation of the device when it reached a HBSI value of 3%. As seen in Fig. 7, the linear trend of the HBSI value easily allowed the user to predict that this level was expected to be reached after a total usage time of 18 months, as the application conditions would remain constant. In this manner he could predict and plan his upcoming maintenance need and avoid previously recurring device checks and the associated downtime.

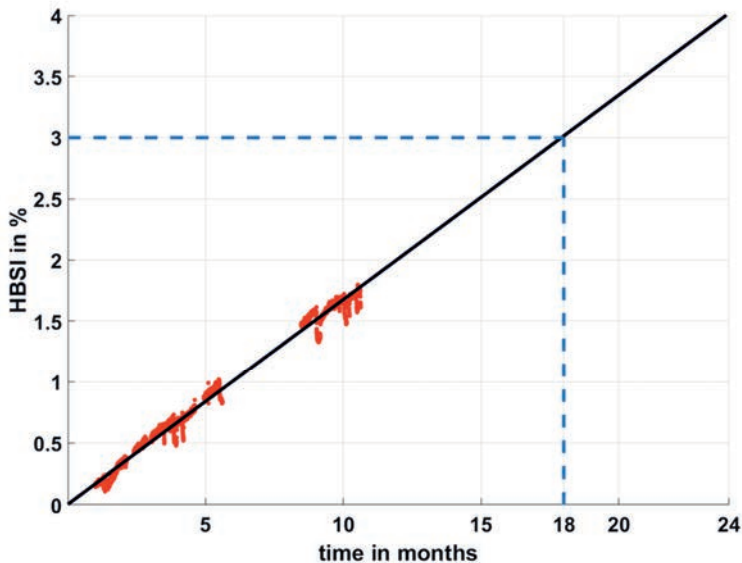


Fig. 7: The black line indicates the linear increase of the measured HBSI value (red dots) caused by continuous abrasion of the measuring tubes. The blue dashed lines indicate when the HBSI value is predicted to reach 3%.

4.3 Detecting an unexpected wear mechanism

A Coriolis flowmeter conFig.d with the Heartbeat Technology option was used in an application where no wear mechanism was expected to occur under normal operation. To confirm this and consequently increase his confidence in the application performance, the user continuously monitored the HBSI parameter.

At first, the HBSI value remained at its initial low level as expected, indicating that no deteriorating mechanism was present in the application. However, after a while the user observed a slight drift of the HBSI parameter over time. When analyzing the matter in detail, it could be established that the drift in HBSI started at the same time as a new cleaning agent was used in the weekly occurring cleaning process.

When looking at the data in even higher resolution, it could be established that the HBSI value was stable during normal operation, but each cleaning cycle increased the HBSI value with 0.1%, producing a step-like pattern over time (see Fig. 8). Thus, providing a clear indication that the slight wear of the sensor was only occurring during the cleaning process and not in normal operation.

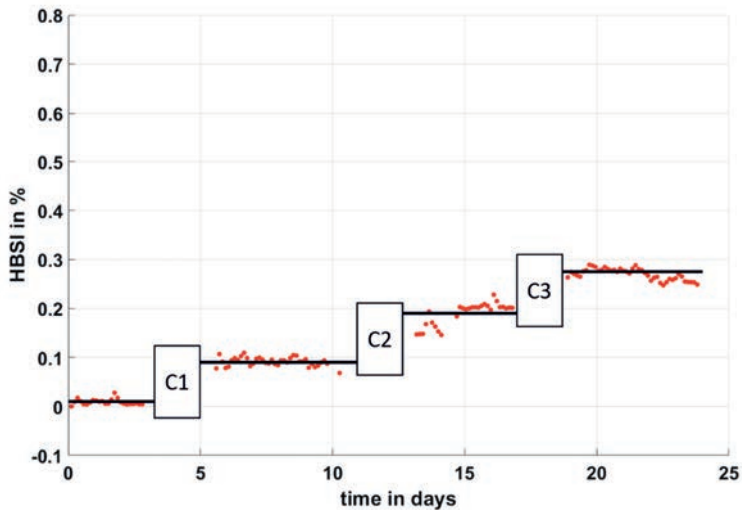


Fig. 8: Monitoring of the HBSI parameter during normal operation at constant process conditions (red dots with black trend lines). The labels C1, C2 and C3 indicate when the cleaning process was performed.

When comparing the composition of the new cleaning agent with non-corrosive cleaning agents, the reason for the corrosion could be pinpointed to a specific component in the new cleaning agent that was not chemically compatible to the flowmeter measuring tube material. This was also confirmed in laboratory tests (see Fig. 9).

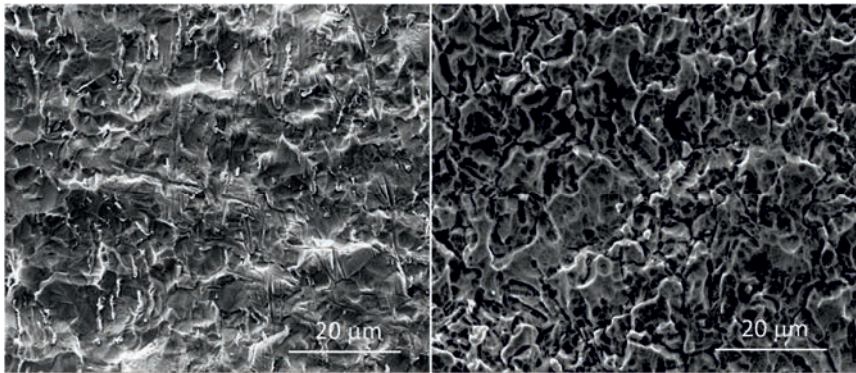


Fig. 9: Scanning Electron Microscope (SEM) images of the interior surface of a measuring tube before (left) and after (right) being exposed to the aggressive, non-compatible cleaning agent for a duration equivalent to 32 cleaning cycles. The same chemicals, concentrations, and temperature as in the user's process were used in the laboratory tests. The corrosive effect on the measuring tube can clearly be identified when comparing the original (left) with the exposed (right) surface.

Based on the results of these investigations, the user adapted his cleaning process and significantly extended the device lifetime by stopping the corrosion. The monitoring of the HBSI parameter continues as it increases the confidence in the application performance and consequently reduces the maintenance effort as the device does not need to be checked as frequently.

5. Conclusions

It has been demonstrated how the Heartbeat Sensor Integrity (HBSI) parameter, available on all Endress+Hauser Coriolis flowmeters featuring Heartbeat Technology, can be monitored to enable early detection of any corrosive or abrasive wear mechanisms in the process. This capability provides the user insight into the process behavior and increases his confidence in the device and application performance. Operational safety and efficiency are improved as process downtime due to unpredicted sensor wear can be avoided. Finally, the HBSI parameter may also help to reduce operational expenses, since maintenance needs can be predicted and maintenance activities planned accordingly.

References

- [1] Baker R.C. (2016): Flow Measurement Handbook, "Industrial Designs, Operating Principles, Performance, and Applications", 2nd Edition, 560–602, Cambridge University Press.
- [2] Anklin M., Drahm W., Rieder A. (2006): Flow Measurement and Instrumentation, "Coriolis mass flowmeters: Overview of the current state of the art and latest research", Volume 17, Issue 6, 317–323.
- [3] Patent EP 2638367 B1 (2020): "Measuring system having a vibration-type measuring transducer", Endress+Hauser Flowtec AG.
- [4] Patent EP 1938052 B1 (2020): "In-line measuring device and method for monitoring changes of a pipe wall", Endress+Hauser Flowtec AG.

Selbstorganisation und Co-Management verteilter Automatisierungssysteme mit dem Kommunikationssystem

Self-Organization and Co-Management of Distributed Automation Systems with the Communication System

Dipl.-Ing. **T. Zeun**, Dr.-Ing. **A. Frotzscher**,
Fraunhofer Institut für Integrierte Schaltungen IIS,
Institutsteil Entwicklung Adaptiver Systeme EAS, Dresden;
A. Nicklas, M.Sc.,
Lehrstuhl für Robotik und Systemintelligenz, Technische Universität
München;
E. Schmitt, M.Sc.,
Vodafone Chair Mobile Communications Systems, Technische
Universität Dresden;
Dr.-Ing. **D. Schulze-Zipper**, Dipl.-Ing. **H. Zipper**, ifak e.V., Magdeburg

Kurzfassung

Verteilte Automatisierungssysteme werden mit steigender Anzahl von Subsystemen immer komplexer. Die zentrale Steuerung in einem solchen verteilten System stellt einen Single-Point-of-Failure dar und das zumeist drahtlose Kommunikationssystem wird bei schwankenden Anforderungen der Subsysteme zeitweise überlastet. Zur Lösung schlagen die Autoren vor, Methoden des fähigkeitenbasierten Engineerings zur Selbstorganisation und Kollaboration der verteilten Anwendungen und ein Co-Management der verteilten Anwendungen mit dem Kommunikationssystem zu nutzen. In diesem Beitrag werden ein dafür zugeschnittenes Systemarchitekturmodell und dessen Umsetzung in einer Systemsimulation mithilfe des Robot Operating System (ROS) vorgestellt.

Abstract

Distributed automation systems become more and more complex with a growing number of subsystems. A central controlling in such a system is a single point of failure and the mostly used wireless communication systems may be overloaded by time-varying demands of the subsystems. To overcome these problems, the authors propose to use methods of skill-based

engineering for self-organization and collaboration together with a co-management between the distributed application and the communication system. In this paper a system architecture model and its implementation in a system simulation using the Robot Operating System (ROS) is presented.

1. Introduction

Distributed control systems are already in use in the industry in several business-critical applications. Automotive manufacturers use automated guided vehicles (AGVs) to supply their assembly lines with material or even to transport the car bodies between assembly isles [1]. Warehouses mobile robots collaboratively pick goods in three-dimensional grid systems [2] or transport mobile shelves to the workplaces of logistics employees. Collaborative robots are used in assembly processes. In all cases, complex process tasks are executed by numerous cooperating or collaborating subsystems. Currently, the distributed mobile subsystems are mainly controlled by a central or at least subordinate local management. This system approach entails two main problems: First, the reliability of the overall system depends strongly on the communication network, especially the wireless segment for connecting the mobile assets. Second, as the number of collaborating assets increases, the complexity of the software modules in the centralized system control become unmanageable, representing a single point of failure. This limits the achievable degree of flexibility and versatility of the system. To overcome these, the combination of two main concepts is proposed:

- 1) Self-organization and collaboration of production assets: The production assets organize themselves to execute orders from the Enterprise Resource Planning (ERP) level, compensate for failures of individual assets and identify and exploit new collaboration potentials between each other.
- 2) Co-design and co-management of the communication network and distributed control application: The management units of both domains collaborate closely together. The network configuration is optimized to the demand of the control application, using its own degree of flexibility to adopt its demands to the current network performance capabilities.

2. State of the Art

To extend the current usage of distributed automation systems, recently research is conducted towards the mapping of requirements of industrial control application to distributed systems as well as the behavior of networked control systems. Currently, industry 4.0 related standardization is a big driver towards the application of distributed control systems in industry.

With the proposals from [3] and [4] distributed control systems are designed for flexible and adaptive production. Communication protocols used in industrial environments evolve from protocols for which the engineer needs to know the meaning of every bit to semantic protocols which describe tasks and assets based on their requirements, properties and skills. In [5], such a protocol using an ontology information model is shown. In [6] semantic protocols for skill-based control engineering of robots have been shown. Ontologies are used to describe the data such as tasks, grippers and 3D geometries. BaSys 4.0 [7] aims to be the operating system for industry 4.0. Its main goals are the standardization of various aspects related to industry 4.0 such as the asset administration shell for data exchange and device representation, ontologies to describe skills and capabilities as well as a common communication infrastructure. To overcome the limitations of using central controllers Networked Distributed Control Systems (NDCSS) were developed that consist of many subsystems that are controlled by local controllers individually. While interacting over the plant, information on the states and control signals are exchanged between the subsystems through a communication network. Recent contributions in science literature contain the stabilization of NDCSS by adaptive control schemes in the presence of delays [8] or determine the addition of an interaction estimator in each subsystem that consist of models of the other subsystems to overcome delays and packet dropouts [9]. Meanwhile, the design of NDCSS that simultaneously take network and controller design into account remains an open field [10]. Within the TACNET 4.0 project, a communications architecture addressing the needs of future industrial IoT networks was developed [11]. Based on 5G, the detailed architecture proposes several improvements over conventional networks including radio technology management to enable the parallel usage of heterogeneous technologies, novel self-organization functionalities and multi-domain management and orchestration. However, the emphasis of TACNET 4.0 was open interface design and ultra-reliable low latency communications (URLLC) in industrial use cases. This paper aims for a paradigm change from application agnostic network adaptation through URLLC to application reliability. Therefore, a novel industrial architecture model is proposed that enables a full co-design and co-management between communication system and control application as well as advanced self-organization capabilities in both domains.

3. System Architecture Model

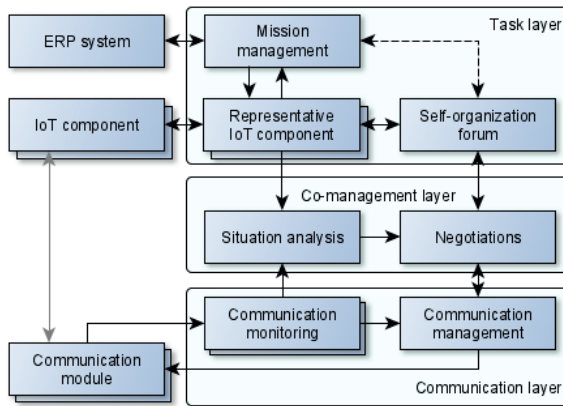


Fig. 1: System Architecture Model

The proposed architecture model as shown in Fig. 1 consists of three layers: task layer, co-management layer and communication layer, representing the management entities. The components on the left side of Fig. 1 represent the physical devices or modules of state of the art systems. Each IoT component contains a communication module. These modules provide a communication protocol stack from OSI layer one to four.

The mission management receives a high-level task from the ERP system and divides it into subtasks which may have dependencies on each other. It offers the subtasks to the representatives of the IoT components and collects the status updates for each subtask, e.g. “in progress” or “completed”. The mission management itself gives status updates about the mission to the ERP system. Each IoT component has its own representative in the task layer comparable to the administration shell in industry 4.0 [12]. The representative gets a subtasks from the mission management and gives status updates about its subtask. It divides the subtask into steps according to the skills of the IoT component while the IoT component does the actual controlling. The representatives also compute quality parameters about controlling and collaboration and provide them to the situation analysis in the co-management layer. The self-organization forum is the platform where the representatives coordinate the execution of their subtasks with each other. They coordinate movements to avoid conflicts and collisions and can further search for partners if an IoT component lacks certain skills needed in its subtask, e.g. an AGV delivering goods needs help of a robot arm for unloading to the line store.

The communication monitoring collects relevant information about the communication, e.g. link packet loss rates or link capacity from the communication modules. Based on this information, the monitoring computes quality parameters and provides them to the communication management and the situation analysis in the co-management layer. The communication management evaluates the provided information, may derive improvement possibilities and forward them to the negotiation module. The result of the negotiations is then reassembled by the communication management to initiate specific changes by the different communication modules.

In the co-management layer the situation analysis collects the quality parameters from the communication monitoring and those from the representatives and creates an overview about the current situation. If one or more parameters exceed their specified limits, it triggers negotiations between the task layer and the communication layer. The negotiation module then tries to find a solution taking into account the information provided by the communication management and the self-organization forum. One solution could be scheduling more communication resources or redundant paths to the affected link. Another one could be changing the degree of collaboration to a less collaborative scheme with less communication effort.

4. Self-Organization

Based on the assessment of the state of the art from Section 2, the overall design goals for self-organization are:

- Skill-based workflow for the engineering of the control task
- Properties represent parameters to skills
- Cooperation or even collaboration between multiple entities in the distributed system is organized based on negotiation

The control program of the plant will not be static. Instead, a control program will be created ad-hoc based on the required tasks and available resources such as manufacturing equipment and communication capacity. Also, the control program can change whenever errors happen or changes in the aforementioned resources occur. Fig. 2 shows our approach to self-organization.

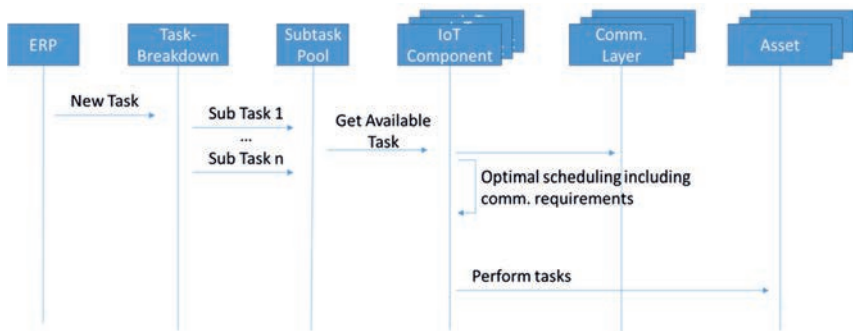


Fig. 2: Self-organization approach

Based on the description above, we implement the following strategy to realize the self-organization:

- A domain specific algorithm breaks down tasks from ERP into sub tasks
- Each IoT component within the system may handle an open sub task.
- We use the concept of capabilities and skills as basis for self-organization. The input for the self-organization algorithm may come from the asset administration shells of industry 4.0 [12]. Because the industry 4.0 concept utilizes ontologies, manufacturer specific skill and capability definitions may be supported.
- Depending on the requirements of the sub task, the handling IoT component asks other components in the system for skills to fulfil this sub task.
- From the skill, capability and sub task definitions, self-organization is stated as an optimization problem based on [13]. Therefore, an SMT (satisfiability modulo theories) is stated that allows not only to find the most efficient scheduling for one sub task, but also allows to include communication related properties in the cost function.
- The optimal scheduling calculated from the SMT problem results in a list of jobs assigned to the specific components.

5. Co-Design and Co-Management

A main goal of the research work is the co-design and co-management between the control applications and the communications network. This central idea of adding a separate co-management layer in the architecture is presented in section 3. In this regard, co-design means the offline methodology that allows the adaptation of control and communications parameters per design, whereas co-management is used during operation of the system and continuously

optimizes the control strategy for the parameters explained in the following. Co-design aims to answer several questions: Which parameters can be modified in control and communications? Which are suitable performance indices? What are the error bounds that can be tolerated? To implement the co-design and co-management approaches, first, suitable performance parameters to determine the Quality of Service (QoS) from the communications layer and the Quality of Control (QoC) from the control side need to be defined. Packet error rate, transmission time, and modulation and coding scheme (MCS) have been identified as suitable QoS parameters and control deviation/tracking error as well as several occupancy parameters as busy-to-waiting-time ratios for QoC. The aforementioned parameters are determined on a wireless link and device level, respectively. To obtain an impression of the current system state and evaluate how far its operation point is from the capacity bounds, the local information per device is gathered and analytically fused to a low-dimensional parameter set. The result serves as an input to co-management strategies.

Co-design includes the implementation of possibilities for interventions in the communications and application execution. On the communications side, assuming a cellular network architecture, this can be done via the scheduler located at the base station. The use of a scheduler enables an intelligent assignment of radio resource blocks to the different links and thus services of IoT-devices, based on the predicted states of the different physical wireless radio channels and the states of the underlying applications that are serviced through the communications system in a joint policy. The latter, called "State-Aware Resource Allocation" [14], assumes that the application can tolerate a certain number of packet losses before shutting down. Fading and interference prediction of future states of wireless radio channels [15] helps to avoid radio channels that are in outage and distribute the available channels in the best possible manner. The used fading prediction relies on appropriate modeling of the present fading process. Measurement campaigns have deepened the understanding of fading processes in industrial environments as factory halls [16]. Further, the reliability could be increased using routing techniques in the network to avoid channels that exhibit a low SINR for a long time based on spatial characteristics. The presented methods aim to increase the efficiency of the inter-process communications tailored to the needs of the control application.

In addition, similar methods are employed to adapt the control application's demands to the opportunities of the communications system. This includes the adaptation of process speed, the number of active IoT-devices, or management of conflict zones in cooperation scenarios. The objective of the co-management is to balance the different interventions and drive the system to the best possible operation point. In order to design a suitable optimization function for this, approaches from control engineering are examined [17].

In contrast to co-design, co-management is using the intervention possibilities by the co-design during the runtime of the system when necessary. Measurements and preprocessed information variables from the control application and the communications network are collected periodically and evaluated on an event-based and time basis to gain insights about the current health state of the whole system through the described performance parameters. An initial concept is presented in [18]. With the help of a controller-based algorithm, a manipulated variable is then calculated from the collected information. The new control and communications strategy is finally implemented on the system such that the process can continue to run without difficulties even though environmental and/or internal changes have occurred.

Considering the described factory automation use case, adaption of process behavior could be, e. g., the change of an AGV's acceleration, velocity or distance to neighboring agents in case of worsening communications availability. In case of degrading control performance, the usage of radio resources such as the number of parallel links, MCS or routing can be adapted. All adaptations are done in a spatio-temporal manner, to serve individual needs and boost efficiency in resource utilization.

6. System Simulation with ROS

In order to fulfill the easy exchangeability of simulated and real components a flexible framework is needed. Additionally, many different communication-channels need to be created. One simulation framework that fulfills all these requirements is ROS (Robotic operating system [19]) with its physical simulator gazebo [20].

6.1. Concept

The concept of ROS consists of different software components called nodes that can be instantiated either singular or multiple times. This provides the needed flexibility to simulate multiple components.

For a realistic simulation and as there exists a ROS-integration already, the Franka Emika Panda arms [21] are used as manipulator. Fortunately, a kinematic model and a visualization already exists. A model including jacobians, gravity-, coriolis- and mass-matrices is provided with KDL (kinematic dynamic library [22]) using the identified masses of the real hardware [23]. To our knowledge, this is the first usage of torque-controllers of a panda robot in simulation. The controllers lie in the IoT-component itself as they need real time capability. Using an action server client, the different controllers for the different tasks/skills can be loaded during runtime. An action server basically provides the communication channels to proclaim a desired goal, to give feedback on the execution of this goal and give back the result of the goal.

The server itself lies in the IoT-component representative that is not needed to be real time capable and overtakes the communication to the other parts of the task layer and the situation analysis. All these parts are realized as nodes that are communicating over ROS.

6.2. Communications Simulation

In order to simulate the complete distributed system, the wireless communications between the different IoT devices need to be simulated as well, including also its time varying characteristics. However, the application-centered nature of the ROS-framework makes it generally relatively unsuited for simulating wireless communications. The framework already includes a communications stack with an either TCP or UDP transport layer. Under the assumption that all components of the simulated system are running on one processing unit with sufficient resources or have a reliable wired connection with the required bandwidth, the communications simulation can be built on top. Nevertheless, Fig. 3 presents a developed concept for simulating the time varying characteristic of wireless transmission Fig. 3.

In order to limit the simulation complexity, the communications effects of the different OSI-layers are abstracted to latencies and packet losses. Further, independent of the modeled communications architecture (cellular/WLAN/Ad-hoc), only one physical channel is modeled between two devices, i.e. no up-/downlink channels to access points are considered. Every IoT-device that is represented by a set of nodes in the simulation setup is associated with one communications node. The communications node introduces the communications effects, packet losses and latencies, on reception of a message addressed to its corresponding IoT-

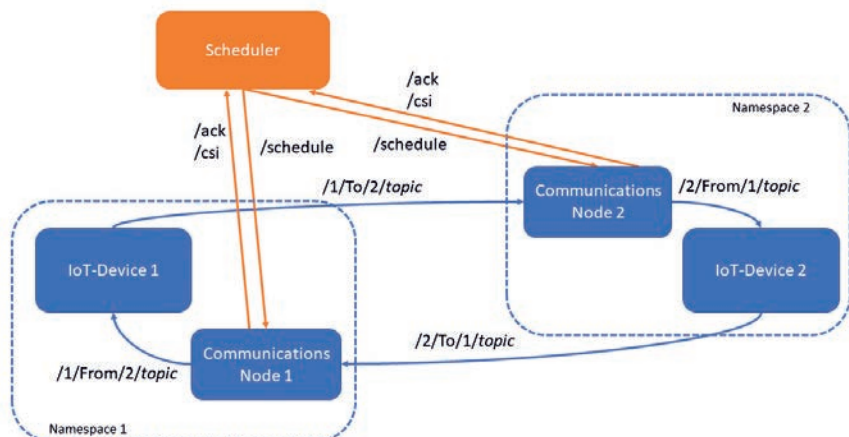


Fig. 3: Communications Simulation

device and republishes it to a new topic, which is subscribed by the IoT-device itself. Outgoing messages from the IoT-device are unaffected. This setup implies the use of one-to-one connections between devices only, which contradicts the topic-based ROS principles. Disturbing action-messages is more complicated since the ROS actionlib expects only one Action Server – with the proposed communications architecture there appear to be two. Heavily increasing the message overhead, the problem can be solved by adding a “dummy-republishing” in the nodes where the respective server and client are running.

To enable the communications-based co-management strategies explained in section 55. Co-Design and Co-Management, a central scheduler is implemented that assigns resource blocks to the communication nodes. A resource block consists of a channel and a time-slot. Together with a small jitter taken from a random process simulated at the communications node, the time-slot defines the latency that is added to the received message. Currently, the loss of a message is determined through the simulated fading process and SNR-threshold. Lost messages are dropped at the communications node.

7. Conclusion and Future Work

Future distributed control systems have a need for self-organization in collaboration and a communication and control co-management and co-design. In this paper the authors proposed a new system architecture model for self-organization and a control and communication co-management in distributed automation systems and showed approaches for self-organization and co-management. The implemented system simulation using ROS was also described and will be used for evaluating different co-management algorithms and improving self-organization algorithms. Further a demonstrator will be implemented in which three or more robot arms will be work together using the explored algorithms.

Acknowledgement

This work was supported by the Federal Ministry of Education and Research of the Federal Republic of Germany (BMBF) within the project VERITAS under the Project Number 01IS18073.

References

- [1] M. Sallinen, T. Heikkilä, T. Salmi, S. Kivikunnas and T. Pulkkinen, "A concept for isles of automation: Ubiquitous robot cell for flexible manufacturing," *IEEE Proceedings of International Conference on Control, Automation and Systems*, pp. 2246 - 2250, 2008.
- [2] "AutoStore. Space Redefined," www.autostoresystem.com.
- [3] Federal Ministry for Economic Affairs and Energy (BMWi), *Details of the Asset Administration Shell - Part 1*, May 2020.
- [4] Federal Ministry for Economic Affairs and Energy (BMWi), *Interaction Model for Industrie 4.0 Components*, March 2016.
- [5] A. Perzylo, J. Grothoff, L. Lucio, M. Weser, S. Malakuti, P. Venet, V. Aravatinos and T. Deppe, "Capability-based semantic interoperability of manufacturing resources: A BaSys 4.0 perspective," *IFAC-PapersOnline*, vol. 52, no. 13, pp. 1590 - 1596, 2019.
- [6] S. Profanter, A. Breitzkreuz, M. Rickert and A. Knoll, "A hardware-agnostic OPC UA skill model for robot manipulators and tools," *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pp. 1061 - 1068, 2019.
- [7] Fraunhofer IESE, "BaSys 4.0. The Middleware for Industrie 4.0," https://www.basys40.de/wp-content/uploads/2019/12/Flyer_BaSys40_english.pdf.
- [8] G. Lympelopoulou and P. Ioannou, "Adaptive control of networked distributed systems with unknown interconnections," *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 3456 - 3461, 2016.
- [9] F. Kazempour and J. Ghaisari, "Stability analysis of model-based networked distributed control systems," *Journal of Process Control*, vol. 23, no. 3, pp. 444 - 452, 2013.
- [10] D. Zhang, P. Shi, Q.-G. Wang and L. Yu, "Analysis and synthesis of networked control systems: A survey of recent advances and challenges," *ISA transactions*, vol. 66, pp. 376 - 392, 2017.
- [11] M. Gundall et al., "5G as Enabler for Industrie 4.0 Use Cases: Challenges and Concepts," *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, pp. 1401 - 1408, 2018.
- [12] Plattform Industrie 4.0, "Describing Capabilities of Industrie 4.0 Components," 2020.
- [13] S. F. Roselli, K. Bengtsson and K. Åkesson, "SMT solvers for job-shop scheduling problems: Models comparison and performance evaluation," *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, pp. 547 - 552, 2018.

- [14] L. Scheuvens, T. Hößler, A. N. Barreto and G. P. Fettweis, "Wireless Control Communications Co-Design via Application-Adaptive Resource Management," *2019 IEEE 5G world Forum (5GWF)*, pp. 298 - 303, 2019.
- [15] A. Traßl, L. Scheuvens, T. Hößler, E. Schmitt, N. Franchi and G. P. Fettweis, "Outage Prediction for URLLC in Rayleigh Fading," *2020 IEEE 29th Conference on Networks and Communications*, 2020.
- [16] F. Burmeister, N. Schwarzenberg, T. Hößler and G. Fettweis, "Measuring Time-Varying Industrial Radio Channels for D2D Communications on AGVs,," in *IEEE Wireless Communications and Networking Conference*, Nanjing, China, 2021.
- [17] F. Allgoewer, R. Findeisen and Z. K. Nagy, "Nonlinear Model Predictive Control: From Theory to Application," *Journal of the Chinese Institute of Chemical Engineers*, vol. 35, no. 3, pp. 299 - 315, 2004.
- [18] D. Schulze and H. Zipper, "A Decentralised Control Algorithm for an Automated Coexistence Management," *2018 IEEE Conference on Decision and Control (CDC)*, pp. 4187-4193, 2018.
- [19] Stanford Artificial Intelligence Laboratory et al., "Robot operating system," <https://www.ros.org>.
- [20] "Gazebo Simulator," [Online]. Available: www.gazebo-sim.org.
- [21] Franka Emika, "Panda," [Online]. Available: www.franka.de/panda.
- [22] "Orocos Kinematics and Dynamics," [Online]. Available: www.orocos.org/kdl.html.
- [23] C. Gaz, M. Cognetti, A. Oliva, P. R. Giordano and A. D. Luca, "Dynamic Identification of the Franka Emika Panda Robot With Retrieval of Feasible Parameters Using Penalty-Based Optimization," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4147-4154, 2019.

Comparison of different spike detection methods in Hall-Héroult cells

R. Kremser, N. Grabowski, Prof. Dr.-Ing. **D. Tutsch**,
Bergische Universität Wuppertal, Wuppertal;
Dr.-Ing. **R. Düssel, A. Mulder**,
TRIMET Aluminium SE, Essen

Abstract

During the process of aluminium smelting, where the major production route is the Hall-Héroult electrolysis process, different types of anomalies can arise. One of these anomalies are called spikes. Spikes are protrusions on the bottom of the anodes and can result in a short circuit between the anode and cathode surface. For that reason they have a negative impact on the current efficiency of a pot and often lead to high temperature excursions. Therefore it is important to detect spikes as early as possible and remove them. State of the art at the plant of TRIMET Aluminium SE Essen (TAE) is a manual detection by the potroom operator. For the purpose of automation in this paper different spike detection methods such as machine learning algorithms and expert systems are developed and compared against each other. Since the formation of spikes is a process of different steps, that can take hours up to days and only the timestamp of a manual detected spike in the historical dataset is known, the main challenge is to deal with this label uncertainty. Furthermore spike formation is a rare event, resulting in a highly imbalanced class distribution of spikes and normal state. The focus of this investigation lays on the earliness and the performance metrics of the classifier models.

1. Introduction

In the aluminium production process smelters experience different types of problems. One type of problem are anomalies affecting the anode. Anodes consist of carbon and are set into the pot and are consumed by the electrolytic reaction. This reaction mainly converts Al_2O_3 to aluminium. After the end of life the anode is removed from the pot and a new anode is set again [1]. During anode life protrusions can be formed on the bottom of the anodes, which can be seen in Fig. 1. These protrusions are often referred as an anodic incident or more precisely as a spike.



Fig. 1: Spike under anode marked with green circle.

Spikes can act as a short circuit between the anode and cathode surface and therefore have a negative impact on the current efficiency and other process variables. Rolofs et al. [2] analysed the effect of anode spike formation on operational performance. It was found that despite a lower average operating voltage, the pots operated at an average 4 °C higher bath temperature. In addition it is said that pots with spikes had twice the number of high temperature excursions (>980 °C) and the current efficiency was approximately 3% lower. Ali et al. [3] also studied the effect of anode spikes on pot line performance in periods before (0.1 spikes/pot/month) and during spike formation (2 spikes/pot/month). It was found that the current efficiency was approximately 1.8% lower while the energy consumption increased by 1.8%. In contrast to [2] the temperature measurements were not affected by the spike problem. Iffert [4] states that in most cases when spikes were found, the superheat was increased and after removing the spike the superheat decreased remarkably. Cahill et al. [5] also observed a 2% current efficiency loss and list cover crust melting, severe airburn, carbon breakage or anode drop-offs, stub damage and/or iron contamination as secondary effects. For that reason it is important to prevent or to detect spikes as early as possible. In this context Cahill et al. [5] studied the root causes of spikes in a meta-analysis. It is stated that the actual cell condition influences the probability of forming a spike, while the main initiator is the act of anode setting. Furthermore a high level of carbon dust and insufficient

cavity cleaning is listed as a root cause, which coincides with the findings by [2], [3], [4] and Fischer et al. [6]. Moreover a low superheat and low anode-to-cathode distance (ACD) operation can lead to spikes, which is also mentioned by Wong et al. [7]. In the literature there have also been attempts to detect spikes. Stam et al. [8] propose to focus on repeatable beam up movements, a stepwise rise of bath temperature ($>15\text{ }^{\circ}\text{C}$) and a reduction of the cumsum of alumina feed ratio to build a reliable detection model. Dion et al. [9] developed an indicator based on measurements of the vanadium content in the aluminium, requiring additional hardware. Majid et al. [10] developed a model by multiway principal component analysis, which was tested on two samples from two different pots. With the introduction of individual anode current measurement (ACM) it was shown by Kolås et al. [11] and Kremser et al. [12] that spikes can be observed by a high current. In the last two years there was a bigger focus on developing spike detection models based on ACM. Martel [13] uses Hidden Markov Models to detect spikes, while LaJambe et al. [14] use multivariate analysis of individual anode current signals. Since this paper deals with data of pots where no ACM is implemented, thus it is based on classic process data, the prediction results of these papers are not discussed in detail. However, Martel [13] describes the process of spike formation, which can be seen in Fig. 2. In step 1 there is an area that is less conductive or insulated. This can be caused by the root causes discussed before. In the second step the spike begins to develop. Waves in the metal pad are able to reach the developing spike, leading to temporary local short circuits or local variations in anode current density. In the third step the spike is fully grown and extends into the metal pad, creating a permanent short circuit. At this stage herein before mentioned negative impacts are the most severe. It is important to note that the described process is simplified and also depends on the shape of the spikes, which are explained in detail in [3] and [5] and can take hours up to days.

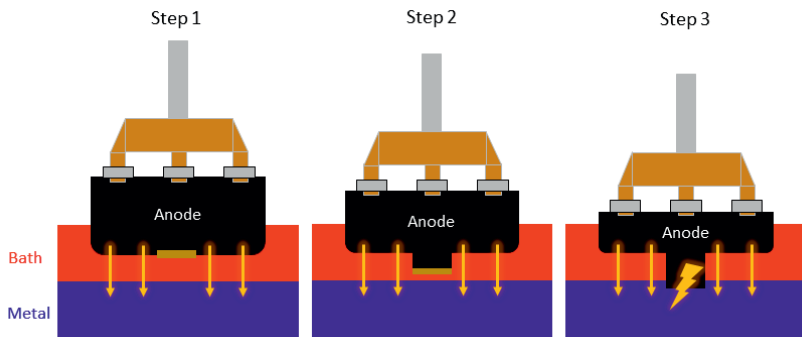


Fig. 2: Steps of spike formation [13].

In summary, it can be stated that some predictions methods were only proposed but never applied on a real test dataset of sufficient size (high number of pots and/or large timespan) or used in a live test scenario. Moreover authors do not provide any evaluation criteria like confusion matrix and F1 score. In this paper different classification models are developed and compared against each other.

2. State of the Art at TAE

The TRIMET Aluminium SE Essen (TAE) plant consists of 360 pots in three production potrooms with EPT-14 PreBaked Point Feeder (PBPF) cells in the first two potrooms and EPT-17 PBPF cells in the third potroom. The prefix EPT stands for “elektronische Prozesssteuerung Tonerde” (computer controlled alumina feed). The number 14 indicates that the potlines originally operated at 140 kA [15]. During train and test time (except current drops), the first potroom was running at an amperage of 172 kA with a standard deviation of around 5 kA, while the second potroom was running at 164 kA with a standard deviation of around 3.7 kA and the third potroom, according to its name, was running at 170 kA with a standard deviation of 3.4 kA. Under normal process conditions a balanced feed strategy with four feed phases is applied, using a derived pseudo-resistance slope calculation to determine when the feed strategy should change from underfeed to overfeed [16]. The bath and liquidus temperature are controlled by the so called 9-Box-Control Concept, described in detail by Iffert [4].

Currently no automatic spike detection system is implemented. There are two main scenarios how spikes are detected: In the first scenario the operating staff is actively involved. By observation of conspicuous process variables like bath temperature, superheat, noise and feeding the operator assumes that there is a spike in a specific pot. There are fixed thresholds that facilitate the decision-making procedure, e.g. 985 °C for bath temperature, 18°C for superheat. Following this, all 26 anodes of the concerned pot are checked with a millivolt fork. If the operator observes a comparatively high voltage the anode is removed from the pot and checked by visual inspection of the anode bottom. If there is a spike, a new or an old anode is set into the pot and the anode with the spike is undergoing a process where the spike is milled off. In the second scenario the spike is found during a regular anode change. The spikes are recorded in a database with the start of shift as timestamp. The goal of this paper is to develop different models for an automatic spike detection. It is important to note that these kind of models are not able to locate the spikes in the pot, so that the checkup with the millivolt fork is still required.

3. Data Selection, Preprocessing and Labeling

In machine learning the whole dataset is usually splitted into at least training and test data. In this paper a 77/23 splitting regarding unique spikes is used, which is shown in Fig. 3. It is important to note that these quantities arise as a result from the preprocessing steps that will be described in the following. According to that the training dataset consists of nearly 24900 rows (82%) with 151 unique spikes, while the test dataset consists of nearly 5400 rows (18%) with 45 unique spikes. If there are multiple spikes on one timestamp, these spikes are counted as one unique spike. The data originates from all 120 pots in potroom 2 to have a consistent dataset from the same pot technology. The data from potroom 1 needed to be excluded because of missing values in beam position and events. Moreover in potroom 3 the bath temperature is measured in steps of two days and therefore it was decided to exclude these pots as well.

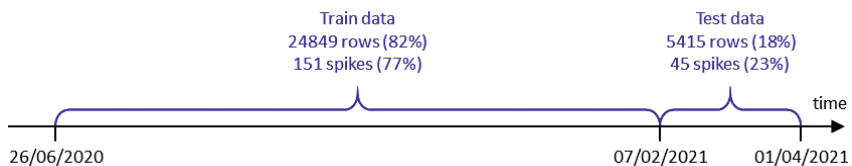


Fig. 3: Splitting of train and test dataset.

The daily measurements are performed manually by the operators. If a measurement is implausible or comparatively high, the measurement is repeated. Therefore measurements have been removed where the difference in bath temperature or superheat exceeds a threshold of 10 °C within an hour to filter implausible measurements. After this step technical implausible values were filtered, e.g. bath temperatures higher than 1050 °C. Furthermore pots with an age lower than 100 days were removed, because these pots belong to the group of early operation pots with different setpoints and a different behaviour.

For a fair comparison all models that are developed get the same test dataset based on daily bath temperature measurements except for different features as input. For this reason labels are matched with the timestamps of the bath temperature measurements. Without any ACM it is a challenging task to create reliable labels. It was decided to focus on step 3 (see Fig. 2) of the spike formation process, because at this stage the negative impacts are the most severe and symptoms like high bath temperature may be significantly more distinct. However, the disadvantage is a late detection. The typical behaviour of the bath temperature before the spike occurrence is illustrated in Fig. 4. The star symbol denotes the timestamp of the beginning of the shift when the spike was found.

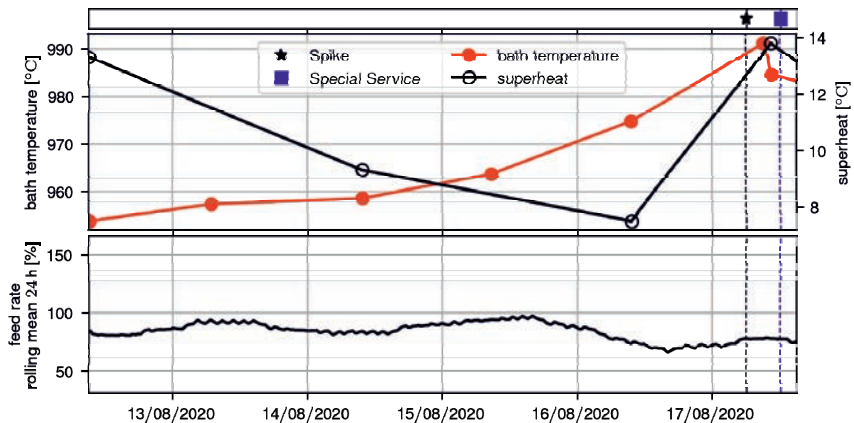


Fig. 4: Temperature measurements and 5-min-data over time before spike occurrence.

The square denotes the timestamp of the special service when the anode was removed for visual inspection. It can be seen that there is a strong increase in bath temperature after August 15. Also the superheat increases, which is usually measured every second day. Furthermore the rolling mean of the feed rate with a window size of 24 h is plotted. The feed

rate originates from the so called 5-min-data that is measured in steps of 5 min as the name already suggests. It can be observed that there is a decrease after August 15. This coincides with the mentioned reduction of the cusum of alumina feed ratio by Stam et al. [8]. The labeling was performed manually regarding a timespan of maximum seven days with respect to the following criterions: samples with (i) a bath temperature $> 977\text{ }^{\circ}\text{C}$ (actual outer control limit of the 9-Box-Control Concept) after a strong increase or (ii) high superheat $> 12\text{ }^{\circ}\text{C}$ without any decrease afterwards are labeled as spike. If the bath temperature is lower than $977\text{ }^{\circ}\text{C}$ and superheat is lower than $12\text{ }^{\circ}\text{C}$ only the latest available sample next to the spike removal was labeled as spike. In some cases the labeling was difficult because the bath temperature was already high for longer than seven days. These samples were labeled in collaboration with TRIMET process experts analysing other process variables like feeding, noise and chemical additions. Furthermore attention was paid to the following point: The backwards labeling was stopped if the change in time between the anode change of the anode that had the spike was less than 24 h. This is because anodes are checked with a millivolt fork 24 h after their change and a spike would have very likely been found at this time. Using the example of Fig. 4 both samples on August 17 were labeled as spike, all other samples are labeled as 'normal'. After labeling it is possible to calculate the prevalence. This yields a prevalence of 1.5% for the train dataset and a slightly higher prevalence of 2% for the test dataset. Another common term is the imbalance ratio (IR), which is defined as the number of negative class samples divided by the number of positive class samples [17]. For the train dataset IR is about 68, respectively 50 for the test dataset. Both these values are in the range of the IRs from the well-known knowledge extraction based on evolutionary learning (KEEL) repository mentioned by Fernandez et al. [17]. Therefore the dataset in this paper belongs to the group of imbalanced datasets. In the following sections the development of three types of prediction models will be described.

4. Baseline Model

The simplest approach to build a prediction model is to use only one feature and a threshold. This kind of model is called baseline model and the purpose is to study if more complex models are able to outperform very simple models. In this case the bath temperature was chosen with a threshold of $986\text{ }^{\circ}\text{C}$ as a result of optimisation using F_1 score in the train dataset. If the bath temperature is higher than this threshold a spike is predicted.

5. Expert Model

Expert systems can be seen as an advanced form of programming and require a knowledge base, an inference engine, a user interface and data as major components. For example a knowledge base can contain rules, usually if-then rules [18]. In this paper the knowledge base is introduced via the definition of seven features by TRIMET process experts after interviewing the operators, shown in Table 1.

Table 1: Features of expert model.

#	Feature	measuring frequency
1	Bath temperature > 975 °C	daily
2	Bath temperature delta (2 days) > 15 °C	daily
3	Superheat > 15 °C	every second day
4	Noise (low frequency) < 0,11 $\mu\Omega$	5 min
5	slowly, continuously and uniformly increasing noise	5 min
6	Overfeeding (OA) cycles < 10	5 min
7	small beam movements in last 6 h > 3	5 min

From September 10 2020 until October 12 2020 456 samples from all potrooms were randomly selected where the bath temperature was higher than 975 °C. With these samples an additional training dataset was created by filling the features from Table 1 with boolean values. Unquantifiable features like #5 and #7 could not be filled automatically and depend on the operator's individual estimation. Altogether 24 different operators were involved with different participation rate. After filling the features every pot was checked with a millivolt fork. In 378 cases a spike was found. To build a model, the last six features of this dataset were used as input. Implementations of Logistic Regression, Passive Aggressive Classifier, Ridge Classifier, Random Forest Classifier and Linear Support Vector Classifier from Scikit-learn (version 0.23.1) [19] and eXtreme Gradient Boosting (XGB) (version 0.90) by Chen and Guestrin [20] were used for training. In the literature there are three main techniques to deal with imbalance in classification. First it is possible to upsample the minority class, second downsample the majority class or third using cost-sensitive machine learning algorithms [21]. Using cost-sensitive machine learning algorithms corresponds to using class weights in Scikit-learn. Therefore the class weights of the models were set to 'balanced', while the minority class weight is used in XGB, which requires to choose the label 'normal' as the positive class. Because 456 rows is comparatively small, there was no need to over- or undersample the dataset, which might also be more challenging because of boolean values.

To find appropriate hyperparameters a gridsearch is performed with the F_1 score for the positive and negative class each.

After training unquantifiable features (#5 and #7) needed to be defined mathematically to apply the models on a test dataset. For that purpose the linear trend of the noise was calculated for windows sizes of 6 h with different end times before spike removal. Since process events like anode change cause a high noise, the noise value was filtered by a threshold. This is called prefiltering. Analogously to find a representation for feature #7, successive differences of the beam position were calculated.

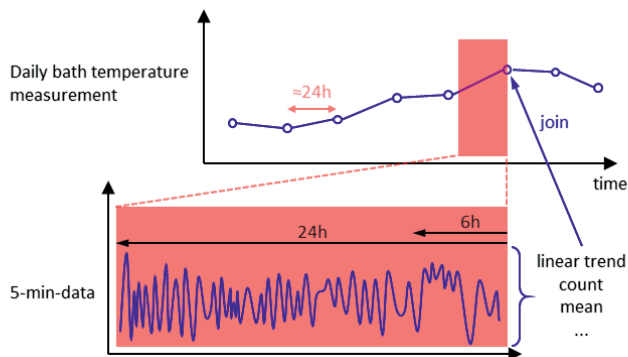


Fig. 5: Feature generation procedure for expert model and machine learning model.

It was necessary to define two thresholds for the prefiltering to specify what a small beam movement is. For both features #5 and #7 different threshold combinations were tested. To decide which combination fits the best, a Logistic Regression model was trained using the binary values as labels. The highest F_1 score is reached if noise values above $5 \mu\Omega$ and beam movements below ± 0.7 mm and above ± 10 mm are removed in the prefiltering step.

The Logistic Regression model was also used to calculate linear decision boundaries for feature #5 and #7, to convert features into boolean values. After these steps the test dataset can be prepared. Since the superheat is usually measured every second day, missing values were filled by forward filling. The feature generation procedure for 5-min-data is shown in Fig. 5. For every timestamp in the bath temperature samples, a corresponding timestamp in the 5-min-data was chosen. Based on this timestamp the last 24 h or 6 h are selected and the appropriate features are calculated if at least 80% of the data is available in this window. Then the results are joined with the bath temperature samples. In summary, it can be stated that these types of models are a hybrid of expert system and machine learning model.

6. Machine Learning Model

Finally the third type of model is a machine learning (ML) approach. In the following the feature generation is described. The complete feature list is shown in Table 2. Except for bath temperature and superheat more daily features like excess aluminium fluoride (xsAlF3), bath ratio, bath height and metal height were included. Since these features are usually measured every second day, missing values were filled by forward filling. Furthermore previous values of daily features are calculated in two ways: If the change in time between two time values is large (>12 h) the corresponding features are annotated with '_large' and vice versa with '_small'. Additionally for the bath temperature successive differences were calculated, see feature #13. Feature #14 corresponds to feature #2 from the expert model. The feature generation procedure for 5-min-data is analogous to the procedure shown in Fig. 5. Numerical features from the expert model #22, #30, #31, #33 were included. In this context '_6h' or '_24' denotes if the window size is 6 h or 24 h. Moreover mean, median and standard deviation (std) were calculated for the generated input heat flow, line current, noise, feed rate and resistance.

As described before the training and test datasets are imbalanced, because the prevalence of spikes is low. To deal with this imbalance a combination of undersampling the dataset and class weights is used. The python package by Lemaître et al. [22] provides different undersampling methods. In this paper NearMiss undersampler is used, which belongs to the group of controlled undersampling techniques that allow to specify class ratios.

Table 2: Features of machine learning model.

#	Feature	#	Feature
1	bathtemperature	20	metalheight_previous_small
2	xsalf3	21	bathtemperature_delta_small
3	bathratio	22	feedphase_combined_count_oa_24h
4	bathheight	23	heat_flow_generation_mean_24h
5	metalheight	24	linecurrent_mean_24h
6	superheat	25	lowfrequencynoise_mean_24h
7	bathtemperature_previous_large	26	lowfrequencynoise_median_24h
8	superheat_previous_large	27	feedrate_mean_24h
9	xsalf3_previous_large	28	resistance_mean_24h
10	bathratio_previous_large	29	resistance_std_24h
11	bathheight_previous_large	30	lowfrequencynoise_linear_trend_6h
12	metalheight_previous_large	31	delta_beamposition1_count_beam_move_6h
13	bathtemperature_delta_large	32	heat_flow_generation_mean_6h
14	bathtemperature_delta_previous_2days	33	lowfrequencynoise_mean_6h
15	bathtemperature_previous_small	34	lowfrequencynoise_median_6h
16	superheat_previous_small	35	feedrate_mean_6h
17	xsalf3_previous_small	36	resistance_mean_6h
18	bathratio_previous_small	37	resistance_std_6h
19	bathheight_previous_small		

Moreover the following cleaning undersampling techniques were tested: Edited-NearestNeighbours, RepeatedEditedNearestNeighbours, AllKNN, CondensedNearest-Neighbour, OneSidedSelection, NeighbourhoodCleaningRule and InstanceHardness-Threshold. Detailed information about these techniques can be found in [22]. The same models that were used for the expert model (Scikit-learn and XGB) were trained to build the machine learning model. To find appropriate hyperparameters a gridsearch (10Fold and Stratified 10Fold) is performed with the F_1 score for the positive class (spike). The hyperparameter space is shown in Table 3.

Table 3: Gridsearch hyperparameter space of machine learning model.

Model	Hyperparameters
Logistic Regression	penalty: [l_1 , l_2], C: [0.001, 0.01, 0.1, 1, 10, 100, 1000]
Passive Aggressive Classifier	C: [0.25, 0.5, 1, 2, 10], fit_intercept: [True, False]
Ridge Classifier	C: [0.25, 0.5, 1, 2, 10], fit_intercept: [True, False]
Random Forest Classifier	n_estimators: [1000], max_features: [0.5, 0.7, 1] max_depth: [10, 20, 30, 40, 50], min_samples_split: [2, 3, 4], min_samples_leaf: [2, 3, 4]
eXtreme Gradient Boosting	learning_rate: [0.001, 0.01, 0.1, 0.3], max_depth: [4, 6, 8, 10], n_estimators: [10, 100, 500, 1000]
Linear Support Vector Classifier	C: [0.001, 0.1, 1, 5, 10, 50, 100]

7. Results

In this section the results of the three types of models, which are applied on the test dataset are described. To evaluate the performance of the models the F_1 score is examined. Beyond that an F_{unique} score is calculated which does not depend on the labels but takes into account, if a spike was detected at all. For evaluating the earliness of the classifiers the mean (denoted with $\emptyset$) lead time is calculated, which is defined as the mean of the changes in time between the earliest true positive and the last labeled sample before the spike occurrence.

All models were optimized using the training dataset and F_1 score. This yields a threshold of 986 °C for the baseline model and 985 °C for the expert model. It is important to note that the expert model was originally intended to be applied on a subset of the dataset with temperatures above 975 °C (see feature #1 in Table 1). The results of the models with the best F_1 and F_{unique} score for each type of model are shown in Table 4. The corresponding confusion matrices are shown in Table 5. The number inside the braces represents the unique true positives and false positives that were used for determining the F_{unique} score. It can be seen that the expert model is slightly better than the baseline model. The machine learning models have a slightly better performance than the baseline and expert model. The machine learning model with best F_1 score was obtained using RepeatedEditedNearestNeighbour ($n_{\text{neighbours}}=10$) in combination with a Stratified 10Fold, while the machine learning model with the best F_{unique} score was obtained using NearMiss-3 ($n_{\text{neighbours}}=12$, sampling strategy=0.1) in combination with a 10Fold. The mean lead time is a little more than one day and almost in the same range for all models.

Table 4: Results on test dataset.

Type of Model	Model	F_1 [%]	F_{unique} [%]	$\emptyset$ lead time [h]
Baseline	Threshold	21.6	14.9	32
Expert	Passive Aggressive Classifier	24.4	16.8	28
ML best F_1	eXtreme Gradient Boosting	25.8	17.4	27
ML best F_{unique}	Random Forest Classifier	22.8	27.2	28

Table 5: Confusion matrices of the best models.

		Baseline		Expert		ML best F_1		ML best F_{unique}	
predicted	spike	30 (15)	141	33 (16)	130	48 (25)	217	17 (11)	25
	normal	77 (30)	5167	74 (29)	5178	59 (20)	5091	90 (34)	5283
		spike	normal	normal	normal	spike	normal	spike	normal
		true		true		true		true	

8. Conclusions

In this paper different types of models were trained to predict spikes in advance. The model with the best F_1 score is an eXtreme Gradient Boosting machine and yields an F_1 score of 25.8%, which is slightly better than a baseline model using a temperature threshold. The poor performance could be caused by the general approach, insufficient features or the label uncertainty. It is unlikely that the models themselves are the reason. The lead time for the model with the best F_1 score is about 27 h. Since it is unknown how many false positives are generated by the manual detection of the operators, there cannot be any comparison to the state of the art. This can only be examined in live tests with an accurate documentation of the operators' actions. Nevertheless, even if only a fourth of the predictions are real spikes, almost half of the spikes are predicted, which helps the operators very much in their actions. For the future it is recommended to focus on feature engineering, e. g. calculating time series characteristics of higher frequency signals. The analysis of these characteristics may also lead to more certain labels. Furthermore the individual anode current measurement could also be used to improve the label certainty and gain more knowledge about the spike formation and its interaction with classic process data.

Funding

This research was funded by Bundesministerium für Bildung und Forschung (03SFK3B1-2).

References

- [1] K. Grjotheim und H. Kvande, Introduction to Aluminium Electrolysis, Düsseldorf: Aluminium-Verlag, 1993.
- [2] B. Rolofs und N. Wai-Poi, „The Effect Of Anode Spike Formation On Operational Performance,“ in *Light Metals 2000*, 2000.
- [3] M. M. Ali und A. M. Omran, „Anode Spike Formation In Prebaked Aluminium Reduction Cells,“ *Al-Azhar Engineering Twelfth International Conference*, 2012.
- [4] M. Iffert, „Aluminium Smelting Cell Control and Optimisation,“ The University of New South Wales, Sydney, Australia, 2007.
- [5] R. A. Cahill und P. Palau, „Common Root Causes for Anode Problems within the Reduction Cell,“ in *Proceedings of the 12th Australasian Aluminium Smelting Technology Conference*, 2018.
- [6] J.-C. Fischer, R. Perruchoud, M. Arlettaz, W. Fischer, R. V. Kaenel, J. Antille und L. Bugnion, „Anode Issues during Smelter Capacity Creep,“ in *International Conference on Non-Ferrous Metal 2015 at Bhubaneswar, Odisha, India*, 2015.
- [7] C.-J. Wong, Y. Yao, J. Bao, M. Skyllas-Kazacos, B. J. Welch und A. Jassim, „Modeling Anode Current Pickup After Setting,“ in *Light Metals 2021*, 2021.
- [8] M. A. Stam, M. P. Taylor, J. J. J. Chen, A. Mulder und R. Rodrigo, „Common Behaviour and Abnormalities in Aluminium Reduction Cells,“ in *Light Metals 2008*, 2008.
- [9] L. Dion, L. I. Kiss, D. Lavoie und J.-P. Arvisais, „Developing a New Process Indicator Based On The Relationship Between An Electrolysis Cell Impurity Balance And Its Incidents,“ in *Light Metals 2014*, 2014.
- [10] N. A. A. Majid, M. P. Taylor, J. J. J. Chen, M. A. Stam, A. Mulder und B. R. Young, „Aluminium process fault detection by Multiway Principal Component Analysis,“ *Elsevier*, 2010.
- [11] S. Kolås, P. McIntosh und A. Solheim, „High Frequency Measurements of Current Through Individual Anodes: Some Results from Measurement Campaigns at Hydro,“ in *Light Metals 2015*, 2015.

- [12] R. Kremser, N. Grabowski, R. Düssel, K. Kessel und D. Tutsch, „Investigation of Different Measurement Techniques for Individual Anode Currents in Hall-Héroult Cells,“ in *Proceedings of the 12th Australasian Aluminium Smelting Technology Conference*, 2018.
- [13] A. Martel, „Spike Detection Using Advanced Analytics and Data Analysis,“ in *Light Metals 2018*, 2018.
- [14] D. LaJambe, É. Poulin, C. Duchesne und J. Tessier, „Anodic Incident Detection through Multivariate Analysis of Individual Anode Current Signals,“ in *Light Metals 2020*, 2020.
- [15] A. Tabereaux, „Prebake cell technology: A global review,“ *JOM*, Bd. 52, p. 23–29, 2000.
- [16] A. Mulder, P. Lavoie und R. Düssel, „Effect of Early Abnormality Detection and Feed Control Improvement on Reduction Cells Performance,“ in *Proceedings of the 12th Australasian Aluminium Smelting Technology Conference*, 2018.
- [17] A. Fernández, S. García, M. Galar, R. C. Prati, F. B. Krawczyk und Herrera, *Learning from Imbalanced Data Sets*, Springer International Publishing, 2018.
- [18] S. D. Hu, *Expert Systems for Software Engineers and Managers*, New York: Chapman and Hall, 1987.
- [19] G. V. F. Pedregosa, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot und E. Duchesnay, „Scikit-learn: Machine Learning in Python,“ *Journal of Machine Learning Research*, Bd. 12, p. 2825–2830, 2011.
- [20] T. Chen und C. Guestrin, „XGBoost: A Scalable Tree Boosting System,“ in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, USA, 2016.
- [21] J. Brownlee, *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning*, Machine Learning Mastery, 2020.
- [22] G. Lemaître, F. Nogueira und C. K. Aridas, „Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning,“ *Journal of Machine Learning Research*, Bd. 18, pp. 1-5, 2017.

Ingenieure wollen immer alles ganz genau wissen. Wie wär's mit einem E-Paper- oder Zeitungs-Abo?



Mehr Meinung. Mehr Orientierung. Mehr Wissen.

Wesentliche Informationen zu neuen Technologien und Märkten.

Das bietet VDI nachrichten, Deutschlands meinungsbildende Wochenzeitung zu Technik, Wirtschaft und Gesellschaft, den Ingenieuren. Sofort abonnieren und lesen.

Donnerstagabends als E-Paper oder freitags als Zeitung.

Jetzt abonnieren: Leser-Service VDI nachrichten, 65341 Eltville

Telefon: +49 6123 9238-201, Telefax: +49 6123 9238-244, vdi-nachrichten@vuservice.de

100
JAHRE

VDI nachrichten

TECHNIK IN
SZENE GESETZT.

www.vdi-nachrichten.com/abo



VDE

Mess- und
Automatisierungstechnik



digital



AUTOMATION

ISBN 978-3-18-092392-5