Fortschritt-Berichte VDI

VDI

Reihe 10

Informatik/
Kommunikation

Nr. 848

Herwig Unger, Wolfgang A. Halang (Eds.)

Autonomous Systems 2016

Proceedings of the 9th GI Conference

FernUniversität in Hagen
Schriften zur Informations- und Kommunikationstechnik

Fortschritt-Berichte VDI

Reihe 10

Informatik/
Kommunikation

Herwig Unger, Wolfgang A. Halang (Eds.)

Nr. 848

Autonomous Systems 2016

Proceedings of the 9th GI Conference

FernUniversität in Hagen
Schriften zur Informations-
und Kommunikationstechnik

Unger, Herwig; Halang, Wolfgang A. (Eds.)

Autonomous Systems 2016

Proceedings of the 9th GI Conference

Fortschr.-Ber. VDI Reihe 10 Nr. 848. Düsseldorf: VDI Verlag 2016.
236 Seiten, 53 Bilder, 33 Tabellen.
ISBN 978-3-18-384810-2, ISSN 0178-9627,
€ 81,00/VDI-Mitgliederpreis € 72,90.

Keywords: Autonomous Systems – Safety-related and Real-time Systems – Networks and Routing – Suppression of Disturbances – Neural and Evolutionary Computing – Natural Language Processing – Education

To meet the expectations raised by the terms Industrie 4.0, Industrial Internet and Internet of Things, real innovations are necessary, which can be brought about by information processing systems working autonomously. Owing to their growing complexity and their embedding in complex environments, their design becomes increasingly critical. Thus, the topics addressed in this book span from verification and validation of safety-related control software and suitable hardware designed for verifiability to be deployed in embedded systems over approaches to suppress electromagnetic interferences to strategies for network routing based on centrality measures and continuous re-authentication in peer-to-peer networks. Methods of neural and evolutionary computing are employed to aid diagnosing retinopathy of prematurity, to invert matrices and to solve non-deterministic polynomial-time hard problems. In natural language processing, interface problems between humans and machines are solved with graph-based text representation and word segmentation. Finally, related aspects of teaching are discussed.

Bibliographische Information der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter http://dnb.ddb.de abrufbar.

Bibliographic information published by the Deutsche Bibliothek
(German National Library)
The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at http://dnb.ddb.de.

Schriften zur Informations- und Kommunikationstechnik
Herausgeber:
Wolfgang A. Halang, Lehrstuhl für Informationstechnik
Herwig Unger, Lehrstuhl für Kommunikationsnetze
FernUniversität in Hagen

Als Manuskript gedruckt. Printed in Germany.

ISSN 0178-9627

ISBN 978-3-18-384810-2

Preface

Currently, the terms Industrie 4.0, Industrial Internet and Internet of Things, the latter one coined by Ashton in 1999 with regard to radio-frequency identification, are frequently heard buzzwords. According to the "New Gartner Hype Cycle for Emergent Technologies"[1] published in 2015, the topic Internet of Things has presently reached its "Peak of Inflated Expectations", which is usually followed by the "Trough of Disillusionment". Such a set-back is often caused by incorporating conventional items into new fashions[2], that becomes visible when euphoria gives way to sober thinking. For a trend to prevail, it must then lead into the "Slope of Enlightenment". To this end, real innovations are necessary, which can, with respect to the Internet of Things, to a large extent be brought about by information processing systems working autonomously on devices and network nodes of any kind as considered in this volume.

To begin with, in its first section on safety-related and real-time systems, well-established techniques for modeling safety requirements are reviewed and compared against essential requirements as provided by the standards prevailing for developing safety-related systems, a set of requirements for computer architectures to be inherently safe is derived from these standards, and verification and validation of robotic control software are discussed. For use in embedded safety-critical systems, a control unit is presented, whose main design criterion was verifiability by widest possible consensus, finally, a concept for an affordable, lossless storage of large amounts of measurement data generated on distributed mobile sensor nodes is worked out.

The section on networks and routing first deals with the use of centrality measures. Applying them and status data of sensor nodes, such as battery charging levels, a novel multi-criteria routing strategy for wireless sensor networks is proposed. In combination with topological analysis and information local in a network, then a routing algorithm aiming to maximise traffic flow and feasibility of selected routes is based on a new centrality measure. To ensure continuous re-authentication of peers in peer-to-peer networks, which were in contact before, finally a mechanism inspired by the dinner cryptographic protocol and the zero knowledge protocol is presented.

[1] http://www.gartner.com/newsroom/id/3114217

[2] P. Mertens and D. Barbian: Digitalisierung und Industrie 4.0 – Trend mit modischer Überhöhung? *Informatik Spektrum* 39, 4, 301–309, 2016

The following section is devoted to two approaches of suppressing disturbances. To this end, the feasibility of applying chaotic carrier frequency modulation for fighting electromagnetic interferences in switching-mode power supplies by spreading the spectra of input and output signals over the entire frequency band is shown, and the structure of a dual-mode passive filter is investigated by bifurcation analysis, which operates in dependence on a control parameter either as a band-pass or a high-pass.

The first contribution in the section on neural and evolutionary computing details a method combining processing retinal images and machine learning with artificial neural networks in order to aid diagnosing retinopathy of prematurity, the most common cause of blindness of premature infants. Employing an echo state network as universal processing element, the concept of a novel matrix inversion system based on black-box-trained reservoir neurocomputing is then presented. Finally, according to the concepts of evolutionary computation and swarm intelligence, an improved quick artificial bee colony algorithm is proposed to solve a non-deterministic polynomial-time hard problem.

In the section on natural language processing a new graph-based method determining centroid terms as text representatives is presented, which allows to calculate semantic similarities between text documents — even if they have no terms in common, and the difficult task of word segmentation in the Thai language is shown to be improved by an algorithm employing automatically reorganising ranking tries and word usage frequency.

The section on teaching presents the unconventional view that interactive-style teaching as usually employed at universities of applied sciences is equivalent — if not even superior — to many modern approaches of teaching, and introduces an autonomous system for online exercises, which allows students to enter their solutions at any time, and which provides automatically generated feedback with a high degree of detail.

With teaching, here a hub facilitating ubiquitous online learning, also deals the first one of the conference presentations, of which only abstracts are provided in this volume. The further ones present a distributed event-triggered control algorithm to achieve consensus in heterogeneous multi-agent systems' outputs, applications of chaos theory for encryption, managing complex networks and suppression of electromagnetical interferences as exemplified for power converters by aperiodic pulse width modulation techniques, a novel approach of combining responses of selected Gabor filters shifted by certain off-set vectors, which are automatically trainable for purposes of pattern recognition, and a method to estimate the values of graph parameters.

We are deeply indebted to Jutta Düring for her effort and care in achieving a consistent and appealing layout for this book, to Barbara Kleine for the administrative preparation of the conference, and to FernUniversität in Hagen for supporting the publication of this volume.

Herwig Unger
Wolfgang A. Halang

Contents

Safety-related and Real-time Systems

Networks and Routing

Suppression of Disturbances

Neural and Evolutionary Computing

Natural Language Processing

Education

Short Contributions

A Comparative Survey of System Specification Techniques for Safety-related Environments

Daniel Koß

FernUniversität in Hagen, Germany

Abstract: Strong specifications are fundamental for systems development in safety-related environments. These specifications can be established by a systematic and sustainable requirements engineering. There are several well-established techniques of modelling requirements, of which an overview will be presented. Starting by describing less formal textual methods of requirements modelling, more and more formal methods are introduced. Finally, the specification techniques shall be compared against essential requirements that are given by standards for developing safety-related environments.

1 Introduction

1.1 The Sense of Specifying

Specifications serve to describe the overall structure and the operational behaviour of a system consisting of hardware and software components. Out of these specifications the system shall be able to be constructed (in the form of a development work) as well as be able to be understood (for maintenance or further development purposes).

Specifications, as a general rule, are more abstract than a system itself. If they were complete, they would contain the system itself. Therefore, specifications can also present a partial, defined view to the system. Specifications only contain the necessary information for a special case of use. A software part of a specification, for example, will probably contain little information about the hardware of the system, except the necessary requirements to run the software. Instead, the hardware specification will probably contain little information about the software part, except the necessary requirements to run the software on it. A system specification, however, will have its focus on the functional behaviour and the presentation of the system to the inner and outer peripherals, which means the components that exist outside the system borders as well as the inner interfaces for software components. The inner details of the functionality of the hardware and software will be reduced to a necessary minimum to understand the overall functionality of the system.

Hardware and software as parts of an embedded system are dependent from each other, because without the one part the other part would hardly work, or, directly spoken, the one part makes no sense without the other part. Considering the software part separated from the hardware part, it could be hard to determine requirements from one part to the other timely and with the adequate attention. Within development phase of the software, for example, it may turn out, that the initially planned memory is not sufficient for the timely execution of the software. Here, a new requirement to the hardware is generated which claims to adjust the memory size.

Therefore, techniques and methods are necessarily needed to depict singular requirements to one part of the software or hardware component as well as relations and dependencies between the components. Especially in safety related environments, there is a strong need to model temporal conditions of the system's behaviour, for example the response time of a system, within which a functional correct answer has to be given. Furthermore, the behaviour of a system in the case of a failure must be specifiable. Within this paper an overview shall be presented of the state of the art of specification techniques as well as an analysis, if the techniques are suitable for specifying safety related requirements.

To achieve this goal, first requirements to specification techniques will be identified out of the standard IEC 61508. After that, well-established specification techniques will be investigated of their conformity against IEC 61508. Finally, a conclusion will be drawn, in what way these specification techniques are suitable to specify systems with safety related constraints.

2 Requirements for Specification Techniques according IEC 61508

The international standard IEC 61508 (entitled by ”Functional Safety of Electrical/Electronic/Programmable Electronic Safety-related Systems”) formulates requirements and instructions for the development of safety related systems with electrical, electronic and/or programmable components.

The parts 2 (”Requirements for electrical/electronic/programmable electronic safety-related systems”) and 3 (”Software requirements”) describe, among requirements on management, documentation, design and validation/verification, requirements on requirement specification techniques. In [1] and [2] the following essential requirements on specification techniques and methods can be found ([1, tables B.1, B.2] as well as [2, table A.1]):

- Structure [A1],
- Visitability/verifiability [A2],
- Use of (semi-) formal methods (see [2, table B.7]) [A3],
- Traceability [A4],
- Computer-based [A5],
- Modularization of components [A6].

These slightly "soft" formulated properties shall on the one hand ensure as much freedom as possible on the selection of specification techniques for individual project needs, on the other hand unambiguously stipulate which fundamental procedures have been well-established and are therefore necessary for specifying safety related systems.

The IEC 61508 even becomes more concrete when describing what a specification has to afford. The following properties have to be expressible by a specification technique (compare [1, p.17] and [2, p.22]):

- Requirements to hardware and software components [B1],
- Internal and external interfaces [B2],
- Throughput and real-time characteristics, that means, time constraints [B3],
- Accuracy and robustness of subcomponents [B4],
- Failure behaviour, faulty behaviour and start-up behaviour [B5],
- Composition and architecture of subcomponents [B6].

With these properties the overall structure of a specification aswell as the least mappable content is set.

Furthermore, according to the standard, the following aspects have to be considered to support the design, development and later comprehensibility of a system consisting of hardware and software components:

- Support of a development life-cycle (compare [1], p.12 as well as [2], p.16) [C1],
- Clarity, precision, unambiguity, verifiability, testability, maintainability, executability (compare [1], p.17) [C2].

The requirements, that have been withdrawn from the standard, have been divided into requirements categories. The requirements marked by [Ax] are considered as necessary requirements, which presence is explicitly called. Requirements marked by [Bx] are less explicitly called, because there may be a strong dependence to the concrete project. Possibly, the one or the other requirement can be left out if there is no need to apply. But this proceeding, that is what the standard says, needs to be reasoned. Last but not least, the requirements marked by [Cx] serve as supportive methods, because their availability is hard to impose and hard to measure. These methods ultimately aim on a guideline for a defined structure and not for a defined matter.

3 Specification and Modelling Methods

There can be a rough differentiation made between unstructured, structured and formal specification techniques. This classification is reasoned on a contrast of freedom on the one hand (which means, to model requirements as you want to in the actual context) and precision on the other hand (which means, to model requirements so that they are unambiguous).

Freedom and precision cannot be completely fulfilled simultaneously, therefore, depending on which development phase you are aiming or which case you are looking at, one has to decide, whether freedom or precision is more important. Naturally, the grade of freedom is continuously decreasing during ongoing development, because there have decisions to be made which reduce the remaining freedom. Simultaneously, the grade of precision increases continuously because of decisions that have been made and details that have to be established (see Figure 1).

Specifications firstly serve for supporting communication among people. Different instances of people have to formulate their thoughts and deliver them to other people, so that they are able to work with them or continue working on them. Because the receiving instance may belong to another company or even to another culture, it is necessary to communicate free of misunderstandings and standardized. Secondly, specifications serve for the construction of systems. This means, that the generated documents are the foundation for the

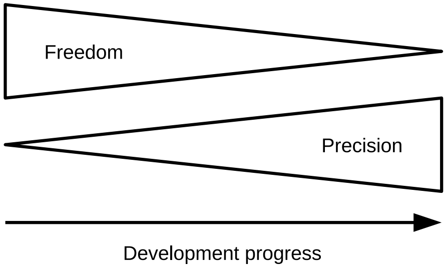


Fig. 1: Levels of detail of freedom and precision in increasing development progress

understanding and the implementation of the target system. For the implementation, specifications should have an adequate grade of detail, so that there is no possibility for interpretation. Therefore, we face the following conflicting objectives:

- Comprehensibility versus formalism.
- Clarity versus completeness.
- Communicability versus automatism.

These conflicting objectives are mainly a detailed list of the conflict of freedom versus precision, which has been shown in Figure 1. For specification techniques the conflicting objectives mean, that both of them could not simultaneously be fulfilled in the same quality. There are only two possibilities to resolve the conflicts:

- Using different specification techniques depending on the case of use,
- Using specification techniques, that are able to scale the level of detail depending on the case of use.

The surveyed specification techniques shall be analysed to what extent they are suitably usable within safety-related real-time data processing.

3.1 Unstructured Specification Techniques

The main characteristics of unstructured specification techniques are, that there are no formal, structural or methodical restrictions for writing them. Therefore, a relatively large degree of freedom is given which implies initial creativity and supports finding ideas. A disadvantage of this freedom is, that understanding a specification made this way requires a large amount of interpretation. This interpretation space can have a wide spread for project members that have different backgrounds, either organizational or even cultural. Eventually, the receiving user of the specification misses essential information, that would have been forced to give by the writing within a given structure. Therefore, an unstructured specification technique, like the ones presented as follows, can only be established with an external definition of a proceeding regulation.

Free Text

The probably most natural method to specify requirements is to write them down the way the come in mind, for example as a free text. An example for a customer requirement in a free text manner:

> "I wish that the system would not freeze that often."

A requirement, formulated in such an informal way, misses the following characteristics:

- Bindingness: it is not possible to determine, of which urgency this requirement is.
- Unambiguousness: there is a definition missing, what is meant by "freeze that often", that is, a quantification is missing.
- Precision: there is a definition missing, what the abstract term "freeze" exactly means and how this behaviour can be measured.

In a free text of a system specification, like the example above, it first must be identified what the real requirements to the system are and how it will later be possible to validate them. Therefore, a free text can only be the first step to collect and describe basic ideas and write down and structure first concepts.

It is not reasonably possible to use a free text as a specification document for suppliers or corresponding developers. For a requirement formulated like this, a transformation must be done into a format that allows a structured processing first.

Informal Visualization

The most familiar informal visualization technique is mind mapping. Mind maps allow structuring ideas and thoughts roughly and make it possible to connect them with lines and arrows to visualize relations among each other. See Figure 2 for a simple example of a mind map.

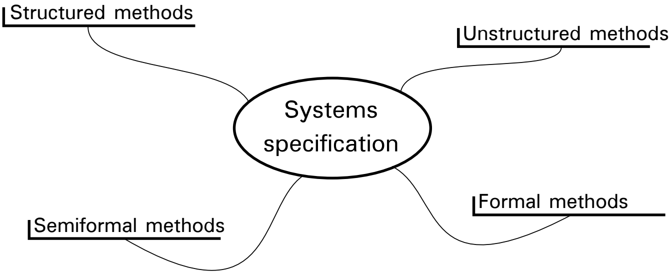


Fig. 2: Simple example of a mind map for structuring ideas

The advantage of having the freedom to paint roughly and quickly what is going through one's mind is also the biggest disadvantage of this method. Though mind maps and other informal visual specification techniques allow to bring a first structure into a specific topic, these methods are missing a structural proceeding. Therefore, their applicability for safety-related systems is normally not suitable. Every mind map can have a different structural appearance and there is much interpretative effort needed to extract relevant information out of a mind map. Normally, there is additional information needed to interpret the mind maps significance and meaning, for a example by a template.

3.2 Structured Specification Techniques

Formal Text

More purposeful than describing requirements in a free text manner is by undergoing the text to a formalism. This can be achieved, for example, with a text template wherein the composition of a requirement's structure is specified so that it contains all relevant information for the design as well as for the verification of the system. An example for a customer requirement in a formal text manner:

> ″The system shall answer correctly on a signal income within 100 milliseconds.″

This requirement fulfils the following formalism:

1. Who has to do something?
2. What has to be done?
3. How binding is the requirement?

In the example at hand the questions can be answered as follows:

1. The system has to do something.
2. Answering correctly after a certain time from a signal income.
3. It has to be fulfilled under all circumstances.

The system behaviour can be validated clearly and under several circumstances against the requirement, for example under average load or high load. The result is then a binary statement, with which it is possible to say the system passed or failed under the specified constraints.

An extension to the principle of text templates is the principle of document templates. With document templates the whole structure of a specification document is given. This affects the layout of the chapters as well as the proceeding of the document creation, for example by requirements elicitation, review techniques and release processes. A document created by this process is easier to understand, because not only the content, but also the way of creating the document is standardized. Furthermore, with document templates it is easier to communicate across organizational borders. Every participating organisation just needs to know the template to follow the structure of the document and is easily able to analyse the content. Further information is given in the standard ISO/IEC/IEEE 29148 (see [3]), which standardizes life cycles of system specification documents.

EPOS

The computer based system EPOS (Engineering and Project management Oriented Support system) serves for the holistic view on project planning of software development and supports every development phase of a software development process [4, p.67]. Thereof:

- Requirements elicitation,

- Project management,
- System design,
- Quality management.

To support the single project steps, EPOS is divided into different mark-up languages and supporting tools which are:

- EPOS-R: language for requirements elicitation,
- EPOS-S: language for system design specification,
- EPOS-A: analysis tool,
- EPOS-D: documentation tool,
- EPOS-P: project management language and
- EPOS-M: general management tool.

A requirement in EPOS-R looks like as follows:

```
REQUIREMENT 0815 (0):
<SafeState>
"Within safe state, the system
has to turn off all of its outputs."
```

An event on the systems layer, described in EPOS-S, looks like as follows:

```
EVENT CriticalFailure
DESCRIPTION:
      PURPOSE: "When a critical failure occurs,
        get into safe state."
      DESCRIPTIONEND.
INTERRUPT FROM RegularExecution
EVENTEND
```

A project resource can be described with EPOS-P in the following way:

```
TEAMMEMBER Meyer–M.
CATEGORY: 'Test', 'Project'
FUNCTION: "Testmanager"
RESPONSIBILITIES: Testing
ASSIGNMENTS:
      Testing 80%,
```

Requirements 20%
TEAMMEMBEREND

An advantage of this method is, that every project-related requirements and constraints can be integrated and linked together and then be accessed in a single system without changing the environment. Because every project member uses the EPOS tools with his or her own access rights it is guaranteed, that every member works on the same, consistent database. Additionally, requirements can be traced from design over implementation to test case. That simplifies project work tremendously.

It cannot be denied that a disadvantage of EPOS is the maintenance of the system. Working with EPOS means that every project member needs to spend additional effort continuously throughout the project lifetime and this has to be done with discipline and sustainability. Because once the database contains inconsistent data, every project member works with this inconsistent data, what provokes unforeseeable consequences for the project.

Specification Languages

TTCN-3

Another structured specification technique is the specification of requirements by test cases. This method is also known as "test driven development" and is aiming at the circumstance, that the correct implementation of a system is only verifiable by testing. If the test cases are given at the beginning of development, it is possible to develop in such a way to pass the test cases. At the same time the developer has the proof, that he or she worked correctly if the test cases passed.

For the test driven development one can use the text-based technique TTCN-3 (Testing and Test Control Notation version 3) which is a language specifically designed for the specification of test cases and test scenarios [6]. TTCN-3 syntactically follows well-known programming languages like C/C++ and Java.

As an example, a test case for a server application shall be developed which initially sends a Ping request to a server and awaits an Echo answer from the server within a defined time interval (follows [6, Listings 1-6]). The test description is structured as follows. First, the module description has to be defined, which defines interfaces and data types for the test case:

```
module pingTest{
    type record urlType {
        charstring protocol,
        charstring host
    }
    template urlType urlTemplate := {
        protocol := "ping",
        host := "www.fernuni-hagen.de"
    }
...
```

Next, the component description takes places, in which the expected answer from the server is described:

```
...
type component ptcType {
    port httpPortType httpPort;
    timer localTimer := 3.0;
}
type port httpPortType message {
    out urlType;
    in echo;
}
type component mtcType {}
type component systemType {
    port httpPortType httpPortMsg;
}
...
```

At last, the implementation of the test case itself is implemented, in which the modules and components are initialized and executed (with PTC = parallel test component, MTC = main test component):

```
...
testcase DoPing1 ( )
runs on mtcType system systemType {
    var ptcType ptcTester;
    ptcTester := ptcType.create;
    map(ptcTester : httpPort, system : httpPortMsg);
    ptcTester.start(ptcBehaviour());
    all component.done;
}
...
```

Additionally, sub functions, that are needed and called by the test case, are specified as follows:

```
...
function ptcBehaviour() runs on ptcType {
    httpPort.send(urlTemplate);
    localTimer.start;
    alt {
        []httpPort.receive(echo) {
            localTimer.stop;
            setverdict(pass);
        }
        []httpPort.receive {
            localTimer.stop;
            setverdict(fail);
        }
        []localTimer.timeout {
            setverdict(fail);
        }
    }
}
```

Finally, on top hierarchical layer, the functionality itself is executed by a control call. The test case in the example above is passed, if the Echo response is received within 3 seconds after the Ping request. Else, the test fails.

Furthermore, the programmed test scenarios can be viewed as graphical representations. There is no transformation of the textual description needed, because both representations are equivalent. The graphical representation can be used for communication purposes with other project members, for example with less technical versed ones.

Advantageous of working with TTCN-3 is the strict formalism of this specification technique as well as the proof, that the system has been developed correctly, when the test cases pass. Disadvantageous is, that TTCN-3 is only of limited usability for communicating between project members, because more or less expert knowledge is needed to understand the specified test environment.

Furthermore, dependencies between different states of the targeted system are hard to model, because there is a test case needed and needs to be maintained for every state and every state transition of the system.

Depending on the targeted project it can be an advantage or a disadvantage that test cases are specified only on top systems layer. There is no demand given if a test case requires implementation in software or hardware. This potentially can have disadvantages for the later reuse of single subcomponents.

It can be stated positively that a system specification with this technique explicitly allows to give time constraints and time-outs. This is an advantage for specifying real-time systems, because at any time the given time constraints are known and can be tested thoroughly.

SystemC

SystemC is a system modelling and description language which is an extension of the programming language C++ with components for hardware description. SystemC is especially useful to specify the following constructs on systems layer:

- Parallelism,
- Synchronism,
- Interprocess communication.

SystemC has special qualities in verifying and simulating specifications that have been designed in this language, because SystemC constructs are compilable and executable. Thereby, it is possible to verify components on a relatively high abstraction layer, with which statements of performance and robustness are possible early in development. A disadvantage of SystemC is that for interpreting a system description as a specification, deeper knowledge in C++ as well as additional knowledge in SystemC is mandatory. This means, that SystemC is less reasonable for requirements elicitation, but is more reasonable to design the system itself.

Beneath, a simple rising edge triggered flip flop is described as an exemplification how digital logic can be described in SystemC (from [5, p.12]):

```
// dff_pos_edge.h
#include "systemc.h"
```

```
SC_MODULE(dff_pos_edge){
    sc_in<bool> clk;
    sc_in<bool> din;
    sc_out<bool> dout;
    void doit();
    SC_CTOR(dff_pos_edge){
        SC_METHOD(doit);
        sensitive_pos << clk;
    }
};

// dff_pos_edge.cpp
#include "systemc.h"
#include "dff_pos_edge.h"

void dff_pos_edge::doit(){
    dout = din;
}
```

S-PEARL

S-PEARL (Specification Process and Experiment Automation Language) is a concept, which is an extension of the programming language PEARL and which uses a similar syntax and structure. PEARL is especially suitable to specify concurrency and time constraints in relation to real-time conditions, which is not intuitively supported by other programming languages. PEARL was especially designed for real-time data processing and supports time constraints within its fundamental functionality. S-PEARL picks up the concept of modularity and the overall structure of PEARL and can be used to specify hardware and software parts simultaneously (see [9]).

A processing unit, which processes an input sensor signal and emits an actor control signal, can be described as below (follows [9, p.13]):

```
ARCHITECTURE:
    STATIONS;
        NAMES: STAT;
        PROCTYPE: MSP430 AT 20 MHz;
         ...
        INTERFACE: MSP_IO
```

```
                (DRIVER: MSP_INOUT; DIRECTION: INOUT);
        STATIONTYPE: NORMAL;
        TICK: 1E-6 SEC;

    SYSTEM;
        NAMES: STAT;
            STAT.MSP_IO INPUT;
        NAMES: SENSOR;
            SENSOR.OUT OUTPUT;
        NAMES: ACTUATOR;
            ACTUATOR.IN INPUT;
    SYSEND;

    CONFIGURATION:
        COLLECTION MSP_PER;
            PORTS MSP_1, MSP_2;
            CONNECT MSP_PER.MSP_1
                IN SENSOR.OUT VIA STAT.MPS_IO;
            CONNECT MSP_PER.MSP_2
                OUT ACTUATOR.IN VIA STAT.MPS_IO;
        COLEND;

        MODULES MSP_Mod1;
            EXPORTS(T1);
            TASK T1
                TRIGGER PORT MSP_1;
                ...
                OUTPUT PORT MSP_2;
            TASKEND;
        MODEND;
    CONFEND;
ARCHEND;

NET;
    STAT.MSP_IO <-> SENSOR.OUT;
    STAT.MSP_IO <-> ACTUATOR.IN;
NETEND;
```

3.3 Semi Formal Modelling Methods

Systems Modelling Language

The Systems Modelling Language (SysML) is a modelling language designed for modelling systems, that consist of hardware and software components. It has arisen from the Universal Modelling Language (UML)which is a generic, graphical specification and design technique that has mainly been designed for software systems design [8, p.8]. Its diagrams can be subdivided into two categories with seven diagram types, respectively:

1. Structure diagrams,
2. Behaviour diagrams.

The following diagram types have been borrowed for SysML out of UML:

- Structure diagrams:
  - Use case diagram (allows modelling a system, its actors and system borders),
  - Sequence diagram (allows modelling a sequence of events),
  - State diagram (allows modelling states and transitions between the states),
  - Activity diagram (allows modelling elementary actions and their interconnections),
- Behaviour diagrams:
  - Package diagram (allows modelling logical groupings and their dependency relations),
  - Internal block diagram (allows modelling architectures and parts of it, for example subsystems),
  - Block definition diagram (allows modelling blocks and their internal interconnections as well as block interfaces).

The most important SysML specific diagram types are the following:

- Requirement diagram (behaviour diagram type, allows modelling system requirements).

- Parametric diagram (structure diagram type, allows modelling constraints and assurances to a continuously working (physical) system).

The diagram types exist hierarchically in parallel than hierarchically vertical. That means, that there is no specification which diagram type has to be used in a certain development phase. Different diagrams just allow different views onto the same system. Use case diagrams, for example, describe the system and its borders to the surrounding environment and its actors, respectively. State diagrams are mainly used to describe the internal behaviour of a system. Furthermore, diagrams are scalable nearly endlessly. Parts of a diagram can, for example, refer to more detailed sub diagrams, whereby the complexity of a diagram is significantly manageable. For a certain demand it is possible to dive into special sub diagrams that give a view onto certain partial functionality without losing the overall context. At the same time, a developer will not be overburdened by too much information that is not necessary for his or her actual work.

Because UML and SysML, respectively, were not primarily designed for the use within safety-relevant real-time environments, trade-offâs or modifications have to be made. It is the nature of these modelling languages that the can be customized by profiles domain specifically, so that they are tailored for the required design purpose.

An example for a profile that allows modelling real-time constraints is the MARTE profile (Modelling and Analysis of Real-Time Embedded Systems) [7] which extends UML. With this profile there is a special aspect of real-time modelling highlighted, that is, time. It especially allows defining time periods and time limits, respectively, within which a system has to provide a functional correct answer. MARTE distinguishes between the following flavours of time: [7, p.57]:

- Asynchronous, in the form of a temporal order (causal),
- Synchronous, in relation to a time basis (discrete) and
- Physical, that means expressing time periods by units.

As a practical example, a system consisting of a hardware and a software component shall be modelled. Therefore, the following requirements are elicited:

- A sensor input shall be read,
- An actor output shall be set,
- The sensor signal shall be transformed in a certain way,

- After the input signal comes up, the output of the actor signal shall be finished within a defined time limit.

The following diagrams will be created:

- use case diagram, for separating the system from its environment and defining the actors (Figure 3),
- sequence diagram, for modelling the chronological sequence of the activities (Figure 4).

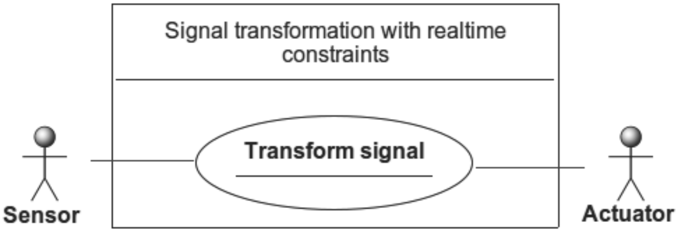


Fig. 3: Use case diagram with system boundaries

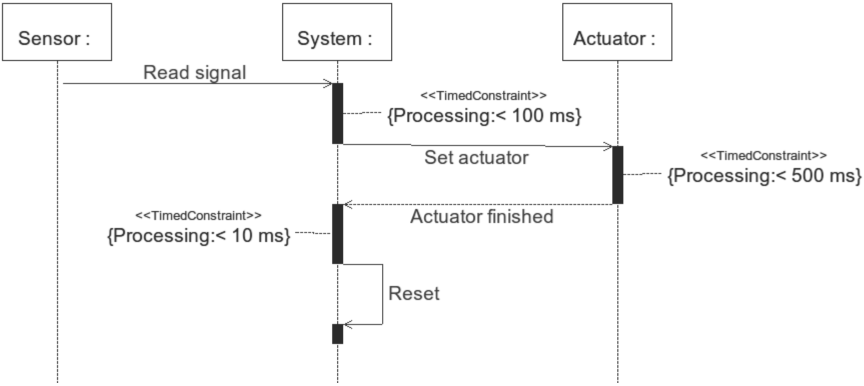


Fig. 4: Sequence diagram of communication interaction with timed constraints

3.4 Formal Modelling Methods

Formal modelling methods mean to be correct by proof. That means, that one is able to examine the correctness of a system against its formal specification. If the examination is successful, the proof of correctness is adduced. The nature of formal methods is their formalism, that means strict notation and absolute precision. This makes them unambiguous, but maybe hard to understand and less flexible. One example for such a formal method is the concept of petri nets, with which concurrent systems can be described formally. The most important characteristics are (compare [4, p.351]):

- Clarity,
- Modelling of dynamic behaviour,
- Abstraction,
- Verifiability,
- Computer-based.

A (marked) petri net is defined by PM = (P, T, Pre, Post, m_0) with

- P: a finite set of places (states),
- T: a finite set of transitions,
- $P \cap T = \emptyset$, that means, that places cannot be transitions and vice versa,
- Pre: an input arc function with $P \times T \rightarrow \mathbb{N}$,
- Post: an output arc function with $T \times P \rightarrow \mathbb{N}$,
- m_0: initial marking.

As an example, a flip flop element shall be modelled as a petri net. The flip flop shall be a 4-tuple of the form (P, T, Pre, Post). Then the following shall apply:

$$P = \{p_1, p_2, p_3, p_4, p_5, p_6\}$$

$$T = \{t_1, t_2\}, \text{mit } P \cap T = \emptyset$$

$$Pre(p_1, t_1) = Pre(p_2, t_1) = Pre(p_3, t_2) = Pre(p_4, t_2) = 1$$

$$Post(t_1, p_5) = Post(t_1, p_3) = Post(t_2, p_2) = Post(t_2, p_6) = 1$$

Where $Pre(p_n, t_m)$ is an arc from place P to transition T and $Post(t_m, p_n)$ is an arc from transition T to place P. All input and output arcs that are equal to zero have been omitted because of clarity. A visualization of this petri net is given in Figure 5. It is obvious, that the presentation of petri nets in formulas can be confusing, if complex behaviour with several external influences has to be modelled. At the same time, error potential increases if petri nets are not modelled computer-based or with automatic validation methods. The advantage of graphical representations of petri nets is, that complex interactions and relations can be perceived quickly and dynamic behaviour can be presented clearly.

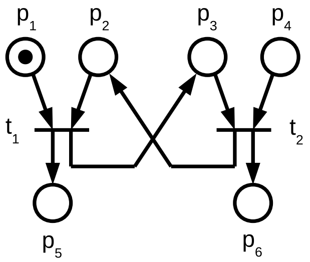


Fig. 5: A flip flop logical circuit modelled by a graphical petri net with initial marking of the set input

4 Comparison

Finally, the presented specification and modelling techniques shall be compared against to the requirements A1 - A6 out of chapter 2.

Method	Structural	Verifiable	Semiformal	Traceable	Computerized	Modular
Free text	-	-	-	-	-	o
Mindmaps	o	-	o	-	+	+
Formal text	+	o	-	-	o	o
EPOS	+	o	o	+	+	o
TTCN-3	+	+	+	o	+	+
SystemC	+	+	o	+	+	+
S-PEARL	+	+	o	+	+	+
SysML	+	o	+	-	+	+
Petri nets	+	+	o	-	o	o

Furthermore, a classification of its level of detail is given in Figure 6 in comparison to Figure 1. In this diagram one can see which specification techniques might be more useful at a specific step in the process of system development, though it is not forbidden to use a technique out of the order when it makes sense. The use of a specification technique strongly depends on the specific project needs.

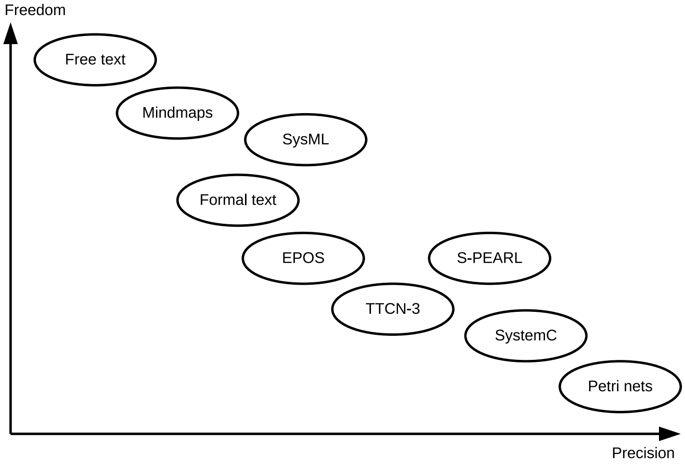


Fig. 6: Qualitative classification of the techniques' level of detail

5 Conclusion

Unstructured, non-traceable techniques for the specification of safety-related embedded systems have become obsolete. The use of structural methods, that allow a traceability through the whole development life cycle, is state of the art.

A huge gain of value is provided by specification methods, that allow an implementation directly out of the specification, like SystemC or S-PEARL do. With this proceeding the effort that needs to be spent, in contrast to using different specification techniques, is less. Unfortunately these methods are hard to understand for personnel without deeper technical background. Therefore, these methods, at their actual function volume, can at the earliest be used in detailed specification phase, and not likely in the requirements elicitation at the beginning. Here, UML and formal text methods show benefits at early concept phase, where SystemC, TTCN-3 and S-PEARL have advantages at late specification phase until early implementation phase.

Future investigations could go into the direction of an automatized tool chain that supports handling requirements from the elicitation over implementation until testing and releasing, where human interference is driven down to an unavoidable minimum.

References

[1] International Electrotechnical Commission: IEC 61508-2 - Part 2: Requirements for electrical/electronic/programmable electronic safety-related systems. 2010

[2] International Electrotechnical Commission: IEC 61508-3 - Part 3: Software requirements. 2010

[3] International Standardization Organization et al.: ISO/IEC/IEEE 29148 - Systems and software engineering – Life cycle processes – Requirements engineering. 2011

[4] Halang, W.A., Li, Z.: Prozessautomatisierung/Echtzeitsysteme II. FernUniversität in Hagen, 2009

[5] Muhr, H.: Einsatz von SystemC im Hardware/Software-Codesign. Diplomarbeit an der Technischen Universität Wien, Fakultät für Elektrotechnik, 2000

[6] Schieferdecker, I. et al.: The Test Technology TTCN-3. In: Hierons, R.M. et al. (Eds.): Formal Methods and Testing, Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2008

[7] Object Management Group: UML Profile for MARTE: Modeling and Analysis of Real-Time Embedded Systems Version 1.1. OMG document number: formal/2011-06-02. `http://www.omg.org/spec/MARTE/1.1`

[8] Korff, A.: Modellierung von eingebetteten Systemen mit UML und SysML. Spektrum Akademischer Verlag. Heidelberg, 2008

[9] Gumzej, R.: Engineering Safe and Secure Cyber-Physical Systems. Springer International Publishing. Switzerland, 2016

Requirements for Safe Computer Architectures

Stefan Widmann

Chair of Computer Engineering, FernUniversität in Hagen, Germany

Abstract: Embedded systems are used in rising numbers of safety-related applications, e.g. steer-by-wire and brake-by-wire in automotive applications. At the same time, complexity of hard- and software of such systems is increasing and the reduction of the minimum feature sizes of used integrated circuits makes them more sensitive to environmental influences, resulting in a rising number of software and hardware errors. Instead of trying to detect those errors by even more complex software, 23 requirements for safe computer architecture hardware are presented that have been derived from the standard IEC 61508 parts 2 and 3 and various typical software errors. The most commonly used conventional architectures x86 and ARM are evaluated based on these requirements.

1 Introduction

Safety-related applications that have been realized using pure mechanical or electromechanical systems before are more and more realized by programmable systems today. Good examples are automotive and avionic applications like braking, steering and flying: brake-by-wire, steer-by-wire and fly-by-wire replace the proven and reliable mechanical systems by electrical and electronic sensors and actors, the sensors' signals being processed by microprocessors to calculate control signals for the actors [26–28].

At the same time the complexity of software used in embedded systems is rising. In 2006, Broy states in [29], that a car contains about 10 million lines of code and that he expects a rise by the factor ten with each car generation. Three years later, in 2009, Broy is cited in [31], stating that the number of lines of code has risen to 100 million. In 2008, a future rise to 200 to 300 millions was estimated in [30].

Complexity of hardware is rising, too. Over 70 embedded systems were used in a car in 2006 [29] and up to 100 in 2009 [31]. On chip level, the number transistors per die has reached 1.3 billion in 2015 [34]. The integration of growing numbers of components in integrated circuits is only possible due to a continuous reduction of the minimum feature size, making them more sensitive to environmental influences like radiation, especially neutrons, even on ground level [32, 33].

All of this results in rising numbers of software and hardware errors. Since commercial-off-the-shelf (COTS) microprocessors commonly used in such systems are not designed to provide effective means for error prevention and detection on hardware level, more and more complex software is used to detect errors during runtime, e.g. by applying arithmetic coding in the form of Software Encoded Processing (SEP) and Compiler Encoded Processing (CEP) [25]. This additional software complexity results in additional software errors.

Instead of continuing to focus on error detection on a software level, 23 requirements for the hardware of a safe computer architecture have been derived from the international standard for safety-related system, the IEC 61508, and various descriptions of frequent and dangerous software errors. By applying these requirements to the hardware of a computer system, they enforce error prevention and detection on a superior level.

The most commonly used conventional computer architectures – x86 and ARM [21] – are evaluated regarding their conformance to these requirements.

2 Requirements for safe computer architectures

Hardware is designed, implemented and tested once, while software goes through its complete lifecycle in every new project. By applying requirements for software to the underlying hardware, software conformance to the requirements can be enforced on hardware level and a superior overall error prevention and detection can be achieved.

2.1 Requirements derived from IEC 61508

The most important source regarding requirements for the design of safety-related systems is the international standard IEC 61508, consisting of seven parts, -1 to -7. Part 2 describes the requirements for the hardware design of safety-related systems, part 3 does the same for software design and part 7 contains extensive details for both parts. All of the requirements listed below are derived from the IEC 61508 parts 2, 3 and 7 [5–7].

Detection, reporting and handling of errors

For the domains software, hardware, sensors and actors there shall be means to detect and handle errors. It is also necessary to report detected errors.

R 1: A safe architecture shall provide features to detect, report and handle errors in software, hardware, sensors and actors.

Means to enter and stay in a safe state

During design of software and hardware means shall be realized allowing the system to enter and stay in a safe state in case of errors, if such a safe state exists in the specific system.

R 2: A safe architecture shall provide means to enter and stay in a safe state.

Means for error detection

Redundant configurations and especially diverse redundancy shall be possible for error detection.

R 3: A safe architecture shall be able to provide (diverse) redundant configurations.

To detect errors in volatile and non-volatile memories, in registers and during transmission of data, error detecting Hamming or polynomial codes shall be used.

R 4: A safe architecture shall apply integrity checks like Hamming or polynomial codes for error detection.

If Hamming codes are used for error detection, it shall be realized as Extended-Hamming-Code, increasing the Minimum Hamming Distance to 4 by adding an additional parity bit.

R 5: If using a Hamming Code for error detection, it shall be realized as Extended-Hamming-Code with a Minimum Hamming Distance of four.

Even if the integrity checking means allow error correction, data that has been identified as corrupted shall not be corrected, since – depending on the number of erroneous bits – a correction attempt could result in a valid, yet wrong data word, without the possibility to detect this faulty correction on a higher level.

R 6: A safe architecture shall not try to correct data identified as faulty, but reject it instead.

Error handling

A safe architecture shall provide means to handle occurring errors and shall allow graceful degradation, e.g. by using redundant units or running less time consuming data processing functions with less accuracy.

R 7: A safe architecture shall provide means to handle occurring errors and shall allow graceful degradation.

Error prevention and good maintainability

The development of safety-related software requires error prevention by application of measures to keep complexity as low as possible. The proposed way to achieve this are modularity, encapsulation and information hiding, providing software maintainability with a low probability of introducing additional errors.

R 8: A safe architecture shall support modularity of software at hardware level by providing means for encapsulation and information hiding.

Realization of time driven software architectures

A cyclic and absolutely deterministic behavior of software is required, with guaranteed response times. This shall be achieved by realization of time driven architectures with fixed execution time slots for the different software parts. This leads to the requirement for only supporting synchronous programming.

R 9: A safe architecture shall support synchronous programming and time driven software architectures by hardware means and shall not support asynchronous programming.

Avoidance of dynamic objects

It is a frequent software error, not to free all dynamically allocated memory after being used (so called memory leaks), resulting in a rising memory consumption over runtime. Such errors are hard to find during testing. If a system runs out of memory, handling these type of error is very complicated in an embedded system.

R 10: A safe architecture shall not support dynamic objects.

Strong typing

Strongly-typed programming languages are to be used to prevent data handling errors without any implicit type conversions. The same requirement can be applied to the underlying hardware:

R 11: A safe architecture shall provide strong typing without implicit type conversions on a hardware level.

When doing explicit type conversions, strict rules shall be applied.

R 12: A safe architecture shall apply a strict set of rules for data type conversions.

Limitation of language features

If there are unsafe or hard-to-test language features, coding guidelines shall forbid the use of such features. A safe architecture can support this requirement by preventing the realization of problematic features.

R 13: A safe architecture shall prevent the realization of unsafe language features and thus limit them to a safe and easy-to-test subset.

Structured programming

For improved error prevention and increased maintainability structured programming shall be used. Especially, this is important when using low-level languages like assemblers. A safe architecture shall enforce structured programming to support this requirement.

R 14: A safe architecture shall enforce the use of structured programming.

Avoidance of pointers

Pointers are being provided by many programming languages for easy access of data. They are especially used when handling dynamic objects in memory. Common errors regarding pointers are usage of Null pointers, accessing data outside of arrays, structures and objects and accessing freed dynamic objects. These types of errors are hard to uncover during testing, hard to analyze and do often show up after long time.

R 15: A safe architecture shall not support the use of pointers, but provide means to realize safe array handling, including detection of out-of-bounds accesses as errors.

Avoidance of recursions

Recursive programming produces hard-to-test programs, since correctness is hard to prove. This applies especially for the prove of convergence over the whole range of input values. An additional problem is the dependency of stack memory consumption on the input values.

R 16: A safe architecture shall not support recursion, leading to a reduction of realizable language features (see R 13, too).

Application of special rules regarding floating point arithmetics

When using floating point data types, serious programming errors can be implemented regarding sequence of evaluation and checking equality of floating point values. While the sequence of evaluation can only be influenced on a software basis, a safe architecture shall not provide means to test for equality of floating point data types.

R 17: A safe architecture shall not provide instructions to check equality of floating point data types.

Avoidance of interrupts

Interrupts weaken the determinism of temporal behavior of software. They can be avoided, if events are polled instead, making their handling absolutely deterministic.

R 18: A safe architecture shall not support interrupts.

2.2 Requirements derived from typical software errors

Further requirements are derived from typical software errors based on [3, 10–15] and directed towards the hardware of safe computer architectures.

Usage of uninitialized variables

Usage of data without assigning a meaningful value first is a frequent programming error [13]. Based on the initialization of the runtime environment of the used programming language and maybe an operating system, this can result in the usage of wrong or random values.

R 19: A safe architecture shall detect usage of uninitialized variables as error.

Out-of-bounds accesses in arrays

Out-of-bounds accesses – accessing data outside of an array, structure or object – are frequent programming errors [11, 14, 15], which can cause memory corruption, information disclosure vulnerabilities and wrong computation results. In [15], this type of error is on the 3rd place of the 25 most dangerous programming errors. Since arrays are vital to most programming languages and implementations, a safe architecture shall provide means for safe array handling, as already specified in R 15.

Pointer related errors

As already mentioned before, pointers bare the risk of dangerous and hard to detect errors [13, 14]. This results in the already formulated requirement R 15.

Errors regarding data type interpretation and type conversion

Errors regarding interpretation of variable's contents and type conversion errors have caused huge damage [10, 12, 14]. A corresponding requirement has already been formulated in R 11.

Arithmetic under- and overflows

Under- and overflows can happen during arithmetic operations and are not always detected and handled properly [12, 15]. This type of error is on the 24th place of the 25 most dangerous programming errors in [15].

R 20: A safe architecture shall provide means to realize safe arithmetics, being able to detect under- and overflows as errors.

Division by zero

A division by zero is undefined and results in raising an exception on most architectures [3, 12].

R 21: A safe architecture shall detect divisions by zero as errors.

Missing checks of function parameters and return values

Many typical software vulnerabilities could be avoided if strict checks would be applied to input and output data [11]. This applies to a system's input and output data as well as on parameters passed to a function and its return values on the lowest level.

R 22: A safe architecture shall provide means to check function parameters and return values.

2.3 Requirement regarding avoidance of arithmetic registers

Nearly all conventional architectures provide a set of registers, some of them dedicated to special functions, others for general purpose. Tanenbaum and Goodman demand in [17], that powerful architectures shall have at least 32 general purpose registers. Management of those registers by compilers is called register allocation and is complex to realize [4, 18]. To determine the validity of the contents of registers, the variables those contents correspond to and how long they must stay valid, highly complex algorithms are used, called live variable analysis [18]. In [1] an architecture without arithmetic registers is introduced, resulting in less complex compilers and thus a less probability of compiler errors.

R 23: A safe architecture shall not provide arithmetic registers.

3 Conventional architectures x86 and ARM

The most commonly used conventional architectures are x86 and ARM [21]. While the x86 architecture is typically used in desktop PCs and industrial computers, ARM processors are predominantly used in mobile devices such as smartphones and tablets, due to their low power consumption.

x86 architecture

Although first steps towards advanced protection features were implemented in the Intel 80286, the Intel 80386 was the first x86 processor that provided a fully featured safe operation mode called protected mode [3], which provides two different types of protection, segmentation and paging.

In segmentation data structures called descriptors are used, specifying a segment's base address, the segment's size and its properties. There are code and data descriptors, identifying code and data segments as such. The contents of code segments are always non-writable and the contents of data segments are always non-executable. By selecting the appropriate properties in the descriptor of a segment, a code segment can be made non-readable and a data segment can be flagged as read-only. Any violation of these rules will be detected by the processor and a general protection fault is raised [3]. Another important property of a descriptor is the specification of its privilege level, expressing the least privilege needed to access the contents of a corresponding segment. The 80386 provides 4 privilege levels and they are often called rings. Code and data running in a ring 0 segment has or requires the highest privileges, code and data in ring 3 the least. In practice, only two of the four privilege levels are used, e.g. in the operating system Linux [19]: ring 0 is used by the operating system kernel and drivers and ring 3 is used by user code.

The paging mechanisms, hierarchically located below segmentation, provide two levels of privilege, supervisor and user, and specify, whether a page is read-only or read- and writable. There was no possibility to mark pages as non-executable until AMD introduced the NX bit in the page table entries in the Athlon 64 processor [24]. That means, prior to that extension, there was no way to prevent execution of data if segmentation was not used to the full extent.

Although segmentation provides strong safety features, in practice the two most commonly used operating system Microsoft Windows [8] and Linux [19] only use a flat memory model, which was proposed in [3]. In a flat memory model, code and data segments of a program are set to the same base addresses and the length of the segment is set to maximum, making it easy for the operating system to access a program's whole memory. This disables nearly all protection features offered by segmentation, leaving mainly only those provided by paging. Segmentation and its protection features aren't available in the modern 64 Bit mode of x86 processors [9], so the only available protection features are those of the paging functionality provided by the Memory Management Unit MMU.

ARM architecture

The company behind the ARM architecture, ARM Ltd., does not produce processors itself. Instead it is granting licenses to other companies to integrate ARM processor designs into their own products [23]. Those designs are optimized towards low power consumption and are widely used in embedded devices such as smartphones and tablets [21].

An ARM processor offers up to nine different modes of operation, of which the three relevant modes are supervisor mode for operating systems, system mode for privileged applications and user mode for standard user applications [2, 20].

In contrast to the x86, ARM processors do not provide segmentation as a protection feature. All encapsulation is to be done using the paging mechanisms of the integrated Memory Management Unit MMU [20]. First implemented in ARMv6, the XN bit (eXecute Never), similar to the NX bit of the x86, allows the software to mark memory pages as non-executable to protect data from being interpreted as code. Not all ARM processor designs raise exceptions on a division by zero, some just return zero as the result [20]. The memory architecture can be – depending on the selected processor design – realized as von-Neumann or Harvard architecture [22].

4 Evaluation of x86 and ARM architectures

Based on their conformance to the 23 requirements for safe computer architectures, both conventional architectures x86 and ARM shall be evaluated. Each requirement is judged based on the level of conformance to the requirement each of the architectures is providing: full conformance, partial conformance and no conformance to a requirement.

R 1: A safe architecture shall provide features to detect, report and handle errors in software, hardware, sensors and actors

x86 and ARM provide full conformance: Both conventional architectures provide means for error detection, signalization and handling, e.g. in the form of exceptions.

R 2: A safe architecture shall provide means to enter and stay in a safe state

x86 and ARM provide partial conformance: Entering a safe state must be requested by the software itself in the corresponding exception handlers, since there is no dedicated supervisory hardware instance.

R 3: A safe architecture shall be able to provide (diverse) redundant configurations

R 7: A safe architecture shall provide means to handle occurring errors and shall allow graceful degradation

x86 and ARM provide full conformance: Redundant configurations are possible in x86 and ARM based systems, as well as graceful degradation, but such configurations must be handled by software.

R 4: A safe architecture shall apply integrity checks like Hamming or polynomial codes for error detection

R 5: If using a Hamming Code for error detection, it shall be realized as Extended Hamming Code with a Minimum Hamming Distance of four

R 6: A safe architecture shall not try to correct data identified as faulty, but reject it instead

x86 and ARM provide partial conformance: Some systems – especially servers – based on x86 and ARM provide error correcting codes ECC. This ECC is usually realized using an Extended-(72,64)-Hamming-Code [16], but it only covers the contents of the external memory and the connection to the memory controller. A comprehensive check of all data paths throughout the system including all of the processor's internal data paths and registers isn't provided. Detected errors are corrected if possible, which violates requirement R 6.

R 8: A safe architecture shall support modularity of software at hardware level by providing means for encapsulation and information hiding

x86 provides full conformance: The x86 provides isolation features in the Protected Mode based on segmentation and paging. Both mechanisms are complex and must be managed by software, e.g. the operating system. Using the segmentation in full extent is omitted in Windows and Linux [8, 19], and only the protection features of paging come into effect. The x86 cannot provide protection on a single data word level, but requirement R 8 must still be considered as fulfilled using segmentation and paging.

ARM provides full conformance, too: The ARM architecture does not provide segmentation, but provides isolation measures based on paging. It doesn't provide protection on data word level, but requirement R 8 must still be considered as fulfilled using paging.

R 9: A safe architecture shall support synchronous programming and time driven software architectures by hardware means and shall not support asynchronous programming

x86 and ARM provide no conformance: Both architectures provide means to support asynchronous programming, making them non-conforming to requirement R 9.

R 10: A safe architecture shall not support dynamic objects

x86 and ARM provide no conformance: The conventional architectures x86 and ARM do not limit the software in matters of memory handling, making it easy to implement and use dynamic objects.

R 11: A safe architecture shall provide strong typing without implicit type conversions on a hardware level

R 12: A safe architecture shall apply a strict set of rules towards data type conversions

X86 and ARM provide no conformance: Neither x86 nor ARM architectures provide means to identify data types of memory contents. The same applies to the differentiation of code and data. The interpretation of memory contents is based on how the software tries to access them, e.g. trying to execute them or apply specific instructions.

R 13: A safe architecture shall prevent the realization of unsafe language features and thus limit them to a safe and easy-to-test subset

R 16: A safe architecture shall not support recursion, leading to a reduction of realizable language features

x86 and ARM provide no conformance: Both architectures don't restrict the usable language features, and recursive function calls are not detected as errors.

R 14: A safe architecture shall enforce the use of structured programming

x86 and ARM provide no conformance: The conventional architectures x86 and ARM don't enforce usage of structured language features.

R 15: A safe architecture shall not support the use of pointers, but provide means to realize safe array handling, including detection of out-of-bounds accesses as errors

x86 and ARM provide no conformance: Pointers are supported in both architectures and explicitly supported by corresponding addressing modes. Checking of pointers and range checks when accessing arrays must be implemented in software.

R 17: A safe architecture shall not provide instructions to check equality of floating point data types

x86 and ARM provide no conformance: If a Floating Point Unit FPU exists, both architectures provide explicit instructions to check equality of floating point values.

R 18: A safe architecture shall not support interrupts

x86 and ARM provide no conformance: Interrupts are supported by both architectures. Applications that do not use interrupts but poll for occurred events instead can be realized, but the hardware does not enforce polling. thus the requirement cannot be considered as fulfilled.

R 19: A safe architecture shall detect usage of uninitialized variables as error

x86 and ARM provide no conformance: Data in memory cannot be flagged as uninitialized on x86 and ARM architectures. The protection features of paging can be used to raise exceptions on read accesses to uninitialized memory pages with coarse granularity, but there are no means to realize this on a single data word level. This requirement cannot be considered as fulfilled.

R 20: A safe architecture shall provide means to realize safe arithmetics, being able to detect under- and overflows as errors

x86 provides no conformance: The x86 architecture does not detect arithmetic under- or overflows as errors, software must explicitly check for them.

ARM provides partial conformance: Some ARM designs support saturating additions and subtractions. This leads to a partial conformance to requirement R 20.

R 21: A safe architecture shall detect divisions by zero as errors

x86 provides full conformance: The x86 generates a division by zero exception in case software tries to divide by zero.

ARM provides only partial conformance: Not all ARM designs raise an exception on trying to divide by zero. Some return zero as the result [20], which cannot be distinguished from a correct result. Therefore this requirement can only be considered as partially fulfilled.

R 22: A safe architecture shall provide means to check function parameters and return values

x86 and ARM provide no conformance: No checks are applied to function parameters and return values unless the checks are implemented in software.

R 23: A safe architecture shall not provide arithmetic registers

x86 and ARM provide no conformance: Both architectures provide general purpose registers that can be used for arithmetic operations.

Evaluation results

The results of the evaluation are shown in table 1, which shows the low conformance of both conventional architectures x86 and ARM. While the x86 architecture conforms to 5 requirements, the ARM architecture conforms to only 4 of them. There are 14 requirements the x86 and 13 the ARM architecture does not satisfy at all. In contrast, the Inherently Safe Microprocessor Architecture ISMA, which was introduced in [35], provides full conformance to 22 and partial conformance to 1 of the requirements.

References

[1] H. Stieger, W. A. Halang: Eine hochsprachenorientierte Rechnerarchitektur ohne arithmetische Register; 1st edition, 2003; IFB Verlag Paderborn; ISBN 3-931263-39-8

[2] U. Brinkschulte, T. Ungerer: Mikrocontroller und Mikroprozessoren; 3rd edition, 2010; Springer Verlag; ISBN 978-3-642-05397

[3] Intel: 80386 System Software Writer's Guide; 1991; ISBN 1-55512-023-7

[4] R. Güting, M. Erwig: Übersetzerbau; 1999; Springer Verlag; ISBN 978-3-540-65389-9

Table 1: Evaluation and comparison of x86, ARM and ISMA

Requirement	x86	ARM	ISMA
R 1	+	+	+
R 2	(+)	(+)	+
R 3	+	+	+
R 4	(+)	(+)	+
R 5	(+)	(+)	+
R 6	(+)	(+)	+
R 7	+	+	+
R 8	+	+	+
R 9	-	-	+
R 10	-	-	+
R 11	-	-	+
R 12	-	-	+
R 13	-	-	+
R 14	-	-	+
R 15	-	-	+
R 16	-	-	+
R 17	-	-	+
R 18	-	-	+
R 19	-	-	+
R 20	-	(+)	+
R 21	+	+	+
R 22	-	-	(+)
R 23	-	-	+

+: full conformance, (+): partial conformance, -: no conformance

[5] IEC 61508-2:2010: Functional safety of electrical / electronic / programmable electronic safety-related systems - Part 2: Requirements for electrical / electronic / programmable electronic safety-related systems (Edition 2.0, 2010-04)

[6] IEC 61508-3:2010: Functional safety of electrical / electronic / programmable electronic safety-related systems - Part 3: Software requirements (Edition 2.0, 2010-04)

[7] IEC 61508-7:2010: Functional safety of electrical / electronic / programmable electronic safety-related systems - Part 7: Overview of techniques and measures (Edition 2.0, 2010-04)

[8] H.-P. Messmer: PC-Hardwarebuch; 6th edition, 2000; Addison-Wesley Verlag; ISBN 3-8273-1461-5

[9] AMD: AMD64 Architecture Programmer's Manual Vol. 2: System Programming; Publ. No. 24593; Rev. 3.23; 2013

[10] J.-L. Lions et al.: Ariane 501 Inquiry Board report; 1996; `http://esamultimedia.esa.int/docs/esa-x-1819eng.pdf`

[11] SAFECode, S. Simpson et al.: Fundamental Practices for Secure Software Development; 2nd edition, 2011; `http://www.safecode.org/publications/SAFECode_Dev_Practices0211.pdf`

[12] R. B. Dannenberg, W. Dormann, D. Keaton, R. C. Seacord, D. Svoboda, A. Volkovitsky, T. Wilson, T. Plum: As-if Infinitely Ranged Integer Model; 2010 IEEE 21st International Symposium on Software Reliability Engineering; pp. 91–100; 2010

[13] Unknown author (TU Munich): Programmierfehler und ihre Behebung; `http://www.in.tum.de/fileadmin/user_upload/Lehrstuehle/Lehrstuhl_XV/Teaching/WS10_11/Einführung_in_die_Informatik/16.pdf`

[14] M. Kompf: Die 12 häufigsten Programmierfehler; `http://cplus.kompf.de/artikel/errc.html`; retrived 18.01.2014

[15] MITRE Corporation: 2011 CWE/SANS Top 25 Most Dangerous Software Errors; `http://cwe.mitre.org/top25/`; retrieved 18.01.2014

[16] I. Koren, C. Krishna: Fault-Tolerant Systems; 1st edition, 2007; Morgan Kaufmann Verlag; ISBN 978-0-12-088525-1

[17] A. Tanenbaum, J. Goodman: Computerarchitektur; 1st edition, 2001; Pearson Studium; ISBN 3-8273-7016-7

[18] H. Falk: Skript Compilerbau, Kapitel 7 Register-Allokation; `http://ls12-www.cs.tu-dortmund.de/daes/media/documents/teaching/courses/ws0910/cb/cb-falk-7.pdf`

[19] D. Bovet, M. Cesati: Understanding the Linux Kernel; 1st edition, 2000; O'Reilly Verlag; ISBN 0-596-00002-2

[20] ARM Limited: Migrating from IA-32 to ARM; Application Note 274; ARM DAI 0274; 2011

[21] IC Insights: Qualcomm and Samsung Pass AMD in MPU Ranking; `http://www.icinsights.com/news/bulletins/Qualcomm-And-Samsung-Pass-AMD-In-MPU-Ranking/`; retrieved 03.03.2014

[22] H.-C. Chi: ARM Processor Architecture; CSIE34600 Introduction to Embedded System Design; `http://soc.csie.ndhu.edu.tw/source/introemb-13/unit2.ppt`

[23] H.-C. Chi: ARM Instructions; CSIE34600 Introduction to Embedded System Design; `http://soc.csie.ndhu.edu.tw/source/introemb-13/unit1.ppt`

[24] J. Breeden II: 'No Execute' Flag Waves Off Buffer Attacks; Washington Post, 2005; `http://www.washingtonpost.com/wp-dyn/articles/A55209-2005Feb26.html`

[25] U. Schiffel: Hardware Error Detection Using AN-Codes; 2011; PhD thesis; Technische Universität Dresden

[26] AIRBUS: Fly-by-wire; `http://www.airbus.com/innovation/proven-concepts/in-design/fly-by-wire/`

[27] NISSAN: Nissan Pivo Concept Press Kit: Overview `http://nissannews.com/en-US/nissan/usa/releases/435dd488-658e-433a-a57a-cd0184e4b51c`

[28] N. Shimizu: Nissan Puts Steer-by-Wire on the Road: An In-Depth Look at the Technology; Nikkei BP Japan Technology Report / A1403-058-005

[29] M. Broy: Challenges in Automotive Software Engineering; ICSE '06 Proceedings of the 28th international conference on Software engineering, pp. 33–42; 2006

[30] S. Ramesh: Software's Significant Impact on the Automotive Industry; Frost & Sullivan Market Insight; 2008

[31] R. N. Charette: This Car Runs on Code; `http://spectrum.ieee.org/transportation/systems/this-car-runs-on-code`; 2009

[32] R. C. Baumann, E. B. Smith: Neutron-Induced Boron Fission as a Major Source of Soft Errors in Deep Submicron SRAM Devices; Reliability Physics Symposium, 2000. Proceedings, 38th Annual 2000 IEEE International; pp. 152–157; 2000

[33] E. Normand: Single Event Upset at Ground Level; IEEE Transactions on Nuclear Science, Vol. 43, Issue 6; pp. 2742–2750; 1996

[34] SEMI: Why Moore Matters; `http://semi.org/en/node/55026`; 2015

[35] S. Widmann: An Inherently Safe Microprocessor Architecture; Autonomous Systems 2014; Proceedings of the 7th GI Workshop; pp. 12–23

Systematic and Probabilistic Testing of Autonomous Mobile Robots

Francesca Saglietti

Chair of Software Engineering
University of Erlangen-Nuremberg, Germany

Abstract: This article intends to offer a brief overview on research activities devoted to the transfer of well-founded verification and validation approaches from the body of knowledge underlying modern software engineering to the domain of robotic applications involving cooperating agents. It aims at illustrating how a roadmap envisioned some years ago could be successfully instantiated within two subsequent European co-operations, in particular by highlighting the most relevant results achieved on the way.

1 Introduction

Fully autonomous mobile robots are supposed to be able to operate in a common environment by taking individual decisions solely based on their individual sensing and perception capabilities and such as to fulfil pre-established rules concerning safe co-existence and effective co-operation. Other than usual for component-based systems, the composing parts often differ not only with respect to their origins, i. e. development time and place, but especially in terms of their functional purpose and of the quality demands for which they were developed.

In addition to the individual functional capabilities, supplementary behaviour may emerge from the resulting interplay; special verification and validation (V&V) techniques are therefore required in order to ensure that any emergent behaviour may only provide benefits and will never contribute to hazardous situations. For the purpose of identifying appropriate V&V approaches, the know-how meanwhile gained in software engineering can be generalized to provide a framework adaptable to address cooperation of autonomous agents. A roadmap offering guidance on how to proceed was proposed at the 4th Workshop on Autonomous Systems 2011 [7]. The present occasion of the 9th GI Conference on

Autonomous Systems offers the opportunity for providing an overview on what could be achieved since then by proceeding along that roadmap. For details, the reader is kindly referred to the corresponding publications.

2 Software Testing

As generally known, software engineers distinguish between two major purposes of software testing:

Early Testing for Fault Detection Early verification and validation activities involve the execution of test cases such as to maximize the chances of revealing the existence of software faults by observing incorrect behaviour. Usually, testing strategies aimed at this purpose rely on a systematic selection of test data allowing to cover as much as possible the functional requirements (black-box, functional testing) resp. the program control and / or data flow structure (white-box, structural testing). Sometimes, an intermediate level of abstraction is taken as a compromise (grey-box, model-based testing).

Late Testing for Reliability Assessment Late verification and validation activities, on the other hand, address the possibility of deriving probabilistic reliability figures from predominantly correct test behaviour. As in general systematic coverage-based test data selection does not relate to the expected operational usage, new operationally representative test scenarios are required to anticipate the expected usage behaviour. In order to derive reliability estimations by statistical sampling theory [10] such test scenarios must be stochastically independent both in terms of their data and of their execution. In other words, neither the definition nor the observation of a test scenario must influence further test scenario definitions resp. executions. Under such assumptions, extensive samples of independent and correct runs allow to apply statistical sampling theory in order to derive conservative claims concerning software reliability.

3 Testing the Cooperation of Mobile Autonomous Robots

Verifying and validating the behaviour of cooperating robots present analogous challenges to those posed by software V&V. In fact, the underlying coexistence and cooperation rules to be obeyed by the agents represent functional demands required to be fulfilled under any circumstance by the resulting system-of-systems. Such functional demands can be captured via compact behavioural models. For this purpose, Coloured Petri Nets (CPNs) [2] have revealed to provide a particularly useful notation [3]. Like classical Petri Nets, also CPNs consist of (marked or unmarked) places linked to transitions and vice versa; via transition firing they also offer the expressive power required to represent concurrency situations, including synchronization and conflicts. In addition, CPNs are enriched by type-specific tokens allowing to instantiate generic actions under scenario-specific conditions: this is achieved by capturing pre- and post-states via data-specific place markings characterizing the system before and after the occurrence of the particular action considered. Technically, this is supported by input arc expressions indicating under which input place markings a transition may be fired, assuming its guard is fulfilled. In addition, input and output arc expressions indicate how input and output place markings change to reflect the new system state. Thus, each state transition is triggered by a data-specific transition firing denoted as a CPN event. Among the major benefits offered by CPNs is the support of scalability, in particular permitting to increase the number of agents without having to extend the underlying net structure.

On the basis of such a model-based specification, the testing strategies summarized above can be transferred from software engineering to robot engineering yielding the following testing paradigms for autonomous robotic agents.

Systematic Cooperation Testing In order to maximize the chances of identifying faulty cooperative behaviour among the agents, systematic functional testing requires to check as systematically as possible the functional behaviour captured by the CPN model. In other words, the CPN structure is to be covered as exhaustively as possible. For this purpose, model coverage criteria based on the elementary notions of CPN entities, namely transitions, events and states, were hierarchically organized [4] such as to allow objectively reproducible measures of the relative amount of cooperative functionality actually observed in the robotic environment of the application domain considered. In addition, automatic test data generation mechanisms were developed in order to ensure the practicality of the testing approach.

Statistical Cooperation Testing On the other hand, to derive reliability estimates, the functional behaviour of cooperating robots must be extensively observed under regular and anomalous conditions, in particular including both the malfunction of individual robots (e. g. due to lack of resources or to physical failures) and the effects of environmental hazards possibly jeopardizing operation (e. g. in the presence of obstacles or of risky pavement conditions). According to pre-defined probabilities for intended and unintended events, appropriate test scenarios could be derived from the underlying CPN model by random manipulation of markings such as to simulate regular and anomalous behaviour.

4 Systematic Cooperation Testing: Results

For the purpose of defining appropriate functional coverage criteria reflecting testing progress, an extensive hierarchy of coverage metrics was illustrated in [4]. It is essentially based on the three basic CPN model entities, namely transitions, events and places. Such a subsumption hierarchy may be helpful in allowing to distinguish between more modest and more demanding criteria in terms of the amount of testing effort required to fulfil them. On the other hand, apart from quantitative economic aspects, the tester may also need guidance to establish a relation between formal coverage criteria and behavioural richness addressed. Such guidance is provided by realizing the following correspondences.

CPN Transition Coverage Covering all CPN transitions requires data-unspecific firing of any transition; therefore, testing according to transition-based coverage criteria is limited to one-time checks of generic actions without distinction of the contextual variability in which they are carried out. Examples are generic checks concerning forward movement.

CPN Event Coverage Covering all CPN events, on the other hand, requires to enrich the information about transition firing by the specific data actually involved in the firing. This means that testing according to event-based coverage demands requires to take into account also the details of the local context in which the action is carried out, e. g. by distinguishing between forward movement on a flat pavement from forward movement on a steep ground.

CPN State Pair Coverage Finally, covering all CPN state pairs requires to reach for any event any potential post-state from any potential pre-state. To do so, the tester must extend local knowledge on event occurrence by global knowledge on the whole system state before and after the event. In case of the aforementioned example addressing forward movement, this would require to include into test observation also any other agent possibly influencing the immediate future of the robot moving forward.

Heuristic and analytical approaches to the automatic generation of test cases to predefined coverage criteria were described in [5] and [6].

Further results achieved in the area of systematic verification of robot cooperation concern testing of reconfigurable robot behaviour which includes regular scenarios as well as exceptional behaviour to be autonomously initiated under anomalous or exceptional conditions. Among classical reconfiguration strategies are re-routing for the purpose of collision avoidance or of energy recharging as well as platoon formation [1] for the purpose of efficiency and safety. In case reconfiguration strategies can be modelled as part of the behavioural model, testing can proceed incrementally by starting with a model kernel representing regular operation to be stepwise extended in order to capture increasing levels of exceptional conditions [9].

5 Statistical Cooperation Testing: Results

For the purpose of implementing CPN-based statistical testing a procedure was developed in [8]. It defines a number of random operators acting on a CPN model by repeated manipulation of its current state to simulate both intended missions according to an expected usage profile and exceptional events according to an estimated anomaly profile.

The approach was applied to an example inspired by a system consisting of linen-carrying trolleys in a hospital environment. The representative and independent scenarios generated can be used for the purpose of testing in a real environment such as to derive conservative reliability measures; in addition, assuming model accuracy, model-based test case simulation can also be used for the purpose of an early qualitative behavioural analysis.

6 Conclusion

The present article summarizes some of the main results gained within two subsequent European co-operations and concerning verification and validation of cooperating autonomous robots. The analysis includes systematic testing aimed at fault detection and probabilistic testing aimed at reliability assessment. In both cases a model-based approach using Coloured Petri Nets revealed as particularly helpful to provide a compact, scalable and incrementally extendable representation of cooperative and reconfigurable robotic behaviour.

Acknowledgements

This article summarizes activities carried out and results achieved within two ARTEMIS projects, namely R3-COP (Resilient Reasoning Robotic Co-operating Systems) and R5-COP (Reconfigurable ROS-based Resilient Reasoning Robotic Co-operating Systems). The support provided by the European Programme ARTEMIS (Advanced Research and Technology for Embedded Intelligence and Systems) and by the German Federal Ministry of Education and Research (BMBF) is gratefully acknowledged.

References

[1] Bergenhem, C., Shladover, S., Coelingh, E.: Overview of Platooning Systems. *Proc. 19th World Congress on Intelligent Transportation Systems*, 2012.

[2] Jensen, K., Kristensen, L. M.: *Coloured Petri Nets*. Springer, 2009.

[3] Lill, R., Saglietti, F.: Model-based Testing of Autonomous Systems based on Coloured Petri Nets. *Proc. ARCS 2012 Workshops, Lecture Notes in Informatics*, Vol. 200, Gesellschaft für Informatik, IEEE Xplore, 2012.

[4] Lill, R., Saglietti, F.: Model-based Testing of Cooperating Robotic Systems using Coloured Petri Nets. *Proc. SAFECOMP Workshop on Dependable Embedded and Cyber-physical Systems*, HAL open access archive, 2013.

[5] Lill, R., Saglietti, F.: Testing the Cooperation of Autonomous Robotic Agents. *Proc. 9th International Conference on Software Engineering and Applications*, Scitepress Digital Library, 2014.

[6] Saglietti, F., Föhrweiser, D., Winzinger, S., Lill, R.: Model-based Design and Testing of Decisional Autonomy and Cooperation in Cyber-physical Systems. *Proc. Int. Conference on Software Engineering and Advanced Applications*, IEEE Xplore, 2015.

[7] Saglietti, F., Söhnlein, S., Lill, R.: Evolution of Verification Techniques by Increasing Autonomy of Cooperating Agents. *Proc. 4th Workshop Autonomous Systems 2011*, Studies in Computational Intelligence, Vol. 391, Springer, 2011.

[8] Saglietti, F., Spengler, R., Meitner, M.: Quantitative Reliability Assessment for Mobile Cooperative Systems. *Proc. SAFECOMP 2016 Workshops*, Vol. LNCS 9923, Springer, 2016.

[9] Saglietti, F., Winzinger, S., Lill, R.: Reconfiguration Testing for Cooperative Autonomous Agents. *Proc. SAFECOMP Workshop on Dependable Embedded and Cyber-physical Systems*, Vol. LNCS 9338, Springer, 2015.

[10] Störmer, H.: *Mathematische Theorie der Zuverlässigkeit*. Oldenbourg, 1983.

On the Construction of a Crowd-verifiable Microprocessor

Marcel Schaible

FernUniversität in Hagen, Germany
Faculty of Mathematics and Computer Science
Chair of Computer Engineering

Abstract: Contemporary microprocessor designs are in quest for increasing the instruction throughput and in the advent of battery powered devices minimising the power consumption. To achieve this kind of features complex algorithms must be applied, which in turn cumber the verification. However, safety-critical systems are demanding for correctness, reliability, availability and a deterministic time behaviour and therefore for verifiable designs. In this paper a consensus-oriented and crowd-verifiable control unit is presented, which is designed for verifiability and concludes with the outline of a consensus-oriented verification methodology.

1 Introduction

The major controlling part of a microprocessor is the control unit. The control unit initiates sequences of controlling signals in a predefined order to coordinate the signal flow of the various functional units. Because the instruction set of a microprocessor is finite, the number of micro-operations is therefore finite and the overall complexity of a microprocessor correlates with the number of the micro-operations. Traditionally control units are designed with hardwired or microprogrammed control logic [26]. Hardwired control units, especially with a sufficiently large amount of instructions, are tedious to verify, because of their hardly comprehensible and understandable wired connections between different parts. However microprogrammed control units model traditionally most of their logic in a read-only control store. The sequence of signal transitions is stored in the control store (see figure 2) as a two-dimensional table containing micro-operations. Each micro-operation is represented by a set of control signals. Especially in the domain of safety-critical systems, which are demanding for correctness, reliability, availability and deterministic time behaviour [19, 20], comprehensible and human verifiable designs are inevitable. The major requirements of the control unit presented in chapter 2 are *consensus-oriented* and *crowd-verifiable* [5, 12]. Both terms are defined in chapter 3 and demanding for a clearness of concepts and reduced feature set.

2 Design of the Control Unit

The presented microprocessor architecture in figure 1 includes the following properties:

- The execution of micro-instructions is strictly sequential.
- Code and data memory is housed on-chip.
- Code and data memory is completely separated for safety reasons (Harvard Architecture [11]) .
- Code memory is read-only during execution time and is assumed in the following description as read-only memory (ROM).
- The processor core is register less [22, 23] in the sense that the programmer cannot access any registers directly. Operands are always read from and results are written to data memory.
- The table-driven control unit is designed to guarantee deterministic execution time per instruction.
- The instruction decoding is simplified by defining all instructions equally long.
- Each instruction op-code is expanded with a micro-address and points directly onto a distinct row of the control store.

Figure 1 outlines the general execution sequence of the control unit:

1. Fetch cycle: The program counter (PC) points to next instruction and is transferred into the program address register (PAR). This initiates a read of the memory content addressed by PAR. After completion the memory content is available in the instruction register (IR).
2. Decode cycle: Copy the op-code of the instruction in IR into the upper part of the control store address register (CSAR) (see fig. 2) and zero out the lower portion of it.

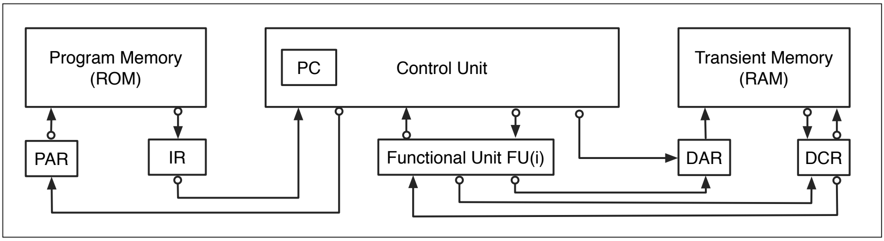


Fig. 1: Architecture Overview

3. Execute cycle: If necessary read additional operands from the transient memory by placing the address into the data address register (DAR). The memory content will be available in the data content register (DCR). Activate the functional unit (FU) like e.g. an adder and execute its operation.

4. Write-back cycle: If necessary write back the result of the operation into the transient memory.

5. Set the PC to the next instruction, which is dependent of the op-code class (non-branching vs. branching op-code) and jump to 1.

All unused operation codes will immediately lead to a processor halt when loaded into the CSAR register of the control unit.

The instruction set is based on the studies of [17, 24], which contain the most-commonly used constructs in high-level programming languages.

In fig. 2 the schematic of the control store is sketched:

- The control store consists of a two dimensional read-only table.
- Each row belongs exactly to one op-code.
- The columns are structured into groups of control signals (CS), micro instruction register (MIR), condition selection (CSEL) and an optional branch address (BRA).
- CS are enabling certain parts of the data paths.
- MIR stores the operation class like e.g. non-branching vs. branching op-code.
- The condition selector (CSEL) determines which condition check like *jump on equal* is enabled.

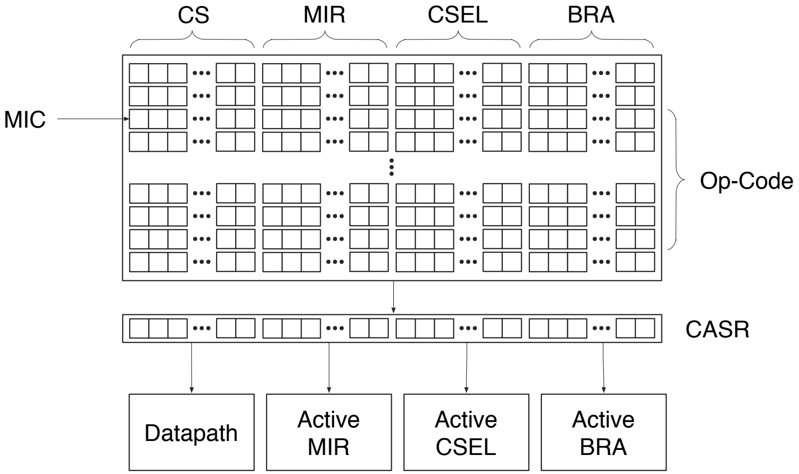


Fig. 2: Block-Diagram of Control Unit

- BRA holds an optional branch address. If BRA is not needed it is zeroed-out.
- The control store address register (CSAR) points the currently active row of the control store.
- The control store data register (CSDR) holds the currently active CS, MIR, CSEL and BRA.

The generation of the next to be executed instruction is outlined in fig. 3. If the current instruction belongs into the class of non-branching op-codes the multiplexer (MUX) enables the incrementer (+1) and stores the result back into the lower part of the CSAR. In case of a branching instruction the selector of the MUX obeys the current CSEL and if fired stores BRA into the lower part of the CSAR.

The architecture including the memory model and data paths are described in more depth in [21].

3 Verification

Formal verification methodologies are utilised in modern cpu manufacturing for risk minimisation. They are making use of mathematical representations

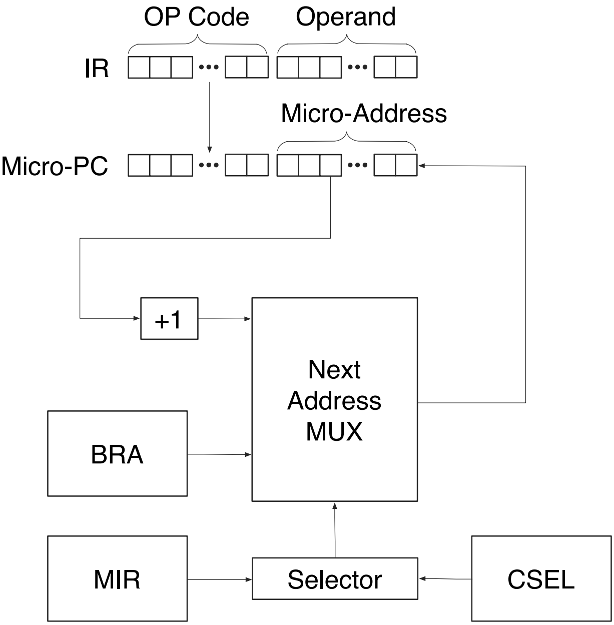


Fig. 3: Block-Diagram of Address Generation

of technical facts and circumstances. Notwithstanding that these mathematical techniques are manageable by experts, the proof of a sufficient complex system like a microprocessor demands for automated theorem proving systems. Automated theorem proving systems (ATPS) [2, 4, 13, 27] tend to generate large proofs, which cannot be examined in detail by even the field experts. It is the state of the art the sufficient complex software programs like ATPS are unlikely to be error-free. This implies that errors or misconceptions regarding the specification will remain undetected. But even with the utmost efforts there is no guarantee for a correct chip-design [14]. In contrast safety-critical systems are demanding for well understood proofs.

The more general question to answer here is:

How do we gain confidence and trust in the validation of technical systems?

The key is *intelligibility* and as a consequence *simplicity* of the architecture. Designs must be make accessible to a broad community of field and non-field experts.

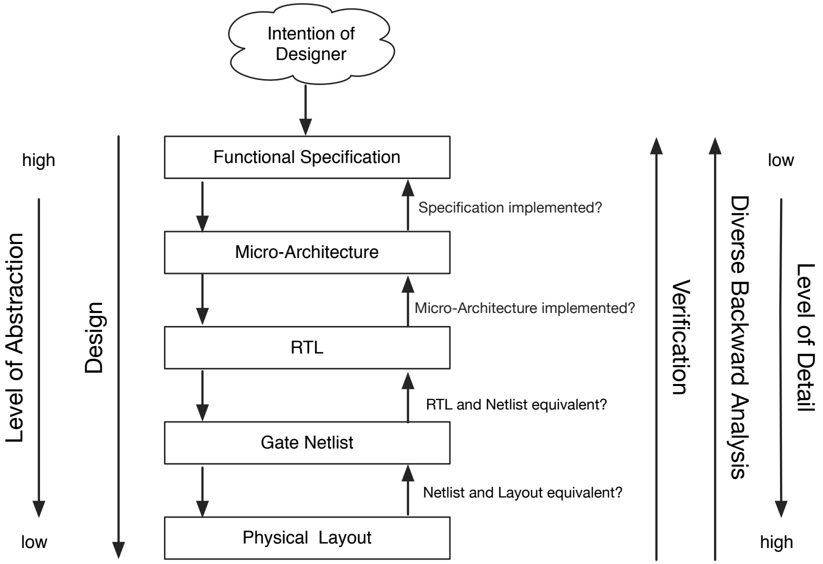


Fig. 4: Verification Model

In fig. 4 starting with the intention of a designer a functional specification is produced, which leads from a micro-architecture over Register-Transfer-Level description to a netlist and to a physical layout on an integrated circuit. In each step the correctness of the transformations must be proved.

Because of the clean and straightforward design of the above-described control unit it is feasible to perform a consensus-oriented and crowd-based verification [12]. Consensus-oriented verification is the process of examination of a community (crowd-based), which comes to a common understanding and agreement that a design can be considered correct. The more persons with distinct expertise are examining the specification and the concrete realisation, the merrier and much more likely design and implementation flaws are revealed.

The following properties of the control unit have to be considered for verification:

- The supporting schematics like e.g. the address generation logic and reset logic can be verified by comparing the design with the requirement specification and check off each wire.

- Since each op-code consists of a consecutive number of micro-operations (rows in the control store) each row can be compared with the requirement specification and ticked off.

But because of the inherent design properties (comprehensible and understandable) of the presented architecture it is feasible to apply a reversed verification process: Starting from the physical layout the verification can be done backwards by a broad range of humans to the netlist all the way up to the specification. This approach was first applied as a software verification methodology by the TÜV Rheinland [18] to acquire permission for commissioning of the nuclear power plant in Halden (Norway). Although this method was used for software validation it can be adopted to hardware verification. An important ancillary effect of the diverse backward analysis methodology [10] is the intrinsic verification of the correctness of the applied transformation tools.

The core characteristics (see [10, pp. 152-153]) of the diverse backward analysis methodology are

1. The process of verification is diverse in respect of the implementation.
2. The process of verification has the character of a proof.
3. Each step of the verification is strictly documented and checkable.
4. The verifier is not defencelessly delivered to potential systematic errors of the verification process.

4 Summary

The objective of this work is to develop a verifiable control unit for usage in safety-critical systems. The described control unit architecture with its table-driven composition and strict sequential execution logic provides a coherent architecture which can be verified by field and non-field experts. Due to the characteristics of the table-driven and reduced complexity design the verification can be literally performed by checking each set of generated control signals. The crowd-based verification can be performed both forward and backward in a consensus-oriented discourse with the objective to gain confidence and trust in the correctness of the examined design. The diverse backward analysis as a powerful verification methodology, which establishes a strong confidentialness not only in the correctness of the design itself but rather in the tools used to generate it, can be applied.

References

[1] Andersen, B. Scott and Romanski, George: Verification of Safety-critical Software, In: *Queue, ACM*, 9, 8, pp. 50–59, 2011

[2] Berg, C. and Beyer, S. and Jacobi, C. and Kröning, D. and Leinenbach, D.: Formal Verification of the VAMP Microprocessor (Project Status), Symposium on the Effectiveness of Logic in Computer Science (ELICS02), pp. 31–36, 2002

[3] Berg, C. and Jacobi, C. and Kröning, D.: Formal Verification of a Basic Circuits Library, In: *Proc. of the IASTED International Conference on Applied Informatics, Innsbruck (AI 2001)*, pp. 31–36, 2001

[4] Beyer, S. and Jacobi, C. and Kroening, D. and Leinenbach, D. and Paul, W. J. : Putting it all together - Formal Verification of the VAMP, In: *Software Tools for Technology Transfer (STTT), Special Section on Recent Advances in Hardware Verification* , 8, pp. 411–430, 2006

[5] Brabham, Daren C.: Crowdsourcing, *The MIT Press*, 2013

[6] Cohn, A: A proof of correctness of the viper microprocessor: the first level, In: *University of Cambridge, Computer Laboratory*, 1987

[7] Dijkstra, E. W.: The next fifty years, 1996

[8] Halang, W. A.: On methods for direct memory access without cycle stealing, In: *Microprocess. Microprogram.*, 17, 5, pp. 277–283, 1986

[9] Halang, W. A. and Jung, S.: A function oriented architecture for process control systems minimizing internal data transfer costs, In: *Microprocess. Microprogram.*, 28, pp. 123–128, 1990

[10] Halang, W. A. and Konakovsky, R.: Sicherheitsgerichtete Echtzeitsysteme, *Oldenbourg*, pp. 150–152, 1999

[11] Hennessy, J. and Patterson, D.: Computer Architecture - A Quantitative Approach, Morgan Kaufmann, 2011

[12] Howe, Jeff: Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business, *Crown Business*, 2008

[13] Hunt, W. A. J.:FM8501: A Verified Microprocessor, In: *Lecture Notes in Computer Science / Lecture Notes in Artificial Intelligence*, Springer, 1994

[14] Intel Corp.: *Intel Xeon Processor E3-1200, v3 Product Family, Specification Update*, 2015

[15] Joyce, J.J.: Formal specification and verification of microprocessor systems, In: *Technical report (University of Cambridge. Computer Laboratory)*, 1998

[16] von Kaenel, P. A.: Designing and testing a control unit, In: *J. Comput. Sci. Coll.*, 19, 5, pp. 228–237, 2004

[17] Knuth, D. E.: An Empirical Study of FORTRAN Programs, In: *Softw., Pract. Exper.*, 1, 2, pp. 105–133, 1971

[18] Krebs, H. and Haspel, U.: Ein Verfahren zur Software-Verifikation. In: *Regelungstechnische Praxis*, rtp 26, pp. 73â78, 1984

[19] Laprie, J. C., Avizienis, A. and Kopetz, H.: Dependability: Basic Concepts and Terminology, Springer, 1992

[20] Laprie, J.-C. and Béounes, C. and Kanoun, K.: Definition and Analysis of Hardware- and Software-Fault-Tolerant Architectures, In: *Computer, IEEE Computer Society Press*, 23, 7, pp. 39–51, 1990

[21] Schaible M.: A Consensus-oriented Crowd-verifiable Microprocessor Architecture, In: *Proceedings of the 7th GI Workshop Autonomous Systems 2014*, Vol. 835, pp. 209–220, 2014

[22] Stieger, H. and Halang W. A.: Eine hochsprachenorientierte Rechnerarchitektur ohne arithmetische Register, *IFB Verlag*, 2003

[23] Suresh, P. and Moona, R.: PERL; A Registerless Processor, In: *Proceedings of the Fifth International Conference on High Performance Computing*, pp. 33–, 1998

[24] Tanenbaum, A. S.: Implications of structured programming for machine architecture, In: *Commun. ACM*, 21, 3, pp. 237–246, 1978

[25] Tanner Consulting & Engineering Services: MTSMS035DL Digital Low Power Standard Cell Library For TSMC 0.35 Sub-micron Process. Revision A., 1999

[26] Wilkes, M. V. and Stringer, J. B.: Micro-programming and the design of the control circuits in an electronic digital computer, In: *Mathematical Proceedings of the Cambridge Philosophical Society*, 2, pp. 230–238, 1953

[27] Phillip J. Windley: Formal Modeling and Verification of Microprocessors, In: *IEEE Transactions on Computers*, 44, 1, pp. 54–72, 1995

Affordable High-bandwidth Real-time Mass-storage Architecture for Distributed Sensor Nodes

Michael Kirchhoff, Christoph W. Wagner, Ralf Herrmann and Reiner S. Thomä

Electronic Measurement Engineering Group
Technische Universität Ilmenau, Germany

Abstract: This paper presents some of the authors' research work in the fields of embedded real-time systems and ultra-wideband radio sensor networks. We introduce a real-time mass-storage architecture with the ability of lossless storage of a large amount of measurement data generated in distributed mobile sensor nodes. After pointing out major requirements for this objective and analysing available storage technologies, we develop an affordable architecture using an FPGA-board and a solid state drive (SSD). As proof of concept, experiments will show the real-time characteristics of our architecture on a representative hardware system.

1 Introduction

State-of-the-art electronic components open up new applications that were hard or even impossible to achieve just a few years ago. In this paper, an example of integrating modern programmable logic components with non-volatile mass storage for uninterrupted data acquisition in a demanding sensor application will be described.

This application is a distributed network of ultra-wideband (UWB) radio sensors. UWB sensing technology allows - among other things - remote and non-destructive range measurements with very high spatial resolution over short distances (typically a few meters). A comprehensive overview of this technology and its use cases can be found in [7]. UWB sensor nodes can be combined into a distributed sensor network and the ranges measured by each node can be used for precise localisation of, e.g., humans or even analysed to extract vitality information such as breathing and heart beat rates [8]. Due to the extremely wide signal bandwidth in UWB and the need to collect data about dynamic test objects in real-time, the amount of raw measurement data coming from the receivers is unusually high compared to traditional radio sensing technologies [7]. The EME Group currently uses a high-speed UWB receiver with the following specifications in each sensor node for sounding applications: two receive channels with up to 16 bits per sample and an acquisition rate of roughly 72 megasamples/s (Ms/s). The resulting raw data stream is a staggering 288 MB/s (ca. 275 MiB/s). Channel-sounding like in [6] and similar tasks often require the raw data to be stored, because it will be analysed offline using different and complex algorithms. During a measurement, such a large stream must be stored continuously for several minutes in a row. Aside from the real-time aspect of uninterrupted storage, this poses additional challenges for a mobile high-bandwidth mass-storage solution embedded in the sensor node. Physical size and power consumption should be limited in order to allow easy sensor deployment and measurement setup in crowded environments such as offices or production halls. The cost of a sensor node is another important factor, especially when using many nodes for multiple-input/multiple-output (MIMO) scenarios.

In this paper we propose the use of FPGAs (Field Programmable Gate Arrays) in conjunction with modern solid-state drives (SSDs) to accomplish this demanding task.

Today FPGAs are one of the most important technologies to construct high-speed real time systems. With the addition of specialised high-speed I/O capabilities such as multi-gigabit serial transceivers available in modern FPGAs [1, 13], different high-bandwidth digital buses can be integrated using the same chip without a need for additional hardware. Therefore, many problems that used to required expensive or large dedicated hardware solutions are now solvable in a quicker and more affordable way - often even with the use of off-the-shelf hardware.

The biggest advantage of FPGAs, compared to other possible technologies like ASICs (Application Specific Integrated Circuits) or ASSPs (Application Specific Standard Products) is the ability to reconfigure the complete logic in a few seconds or parts of its logic function within milliseconds. Thanks to this characteristic, it is possible to quickly handle the increasing design and system complexity that comes along with higher density of usable logic resources or the mentioned integration of new (and in itself complex) design elements. A trade off can be made between the added costs and power consumption of the programmable logic and the initial effort for design, verification, and production of

ASICs. In our example application as well as other research areas, the number of required units is small to medium making a pure FPGA development a reasonable choice. Consequently, our last important objective is to minimize the overall development time by means of modularisation and component reuse.

For the objective target of an FPGA-based high-bandwidth mass-storage solution for the UWB sensor node described above, section 2 discusses the choice of suitable mass storage technologies. With this analysis, section 3 explains the architecture and design of the selected solution in detail and explores some of its special challenges. In the last section, a test setup with an example implementation as a proof-of-concept is described and some benchmark results are given. A brief summary and outlook will conclude this contribution.

2 Selection of Mass Storage Technology

Before analysing the detailed data flow from the sensors receiver to the storage memory, the choice of the right mass storage technology has to be made. Criteria have already been mentioned in the introduction. The storage solution must be able to handle a sustained sequential transfer rate of at least 275 MiB/s. This results in a data volume of about 16.1 GiB/min, and since a measurement for several minutes should be possible, the total amount of storage must exceed several tens of GiB. Other softer criteria are low in weight and small in size (being able to mount the whole sensor node on scanners or in crowded spots) as well as a low power consumption. Other very important metrics have not yet been discussed - various latencies of the storage devices and their digital interfaces must be taken into account when continuous storage should be achieved. They play a decisive role in the properties and effort put into the FPGA board and design which must provide enough buffer memory to compensate for all latencies.

A general overview of modern mass-storage technologies has been given in [12]. Three main classes for mass-storage have been identified: magnetical, optical and electronic (i.e. solid state) storage. Nowadays, they all provide enough storage capacity but still greatly differ in terms of sustained data rates and storage latencies. These two criteria already make optical storage very unattractive - a single BluRay Disc manages about 54 MB/s transfer speed and access latencies are in the order of several 100 ms. An array of at least 6 parallel writers would be needed to achieve the target rate. Size and power consumption would also become significantly high.

Magnetic storage comes in the form of either hard disc drives (HDDs) or tape recorders. While the latter are indeed optimized for sequential storage of high amounts of data, their primary use are backups and latencies can be very high, e.g. when the tape must be started at the beginning of a write burst or when a seek to another part of the tape must be done. HDDs technically also work well with sequential write patterns and do not have such a high initial latency since the platters are spinning all the time. However, since the medium is realised as discs rotating at a constant speed, the outer tracks provide higher data rates than the inner ones, so the sustained rate is dictated by the slowest regions that should be written to (one could decide to not use the innermost tracks and leave some storage capacity unused, of course). Typical average latencies are in the order of 10 ms [12] but maximum latencies can also be much longer. Both sustained data rates and access latencies greatly depend on areal density and rotational speed while especially the latter property has a big influence on drive power consumption. There are currently only very specialised HDDs that could store 275 MiB/s as a single drive, so at least an array of two discs would be needed. Form factors of high-speed drives include 3.5″ and 2.5″ sizes which may be suitable for a mobile sensor node. In conclusion, HDDs can meet the required criteria but are less optimal for the softer criteria of our objective.

The last mass storage class – solid state electronic storage – comes in a number of implementations [12]. While raw NAND-flash chips could be integrated on an FPGA board, this would require additional complex hardware- (high-speed printed circuit board) and software design (of a flash controller) and run counter to our goal of limiting overall design effort. Fortunately, flash-based storage also comes in a number of ready-to-use devices. Some examples are flash cards often found in digital cameras or solid state drives (SSDs) increasingly found in personal computers or mobile devices. As pointed out in [12], the principal characteristics of flash cards and SSDs are well suited to our application. Very low average access latencies, high sustained sequential write speeds and small size/power consumption are provided by many products on the market. When looking at write speeds, it becomes clear that in the case of flash cards, an array of two to four high-end cards would still be needed to accommodate 275 MiB/s and reach a reasonable total amount of storage. While size and power consumption remain very low even with such an array, costs increase significantly when using multiple very fast high-end cards. Even though the smaller size may favour flash-cards over an SSD, the latter technology provides a much broader choice of suitable products. Typical modern SSDs can write more than 400 MiB/S as a single drive, provide several 100 GiB of stor-

Table 1: Comparison of Performance Aspects for Storage Devices, excerpt from [3], Fig. 3

Storage Device	**Seq. Throughput**	**Latency**	
Class, Capacity	**Write**	**Avg.**	**Max.**
15000 rpm HDD, 80 GB	84 MB/s	5.39 ms	97.28 ms
Client SSD 1, 256 GB	240 MB/s	0.51 ms	1218.45 ms
Enterprise SSD, 400 GB	393 MB/s	0.05 ms	19.6 ms

age, and come in a number of small form factors (e.g. 1.8″ or 2.5″ HDD form factors and even smaller). They also typically use much less power than high-performance HDDs.

Compared to HDDs, SSDs or flash card arrays are a better overall fit to the requirements. A final decision was made in favour of SSDs due to the price advantage and a lower effort of hardware integration with an FPGA board. The SSD market is very versatile nowadays, but it is still challenging to select the right product for our demanding application. The biggest issue is the lack of complete technical specification from most manufacturers, i.e. the advertised performance figures do not give a complete picture of the drive's behaviour. An interesting whitepaper with a comparative study (using anonymised brand names) can be found in [3]. Fig. 3 from [3] compares representative performance data of a variety of drives and is partially replicated here in Table 1 to highlight some prime aspects that need to be considered in the FPGA design. The overall results of HDDs compared to SSDs support our reasoning as presented above. It is also apparent, especially when looking at maximum latencies, that there can be significant differences between drive models especially in the client SSD space whereas enterprise products are usually engineered towards a more consistent performance. There are, however, client SSDs (often labeled "pro") that can compete with the much more expensive enterprise drives for sequential write use cases. Since there are usually two orders of magnitude between average and maximum latencies, a probability distribution would be needed to get a robust understanding of how much data must be buffered to avoid loss. Such latency distributions are rarely published or investigated. As a result, some estimations and safety margins must be included in the requirement analysis for the buffering solution and the choice of a specific SSD model must be backed by real life tests to guarantee stable and sustained data storage. For our FPGA storage management design, we therefore define the following requirements with respect to latencies: an average latency of 1 ms and a maximum latency of at least 100 ms must be accounted for via buffering.

3 FPGA Real-time Storage Architecture

3.1 Data Flow and Buffer Performance

After fixing the storage technology to an (at least) pro-grade SSD with SATA interface and setting some related design goals, further requirements for the FPGA architecture are dictated by the actual data flow. The UWB receivers in our sensor nodes are connected to high-speed analogue-digital converters (ADCs) which are driven by an external sampling clock source. The same clock is available in the FPGA, but must be generally considered independent of any core or reference clock needed for realisation of the storage interface. Data therefore flows into the FPGA in the ADC clock domain and must be buffered in some form of memory. A dual-port memory buffer would compensate any storage latencies and handle clock domain crossing into the storage clock domain at the same time. The requirements for such a theoretical monolithic storage buffer can be derived form the previous design goals as follows.

The first property is the minimum size of the buffer. It can be estimated from the maximum storage latency which has been defined to be 100 ms above. With a total of 2x 16 bits/sample and a sampling rate of 72 Ms/s, the buffer size must be at least 27.5 MiB. This estimate disregards the unknown distribution of reoccurring higher latencies over time which means the actual buffer size should potentially be larger. This aspect is related to the following discussion and will also be investigated further in section 4 by real-world tests.

The second property is buffer interface bandwidth. At the input side, the requirement is that the 275 MiB/s from the ADCs can be stored in a cycle-accurate manner, i.e. data must be stored with each clock cycle and no interruption can be tolerated. This rules out burst-oriented memory technologies such as SDRAM at the buffer input, even if modern DDR3 interfaces easily provide bandwidths of several GiB/s. The required average bandwidth at the storage buffer output is not immediately obvious because storage block size and average latency must be taken into account. The average latency theoretically occurs for every data block sent to the SSD and since there is no data drained from the buffer during that time, the burst buffer output bandwidth as well as write bandwidth of the SSD must be higher than 275 MiB/s. The simplified relationship is given in (1) where BW_{burst} is the required burst bandwidth at the buffer output in MiB/s, $DataSize$ is the amount of data at the buffer input in MiB that accumulates within one second, and $BlockSize$ in MiB is the size of each single request to the SSD. As can be seen, the required burst bandwidth BW_{burst} decreases with increasing block sizes and decreasing average storage latency Lat_{avg}. Considering the superior average latency of the enterprise SSD of Table 1, the burst bandwidth demand at the same 1 MiB block size would reduce to 279 MiB/s. However, since we have already fixed the average latency to 1 ms as a design goal, burst bandwidth requirements can only be relaxed by larger block sizes per storage command.

$$BW_{burst} = \frac{DataSize}{1 - \frac{DataSize}{Blocksize} \cdot Lat_{avg}} \tag{1}$$

If the buffer size is large enough to accommodate a maximum latency stall and the buffer output bandwidth adheres to (1), an overflow cannot occur.

3.2 Memory Buffer Architecture in an Example FPGA

For the implementation, we used a Xilinx Artix-7 XC7A100T FPGA available in the UWB sensor nodes. This FPGA comes with 4.860 KiBit of integrated block-RAM (BRAM) that can be configured as dual-port RAM with each port working in a different clock domain. It could thus be used as a monolithic buffer crossing clock domains and allowing single-cycle access on both read and write sides. However, some of the other modules in the FPGA design also require BRAM and the overall buffer size of roughly 600 KiB is too small for our maximum latency requirement anyway.

The latter argument even holds for larger FPGAs of the same or other families. Consequently, FPGA-external memory is required in the storage architecture which could come in the form of more SRAM or SDRAM. The latter option is found on many off-the-shelf FPGA boards and is also available as 1 GiB DDR3 RAM in the sensor nodes. When using this memory as a buffer (and only this), it is possible to store 3728.2 ms of the data stream before any data would get lost. Unfortunately, the SDRAM also cannot be used as a monolithic buffer solution for our task. Firstly, it is not feasible to run it in the ADC clock-domain because the DDR3 memory interface requires certain fixed reference clocks when used according to the standard. Secondly, the DDR3 interface uses burst mode transfers and therefore has more than one clock cycle latency. Lastly, because of the characteristics of the SDRAM, it is not possible to read and write in parallel and it also takes several clock cycles to switch between these two operation modes.

The solution is to combine the two RAM options to benefit from their respective advantages while eliminating the unwanted characteristics. The resulting architecture is shown in Fig. 1. The large-size SDRAM is organized as a ring buffer with sequential read and write access. This also means that the accumulated bandwith at the DDR3 interface must exceed the sum of the storage buffer input bandwidth and the worst case storage buffer output bandwidth (275+380 MiB/s as derived in section 3.1) in order to accommodate DDR3 protocol latencies. To cross clock-domains and realise a single-cycle latency at the storage buffer input, the SDRAM controller is surrounded by two FIFOs utilising dual-port BRAM. Each FIFO translates clock domains and data bus widths between ADC, DDR3, and SATA modules, respectively.

The DDR3 physical controller as well as the FIFOs are proven integrated cores of the Xilinx design suite and can be used free of charge. The SDRAM-ctrl block in Fig. 1 contains a custom ring-buffer organizer that manages the addresses and commands for reading and writing data using the DDR3 memory controller. Since the FIFO buffer input provides single-cycle access for the ADC data, the

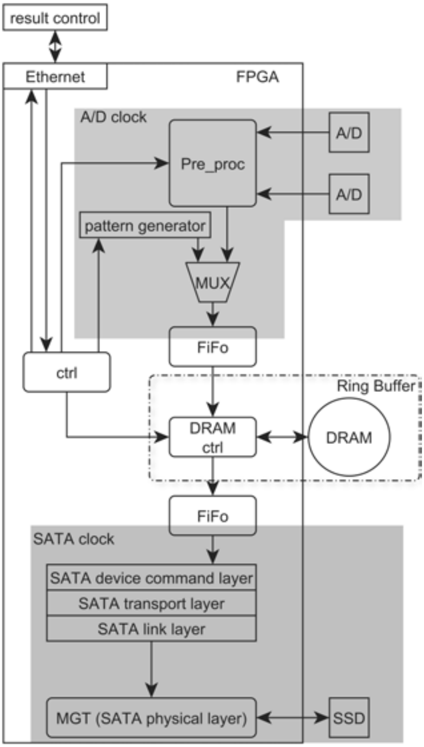


Fig. 1: Real-time mass-storage architecture for FPGA implementation

DDR3 protocol latencies are easily compensated. To use the DDR3 interface bandwidth in an efficient manner, write access to the ring buffer is done in long bursts. Reading from the SDRAM is done in bursts as well, but the burst length is increased 8 fold. The total DDR3 bandwidth using such long bursts is well over 2 GiB/s and the burst asymmetry between reads and writes ensures that the ring buffer can be drained much faster than it is re-filled in the case some data has accumulated due to longer SSD latencies. The input FIFO had an initial size of 8192 words (16 KiB) because it had to have some buffering margin for incoming data even when the ring buffer was operating in DDR3 read mode. The size of the output FIFO was selected to be four times the maximum size of SATA data frames (SATA data FIS (Frame Information Structure): 8 KiB [10]) such that SATA data can be sent with minimal protocol interruption as well.

The simplified SATA protocol stack indicated in Fig. 1 has been adopted from previous work described in [2], which is based on an open-source core [5]. This previous work provides a very good starting point for research and test implementations but could not be used out-of-the-box. The Artix7 family is different from the older generation FPGAs used in the references and consequently the high-speed serial transceiver module which provides the physical SATA link had to be reworked almost from scratch. Furthermore, the physical layer was extended to use SATA3 protocol with a raw bandwidth of 600 MB/s which required minor modifications to the other layers. Finally, the command layer was optimized for our application using large block writes. An additional state machine was added between the storage ring buffer and the command layer to act as an application to the SATA backend. It observes the output FIFO of the ring buffer for new data and triggers write activity on the SATA side when enough data is available. The block size for write commands as discussed in section 3.1 can be setup at design time to adopt to different SSD models. The Ethernet and ctrl blocks allow communication with a PC and the pattern generator will be used in section 4 to test the realised storage buffer.

Fig. 2 provides a more detailed view of the data flow through the storage buffer. As mentioned before, both external interfaces of the buffer (FIFOs) as well as the buffer itself run at different clock rates. The different clock domains are necessary for achieving real-time ability at the input and interface compatibility as well as bandwidth expansion at the output. The buffer's clock rate is fixed at 100 MHz with a memory width of 256 bit (which equals 16 words) due to requirements of the DDR3 memory interface. The burst bandwidth to either FIFO is therefore 3.2 GB/s which equals that of the DDR3 memory interface. This design ensures that the ring buffer controller never becomes a bottleneck for the

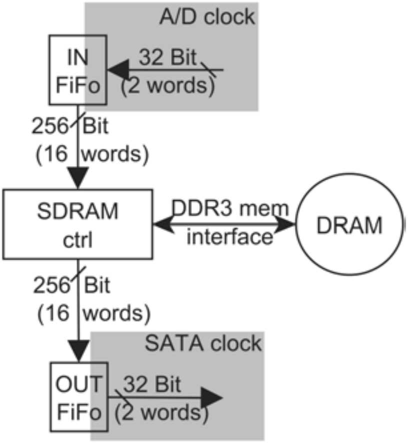


Fig. 2: Storage buffer data flow (detail)

data flow. The SATA storage module runs at 150 MHz for SATA3 (or 75 MHz for SATA2) and has a 32 bit wide data interface provided by the buffer's output FIFO. The maximum bandwidth is 600 MB/s which is also the raw data rate at the SATA3 physical interface. This provides a margin for the higher burst write speeds as required by equation 1. In real implementations, the actual SSD is likely to limit write speed to a lower value. With typical sequential write speeds beyond 400 MB/s for pro-grade SSDs there is still enough headroom for fast buffer draining after data has accumulated due to previous longer latencies.

4 Real-time Experiments

4.1 Test Setup

As shown in Fig. 1, there are some additional modules in the FPGA design. The communication with a host PC for control and data exchange is realised via an Ethernet interface with TCP/IP functionality. The ctrl-block is responsible for setup of data processing in the UWB sensors and control of measurement runs. For example, it is possible to continuously receive snapshots of the measured data in the host computer so that the user can check if every sensor works as expected. Furthermore, one can switch between pre-processed ADC data and a test pattern generator as a source for the storage buffer and use the control block for testing purposes. As mentioned before, the FPGA is an Artix7 XC7A100T from Xilinx. The chip itself does not have built-in SDRAM or Ethernet capability. These functions are added as external components on an off-the-shelf module TE0712 from Trenz Electronic GmbH [11] (other similar modules from different vendors also exist). The TE0712 is integrated with the UWB electronics and ADCs (both unused for the real-time experiments presented here) on a proprietary baseboard. Additionally, the baseboard provides a power supply and a standard SATA connector for the SSD. For testing we chose a recent SATA3 SSD - the Samsung 850 PRO 512GB [9]. According to its specs, it can write at 520 MB/s sequentially and has an average latency of around 0.1 ms (maximum latency not published). Moreover, it is targeted at write-intensive applications. When we conducted short burst tests writing a few GB each time at the full SATA3 rate for testing of the SATA stack, the drive indeed delivered its specified speed. Other SSDs may be suitable as well but should also be tested before using them in critical measurements. For testing of the storage buffer, only the test pattern generator driven by the ADC clock has been used. This was done to verify correctness of the data stored on the SSD by comparing them to the known pattern and also enabled checking of uninterrupted storage down to a single sample.

4.2 Test Objectives and Results

A major objective of testing was to ensure real-time capability of the whole storage architecture. A PC application is used to sent a command to the ctrl-block in Fig. 1. This command switches the MUX to the the pattern generator and then sends the amount of data to be generated. The pattern generator simply realizes a counter that counts up until the requested data volume is transferred. The storage buffer controller monitors the input FIFO and starts to transfer data to the SDRAM buffer when enough values are available. Data in the SDRAM is subsequently forwarded to the output FIFO and finally the SATA application generates the required triggers in the command layer to start writing to the actual SSD. As soon as the control unit is signalled that all data has been generated, it sends a signal to the PC application. The PC can now read the data available on the SSD as well as retrieve various performance and statistical metrics collected during the test.

This simple method of saving counting numbers allows the user to check if every generated sample was successfully stored on the SSD. In addition to that, the storage buffer logs information about buffer usage for both FIFOs and the SDRAM: the average and maximum amount of data buffered in each of the components. Test runs were done with increasing total test sizes to simulate mea-

Table 2: Maximum Occupancy of Different Buffer Components

Test Size	IN FIFO	SDRAM	OUT FIFO
GiByte	Byte	KiB	Byte
0.5	992	6144	31392
1.0	992	6144	30656
2.0	1024	8192	30656
10	1024	6144	31456
20	1024	6167	31776
50	1024	7167	31840
100	1024	7167	31776
470	1024	7167	31776

surements of different durations. The maximum test length was 470 GiB taking more than 20 min and filling the drive almost completely at once. It should be noted that we did not condition the SSD in any way between tests, e.g. there was no secure-erase or trimming in order to get worst case results for our specific application. The tests were run multiple times, dependent on the test size up to 20 times each.

As a first result, the storage system worked correctly, i. e. we did not observe missing samples in any of the tests. Furthermore, it turned out that results were quite reproducible between repetitions with little variance. This indicates a stable implementation of all timing critical paths as well as a surprisingly steady performance of the SSD (our use model does not really seem to challenge modern pro-grade SSD architectures since we only write sequentially and in large blocks). The results of the maximum level of data buffered in some of the tests can be seen in Table 2. The values were translated into total bytes or KiB, respectively. Average values are not reported here since they only show that each FIFO buffer was nearly empty on average. The input FIFO is read out in bursts of 16 entries (512 bytes) at a time. The maximum value in Table 2 indicated that at most two bursts worth of data were ever stored in this FIFO - i.e. the ring buffer controller could empty the input FIFO almost as soon as enough data was present despite the required read cycles to forward data to the output FIFO. From the already quite limited 16 KiB space provided at the input, only 1 KiB was used at most. The input FIFO size could be reduced to further save BRAM resources. The output FIFO shows equally unremarkable results. Maximum values show that this FIFO was almost full at some point during the test. This is no reason for concern because this FIFO is filled by large and fast bursts from the SDRAM and holds this data until the SATA write command has been prepared as well as the SSD latency has passed. It is not possible that this FIFO can ever overflow, because data is only sent from the SDRAM to the FIFO by the ring buffer organizer when enough space is available. Additionally, as mentioned earlier, this FIFO is almost empty on average which means data transport works efficiently after the initial start of a block write command. The SDRAM results have been very consistent throughout all tests. The maximum level observed was 8 MiB. This equates to a utilization of 0.78% of the available 1 GiB, which means that 99.22% of the SDRAM remain unused even in the worst case. This value is quite low compared to the assumed worst-case of more than 27 MiB with the projected maximum SSD latency of 100 ms. It is safe to say that this worst case has not been reached - not even close. Additionally, average values were much lower as was the case with the FIFOs. However, we observed an interesting effect while doing the tests. Initially, our test FPGA design has been translated by the Xilinx tools as an unpartitioned design. Especially the SDRAM physical controller core took a long time to process. In order to save compilation time, we put the storage controller including the FIFOs into a separate partition after it was verified which could be reused in subsequent developments. To verify the partitioned design, we repeated some of the tests and ended up with similar results for maximum occupancy levels. Surprisingly, the average SDRAM level dropped by 33% with the partitioned design compared to the unpartitioned one. This means that the ring buffer controller managed to empty the SDRAM faster on average. So far, it is unclear what caused this discrepancy. Our best guess so far is, that due to the different clock domains involved, occupancy information of the FIFOs (which is used by various state machines to trigger data transfer) propagates with different delays between the two implementations. This underlines the importance of testing and verification of such complex FPGA designs and that (limited) effort to help the automated implementation tools by defining partitions can indeed improve results in high-bandwidth data-flow designs.

5 Conclusions and Future Work

A new high-bandwidth mass-storage architecture for FPGA implementation has been presented. It allows real-time capturing in mobile UWB sensor nodes in an efficient and economical way. With this architecture, it is possible to store large data streams of almost 300 MiB/s continuously over several minutes without any loss of data. This architecture makes use of affordable off-the-shelf components like FPGA module boards and SATA SSDs. Its special design features and hardware requirements were explained based on primary and secondary design objectives as defined by the UWB application. The solution of the storage buffering problem to compensate bus protocol and hardware latencies has been discussed in detail and it was shown that different non-real-time technologies like SSDs and SDRAM can be utilised while guaranteeing real-time capability with respect to UWB data acquisition. To prove this hypothesis, experiments were performed on real hardware and the results were presented. Aside from a a successful functional verification, analysis of the buffer occupancy during long test runs showed that the actual hardware requirements and FPGA utilisation could even be lower than initially estimated.

With the help of these experiments it can be seen that the efforts to built a real-time high-bandwidth mass-storage architecture for distributed UWB sensor nodes were successful. With the presented results it also becomes clear that the design still contains some unused margins and even more demanding storage needs could be handled this way. For example, it is possible to use more ADC channels and still guarantee that no data is going to be lost. The presented ring buffer based on SDRAM combined with BRAM is capable of accommodating much higher data rates than used here. One possible bottleneck could then be the single SSD. However, the Artix7 FPGA used for testing provides four MGTs, i.e. up to four SATA3 SSDs could be used in parallel without much added design effort. On the other hand, it would be possible to select a smaller FPGA and lower SDRAM capacity to drive down costs of the UWB sensor nodes even further. Another possible enhancement would be the possibility to evaluate the measured data on-the-fly between the sensor nodes and the SSD. This would open, depending on the special usage, many new options. Most of the problems require expensive calculations on the measured raw data. For example, the amount of stored data could be decreased if a highly effective softcore processor that can guarantee the real-time needs would perform the required calculations on-the-fly and stores only the results and not the raw data. Such a softcore processor is presented in [4].

References

[1] Altera Corp. (2015) Transceiver Technology. [Online]. Available: `https://www.altera.com/solutions/technology/transceiver/overview.html`

[2] C. Gorman, P. Siqueira, and R. Tessier, “An open-source SATA core for Virtex-4 FPGAs,” in *Field-Programmable Technology (FPT), 2013 International Conference on*, Dec. 2013, pp. 454–457.

[3] E. Kim. (2013, Aug.) SSD Performance - A Primer - An Introduction to Solid State Drive Performance, Evaluation and Test. SNIASSSI.SSDPerformance-APrimer2013.pdf. [Online]. Available: `http://www.snia.org/forums/sssi/knowledge/education`

[4] M. Kirchhoff and W. Fengler, "Realization of an embedded hard realtime softcore processor," in *Autonomous Systems 2014 Proceedings of the 7 th GI Workshop*, Oct. 2014, pp. 33–42.

[5] A. Mendon, B. Huang, and R. Sass, "A high performance, open source SATA2 core," in *Field Programmable Logic and Applications (FPL), 2012 22nd International Conference on*, Aug. 2012, pp. 421–428.

[6] R. Müller, D. A. Dupleich, C. Schneider, R. Herrmann, and R. S. Thomä, "Ultrawideband 3d mmwave channel sounding for 5g," in *General Assembly and Scientific Symposium (URSI GASS), 2014 XXXIth URSI*, Aug. 2014, pp. 1–4.

[7] J. Sachs, *Handbook of ultra-wideband short-range sensing - theory, sensors, applications*. Berlin: Wiley-VCH, 2012.

[8] J. Sachs and R. Herrmann, *M-sequence-based ultra-wideband sensor network for vitality monitoring of elders at home*, ser. IET radar, sonar & navigation. London : The Institution of Engineering and Technology, 2015. [Online]. Available: `http://dx.doi.org/10.1049/iet-rsn.2014.0214`

[9] Samsung Electronics. (2015) Samsung 850 PRO SATA3 2,5" SSD. [Online]. Available: `http://www.samsung.com/de/consumer/memory-storage/ssd/850-pro/MZ-7KE512BW`

[10] SATA-IO, "Serial ATA Revision 3.2 Specification," The Serial ATA International Organization, Standard specification, Aug. 2013. [Online]. Available: `https://www.sata-io.org/sata-revision-32`

[11] Trenz Electronic GmbH. (2015) Trenz Electronic TE0712 series FPGA module. [Online]. Available: `http://www.trenz-electronic.de/products/fpga-boards/trenz-electronic/te0712.html`

[12] C. W. Wagner, "Betrachtung aktueller Datenspeicher zum Einsatz in verteilten UWB-Echtzeit-Messsystemen," Technische Universität Ilmenau, Hauptseminar, Jul. 2015.

[13] Xilinx Inc. (2015) High Speed Serial - Xilinx Transceiver Offerings. [Online]. Available: `http://www.xilinx.com/products/technology/high-speed-serial.html`

Applying Centrality Measures in a Multi-criteria Routing Strategy for Wireless Sensor Networks

Sunantha Sodsee[1] and Maytiyanin Komkhao[2]

[1]Department of Data Communication and Networking
King Mongkut's University of Technology North Bangkok, Thailand

[2]Department of Computer Science, Faculty of Science and Technology
Rajamangala University of Technology Phra Nakhon, Thailand

Abstract: A novel multi-criteria routing strategy for wireless sensor networks is proposed. The networks are represented by undirected graphs. The ranges of the radio signals emitted by the sensors are assumed to be circular and represented by virtual links. With centrality measures and status data of sensor nodes, such as battery charging levels, probabilities of transitions to neighbouring nodes in multi-hop data paths are determined. In doing so, it is aimed to minimise travelling times between source and sink nodes, and to use as intermediate nodes the ones with the highest battery charging levels. Employing three different centrality measures, the routing strategy is evaluated by simulations on two network structures, viz. the grid and the triangle topologies. Effects on routing efficiency are found for two-hop connections, only.

1 Introduction

In the era of the Internet of Things (IoT) data of various kinds are increasingly monitored via wireless telecommunication. A prominent example for this are wireless sensors. Wireless sensor networks (WSNs) are utilised in numerous applications such as environmental monitoring, health-care, transportation, military and commerce [1, 2]. A WSN consists of a number of individual sensor nodes which are spatially deployed to the specific locations of data monitoring [1]. Neighbouring wireless sensor nodes must be placed within the ranges of their radio signals. In general, data are forwarded from source nodes to sink or destination nodes in form of single or multiple hops between network nodes. This intra-network data communication faces certain limitations with respect to data transmission rate, power consumption and network topology. Focusing on topology as the backbone of communication, a multi-hop network topology called indirected delivery communication was devised [4, 5], which is more effective than the directed one. In order to identify optimal routes for sending data from source to destination nodes, effective routing approaches are needed in multi-hop data paths [3].

As mentioned above, a novel multi-criteria routing strategy for wireless sensor networks is presented in this paper. It is based on a concept of multi-hop data paths for communication, and on multiple criteria for the identification of optimal routes. In the next section, a graph-theoretical representation of WSNs is presented. Section 3 details the proposed multi-criteria routing strategy and Section 4 presents simulation results.

2 Representation of Wireless Sensor Networks

A representation of wireless sensor networks (WSNs) using a graph theory is presented now. Any such network is assumed to consist of many wireless sensor nodes and a single sink node (see Figure 1). Typically, the sink node is central collection point gathering data from sensors. Each sensor node has its communication radius, i.e. the circular range reached by the radio signals it emits, within which it can wirelessly transmit data to other nodes. If need be, data are sent from a sensor node via other nodes before reaching the sink node, constituting multi-hop data paths.

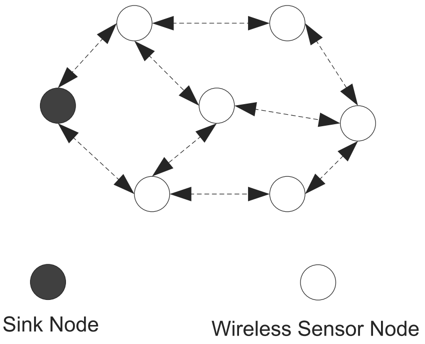
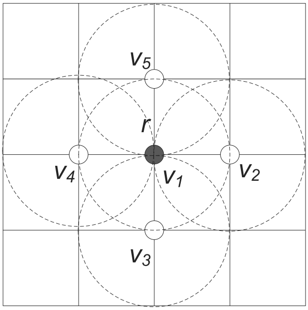


Fig. 1: Representation of wireless sensor networks



r = Communication Radius

Fig. 2: Circular ranges of sensor nodes' radio signals

We describe the topology of a network with $n \in \mathbb{N}$ wireless sensors by an undirected graph $G_n = (V_n, E_n)$, where V_n is a set of vertices (viz. these wireless sensors) v_i, and E_n a set of virtual links (called circular ranges of sensor nodes' radio signals, see Figure 2) $\eta_{ij} = (v_i, v_j)$, $i, j = 1, \ldots, n$. The adjacency matrix $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ associated with G_n is defined by

$$a_{ij} = \begin{cases} 1, & \text{if } v_i,\ v_j \in E_n, \\ 0, & \text{otherwise.} \end{cases}$$

For a sensor v_i, its set of neighbouring sensors is defined as $N_i = \{v_j \mid a_{ij} \neq 0,\ i \neq j,\ j = 1, \ldots, n\}$, and its degree as $deg(v_i) = \sum_{j \neq i} a_{ij}$. Thus, $|A|$ equals $\sum deg(v)$.

Hence, the network's graph representation allows to identify the *importance of sensors* by, for instance, centrality measures such as degree centrality (DC) and betweenness centrality (BC). Related to degree, DC indicates the number of a sensor's neighbours within one hop of connection, and is expressed as $DC(v_i) = \frac{deg(v_i)}{n-1}$. On the other hand, BC indicates how often a sensor occurs on the shortest paths between indirectly connected sensors. The betweenness centrality of sensor i is given by $BC(v_i) = \sum \frac{sp_{j,k}(v_i)}{sp_{j,k}}$, where $sp_{j,k}(v_i)$ represents the number of shortest paths connecting sensors j and k and passing through sensor i, and $sp_{j,k}$ is the number of shortest paths connecting sensors j and k.

3 Multi-criteria Routing Strategy for WSNs

The concept of a multi-criteria routing strategy for WSNs as depicted in Figure 3 is now described. It identifies the importance of sensor nodes by network-topological characteristics, using Pajek [6]. According to the objectives minimising the travelling time and maximising the battery charging levels of the sensor nodes along a path, the routing strategy employs inter-node transition probabilities to select the next node towards reaching the sink node in a multi-hop path.

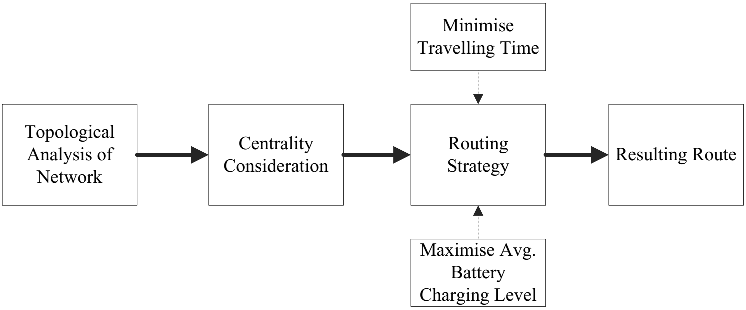


Fig. 3: Concept of routing approach

The probability of transition between two nodes is defined as

$$T_{Prob}(i,j) = p_j s_j C_j^{\frac{p_j}{100}} \tag{1}$$

where $p_j \in \{0\%, 10\%, 20\%, ..., 100\%\}$ is the battery charging level of sensor node j and $s_j \in \{0,1\}$ its status, viz. non-active or active. The centrality C_j of node j is measured by degree centrality ($DC(v_j)$) or betweenness centrality ($BC(v_j)$).

For routing purposes, assume that sensor v_i tries to send data to the sink node (v_s). To this end, v_i first requests from its neighbouring nodes N_i their local information, namely p_j, s_j and C_j. Then, it computes with Eq. (1) the transition probability for each neighbouring node j. The node j with the highest value T_{Prob} is selected to be the next node on the route to the sink node. If the selected next node j is already the sink node v_s, then the data have reached to their destination. Otherwise, the process of next node selection based on transition probability is be repeated.

4 Simulation Results

The routing strategy introduced above and its efficiency is now investigated simulatively for different topologies of sensor networks.

4.1 Simulation Setting

For the simulations two sensor networks with 25 nodes each and structured according to the grid and triangle topologies, respectively, as shown in Figures 4 and 5 are chosen. Values $p_j \in \{0\%, 10\%, 20\%, ..., 100\%\}$ and $s_j \in \{0, 1\}$ are randomly assigned to each sensor node in the networks. In each simulation step, the p_js are reduced by 10%. Simulation runs are carried out for different values of the parameter r within the range of 1 to 4 indicating the maximum number of hops in multi-hop data transmissions. To analyse the routing efficiency, source and sink nodes are randomly selected.

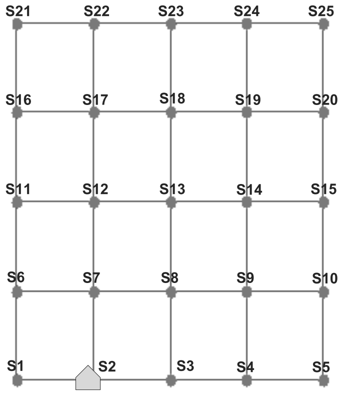


Fig. 4: Grid topology

4.2 Characteristics of Topologies

In the grid topology, any communication reaches its sink after not more than 2 to 4 hops, i.e. the maximum degree of nodes in this network is 4. On the other hand, the triangle topology has a greater maximum degree than the grid, namely 6.

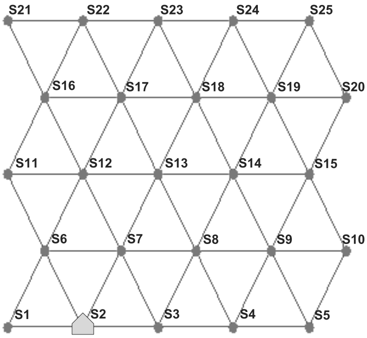


Fig. 5: Triangle topology

Table 1: Characteristics of topologies grid and triangle for $r = 2$

Topology	Density	Average Degree	Average Distance
Grid	0.34	8.16	1.9266
Triangle	0.46	11.04	1.62

In this subsection, these two topologies are analysed in terms of density, average degree and average distance to reach all nodes in the networks. Their characteristics are shown in Tables 1 and 2 for the circular ranges of radio signals $r = 2$ and 4, respectively. The tables show that the triangle topology has greater network density and average degree than the grid topology. In contrast, the former has less average distance, which means that it provides greater accessibility to all nodes in the network than the grid topology.

4.3 Centrality Consideration in Multi-criteria Routing

In this subsection, the effect of centrality measures is evaluated by employing in Eq. (1) the degree (DC) and the betweenness (BC) centralities, respectively. The comparisons will reveal the efficiency of the routing method proposed applied

Table 2: Characteristics of topologies grid and triangle for $r = 4$

Topology	Density	Average Degree	Average Distance
Grid	0.8466	0.32	1.153
Triangle	0.92	22.08	1.08

Table 3: Routing efficiency of 10% active sensor nodes on grid ($r = 2$)

Centrality	Average Travelling Time	Average Battery Charging Level
Degree	3	27.71143
Betweenness	5	28.04967

Table 4: Routing efficiency of 90% active sensor nodes on grid ($r = 2$)

Centrality	Average Travelling Time	Average Battery Charging Level
Degree	2	41.1
Betweenness	2.333	29.4362

to the two topologies considered for variations of the number of active sensor nodes.

Table 3 presents the comparison on for the grid topology when 10% active sensor nodes are available and with $r = 2$. The results show that the measure DC provides for lower average travelling time and average battery charging level than BC.

When the number of active sensor nodes is increased to 90%, the average travelling times and the average battery charging levels for both DC and BC are better than for the case of 10% active sensor nodes. The results are shown in Table 4.

Presenting the routing efficiency for the triangle topology, Table 5 shows that the centrality measure DC yields higher average travelling time and average battery charging level than BC when considering 10% active sensor nodes with $r = 2$.

When the number of active sensor nodes is increased to 90%, the average travelling times are better than in the case of 10% active sensor nodes, and almost equal for DC and BC. The average battery charging level is increased for the measure DC. The results are shown in Table 6.

Table 5: Routing efficiency of 10% active sensor nodes on triangle ($r = 2$)

Centrality	Average Travelling Time	Average Battery Charging Level
Degree	3.667	39.381
Betweenness	2.33	29.4362

Table 6: Routing efficiency of 90% active sensor nodes on triangle ($r = 2$)

Centrality	Average Travelling Time	Average Battery Charging Level
Degree	1.667	34.61
Betweenness	1.667	29.6683

Table 7: Routing efficiency on structure topologies for $r = 4$

Centrality	Average Travelling Time	Average Battery Charging Level
Degree	1	9.9
Betweenness	1	9.9

Increasing the circular range of radio signals r now to 4 results in the average travelling times and the average battery charging levels both to be equal for DC and BC, even though evaluated on different topologies and different number of active sensor nodes, because the great value of r leads to bigger hops and, thus, often facilitates single-hop data paths. The results are shown in Table 7.

To summarise, employing the centrality measure DC provides routing results with fluctuating average travelling times and average battery charging levels when evaluated on different topologies and for different numbers of active sensor nodes. On the other hand, the routing results obtained when using BC show that they depend on the number of active sensor nodes. Comparing the topologies, applying BC on the triangle topology also shows better average travelling times and average battery charging levels for the determined routes than the grid topology.

5 Conclusion

Aiming to minimise travelling time and to maximise the battery charging level of node members along the routes determined, a multi-criteria routing strategy for wireless sensor networks was devised. Based on given aims, it employs centrality measures (DC and BC), states and battery charging levels of sensor nodes to identify transition probabilities in selecting the respective next nodes in multi-hop transmissions. Simulation results reveal that employing BC yields better routing results on the topologies considered than DC. In future work, other parameters of wireless sensor nodes such as data transmission rate or power consumption should be considered to determine the transition probability.

Acknowledgements

We gratefully acknowledge the financial support by the Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Thailand, and of Gesellschaft der Freunde der FernUniversität e. V., Hagen, Germany.

References

[1] Akyidiz, I.F., Su, W., Sankarasubramaniam, Y.: Wireless Sensor Networks: A Survey, *Computer Networks*, 34, 4, 393–422, 2002

[2] Alemdar, H., Ersoy, C.: Wireless Sensor Networks for Healthcare: A Survey, *Computer Networks*, 54, 15, 2688–2710, 2010

[3] Al-Karaki, J.N., Kamal, A.E.: Routing Techniques in Wireless Sensor Networks: A Survey, *IEEE Wireless Communications*, 11, 6, 6–28, 2004

[4] Gengzhong, Z., Qiumei, L.: A Survey on Topology Control in Wireless Sensor Networks, In: *Proc. 2nd Intl. Conf. on Future Networks (ICFN 2010)*, pp. 376–380, 2010

[5] Ramanathan, R., Rosales-Hain, R.: Topology Control of Multihop Wireless Networks Using Transmit Power Adjustments, In: *Proc. IEEE Intl. Conf. on Computer Communication (IEEE INFOCOM 2000)*, pp. 404–413, 2000

[6] Pajek: Program for Large Network Analysis, `http://mrvar.fdv.uni-lj.si/pajek`

Routing-based Topological Analysis on the Road Network in Myanmar

Tun Tun Naing[1] and Sunantha Sodsee[2]

[1]Department of Information Technology
[2]Department of Data Communication and Networking
King Mongkut's University of Technology North Bangkok, Thailand

Abstract: An efficient routing algorithm for the road network of Myanmar is presented. It utilises a concept of topological analysis and local information on the network to identify a routing strategy aiming to maximise traffic flow, feasibility of road selection and low processing time to identify suitable routes. The road network is represented by a weighted undirected graph. A new centrality measure is formed by combining degree and betweenness centralities with the Cantor pairing function, and employed to determine efficient paths between cities. Corresponding simulations reveal that the routes from source to destination cities generated by the algorithm meet the objectives set. It requires the lowest processing times as compared with the current approaches to generate routes, which show high average degrees and betweenness centralities.

1 Introduction

Myanmar is situated in South-East Asia, sharing borders with Bangladesh, India, China, Laos and Thailand. It consists of seven states and seven regions as shown in Fig. 1. The country extends about 925 km from east to west and 2,090 km from north to south. It has a population of around 60 million people living on a surface area of 678,033 km^2 that makes Myanmar one of the largest Asian mainland countries [1].

After having gained independence from Great Britain in January 1948, Myanmar started to develop its transport networks. Economic crises in the past caused investments into the infrastructure to decline. As a result, Myanmar's transport sector is now underdeveloped as compared with other member countries of the Association of Southeast Asian Nations (ASEAN).

In fact, Myanmar is a developing country facing logistic problems due to the low quality of its transportation networks [2]. The overall road density for ASEAN is about 11 km per 1,000 people, whereas Myanmar has just 2 km per 1,000 people [3]. On the other hand, the number of registered vehicles increases continuously and stands now at some 2.5 million [3].

At present, the four Ministries of Construction, of Rail Transport, of Transport and of Border Affairs are responsible to develop or provide transportation services. As shown in the report [4], the total length of roads is 148,689 km consisting of union highways, township roads, major city roads and village roads. Damaged by natural disasters, the use of a considerable part of these roads is limited. As a consequence of this situation, Myanmar needs efficient routing guidelines to avoid using damaged roads.

Several approaches for efficient routing in transportation networks were developed based on shortest-distance considerations and given characteristics of nodes [5]. Later, routing methods based on the topological properties of networks were devised [2, 6–8] producing more efficient paths. In fact, taking traffic congestions and waiting times into consideration, efficient paths may not necessarily be the shortest ones [9].

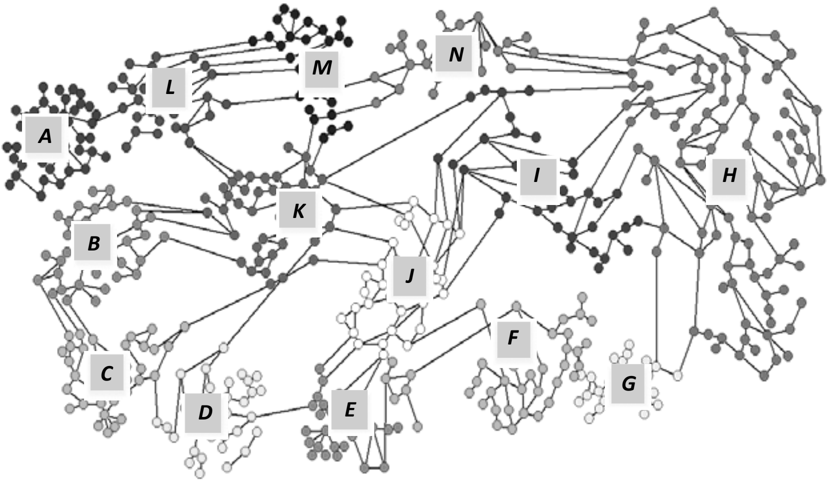
Therefore, in this paper, an efficient routing strategy suitable for the road network in Myanmar is proposed. It employs topological analysis and local information on road networks in order to select feasible roads and to maximise traffic flow. The body of the paper begins with introducing a representation allowing to analyse the road network in Section 2. The topological analysis-based routing approach is then described in Section 3, and simulation results are presented in Section 4.

2 Representation of Road Networks

To facilitate analysis, we represent a road network connecting $n \in \mathbb{N}$ cities by a weighted undirected graph $G_n = (V_n, E_n)$, where V_n is a set of vertices (viz. these cities) v_i, and E_n a set of links (called roads connecting cities) $\eta_{ij} = (v_i, v_j)$, $i, j = 1, \ldots, n$. The weight w_{ij} denotes the length of the road η_{ij} in kilometres. With the adjacency matrix $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ associated to G_n and defined as

$$a_{ij} = \begin{cases} 1, & \text{if } v_i,\ v_j \in E_n, \\ 0, & \text{otherwise.} \end{cases}$$

for a city v_i its set of neighbour cities is defined as $N_i = \{v_j \mid a_{ij} \neq 0,\ i \neq j,\ j = 1, \ldots, n\}$ and its degree as $deg(v_i) = \sum_{j \neq i} a_{ij}$. Thus, $|A|$ equals $\sum deg(v)$. Fig. 1 depicts the graph of the road network in Myanmar classified by its seven states and seven regions. The number n of cities totals 523 and $|\eta_{ij}| = 643$.



A	-	Tanintharyi	F	-	Sagaing	K	-	Pago
B	-	Yangon	G	-	Kachin	L	-	Mon
C	-	Ayeyarwady	H	-	Shan	M	-	Kayin
D	-	Rakhaine	I	-	Mandalay	N	-	Kayah
E	-	Chin	J	-	Magway			

Fig. 1: The road network in Myanmar

The network's graph representation also allows us to identify the *importance of cities* by, for instance, centrality measures such as degree centrality (DC) and betweenness centrality (BC). Related to the degree, DC indicates the number of an city's neighbours within one hop of connection. It is expressed as $DC(v_i) = \frac{deg(v_i)}{n-1}$. On the other hand, BC indicates how often a city occurs on the shortest paths between indirectly connected cities. The betweenness centrality of city i is given by $BC(v_i) = \sum \frac{sp_{j,k}(v_i)}{sp_{j,k}}$, where $sp_{j,k}(v_i)$ represents the number of shortest paths connecting cities j and k and passing through city i, and $sp_{j,k}$ is the number of shortest paths connecting cities j and k.

3 Routing based on Topological Analysis

Now, a routing strategy is to be proposed improving the efficiency of transportation efficiency on the road network in Myanmar. The idea of this approach is to find the best possible paths between source and destination cities by centrality considerations as based on topological analyses of the network (see Section 2) derived by Pajek [10], but also to utilise local information of nodes represented by the transition probabilities ($T_{Prob} \in [0,1]$) for selecting the respective next cities route. This means that for identifying suitable paths is not necessary to know the global structure of the network. Hence, an implementation of the approach will require the lowest amounts of memory and processing time in comparison with the current ones. Optimal paths between source and destination cities are selected by consideration of undisturbed traffic flow, feasibility of roads and of degree and betweenness centralities. To achieve these aims, here we consider the highest centrality values of the cities expressed by combining each pair of degree centrality and betweenness centrality with the Cantor pairing function [11].

3.1 Transition Probabilities

In this section, three transition probabilities will be defined, namely based on degree centrality, on betweenness centrality and on a combination of these two centralities expressed by the Cantor pairing function.

The transition probability based on degree centrality (DC) is defined as

$$T_{Prob}(DC)_i = exp^{-w_{ki}} \frac{DC(v_i)}{\sum_j exp^{-w_{kj}} DC(v_j)} \tag{1}$$

where w_{ki} represents the distance between v_k and v_i, $DC(v_i)$ is the degree centrality of v_i and the $DC(v_j)$s are the degree centralities of all v_k's neighbours, $j = \{1, ..., |N_k|\}$. Moving from v_k, the next node v_i on the route is determined as that neighbour, for which the transition probability assumes its highest value.

Analogously, the transition probability based on betweenness centrality is defined as

$$T_{Prob}(BC)_i = exp^{-w_{ki}} \frac{BC(v_i)}{\sum_j exp^{-w_{kj}} BC(v_j)} \tag{2}$$

where $BC(v_i)$ is the betweenness centrality of v_i and $BC(v_j)$ is the betweenness centrality of all v_k's neighbours.

Combining degree and betweenness centralities with the help of the Cantor pairing function

$$Cantor(v_i) = \frac{1}{2}(DC(v_i) + BC(v_i))(DC(v_i) + BC(v_i) + 1) + DC(v_i) \qquad (3)$$

leads to the analogous definition of a further transition probability:

$$T_{Prob}(DC, BC)_i = exp^{-w_{ki}} \frac{Cantor(v_i)}{\sum_j exp^{-w_{kj}} Cantor(v_j)} \qquad (4)$$

Here, more weight is given to degree centrality than to betweenness centrality in order to favour flow of traffic and feasible road alternatives.

3.2 Routing based on Combined Degree and Betweenness Centralities

The transition probability $T_{Prob}(DC, BC)_i$ defined above is now applied for routing purposes, i.e. to select the next node v_i on a path before reaching the destination node. To this end, the v_i is selected for which the highest value of T_{Prob} is assumed. This consideration gives rise to the following algorithm.

Step 1: Initialise an ordered list **P** with source and destination node of a path to be determined.

Step 2: Let the source node be the current node v_k.

Step 3: Compute the transition probabilities of all neighbours of v_k with Eq. (4).

Step 4: Insert into **P** as next node the v_i for which the highest transition probability was assumed in Step 3.

Step 5: If v_i is the destination node, then terminate.

Step 6: Set $v_k = v_i$ and go to Step 3.

To show the efficiency of routing based on the combinational transition probability introduced by Eq. (4) and expressing the importance of each node in a network, in Fig. 2 an example is given for finding a route from node S to node D, where the path distance between the nodes is equal, i.e. $w_{ij} = 1$. The results are shown in Table 1. The most important node is S, whereas 1, 4, 5 and D are the nodes of lowest importance. Applying the proposed routing approach to this network yields the route $v_S - v_3 - v_8 - v_D$, where $DC(v_3) = 0.33$, $BC(v_3) = 0.5$, $DC(v_8) = 0.33$ and $BC(v_8) = 0.26$. The rout's length is 3 and the averages of DC and BC are 0.33 and 0.38, respectively.

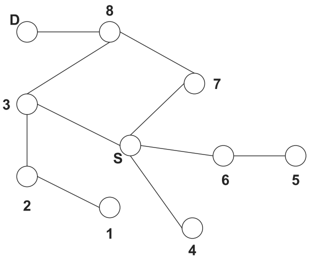


Fig. 2: Example of a network and of routing based on $T_{Prob}(DC, BC)$

Table 1: Defining node importances by $Cantor(v_i)$

Node ID	Value of Cantor func.	Rank
S	1.7	1
1	0.1	6
2	0.5	4
3	1.3	2
4	0.1	6
5	0.1	6
6	0.5	4
7	0.3	5
8	0.7	3
D	0.1	6

4 Simulation Results

The road network of Myanmar is used to validate the approach presented above. There are 523 cities in this network represented by nodes, which are connected by 643 major roads. The network's density is only 0.0047, which means that the number of roads connecting cities is very low. With respect to the states, the highest density of 0.0788 is found in Kayah state. In contrast, the Shan state's density is the lowest one with only 0.0229. Kayah state is important for the network, not only as its location is close to the new capital city Naypyitaw, but also since it has several roads connecting to cities within the state. Regarding cities, Kengtung, Pathein and Loikaw are the most central ones as far as degree centrality is concerned, whereas Hpa Yar Gyi is the most important city with respect to betweenness centrality. On the other hand, Myanmar's biggest city, former capital and most important commercial centre Yangon and the current capital Naypyitaw do not appear as most important cities as measured by centralities, because they have less roads connecting them to their neighbours than other cities in the network. As a consequence, people may not reach these big cities easily and conveniently.

To evaluate the proposed routing approach, three source and destination cities were selected randomly, which are $T1$: routing from Hwei Long to Hsipaw, $T2$: routing from Moke Pa Lin to Kyon Nye, and $T3$: routing from Win Kan to Kawa. The corresponding simulation results listed in Table 4 can be compared with the ones for the routings derived by the degree-based and betweenness-based approaches as shown in Tables 2 and 3, respectively. These results reveal that routing based on the combinational transition probability $T_{Prob}(DC, BC)_i$ renders routes with the best characteristics, viz. high average degree and betweenness centralities, lowest numbers of cities passed enroute, shortest average distances and lowest processing times.

Table 2: Routing results derived from degree-based approach

Route	Avg. DC	Avg. BC	No. Cities	Avg. Dist.	Processing Time
$T1$	3.39	0.0631	142	54.06 km	15.711 ms
$T2$	3.05	0.0793	19	44.25 km	6.220 ms
$T3$	3.28	0.0676	144	49.72 km	16.500 ms

Table 3: Routing results derived from betweenness-based approach

Route	Avg. DC	Avg. BC	No. Cities	Avg. Dist.	Processing Time
$T1$	3.41	0.0944	78	52.45 km	19.647 ms
$T2$	3.30	0.1096	73	54.70 km	9.199 ms
$T3$	3.23	0.0954	80	49.24 km	14.776 ms

Table 4: Routing results derived from the proposed approach

Route	Avg. DC	Avg. BC	No. Cities	Avg. Dist.	Processing Time
$T1$	3.50	0.1244	10	34.43 km	5.4848 ms
$T2$	3.31	0.1411	13	39.32 km	4.6480 ms
$T3$	3.36	0.1119	45	44.57 km	7.7489 ms

5 Conclusion

An efficient routing algorithm aiming to maximise traffic flow and feasibility of road selection was presented, which is based on a combination of the degree and betweenness centralities by the Cantor pairing function. Three transition probability functions were introduced and compared with respect to their performance in routing applications. Simulation results show clearly that the combinational transition probability leads to routes not only of shortest distances, but also of high average degree and betweenness centralities. This proves that the generated routes use the most feasible roads and facilitate better flow of traffic. It is remarkable that the algorithm's execution takes the lowest processing time.

Acknowledgements

We gratefully acknowledge the financial support of Thailand International Cooperation Agency (TICA), of the Faculty of Information Technology at King Mongkut's University of Technology North Bangkok, Thailand, and of Gesellschaft der Freunde der FernUniversität e.V., Hagen, Germany.

References

[1] Ministry of Information, Myanmar, `http://www.moi.gov.mm`

[2] Mohmand, Y.T., Wang, A.: Complex Network Analysis of Pakistan Railways, *Discrete Dynamics in Nature and Society*, pp. 1–5, 2014, Hindawi Publishing Corporation

[3] Road Transport Administration Department, Myanmar, `http://www.myanmarrtad.com`

[4] Ministry of Construction, Myanmar, `http://www.ministry-construction.gov.mm`

[5] Zhang, Z., Jigang, W., Duan, X.: Practical Algorithm for Shortest Path on Transportation Network, *Intl. Conf. on Computer and Information Application (ICCIA 2010)*, pp. 48–51, 2010

[6] Hossain, M., Alam, S., Rees, T., Abbass, H.: Australian Airport Network Robustness Analysis: A Complex Network Approach, *Australasian Transport Research Forum 2013*, pp. 2–4, Brisbane

[7] Kocur-Bera, K.: Scale-Free Network Theory in Studying the Structure of the Road Network in Poland, *Promet–Traffic & Transportation*, 26, 3, 235–242, 2014

[8] Park, K., Yilmaz, A.: A Social Network Analysis Approach to Analyze Road Networks, *ASPRS Annual Conf. 2010*, pp. 26–30, San Diego, CA.

[9] Yan, G., Zhou, T., Hu, B., Fu, Z.Q., Wang, B.-H.: Efficient Routing on Complex Networks, *Physical Review E*, 73, 4, 2006

[10] Pajek: Program for Large Network Analysis, `http://mrvar.fdv.uni-lj.si/pajek`

[11] Cegielski, P., Richard, D.: Decidability of the Theory of the Natural Integers with the Cantor Pairing Function and the Successor, *Theoretical Computer Science*, 257, 1–2, 57–77, 2001

Toward Authentication between familiar Peers in P2P Networking Systems

Fariborz Nassermostofi

Chair of Communication Networks
Faculty of Mathematics and Computer Science
FernUniversität in Hagen, Germany

Abstract: Due to the anonymity of peers in P2P networking systems and absence of a CA, the authentication of peers is the main problem. A secure environment can only be achieved, when peers are sure, that they are communicating with the partner desired. After the first contact peers are known to each other. But there is still the possibility that a malicious peer masquerades itself as another one using a spoofing attack. A mechanism is needed to ensure secure re-authentication of peers, who have met each other once and are known to each other. This work focuses exactly on this problem and proposes a mechanism for familiar peers to get certainty about the identity of their known communication partner. The proposed mechanism combines a cryptographic protocol inspired by the dinner cryptographic protocol and zero knowledge protocol to ensure continuously re-authentication of peers in each new session.

1 Introduction

An important aspect of each application consisting of different participant in a networking system is the matter of security. Participants need to have trusted relationship and be sure about the authenticity of their partners. Various security mechanisms are proposed for use in P2P environments. But still there is not a 100% secure mechanism making this systems secure enough to be used in industry.

In client/server systems only known and registered members are trusted and allowed to gain access to the network resources. In P2P systems, peers are unknown. In client / server environment all clients will contact and communicate with a server which can provide certifications issued by a CA (Center of Authorization). Peers in a P2P environment communicate with each other without any server and they don't have any certification issued by a CA.

On the other side P2P systems are scalable, easy to set up, represent the natural way of social live. These systems are also robust and stable by missing some peers. Setting up a client server with support for many clients will need an organization with policies and required hardware infrastructure. P2P systems don't need any such an organization and could be set up without any effort for hardware infrastructure.

P2P systems are a point of risk for applications, which need a secure environment. A Secure environment covers authentication of participants and needs mechanisms and facilities to establish trustful communication between participants. All of this circumstances are hard to achieve due to the anonymous nature of these systems and also due to the lack of a CA. What is needed is an environment, in which applications could feel confident about security related matters while fulfilling their business requirements.

Overcoming this problem needs a mechanism, which enables familiar peers to re-authenticate each other using kind of additional functionality with the help of some private relation based authentication credentials. It is an essential requirement, that these private information remain secret. Sending these credentials over the net, even if encrypted, is a point of danger. In fact the best secure way is, if they both agree on some credentials without exchanging them even at the moment of the authentication.

1.1 Contribution

In the present work, a mechanism will be proposed, which uses a combination of protocol inspired by dining cryptographic protocol and zero-knowledge protocols to ensure the highest degree of secrecy during the authentication process between two familiar peers.

While dining cryptographic network (DC-nets) are used to ensure the untraceability of users within a community, in the present work a similar approach has been chosen, which ensures, that two parties exchange information and ask for votes using the same shared secret information like in the dc-nets. Using this protocol, two peers will be able to agree about a shared key as secret information without revealing this information over the net by using the Diffie-Hellman key exchange protocol. The shared information will be used then to produce the same result on both sides answering a question issued by one of the peers.

To achieve a 100% privacy within the authentication process, even that result will be kept secret and only the prove of possession of the right result will be sent to the communication partner using the zero-knowledge protocol. The combination of both mentioned mechanisms ensures a 100% secure and trusted authentication between two peers, which have had once the possibility to get know each other, i.e. establishing a first contact and got a first successful authentication using one of established mechanisms like public / private key system.

After the first session and creation a set of private relation based credentials with the above method, they can be sure that they will be able to authenticate each other in further sessions without fearing that they might get victim of a spoofing or a Man in the middle attack.

1.2 Structure of the Paper

The remaining parts of this paper is organized as followed. While section 1 delivered an introductory information about the subject of research and its brief solution, section 2 introduces and discusses shortly established mechanisms used in authentication and cryptographic area.

In section 3 then the proposed mechanism to be used during the authentication phase between familiar peers will be explained.

2 Related Works

Security in networking area covers a wide range of subjects. This wide spectrum begins with trusted environment and secure authentication up to resistivity against special attacks aimed to harm those environments. A secure environment must provide facilities to handle each of the above mentioned issues.

As a special item in the field of security, the authentication is an important subject. As far as one of the basic features of P2P system the anonymity is, there is no authentication in the classic form. There is no registered identity of a real person or user. There is no login mechanism, also there is no On and Off status of members. This implies that member of these networks are not authenticated by an authentication service.

Deduced From The observation above, it will be clear, that in such a system even authentication will not be the same process as in any other networking system.

Communication of peers can be victim of attacks like Man in the middle even after a first authentication during the handshake phase of peers.

2.1 Authentication in P2P Systems

To be able to elaborate the matter of authentication, we need to clarify what we need as authentication functionalities in a P2P environment. It is clear that in P2P systems especially in pure P2P systems, there is no central services for administration tasks. There for one of the major criteria in an authentication system in P2P environment is, that it has to work without a central authentication instance (CA). So the mechanism we look for must be ***decentralized***.

In absence of a CA we need special mechanism to make sure the required credential for identifying a peer is available for all peers. There for they must be dispersed over the network and be available to all peers i.e. they must be ***distributed***. So far we are able to define some criteria for an authentication system based on the discussion above. The secure authentication in a P2P environment must be:

1. Decentralized
2. Distributed

Authentication mechanisms belong either to the challenge/response interactive group or to the one-way authentication methods. The following subsection introduces some of methods used in distributed one time authentication mechanisms.

Threshold Cryptography The (k, n) threshold scheme was first introduced by Shamir in [1]. Within this algorithm a data item such as a secret key is divided into n part. To reconstruct the secret item at least k parts of n is needed. Knowledge about k - 1 part of the whole n parts is not sufficient to compute the original key where n = 2 k - 1. The set of n nodes are also called access structure of the scheme. The draw back of this scheme is that to initiate the shares and build the access structure a known centralized node (sometimes called the dealer) is needed [2].

Certificate Chains The approach is based on public-key certificate distribution. In this pure P2P approach a node updates a sub-graph of the certification graph of the network periodically. Whenever two nodes want to authenticate each other, sub-graphs are merged with or without helper nodes in an attempt to create a certificate chain. [3]. Within this system when a user *u* wants to obtain the public key of another user *v*, she acquires a chain of valid public key certificates. The certificate of node *v* will be the last certificate in the chain. The first certificate of the chain can be directly verified by *u* itself. The draw back of this method is, that the certification chain can be long and depends on other peers holding necessary certificates.

Distributed PK via DHT Within this model the authors of [4] use authentication credentials created during the first registration phase to authenticate each node. A new joining node creates its public and private key pairs and uses the created public key as its own user id within the P2P network environment. The node id i.e. the public key is also used to route requests using a KBR (key based routing) algorithms in DHT tables. Using the Public key to send encrypted requests to each nodes makes a secure authentication and communication at the same time possible. This can be used for a first authentication, but does not prevent the system from being victim of spoofing and Man in the middle attacks in communication phase after the authentication process.

P2P Anonymous Authentication (PPAA) The authors of [5] proposed a credential system in which peers are pseudonymous to one another (that is, two who interact more than once can recognize each other via pseudonyms) but are otherwise anonymous and unlinkable across different peers. The linkability of authentication runs in the scheme. In their solution they use the following set of attributes for the linkability context in P2P environment. LC = {{client-ID, server-ID}, event-ID}. So each authentication attempt is related to a certain event between the server and client. Entities involved in PPAA are the Group Manager (GM) and a set of peer users, or simply peers. The GM is responsible for registering peers. Operations in this model are {setup, registration, authentication and linking}. The draw back of this method is the need of GM as a centralized server to manage the registering of peers while authentication and linking operations are fully distributed and decentralized.

Distributing Security-mediated PKI (SEMs) This approach is based on mediated RSA (mRSA). As in standard RSA, each user has a public key (nu , eu) and a private key du , where n is the product of two large primes. The authors of [6] split the private key du in two parts, parts of a user's secret key are dsem,u and duser,u. The part dsem,u is then held by the SEM. All private key operations require the participation of both the user and the SEM. They use threshold cryptography to distribute the dsem,u between multiple SEMs and certification change technique. Each of the multiple SEMs are considered as a trustworthy islands. A trusted party (e.g., a CA) performs key setup by generating a statistically unique credential for each user. The draw back of this method is it's need for a CA to generate credentials for each user.

Selforganized Public Key Management enables an authentication without any centralized service in an ad-hoc network. In a self-organized public-key management system, all nodes automatically get new public keys from trusted neighbor nodes in an ad hoc network [7]. Nodes create their public / private keys and distribute their public key across their known neighbors. During this publish operation a certificate graph will be created. Collected certificates in certificate graph will be exchanged then with other known nodes. This graphs will be updated continuously with new certificates. This method needs no central instance and is fully distributed but the management of the certificate graph can be a heavy task with the growing the P2P networking system.

Table 1: P2P authentication methods

	Decentralized	Distributed
Threshold Cryptography	No	Yes
Certificate Chains	Yes	Yes
Distributed PK via DHT	Yes	Yes
P2P Anonymous authentication (PPAA)	No	Yes
Distributing Security Mediated PKI	No	Yes
Self organized public key management	Yes	Yes

Table 1 lists the above elaborated algorithms regarded to the criteria defined in this section earlier. While all of the mentioned algorithms are distributed, some of them rely on a central instance to disseminate the information they want to distribute.

2.2 Anonymous Authentication

As it is described in section 1, one of the features of P2P networking system is the anonymous nature of peers behavior. There is no registration and Login process in such a networking systems. The Message exchange performs in an environment, in which normally sender and receiver are anonymous. This is called anonymous broadcast communication. The problem within this type of the communication is clearly the matter of authentication.

The key-point here is, that a sender can communicate with a receiver, without the receiver learning the identity of the sender. There are a few successful system in this area like, Tor, Tarzan, Freenet, and FreeHaven [8].

This area is related the aim of this paper, where two peers which are anonymous and can not deliver any secure identification credentials, due to the lack of a server and a CA, need to authenticate each other. The problem of anonymous broadcasting communication is intensively elaborated and studied in Dining cryptographers problem (DC-nets) [8], which will be discussed in next subsection.

Dining Cryptographers Problem

Dining cryptographers Problem or shortly DC-nets protocol is first introduced by David Chaum in his work in [9]. He illustrated the introduced algorithm with the famous example of the problem, that few friends were going for a dinner and at the end they wanted to know who has paid the bill for that dinner; one of them or NSA as example. If one of them has paid, they did'nt want to know who. Within the proposed protocol by Chaum, every two person share a secret key. The first person votes then with 1 or 0 as her statement whether she has paid for the dinner or not and The result will be sent to the second person encrypted using the shared secret key.

The second person will also vote and performs a Boolean XOR operation on her vote and the vote of the first person. the result determines whether one of them has paid or not. when the result is 1, one of them has voted with 1 and has paid for the dinner. Otherwise no one of the two person has paid for the dinner. The result of this two person will be sent then to the next person in the round and so on. At the end of the round the result is 1 or 0 and they will be able to know whether one of them has paid for the dinner or not without knowing who has paid in the case that the final result is 1.

The secret key will be created between any two communication partners using Diffie-Hellman protocol [10], which ensures the creation of a secret information between two parties without sending it through the communication media, whatever it might be. Within this protocol both parties create a secret key, under the combination of a shared key, known to both of them, and their own private key. At the end of the protocol, both parties are in possession of the same secret key, which is known only by the two parties.

As it is described in [11], DC-nets are normally used for sender-anonymous, point-to-point communication. Especially when there is more than two communication partner. Message packets will be encrypted using the same secret key on both side.

Due to the nature of this protocol and the need to establish a private communication between every two peers, The protocol needs a higher effort Ω(N 2), both computational and communication in large groups [12], which is the reason why this protocol is not used widely.

Zero-knowledge Protocol

As a tool in the area of cryptography, Zero-knowledge protocol plays an important role, where not the desired shared information, but the proof of the possession of that information will be delivered. This protocol was first proposed by Goldwasser et al in [13]. This is an interactive proof system (P,V), where P is a prover and V is a verifier.

A zero-knowledge proof must satisfy three properties [13].

1. Completeness: that it should be possible to ″prove″ a true theorem
2. soundness: that it should not be possible to ″prove″ a false theorem.
3. Zero-knowledge: that is by proofing the theorem, no information about the theorem itself is revealed.

An important factor in a Zero-knowledge protocol is the choice of the mathematical problem as the basis of the protocol. One of the most used mathematical problem is the graph isomorphism problem and hamiltonian cycle graph. The interactive process of proofs between the prover and verifier using the graph isomorphism and hamiltonian cycle graph proceeds normally with the creation of a graph G_0 by the prover The prover knows the hamiltonian cycle graph to graph G_0. Now in each interaction the prover needs to create another distinct isomorph graph of G_0 and show as an example that she knows the hamiltonian cycle of the created isomorph graph, which is then the proof that she knows the hamiltonian cycle of the original graph.

Due to the survey in [14] The protocol is used in various forms, such as Zero Knowledge Password Authentication Protocol or a more sophisticated form like Zero knowledge Password Authentication Protocol with Public key encryption. The protocol is even used in P2P environments for authentication as it is used in [15] to prove the possession of the right authentication credentials. They use a two phase scheme in which in the 1. phase a registration of peer at a management site will be performed. In the second phase peers authenticate themselves without the need of the management peer using Zero-Knowledge protocol. The draw back of their solution is the need for a centralized management peer.

while the ZKP model is an interactive protocol in which messages will be sent to the partner several times, another non-interactive variant of this protocol is proposed in [16]. In their solution all required information of every interaction will be send in one message consisting of different segments. Every segment consists of one representation of an isomorph graph to the original graph. while the first segment is not encrypted, all other segments are encrypted each with different key. The encryption key of each segment depends on the previous segment. Every user who want to decrypt the last segment, must decrypt all other segments.

Both party agree to use a one-way hash function to define the challenge that the receiver must solve on each isomorphic graph. Additionally the same hash function will be used to define the encryption key for each message segment.

As it is introduced in [16] the operations, the prover has to perform to verify the correctness of the proof is as followed:

1. Process the first segment of the message, which is not encrypted.
2. Compute, using the hash function, the challenge that matches the information included in the segment.
3. Check whether the response corresponds to the challenge and the isomorphic graph.
4. From the challenge, compute the key it has to use to decrypt the next segment.
5. Apply Steps 2 through 4 until the last segment, which once deciphered contains the information needed to establish the desired result.

The security level of this protocol depends on the number of segments in the message. The more segments are used in the message the higher is the more complex is to reach the last segment. This model is known as Non-Interactive Zero-knowledge proof (NIZKP).

3 Proposed Authentication Mechanism

The mechanism proposed in this paper uses features of the both above described mechanisms in cryptography to ensure a secure authentication between peers, which are known to each other. Inspired by the DC-nets protocol, a shared key will be created between each two communication parties after the first authentication to be used in the next session using Diffie-Hellman protocol.

When peers establish connection at the time 0 contacting each other and perform a standard authentication by using their Public / private key credentials, from that point they know each other and are familiar. At the end of the first authentication process, both parties can create relation based private credential to be used for creation of further additional authentication criteria. So the authentication process between two familiar peers consists of the two following steps.

1. Producing of relation based Authentication credentials.
2. Proofing the possession of the authentication credentials using NIZKP protocol.

The first component is used to perform two tasks in the process of authentication between familiar peers, which are:

1. Creation of one time secret key using Diffie-Hellman protocol.
2. Creation of one time authentication credential.

As it is mentioned in 2.2 in DC-nets, a secret key will be created between any two peers using Diffie-Hellman protocol. This key is considered to be a one time pad to be used, during the next authentication process between them. So at the end of authentication process of the first session also s_0, the secret key sk_1 has been created and both partners hold them. During the next session, when they contact each other, the secret key sk_1 will be used as a one time pad in the authentication process. At the end of the authentication process in every sessions s_n the secret key for use in the next session also sk_{n+1} will be created.

So for every two peers $P(a)$ and peer $P(b)$, there will be a secret key sk_n for use in session s_n such that following relation exists:

let $\mathtt{S}$ be the set of all authentication sessions and
$\mathtt{SK}$ be the set of all secret keys

$$\forall\ \mathtt{s} \in \mathtt{S}\ \exists\ \mathtt{sk}_n \vdash \mathtt{sk}_n \neq \mathtt{sk}_{n+1}$$

The second step is creation of an authentication credential. The credential for session s_n will be created at the result of a boolean XOR Operation between the sk_n and a known information between the two peers. This Information does not need to be secret. one of them can even send a larg number as the information to the other peer. The secret key of that session also sk_n will be used to perform the boolean XOR operation with that large number. Using this method there will be following relation:

let $\mathtt{S}$ be the set of all authentication sessions and
$\mathtt{SK}$ be the set of all secret keys and
$\mathtt{SAC}$ be the set of all shared authentication credentials and
$\mathtt{I}$ be the set of all known shared Information and

$$\forall\ \mathtt{i} \in \mathtt{I}\ \exists\ \mathtt{i}_n \vdash \mathtt{i}_n \neq \mathtt{i}_{n+1}$$

which denotes that the used information to create the authentication credential for each session changes for every new session and:

$$\forall\ \mathtt{s} \in \mathtt{S}\ \exists\ \mathtt{sac}_n = \mathtt{sk}_n \bigoplus \mathtt{i}_n\ \&\ \mathtt{sac}_n \neq \mathtt{sac}_{n+1}$$

From the above relation we deduce, that authentication credential used for every session is changed and is not equal with the previous one. Now when the proofer has been able to create the expected result, she needs to inform the verifier about the correct result, she has been able to produce. The verifier, on the other side, can produce the same result. Both parties are in the possession of the same one time produced information, which is called in this schema the authentication credential.

Using the NIZKP, the prover will send only the proof that she is in possession of the right information. As it is explained in 2.2 the prover will create a graph representing the authentication credential and produce then the n isomorph graphs corresponding to the original graph. The verifier peer creates also the same representing graph of the authentication credential using the same algorithm.

At this point the prover can prepare the n segmented proof message using NIZKP explained in 2.2 and send it to the verifier, which can verify the received information. The hole schema consists of the following protocol functionality:

1. **create_pk**: which creates the private key hold by each peer.
2. **create_shk**: which creates the shared key initiated by the verifier on both side.
3. **create_sk**: which creates based on the shared key and private key, the secret key.
4. **create_graph**: which creates a graph representing the authentication credential.

The algorithm used to create the secret key for every new session looks like the code here:

```
function create_sk()
{
    pk = create_pk();               \\creating the private key
    shk = create_shk();             \\creating the shared key
    send_msg(shk);                  \\sending the shared key
                                    \\to the partner
    sk_1 = boolean_XOR(pk, shk);    \\performing the boolean OR
                                    \\operation between the private
                                    \\key and the shared key
    send_msg(sk_1);                 \\sending the result to the partner
    sk_2 = receive_msg();           \\receiving the intermediated key
     sk = boolean_XOR(pk,sk_2);     \\creating the secret key
       return(sk);
}
```

The authentication process will then use the following algorithm:

```
function authenticate(number n)
{
    large_number= random();                \\creating a random large number
    send_msg(large_number);                \\creating the private key
    sac = boolean_XOR(sk(n),large_number); \\creating the desired result
                                           \\based on secret key of
                                           \\Session n and the larg_number
     graph = create_graph(sac);            \\creating the representatin
                                           \\graph of sac
    result_p = receive_msg();            \\receiving the proof message
                                         \\from the  partner prepared
                                         \\using NIZKP.
    if (check(graph, result_p));         \\check if the delivered result
                                         \\by the partner
        create_sk(n+1);                  \\corresponts to the result
                                         \\created localy
        return 1;
    else
        return 0;
}
```

although the exchanged information in our example is a randomly created large number, in different environment any other shared information between both parties can be used, such as the number of communication packets exchanged in the last session.

3.1 Proof of Concept

The proposed mechanism provides a way to create private and secret authentication credentials between two familiar peers. In each authentication session, they agree on a privately kept secret key which is not exchanged between them using the Diffie-Hellman protocol. The created secret key changes for every new session and is seen to be used as a one time pad for the authentication purposes. As far as the secret key created on both side using the Diffie-Hellman is not exchanged between the two peers, no third party can have knowledge about it. Additionally this secret key is used as one-time pad and will be recreated for each new session.

The private secret key will be used to produce the required authentication credential every time on demand. So these credentials can not be revealed even by intrusion the systems since they are not stored somewhere as persistence data. Moreover even the reproduced one-time authentication credentials will not be exchanged over the net with the communication partner, but only the proof of possession of the desired right information i. e. the authentication credential will be sent to the verifier.

As far as information required to perform the secure authentication are not exchanged and are not revealed to a third party and the fact, that the actual authentication credential is seen as one time pad and are not stored in persistent forms, the authentication process using this credential will perform secure without the possibility for a third party to interfere in the authentication process. No third party will have the chance to produce the desired and expected message for verifier.

4 Conclusion

In this paper a new mechanism is proposed to perform a secure authentication between familiar peers. Peers which know each other can create dynamically required authentication credentials, which are not exchanged between them and so are not revealed to anyone else. The proposed mechanism uses a protocol inspired by the dining-cryptographers protocol and a non interactive zero-knowledge proof protocol to satisfy the desired functionality. Using the specified proposed mechanism here, two familiar peers can authenticate themselves mutually and be sure that the process of authentication is not interfered and forged by a third party.

5 Acknowledge

My special thanks belongs to Prof. Dr. Herwig Unger and Dr. Hauke Coltzau for their support and helpful ideas which helped to evolve and maturate the proposed mechanism.

References

[1] Ronald L. Rivest, Adi Shamir, and Yael Tauman. How to share a secret. *Communications of the ACM*, vol. 22, pp. 612–613, 1979.

[2] Jun Kurihara, Shinsaku Kiyomoto, Kazuhide Fukushima, and Toshiaki Tanaka. A new (k,n) threshold secret sharing scheme and its extension lncs

5222. *Information Security Lecture Notes in Computer Science*, pp. 455–470, 2008.

[3] Stefan Schwoon Hao, Hao Wang, Somesh Jha, and Thomas W. Reps. Distributed certificate-chain discovery in spki/sdsi. *Computer Security Foundations Workshop, Proceedings. 15th IEEE*, pp. 129–144, 2002.

[4] Kalman Graffi, Patrick Mukherjee, Burkhard Menges, Daniel Hartung, Aleksandra Kovacevic and Ralf Steinmetz. Practical security in p2p-based social networks. *IEEE Society: The 34th Annual IEEE Conference on Local Computer Networks (LCN) October 2009*, pp. 269–272, 2009.

[5] Patrick P. Tsang and Sean W. Smith. Ppaa: Peer-to-peer anonymous authentication. Applied Cryptography and Network Security *Lecture Notes in Computer Science*, vol. 5037, pp. 55–74, 2008.

[6] Gabriel Vanrenen and Sean Smith. Distributing security- mediated pki. *In 1st European PKI Workshop Research and Applications, Springer-Verlag*, pp. 620–634, 2004.

[7] Srdjan Capkun, Levente Buttyán, Jean-Pierre Hubaux. Self-organized public-key management for mobile ad hoc networks.*IEEE Transactions on Mobile Computing*, vol. 2, pp. 52–64, 2003.

[8] Giulia Fanti, Peter Kairouz, Sewoong Oh, and Pramod Viswanath. Spy vs. spy: Rumor source obfuscation. *In Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pp. 271–284. ACM, 2015.

[9] David Chaum. The dining cryptographers problem: Unconditional sender and recipient untraceability. *Journal of Cryptology*, vol. (1), pp. 65–75, 1988.

[10] Whitfield Diffie and Martin Hellman. New directions in cryptography. IEEE transactions on Information Theory, Volume 22(6), pp. 644–654, 1976.

[11] Giulia Fanti and Pramod Viswanath. Algorithmic advances in anonymous c ommunication over networks. In: *2016 Annual Conference on Information Science and Systems (CISS) IEEE*, pp. 133–138, 2016.

[12] Joan Feigenbaum and Bryan Ford. Seeking anonymity in an internet panopticon. *Communications of the ACM*, Volume 58(10), pp. 58–69, 2015.

[13] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof systems.*SIAM Journal on computing*, vol. 18(1), pp. 186–208, 1989.

[14] Jitendra Kurmi and Ankur Sodhi. A survey of zero-knowledge proof for authentication. *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5(1), pp. 494–501, 2015.

[15] Xiyu Pang, Cheng Wang, and Yuhong Zhang. A new p2p identity authentication method based on zero-knowledge under hybrid p2p net-

work. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 11(10), pp. 6187–6192, 2013.

[16] Francisco Martín-Fernández, Pino Caballero-Gil, and Cándido Caballero-Gil. Authentication based on non-interactive zero-knowledge proofs for the internet of things. Sensors, *Multidisciplinary Digital Publishing Institute* vol. 16(1), page 75, 2016.

Feasibility Study of Applying Chaotic Carrier Frequency Modulation in Switching-mode Power Supply

Junying Niu[1,2], Zhong Li[2], Yuhong Song[1,2] and Wolfgang A. Halang[2]

[1]Dept. of Electronic and Information Engineering, Shunde Polytechnic, China

[2]Chair of Computer Engineering, FernUniversität in Hagen, Germany

Abstract: Owing to the pseudo-randomness and the continuous spectrum features of chaos, chaotic carrier frequency modulation (CCFM) technique can be well employed to fight electromagnetic interference (EMI) of switching mode power supplies (SMPS) by spreading the spectra of input and output signals over the entire frequency band. EMI test, output voltage ripple measurement and efficiency measurement have been conducted on the chaos-control SMPSs, leaving the influence of chaos control on the whole system performance to be further investigated. In order to investigate the feasibility of applying CCFM in SMPS, the tests on the electrical characteristic, working condition and EMC performance of SMPS are carried out, and the experiments are conducted to verify that the application of the chaotic modulation in SMPS can reduce EMI without weakening the system performance.

1 Introduction

Due to its high efficiency, a switching mode power supply (SMPS) has been increasingly widely applied in electric industry. However, rapid switching actions of semiconductor devices, which result in high change rates of voltage and current, lead to severe electromagnetic interference (EMI) problems.

Owing to the continuous spectrum feature of chaos, chaotic carrier frequency modulation (CCFM) technique is effective to reduce EMI by spreading the harmonics of the input and output signals over the whole frequency band. Hence, plenty of work has been done to propose chaotic modulation schemes for EMI reduction of SMPSs [1, 2, 4–11, 14], and a lot of experimental research has been done to study the effect of chaos control on the performance of SMPS.

The SMPS has to meet some basic requirements, which include 1) providing stable voltage or current for the output load, 2) satisfying EMC standards and 3) still operating stably after a long time of work. Until now, a lot of experimental research has been done on 1) [7, 8, 13, 17] and 2) [3, 4, 12, 16]. Nevertheless, few work was done on 3), so that whether chaos control affects the operating condition of system or not remains a question. The working condition of SMPS is determined by the key elements, namely the transistor and the high-frequency transformer, which usually heat up as operating. Once their temperature exceeds the safe range, they may be out of order, resulting in a breakdown system. Hence, it is necessary to make the thermal test of the key elements, which has never been done before.

Therefore, to investigate the feasibility of applying CCFM in SMPS, an experimental research, taking a LED driver as an example, will be carried out. Because the flyback topology is most commonly used in medium-low power SMPS, a flyback converter-based LED driver, which is controlled by UC3842, a typical pulse width modulation (PWM) IC, is chosen as one example. The tests on the electrical characteristic, the key element's working condition and EMC performance of the LED driver will be carried out. According to the experimental research, chaos control can be use to reduce EMI of SMPSs, and does not impair the overall systme performance.

2 Implementation of Chaotic Carrier Frequency Modulation

2.1 Working Principle

Generally, the oscillator of frequency programmable PWM IC makes use of an external timing capacitor and an external timing resistor (or only a timing capacitor) to set the switching frequency. As shown in Fig. 1, a typical oscillator, used in UC3842 [15], is composed of the internal circuit (inside the dotted line) and the peripheral components, C_T and R_T. Initially, the voltage of C_T (v_C) is zero and $v_\mathrm{C} < \mathrm{V}_\mathrm{low} < \mathrm{V}_\mathrm{upp}$. Then, the switch S_1 is turned off, and C_T will be charged by a reference voltage V_REF through R_T. Once v_C arrives at or exceeds V_upp, S_1 is turned on, and C_T begins to be discharged through a current source $\mathrm{I}_\mathrm{discharge}$ until $v_\mathrm{C} \leq \mathrm{V}_\mathrm{low}$. Thereafter, C_T is charged and discharged circularly between V_low and V_upp, and the switching period is the summation of the charging and discharging time.

The frequency of the oscillator can be modulated by making the charging time of C_T changing chaotically. Consequently, as shown in Fig. 2, a CCFM module is

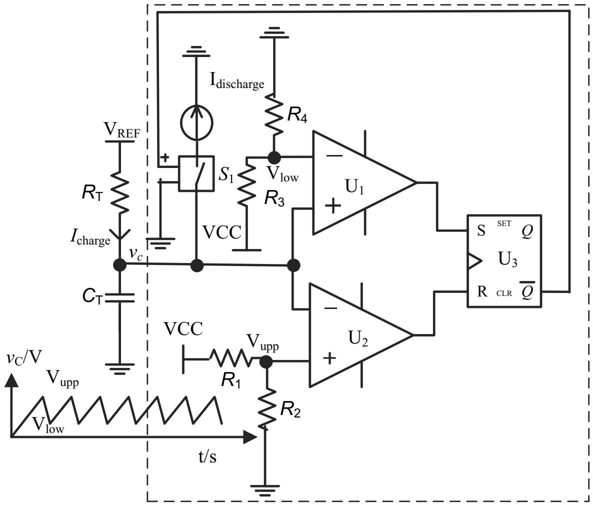


Fig. 1: Oscillator of PWM IC

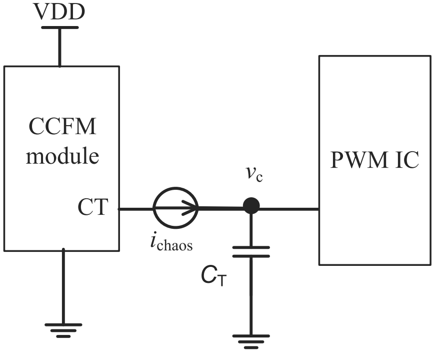


Fig. 2: Interface of the CCFM module

designed to generate a dithering chaotic charging current to C_T. The schematic of the CCFM module is given by Fig. 3, which is composed of Chua's circuit, an amplitude-limiting circuit and a current-limiting resistor. The amplitude-limiting circuit is composed of R_1-R_4 and the amplifier U_1, and R_I is the current-limiting resistor. First of all, a chaotic voltage, v_2, is generated by Chua's circuit, and $v_2 \in [0, \mathrm{VDD}]$. Second, because the chaotic voltage for charging the timing capacitor C_T, namely v_chaos, should be larger than the high threshold voltage of C_T, V_upp, the amplitude-limiting circuit linearly transform v_2 to v_chaos. By adjusting R_1-R_4, v_{chaos} is set to above V_{upp}, and thus provides a chaotic charging current to C_{T} through R_{I}.

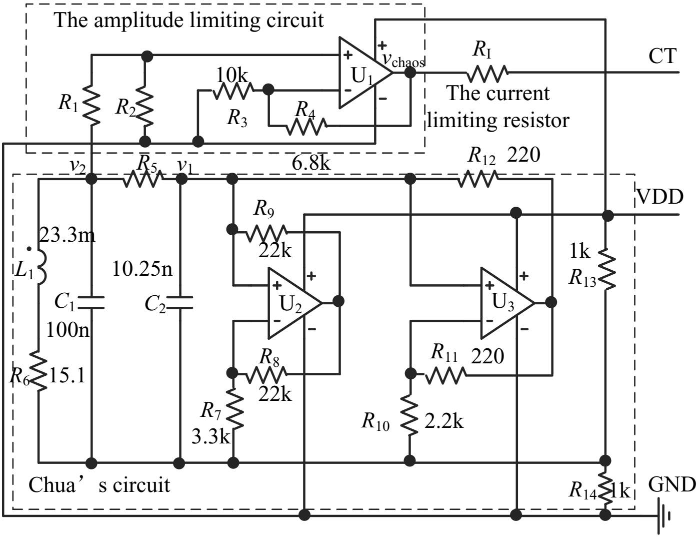


Fig. 3: The schematic of the CCFM module

Hence, during the kth switching cycle, the charging time of C_{T} is calculated as

$$t_{\text{c}_k}(v_{\text{chaos}}, R_{\text{I}}) = \frac{R_{\text{T}}R_{\text{I}}}{R_{\text{T}}+R_{\text{I}}} C_{\text{T}} \ln \frac{\text{V}_{\text{low}} - \dfrac{R_{\text{I}}\text{V}_{\text{REF}} + R_{\text{T}}v_{\text{chaos}}}{R_{\text{T}}+R_{\text{I}}}}{\text{V}_{\text{upp}} - \dfrac{R_{\text{I}}\text{V}_{\text{REF}} + R_{\text{T}}v_{\text{chaos}}}{R_{\text{T}}+R_{\text{I}}}}. \tag{1}$$

Because the discharging time is approximately 0 normally, the switching frequency can be expressed as

$$f(v_{\text{chaos}}, R_{\text{I}}) \quad = \frac{1}{t_{\text{c}_k}(v_{\text{chaos}}, R_{\text{I}})} = \text{F} + \Delta f (\Delta f \in [0, \Delta F]), \tag{2}$$

where F is the minimum of the frequency as v_{chaos} is minimized to v_{min}. $\triangle F$ is the varied range of the frequency, and can be calculated by $\triangle F = f(v_{\text{max}}, R_{\text{I}}) -$

$f(v_{\mathrm{min}}, R_{\mathrm{I}})$, where v_{max} is the maximum of v_{chaos}. It is apparent that the frequency range can be set by adjusting R_{I}, as the parameters of the amplitude limiting circuit are fixed.

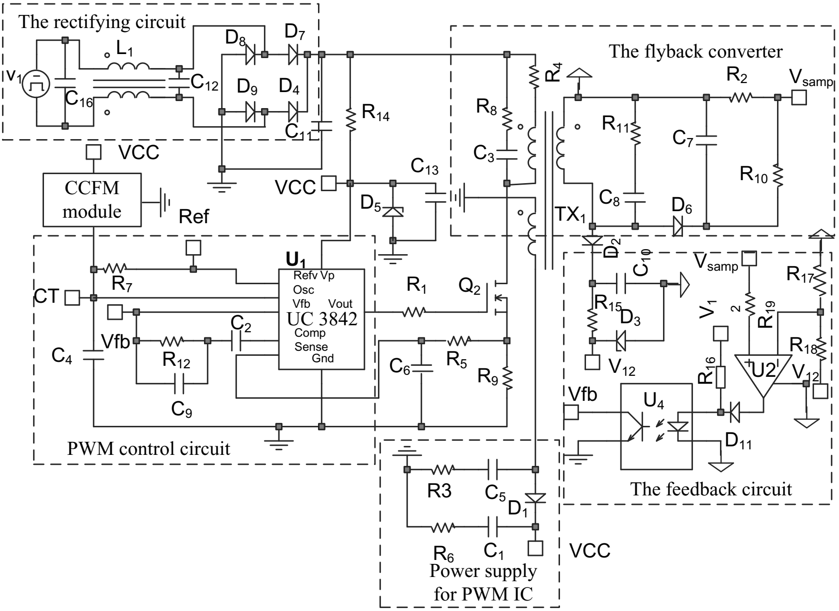


Fig. 4: The schematic of the flyback converter-based LED driver

2.2 System Design

The LED driver, supplied by AC power of 220V and 50Hz, is a constant current source of which the output current is 1A and the maximum rated power is 20W. As shown in Fig. 4, the rectifying circuit converts the AC input power to the DC voltage of about 300V which is periodically conducted to the primary side of TX_1 . When the transistor Q_2 is turned on by PWM control circuit, the primary of the transformer is directly connected to the input voltage source. The primary current and magnetic flux in the transformer increases, storing energy in the transformer. The voltage induced in the secondary winding is negative, so the diode is off, and thus the output capacitor supplies energy to the load.

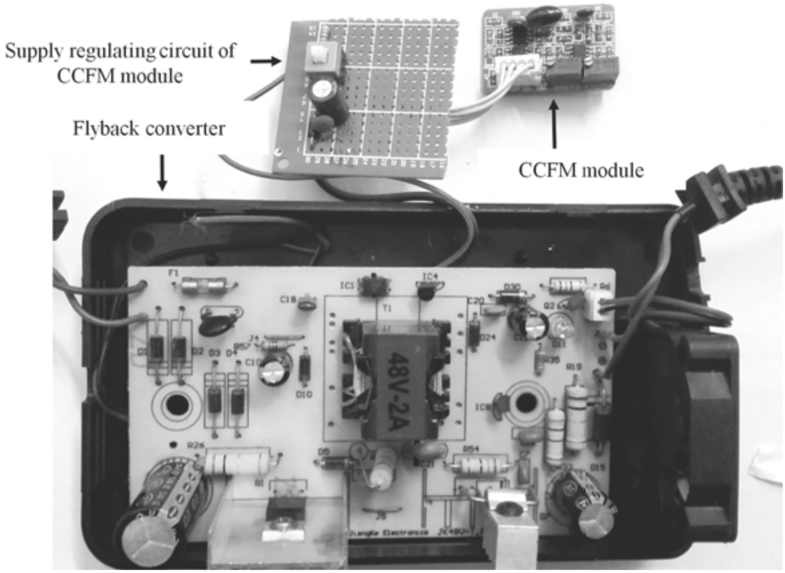


Fig. 5: The entity of the flyback converter-based LED driver with CCFM

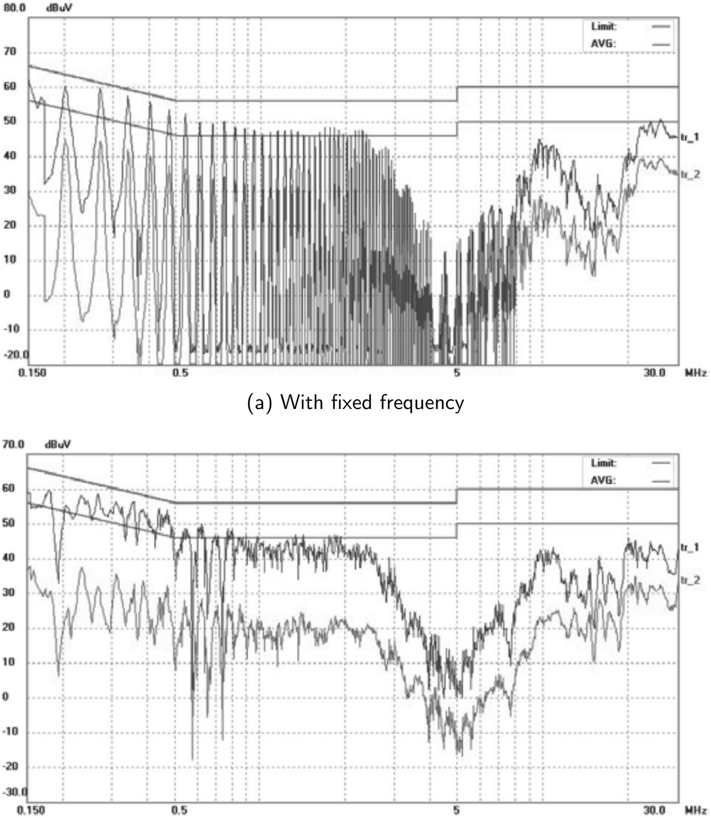
When the transistor is turned off, the primary current and magnetic flux drops. The secondary voltage is positive, turning on the diode, allowing current to flow from the transformer. The energy through the transformer core recharges the capacitor and supplies the load. The output current is sampled, processed, sent to the PWM IC (UC3842) by the feedback circuit, and regulated to keep constant.

The CCFM module is attached to the LED driver, and the entity circuit is shown as Fig. 5. Once a simple power processing circuit is attached to the module for stable power supplying, CCFM is implemented. What follows is to investigate the influence of chaos control to electrical characteristics, EMC and the key elements working temperature of the LED driver.

3 Experiments

3.1 Test for Electromagnetic Interferences

The conducted interference of the LED driver is tested with Rohde&Schwarz instrument according to the standard of CISPR PUB.22 CLASS B. As shown in Fig. 6, the upper line is the quasi-peak (QP) limiting curve, and the lower is the average value (AV) limiting curve. By comparing Fig. 6a and Fig. 6b, it is obvious that both QP and AV peaks existing under the traditional PWM control are reduced by chaos control.



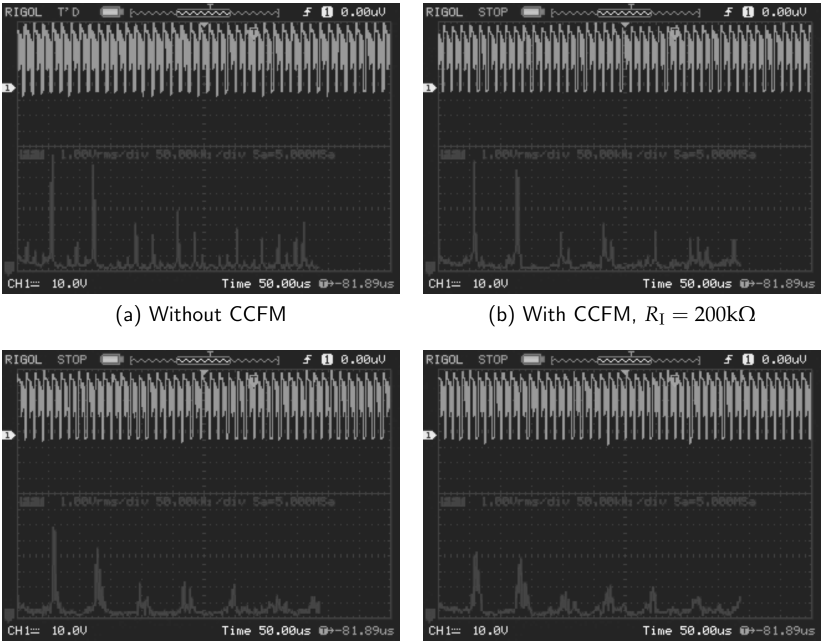
(a) With fixed frequency

(b) With chaotic frequency

Fig. 6: The EMI measurement results (upper: the peak value, lower: the average)

3.2 Electrical Characteristics Measurements

Firstly, the basic requirement for power supply is to provide the stable output voltage or output current, so that the output voltage ripple is measured. For the utilization of the linear load, the voltage ripple can replace the current one, and is tested as the power operates with the max rate output voltage of 20V. Thus, as the LED driver works under the traditional PWM control, the ripple is 68mV. The ripple under chaos control, increasing slightly, is 97mV, 78mV and 69mV as R_I is 100kΩ, 150kΩ and 200kΩ respectively. The ripple increment becomes larger as R_I decreases, but is still acceptable.



(a) Without CCFM

(b) With CCFM, $R_\mathrm{I} = 200\mathrm{k}\Omega$

(c) With CCFM, $R_\mathrm{I} = 150\mathrm{k}\Omega$

(d) With CCFM, $R_\mathrm{I} = 100\mathrm{k}\Omega$

Fig. 7: Waveforms and spectra of the switching voltage

Secondly, the waveform and the spectrum of the switching voltage, the voltage on the transistor's drain electrode, are tested. On the one hand, the transistor working under the turned-off state is needed to endure the huge switching voltage, which is not allowed to exceed the rating to make sure the transistor not damaged. Therefore, its waveform is to detect whether there is an overvoltage on the transistor or not. On the other hand, because the switching actions of the

Table 1: The electrical characteristics measurements

Index	Parameter	Without CCFM			With CCFM		
The voltage regulation	Input voltage (V)	180	220	260	180	220	260
	Input current (A)	0.208	0.189	0.179	0.215	0.193	0.181
	Output voltage (V)	20	20	20	20	20	20
	Output current (A)	1	1	1	1	1	1
Power factor	Power factor	0.65	0.61	0.58	0.64	0.60	0.57
Efficiency (%)	Input power (W)	24.8	25.4	26.5	25.1	25.6	26.8
	Output power	20	20	20	20	20	20
	Efficiency (%)	80.6	78.7	75.4	79.7	78.1	74.6
Load regulation	Resistor load (Ω)	20	15	10	20	15	10
	Output current (A)	1	1	1	1	1	1

transistor is the underlying cause of EMI, the spectrum of the switching voltage is measured to observe the spread-spectrum effectiveness of the CCFM module. Hence, as shown in Fig. 7, there is no obvious distinction between the switching voltage waveform with fixed frequency and that with chaotic frequency. Its harmonics peaks existing under the traditional PWM control are reduced by the chaotic modulation, resulting in EMI reduction, and the EMI suppression effect is improved with a smaller R_{I}.

Finally, the other indexes are given by Tab. 1, while tests under chaos control are carried out as $R_{\mathrm{I}} = 100\mathrm{k}\Omega$. The tests of the output stability indexes, including input regulation rate, the load regulation rate and the output star-up waveform, show that the power supply with CCFM can supply the load as stably as that without CCFM. The test of power factor and the input and output power ratio is to survey the working efficiency of the SMPS. Because the LED driver supplies not only the original load but also the extra CCFM module, the supply's efficiency is impaired very slightly.

3.3 Test of Key Elements' Temperature

The transistor and high-frequency transformer, the necessary elements for SMPS, always work with a large current, resulting in the thermal problem. On the one hand, the high-frequency transformer should work below 80°C, otherwise its saturation magnetic flux density falls to 70% of that at the normal temperature. It falls even more as the temperature increases, and thus the current and the power consumption of the transformer will rise sharply. As a result, the overheat problem is exacerbated, resulting in a vicious circle, until the around

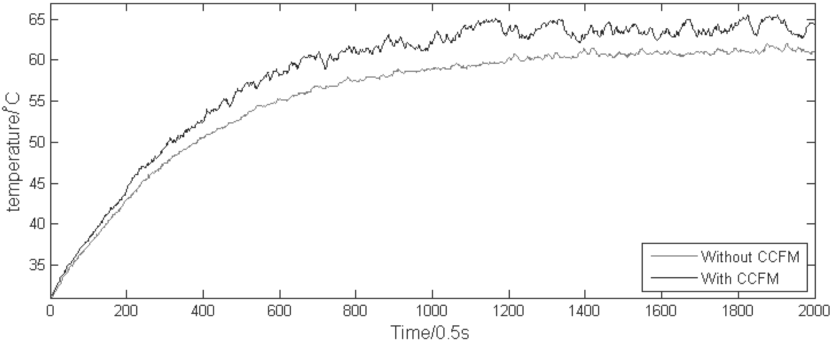


Fig. 8: The thermal curve of the transformer

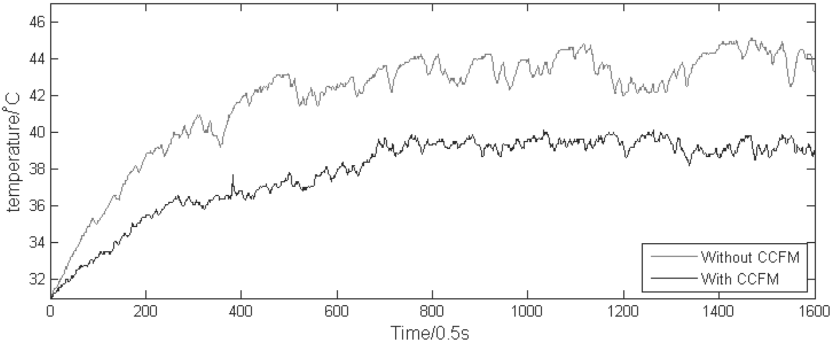
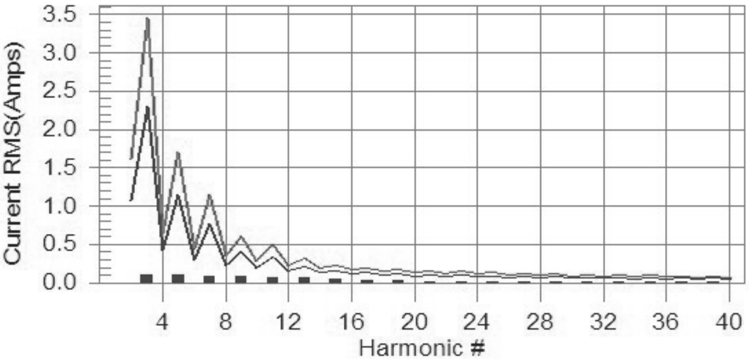


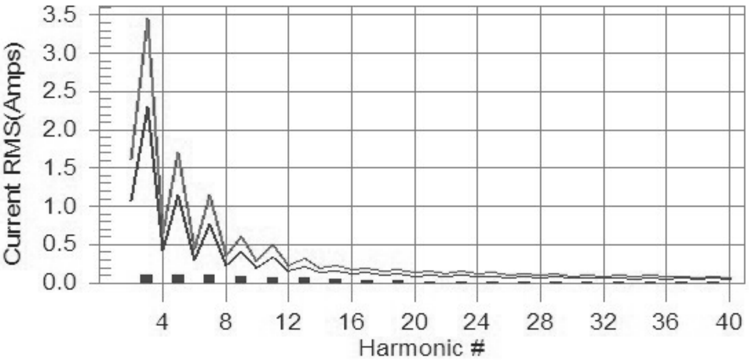
Fig. 9: The thermal curve of the transistor

elements are damaged by a huge current. On the other hand, the overheat of the transistor may impact the reliability of the switching action, leading to a breakdown system.

Hence, it is necessary to perform thermal tests on the transistor and high-frequency transformer of SMPS. As shown in Fig. 8, the transformer's temperature under chaotic modulation arises at a faster speed, and is about 4°C higher than that with fixed frequency, however, still within the safe range. As shown in Fig. 9, the comparison of the transistor's thermal curves indicates that the transistor's temperature under CCFM is 4°C lower than that under the traditional PWM control.



(a) Without CCFM



(b) With CCFM

Fig. 10: The measurement results of input current harmonic

3.4 Harmonics Current Measurement

The harmonic current of the electronic equipment may disturb the power grid, disorder the other devices on the grid, and even cause the grid overload and block the power transmission. Therefore, the standards, such as GB 17625.1 and IEC61000-3-2, are evolved to restrain the harmonic current. Hence, the harmonic current measurement is carried out, and the result under chaos control (Fig. 10b) is just as that under the traditional PWM control (Fig. 10a).

4 Conclusion

To investigate the effect of chaos control on the SMPS's performance, CCFM technique is applied in a LED driver, and the tests of electrical characteristic, working condition and EMC performance are carried out.

First of all, the EMC test verifies that the chaotic modulation is effective to reduce EMI of the power supply. Secondly, the electrical characteristics test shows that only the output ripple under chaos control increases slightly, without any other different indexes from those under the conventional PWM control. Thirdly, the thermal test of the key elements indicates that the temperature of the high-frequency transformer increases several degrees under chaos control, but it doesn't disturb the operation of the system.

The CCFM technique is effective to reduce EMI, and does not weaken the overall performance of the power supplies. To sum up, it is feasible to apply CCFM in SMPS.

References

[1] S. Banerjee, D. Kastha, and S. SenGupta. Minimising EMI Problems With Chaos. In *Electromagnetic Interference and Compatibility, 2001/02. Proceedings of the International Conference on*, pages 162–167. IEEE, 2002.

[2] S. Callegari, R. Rovatti, and G. Setti. Chaotic Modulations Can Outperform Random Ones in Electromagnetic Interference Reduction Tasks. *Electronics Letters*, 38(12):543–544, 2002.

[3] J. Jankovskis, D. Stepins, S. Tjukovs, and D. Pikulins. Examination of Different Spread Spectrum Techniques for EMI Suppression in DC/DC Converters. *Electronics and Electrical Engineering.–Kaunas: Technologija*, (6):86, 2008.

[4] H. Li, Z. Li, B. Zhang, F. Wang, N. Tan, and W. Halang. Design of Analogue Chaotic PWM for EMI Suppression. *Electromagnetic Compatibility, IEEE Transactions on*, 52(4):1001–1007, 2010.

[5] H. Li, Z. Li, B. Zhang, Q. Zheng, and W. Halang. The Stability of a Chaotic PWM Boost Converter. *International Journal of Circuit Theory and Applications*, 39(5):451–460, 2011.

[6] Z. Li, S. Qiu, and Y. Chen. Experimental Study on the Effectiveness of EMI Suppression with Chaotic Modulation for Switching Converters. *Transactions of China Electrotechnical Society*, 21(8):97–102, 2006.

[7] Z. Li, S. Qiu, and Y. Chen. Experimental Study on the Suppressing EMI Level of DC-DC Converter with Chaotic Map. *Proceedings of the CSEE*, 5:012, 2006.

[8] Z. Li, S. Qiu, and L. Zhang. Research on the Enhanced Electromagnetic Compatibility of Switching Converter with Chaotic Frequency Modulation. *Acta Electronica Sinica*, 33(11):1983, 2005.

[9] J. Rodriguez Marrero, J. Font, and G. C. Verghese. Analysis of the Chaotic Regime for DC-DC Converters Under Current-Mode Control. In *Power Electronics Specialists Conference, 1996. PESC'96 Record., 27th Annual IEEE*, volume 2, pages 1477–1483. IEEE, 1996.

[10] S. Santi, R. Rovatti, and G. Setti. Advanced Chaos-Based Frequency Modulations for Clock Signals EMC Tuning. In *Circuits and Systems, 2003. ISCAS'03. Proceedings of the 2003 International Symposium on*, volume 3, pages 111–116. IEEE, 2003.

[11] Y. Song, Z. Li, J. Niu, G. Zhang, W. Halang, and H. Hirsch. Reducing EMI in a PC Power Supply with Chaos Control. In *Foundations and Applications of Intelligent Systems*, pages 231–241. Springer, 2014.

[12] K. Tse, W. Ng, H. S. Chung, and S. Hui. Evaluation of a Chaotic Switching Scheme for Power Converters. In *Power Electronics Specialists Conference, 2000. PESC 00. 2000 IEEE 31st Annual*, volume 1, pages 412–417. IEEE, 2000.

[13] R. Yang. *Chaotification and EMI Suppression of Power Converter*. PhD thesis, South China University of Technology, 2007.

[14] R. Yang, B. Zhang, F. Li, and J. Jiang. Experiment Research of Chaotic PWM Suppressing EMI in Converter. In *Power Electronics and Motion Control Conference, 2006. IPEMC 2006. CES/IEEE 5th International*, volume 1, pages 1–5. IEEE, 2006.

[15] J. Zarębski and K. Górecki. Spice-aided modelling of the uc3842 current mode pwm controller with selfheating taken into account. *Microelectronics Reliability*, 47(7):1145–1152, 2007.

[16] J. Zhang, S. Qiu, and L. Chen. An Investigation of Chaotic Frequency Modulation for Improved EMC of Switching Mode Power Supply. *Electrical Applications*, 25(6):71–74, 2006.

[17] J. Zhang, L. Zhang, and S. Qiu. An Experimental Investigation of EMI Suppression of Off-Line Switching Converter by Chaotic Modulation. *AerOspace Control*, 24(4):87–90, 2006.

Identification of Filter Types by Bifurcation Analysis: Mathematical Modelling and Numerical Simulation

Nkiediel Alain[1], Hughes Bisuta Bieto[2], Jean Chamberlain Chedjou[1]
and Kyandoghere Kyamakya[1]

[1] Institutes of Smart System Technologies, Transportation Informatics
Alpen Adria University, Klagenfurt, Austria

[2] University of Kinshasa, Polytechnic Faculty, DR Congo

Abstract: This paper presents a specific structure of a dual-mode passive filter. It is demonstrated that the filter can be used both as Band-pass and High-pass. The dual-mode of the proposed filter is demonstrated through monitoring of the value of a circuit component/element used as a control parameter (or bifurcation parameter). The bifurcation analysis carried out leads to the derivation/detection of ranges of the control parameter under which the proposed filter can be used as Band-pass and High-pass simultaneously. This corresponds to the functioning of the filter in dual-mode is observed. It is further demonstrated that the monitoring of the control parameter significantly affects both the Quality factor (Q-factor), and the bandwidth of the filter. The bifurcation analysis carried out leads to the derivation of a specific value of the control parameter, above which the dual-mode filter is reduced to a single-mode filter of type High-pass. The pros and cons of using the proposed dual-mode filter and the potential applications in Engineering are discussed. As proof of concepts in order to validate the results obtained, a benchmarking is performed. The analytical results (theory) are compared with numerical results (PSPICE). The outcome of the benchmarking has led to a perfect agreement between the methods.

1 Introduction

1.1 Sample Potential Applications of Filters in Engineering

During the past decades, the analysis of filters (active and/or passive) has been devoted a tremendous attention due to their multiple potential applications in the field of engineering. Some interesting applications are: De-noising in signal processing [1], LF-transmission [2, 3], HF-transmission [2, 3] in telecommunications, (a) frequency-band limiting and equalization, in audio systems [4, 6], (b) frequency-tuning and rejection, in communication systems [4, 6], (c) analog-to-digital conversion (ADC), in signal processing [4, 6], etc. These applications are traditionally insured by specific types of filters: High-pass, Low-pass, Bandpass, and Band-stop, just to name a few.

1.2 Active Filters vs. Passive Filters: Pros and Cons in Engineering

The use of active or passive filters in Engineering can be explained by the fact that each type of filter corresponds to specific applications. Further, each type of filter can lead to specific advantages and/or limitations of instance, one key advantage (amongst many others) of using active filters is the possibility of adding and sustaining energy within the system (or circuit); in contrast, passive filters are energy demanding (i.e. high energy consumption) since/as they are purely dissipative. The main drawback of active filters can be explained by the fact that their frequency range depends on the bandwidth of the analog devices (i.e. Transistors, Operational amplifiers, CMOS, etc.) used. Thus, the performance of active filters is significantly affected at high frequencies. Therefore, active filters appear inappropriate at high frequencies. In contrast the passive filters are appropriate prototypes which are commonly used at high frequencies. Further, the range for the dynamic (or temporal) variation of the voltages within the system (or circuit) depends on the static bias of analog devices. Hence, the dynamic voltages within the active filters must always be below the static bias in order to avoid saturation phenomena within the system (or circuit). Passive filters are not prone to this problem (or limitation) faced by the active filters. Regarding the performances (e.g. power transmission, stability of the filter, etc.) and/or properties/characteristics (e.g. Low-pass, High-pass, Band-pass, Band-stop) of active filters, they are not significantly affected by the load, while in contrast the performances and/or properties of passive filters do significantly depend on the load.

This paper considers the changes observed in the properties/characteristics of passive filters and demonstrates the benefits/advantages resulting to these changes in terms of concrete potential interesting applications in Engineering.

1.3 Key Contribution

The philosophy of the work carried out exploits the dynamic property/characteristic of passive filters. This property is defined as the possibility for a given filter (with a fix structure defined in terms of known circuits components) to behave or perform in a dual-mode (e.g. Dual 1: ″Bandpass and High pas″, Dual 2: ″Low-pass and Band-pass″, etc.). This paper proposes a fixed structure of a passive filter and we use both analytical and numerical methods to provide evidence of the dynamic properties of the proposed filter. The analytical study is concerned with the modelling of the filter and the derivation of the corresponding transfer function. This function is further used in the frequency domain to depict the dynamics/behavior of the filter. We demonstrate that, the property/characteristic of the filter depends on a specific component (i.e. a resistive component of the circuit). This component is monitored as a control parameter in order to depict the possible properties that, the proposed filter is likely to undergo (e.g. High-pass, Low-pass, Band-pass, etc.). The analysis leading to the discovery of the filters properties is called bifurcation analysis. Finally a numerical study is considered. This study is concerned with the PSPICE implementation of the proposed filter. The results provided by PSPICE are compared with the analytical results and a perfect agreement is observed between them. The advantages of using filters in dual-mode are clearly presented with reference to several interesting and concrete applications in Engineering.

1.4 Organization

The remaining of the paper is structured as follows. Section 2 considers the analytical study. A full description of the dual-mode filter is proposed and the mathematical modelling of the filter is carried out. The transfer function of the dual-mode filter is derived in the frequency domain in terms of the filter′s components (i.e. resistors, capacitors, inductors, etc.). Section 3 deals with the numerical study. The implementation of the dual-mode filter in PSPICE is considered and the numerical results are obtained. Section 4 presents the results obtained analytically and numerically. To validate the concepts, a benchmarking is carried out. Some concluding remarks are formulated in section 5 and some ongoing research fronts/avenues are presented as outlooks.

2 Mathematical Modelling of the Dual-mode Filter

Figure 1 is the electrical structure of the proposed dual-mode passive filter. The filter is made up of a resistor R, an inductor L, and two capacitors C_1 and C_2.

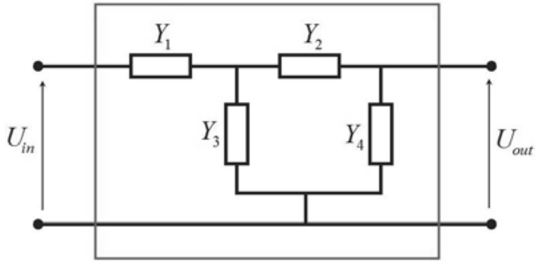


Fig. 1: The Electrical structure of the dual-mode passive filter

In Figure 1, $Y_1 = jC_1\omega$ and $Y_2 = jC_2\omega$ correspond to the admittances of C_1 and C_2, $Y_3 = 1/jL\omega$ is the admittance of L, and $Y_4 = 1/R$ corresponds to the admittance of R.

Using the classical theorems of electrical Engineering (i.e. Kirchhoff- and Thevenin-theorems), the transfer function in Eq. (1) corresponding to the dual-mode filter is derived.

$$T(S) = \frac{LRC_1C_2S^3}{LRC_1C_2S^3 + L(C_1+C_2)S^2 + C_2RS + 1} \tag{1}$$

Where $S = j\omega$ corresponds to the complex expression of the frequency f ($\omega = 2\pi f$).

Equation (1) can be used to evaluate the module of the transfer function in (2) denoted by $|T(S)| = T(S)T(S)^*$, where $T(S)^*$ corresponds to the complex conjugate of $T(S)$.

$$|T(S)| = \frac{a\,\omega^3}{b\,\omega^6 + c\,\omega^4 + d\,\omega^2 + 1} \tag{2}$$

The parameters a, b, c and d are expressed in terms of the circuit components as follows: $a = LRC_1C_2$, $b = a^2$, $c = [L^2(C_1 + C_2)^2 - 2\,LR^2C_1C_2^2]$ and $d = [(C_2R)^2 - 2\,L(C_1 + C_2)]$.

According to Equation (2), the fundamental parameters for 3dB (i.e. cut-off frequencies, central frequency, Q-factor, Bandwidth, etc.) of the dual-mode filter in figure 1 are obtained as solutions of equation (3).

$$\left[\frac{|T(S)|_{Max}}{\sqrt{2}}\right] = \frac{a\,\omega^3}{b\,\omega^6 + c\,\omega^4 + d\,\omega^2 + 1} \tag{3}$$

The quantity ω_c is obtained as solution of Eq. (3) and corresponds to the cut-off frequencies.

It is worth mentioning that the analytical solving of Eq. (3) in order to determine the fundamental parameters (i.e. cut-off frequency, central-frequency, Q-factor, etc.) of the filter at 3dB is very tedious. Thus, the reader should refer to the use of i.e. the Symbolic Calculation toolbox in MATLAB to obtain the analytical solution of Eq. (3). The reader could also use the mathematical ordinary differential equation derived as the model of the dual-mode filter in figure 1 in order to determine the fundamental parameters of the filter. Another method, which could be appropriate to determine the fundamental parameters of the dual-mode filter in figure 1 is the SIMULINK graphical representation (see figure 2), which has been designed to solve equation (4).

$$\frac{d^3U_{out}}{dt^3} + \left(\frac{C_1 + C_2}{RC_1C_2}\right)\frac{d^2U_{out}}{dt^2} + \left(\frac{1}{LC_1}\right)\frac{dU_{out}}{dt} + \left(\frac{1}{LRC_1C_2}\right)U_{out} = \frac{d^3U_{in}}{dt^3} \tag{4}$$

Where U_{out} is the output signal of the filter and the coefficients of equation (4) are expressed in terms of the components (i.e. L, R, C_1, and C_2) of the dual-mode filter.

The analytical solving of equation (4) is straightforward and can be performed by assuming the analytical solution in the form:

$$\begin{cases} U_{out} = & A_{max}\,e^{i\omega t} + cc \\ U_{in} = & B_{max}\,e^{i\omega t} + cc \end{cases} \tag{5}$$

Where A_{max} and B_{max} are the amplitudes of the output and input signals respectively. The expression cc stands for the complex conjugates of the preceding terms.

Substituting equation (5) into equation (4) leads to the derivation of the transfer function in equation (1). Let us mention that due to the range of variation of the components values of the dual-mode filter, the numerical solving of equation (4) requires an appropriate time scaling for the sake of convergence of the numerical scheme (e.g. MATLAB and/or SIMULINK).

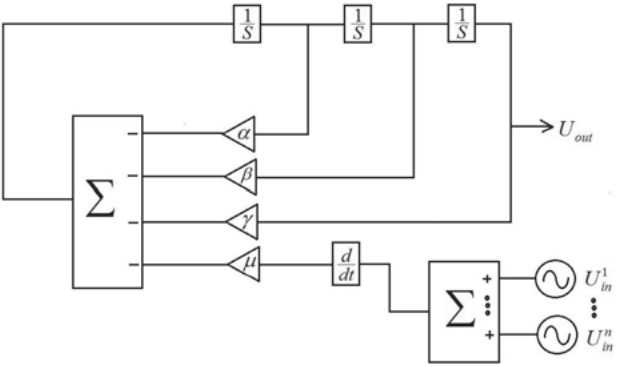


Fig. 2: Design of the dual-mode filter in SIMULINK

Let us mention that the time scaling chosen to solve equation (4) is expressed in the form $t = 10^N \tau$. The integer N is chosen according to the criterion of convergence of the numerical scheme. This criterion depends on the values of the dual-mode filter′s components. The time scaling is also necessary for the design of the dual-mode filter in the SIMULINK representation in figure 2. There, the input signal can be expressed in terms of harmonics according to the summation-module at the shown circuit′s. The summation-module offers the possibility of obtaining an input signal with a large spectral density (i.e. a mixture of LF, HF, RF, etc.).

Overall, our findings (using both analytical and numerical methods) have revealed very interesting results, which are summarized in figure 3.

According to figure 3 the dual-mode filter in figure 1 is a Band-Pass filter at frequencies between $0Hz \leq f \leq 100Hz$, while at radio frequencies (HF) the dual-

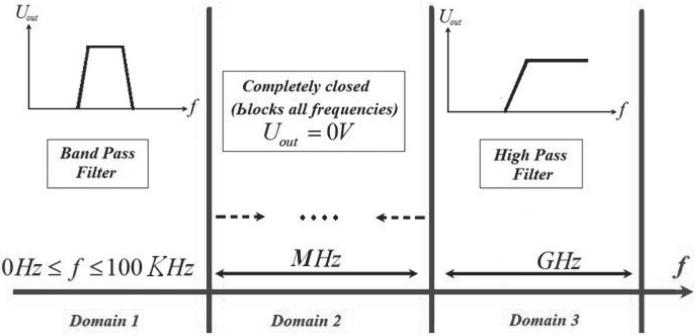


Fig. 3: Illustration of the performances of the dual-mode filter in Fig. 1

mode filter is a High-pass filter. Our achievements have also revealed that when the resistor R increases, the bandwidths of the Bandpass and High-pass filters enlarges while the middle domain (i.e. Domain 2) decreases and finally both filters are merged into a single-mode filter of type High pass. This achievement appears to be very interesting as it could be exploited to separate low frequencies (LF) from high and very high frequencies (HF and RF). The next section (section 3) presents the results obtained when monitoring the resistor R, which is used as control parameter (i.e. bifurcation parameter).

3 Results

This section presents the results of the bifurcation analysis obtained using both analytical and numerical methods. A comparison between the two methods is carried out to validate the concepts. The results are obtained by monitoring the resistor R between the range $[0.1\,K\Omega \quad 100\,M\Omega]$.

The first bifurcation analysis considers a monitoring of the resistor R in the range $[0.1\,K\Omega \quad 1\,M\Omega[$. In this range our investigation has revealed that the filter in figure 1 performs in dual-mode. The results obtained and mentioned in table 1 can be summarized as follows:

- $R = 0.1\,K\Omega$: The filter is highly selective and has the Band-Pass property. The resonant frequency is $f_0 = 87.612\,KHz$ and the cut-off frequencies are $f_{c_{low}} = 87.611\,KHz$ and $f_{c_{high}} = 87.613\,KHz$. However, when the frequency range is of order greater than a Gigahertz ($f > 1\,GHz$), the same filter has the High-Pass property with the cut-off frequency $f_c = 3.1847\,GHz$.

- $R = 1\,K\Omega$: The filter is highly selective and has the Band-Pass property. The resonant frequency is $f_0 = 87.612\,KHz$ and the cut-off frequencies are $f_{c_{low}} = 87.600\,KHz$ and $f_{c_{high}} = 87.624\,KHz$. However, when the frequency range is of order greater than hundred Megahertz ($f > 100\,MHz$), the filter has the High-Pass property with the cut-off frequency $f_c = 318.47\,MHz$.

- $R = 10\,K\Omega$: The filter is highly selective and has the Band-Pass property. The resonant frequency is $f_0 = 87.612\,KHz$ and the cut-off frequencies are $f_{c_{low}} = 87.493\,KHz$ and $f_{c_{high}} = 87.734\,KHz$. However, when the frequency range is of order greater than 10 Megahertz ($f > 10\,MHz$), the filter has the High-Pass property with the cut-off frequency $f_c = 31.846\,MHz$.

- $R = 100\,K\Omega$: The filter is highly selective and has the Band-Pass property. The resonant frequency is $f_0 = 87.612\,KHz$ and the cut-off frequencies are $f_{c_{low}} = 86.507\,KHz$ and $f_{c_{high}} = 88.932\,KHz$. However, when the frequency range is of order greater than 1 Megahertz ($f > 1\,MHz$), the filter has the High-Pass property with the cut-off frequency $f_c = 3.1766\,MHz$.

The first bifurcation analysis performs in the range $[0.1\,K\Omega \quad 1\,M\Omega[$ has revealed that when the resistor R increases, the bandwidths of the dual-mode filter increase and finally the dual-mode filter is merged into a single-mode filter when $R > 1\,M\Omega$. This statement justifies the aim of the second bifurcation analysis.

The second bifurcation analysis has been performed by considering a monitoring of the resistor R in the range $[1\,M\Omega \quad 100\,M\Omega]$. When the resistor R is greater than $1\,M\Omega$ the filter in figure 1 performs in a single-mode with the property of High-Pass. The property of Band-Pass disappears and is transformed into overshoots, which are depicted in table 2. The results obtained and mentioned in table 2 can be summarized as follows:

- $R = 1\,M\Omega$: At this specific value of the bifurcation parameter R, the dual-mode filter is reduced/transformed into a single-mode filter with the property of High-Pass. The cut-off frequency is $f_c = 208.08\,MHz$. The pic (or amplitude) of overshoot is $V_{pp} = 1.1612\,Volts$. The value $R = 1\,M\Omega$ corresponds to the bifurcation value at which the dual-mode filter is transformed into a single-mode filter of type High-Pass.

- $R = 10\,M\Omega$: The filter has a High-Pass property with the cut-off frequency $f_c = 79.7883\,MHz$. The pic (or amplitude) of overshoot is $V_{pp} = 7.80\,Volts$.

Table 1: Results of the bifurcation analysis showing the dual-mode when $0.1\,K\Omega < R < 1\,M\Omega$

	Dual-mode (i.e. Band-Pass and High-Pass)												
	DOMAIN 1 (Bandpass)								DOMAIN 2 (Stop-all)	DOMAIN 3 (High-Pass)			
	T_{max}		$f_{c_{low}}(KHz)$		$f_{c_{high}}(KHz)$		$f_{0_{central}}$ (Resonance)		$U_{out} = 0V$	T_{max}		F_C (GHz)	
R ($K\Omega$)	Analytical & **Numerical**		Analytical & **Numerical**		Analytical & **Numerical**		Analytical & **Numerical**		$U_{out} = 0V$	Analytical & **Numerical**		Analytical & **Numerical**	
0.1	1	**1**	87.655	**87.611**	87.657	**87.613**	87.656	**87.612**	$U_{out} = 0V$	1	**1**	3.1878	**3.184700**
1	1	**1**	87.644	**87.600**	87.668	**87.624**	87.656	**87.612**	$U_{out} = 0V$	1	**1**	0.31864	**0.318470**
10	1	**1**	87.536	**87.493**	87.778	**87.734**	87.656	**87.612**	$U_{out} = 0V$	1	**1**	0.03182	**0.031846**
100	1	**1**	86.551	**86.507**	88.971	**88.932**	87.761	**87.612**	$U_{out} = 0V$	1	**1**	0.00317	**0.003177**

- $R = 100\,M\Omega$: The filter has a High-Pass property with the cut-off frequency $f_c = 77.852\,MHz$. The pic (or amplitude) of overshoot is $V_{pp} = 79.76\,Volts$.

The second bifurcation analysis has also revealed that the bandwidth of the High-Pass filter and the pic of overshoots increase with the increasing value of resistor R. This statement is depicted in figures 8, 9, and 10 and also in table 2.

The results in the first line/row of table 1 (i.e. for $R = 0.1\,K\Omega$) are shown in Figs. 4. The transfer function of the Band-Pass property is depicted in Figs. 4 a) (analytical results (left) and numerical results (right)). In Fig. 4 b) is shown the transfer function for the High-Pass property (analytical results (left) and numerical results (right)). A comparison between the two methods clearly shows a perfect agreement between them.

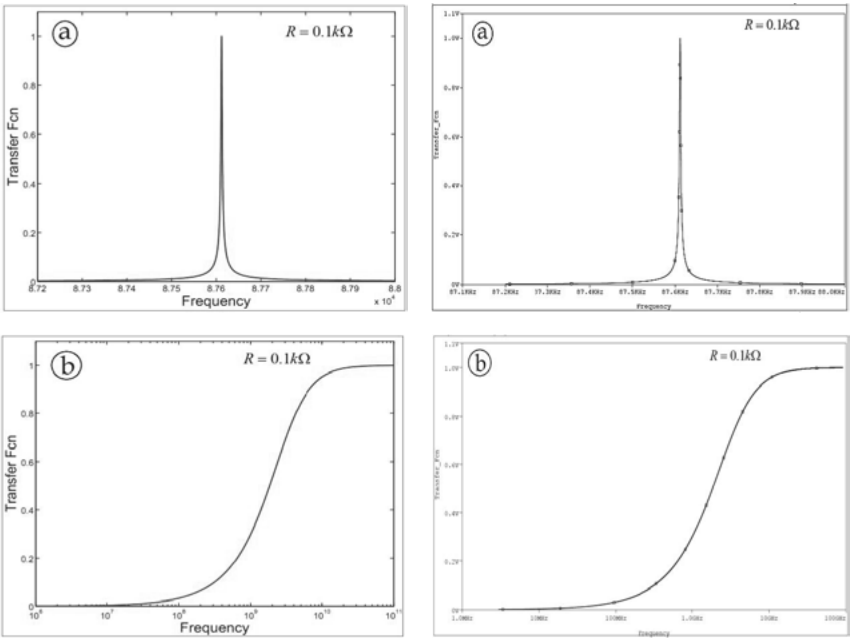


Fig. 4: Transfer functions of the dual-mode filter for $R = 0.1\,K\Omega$

The case $R = 1\,K\Omega$ (see second row in table 1) has led to results in Fig. 5. The transfer function of the Band-Pass property is depicted in Figs. 5 a) (analytical results (left) and numerical results (right)). In Fig. 5 b) is shown the transfer function for the High-Pass property (analytical results (left) and numerical results (right)). A perfect agreement is shown between the methods.

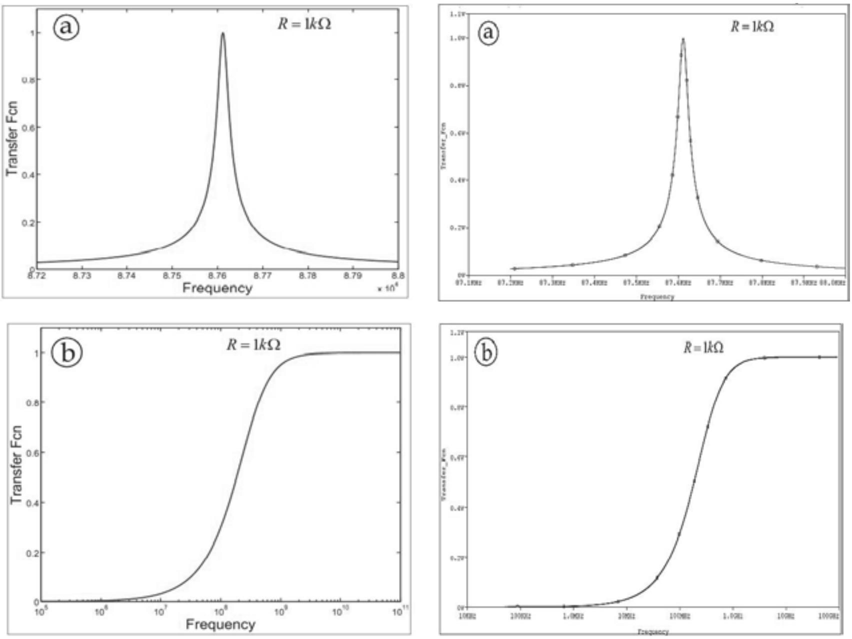


Fig. 5: Transfer functions of the dual-mode filter for $R = 1\,K\Omega$

The case $R = 10\,K\Omega$ (see third row in table 1) leads to results in Fig. 6. The transfer function of the Band-Pass property is depicted in Figs. 6 a) (analytical results (left) and numerical results (right)). In Fig. 6 b) is shown the transfer function for the High-Pass property (analytical results (left) and numerical results (right)). A perfect agreement is shown between the methods.

The case $R = 100\,K\Omega$ (see fourth row in table 1) leads to results in Fig. 7. The transfer function of the Band-Pass property is depicted in Figs. 7 a) (analytical results (left) and numerical results (right)). In Fig. 7 b) is shown the transfer

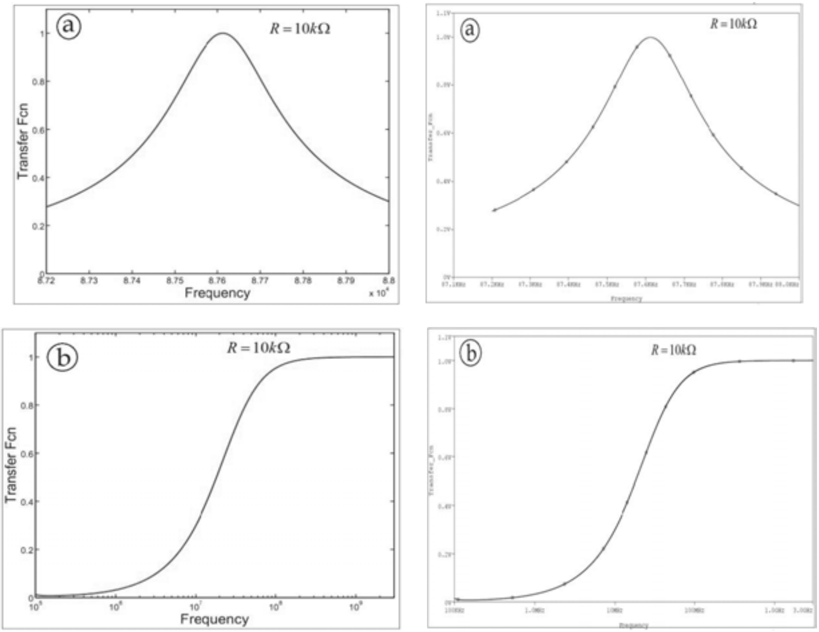


Fig. 6: Transfer functions of the dual-mode filter for $R = 10\,K\Omega$

function for the High-Pass property (analytical results (left) and numerical results (right)). A perfect agreement is shown between the methods.

An interesting comment to be made is related to the range of variation of the bifurcation parameter R. The preceding results have revealed that when the values of R increase, the bandwidths of the dual-mode filter increase. However, a specific value of R has been detected (i.e. $R = 1\,M\Omega$) at which the dual-mode filter is merged (or is transformed) into a single-mode filter of type High-Pass. It has been observed that the High-Pass property is subject to overshoots, which are characterized by a transient and sudden variation of the amplitude of the output signal and finally this amplitude converges to a fixed (or constant value) at long term. This convergence reveals the stationary behavior of the single-mode filter with the property of High-Pass. The results for $1\,M\Omega < R < 100\,M\Omega$ depicted in table 2 confirm the stationary behavior of the single-mode filter with the High-pass property.

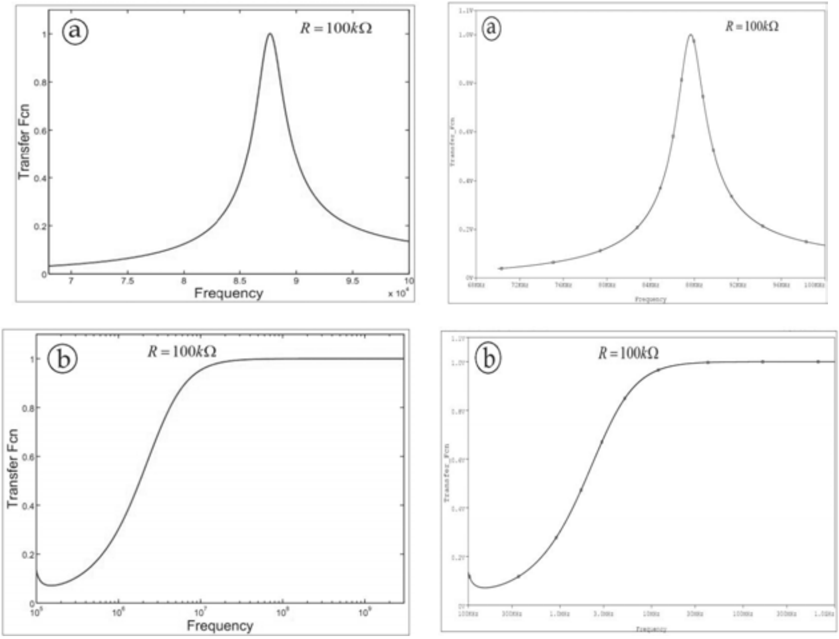


Fig. 7: Transfer functions of the dual-mode filter for $R = 100\,K\Omega$

Table 2: Results of the bifurcation analysis showing the single-mode when $1\,M\Omega < R < 100\,M\Omega$

	Single-mode (i.e. High-Pass)			
	Amplitude of Overshoots (V)	T_{max}	$F_c(KHz)$	
$R(M\Omega)$			Analytical	Numerical
1	1.1612	1	208.08	208.20
10	7.8014	1	79.822	79.783
100	77.852	1	79.783	79.766

The case $R = 1\,M\Omega$ (see first row in table 2) leads to results in Fig. 8. The transfer function of the High-Pass property is depicted in Fig. 8 (analytical result (left) and numerical results (right)). The overshoot appears with an amplitude (or a pic-voltage) of approximately $1.16\,Volt$. A perfect agreement is obtained between the analytic and numerical methods.

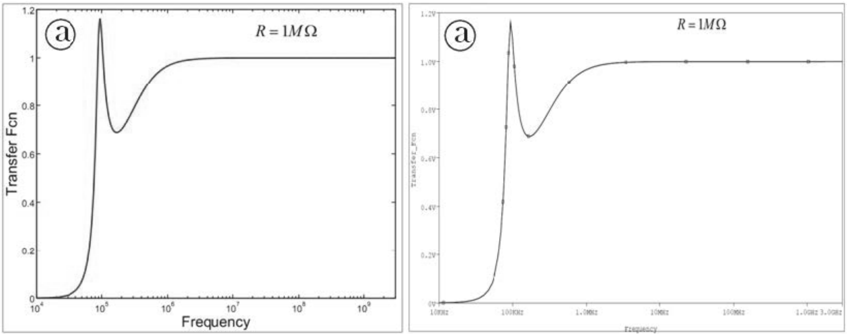


Fig. 8: Transfer functions of the single-mode filter for $R = 1\,M\Omega$

The case $R = 10\,M\Omega$ (see second row in table 2) leads to results in Fig. 9. The transfer function of the High-Pass property is depicted in Fig. 9 (analytical result (left) and numerical results (right)). The overshoot appears with an amplitude (or a pic-voltage) of approximately 7.8 $Volts$. A perfect agreement is obtained between the analytic and numerical methods.

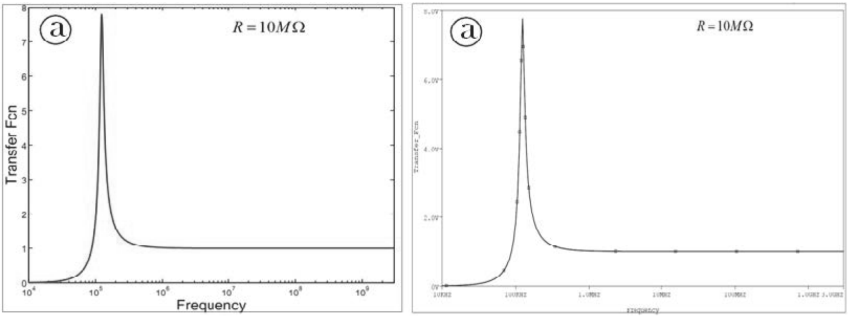


Fig. 9: Transfer functions of the single-mode filter for $R = 10\,M\Omega$

The case $R = 100\,M\Omega$ (see third row in table 2) leads to results in Fig. 10. The transfer function of the High-Pass property is depicted in Fig. 10 (analytical result (left) and numerical results (right)). The overshoot appears with an amplitude (or a pic-voltage) of approximately 77.850 $Volts$. A perfect agreement is obtained between the analytic and numerical methods.

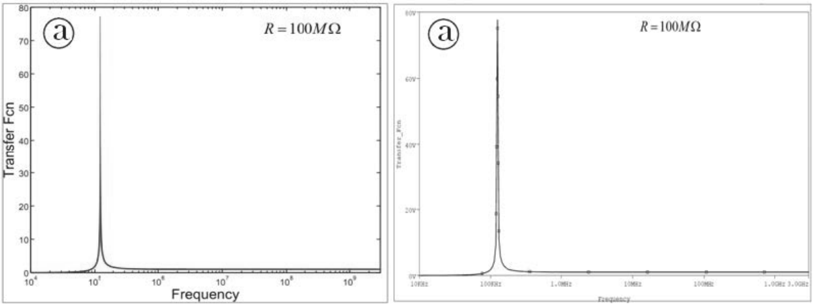


Fig. 10: Transfer functions of the single-mode filter for $R = 100\,M\Omega$

Another interesting comment observed in Figure 8, 9 and 10 is that the amplitudes (or pic) of overshoots increase with the increasing values of R. Overall our various investigations have revealed a perfect agreement (for all values of the bifurcation parameter R) between the analytical and numerical results.

4 Conclusion

This work has considered the analytical and numerical studies of a dual-mode passive filter with the properties of Band-Pass and High-Pass. A bifurcation analysis has been carried out using a dissipative component (i.e. resistor R) as the control or bifurcation parameter. A range/windows of the values of R has been derived, under which the dual-mode property is observed. A specific value corresponding to the values of $R = 1\,M\Omega$ has been derived at which the dual-mode filter merges into a single-mode filter with the property of High-Pass. Further, a range of the values of R has been derived under which the filter is a single-mode with the property of High-Pass. Overall, it has been observed that due to the High-order of the dual-mode filter considered in this work, it is very challenging to derive the fundamental parameters (i.e. cut-off frequencies, central (or resonant) frequencies, Q-factor, etc.) of the dual-mode filter analytically. To tackle this problem, several approaches have been proposed such as the use of the Symbolic toolbox in MATLAB, the mathematical modelling of the dual-mode filter into ordinary differential equations and the design of the dual-mode filter in SIMULINK. As proof of concepts in order to validate the concept developed in this paper, a benchmarking has been carried out leading to a comparison between the analytical and numerical results. This comparison has led to a perfect agreement between the methods.

References

[1] Ce Fu, Hui Wang: optimization of passive filter for Wireless Communication. IEEE computer Society, World Congress on Sofware Engineering. (2009) 483-486.

[2] Lu, J.-H.: A passive filter for 10-Gb/s analog equalizer in 0.18-m CMOS technology. Solid-State Circuits Conference, 2007. ASSCC '07. IEEE Asian. (2007) 404 - 407.

[3] Mishra G. and Gopalakrishna: Design of Passive High Pass Filter for Shunt Active Power Filter Application. International Conference on Circuits, Power and Computing Technologies IEEE. (2013) 17 - 21.

[4] Baker B.: Designing active analog filters in minutes. Analog Applications Journal. 40 (2013) 28-32.

[5] Musa Faisal A., and Carusone Chan A.: A Baud-Rate Timing Recovery Scheme With a Dual-Function Analog Filter. IEEE Transactions on Circuits and Systems II. 53 (2006) 1393-1397.

[6] Rumberg B., and Graham David W.: A Low-Power and High-Precision Programmable Analog Filter Bank. IEEE Transactions on Circuits and Systems II. 59 (2012) 234-238.

Retinopathy of Prematurity Classification based on Image Analysis

Maleerat Sodanil and Phattharachon Thongrit

Faculty of Information Technology
King Mongkut's University of Technology North Bangkok, Thailand

Abstract: Retinopathy of Prematurity (ROP) has been nominated as the most common cause of premature infant blindness. Among those affected, premature infants were found to be the most likely to survive. In Thailand, medical staff still lack ophthalmologist training to screen for Retinopathy found in premature infants. The purpose of this research was to develop an algorithm to aid the diagnosis and screening of patients using a retinal image to classify five stages of Retinopathy of prematurity. The research proposed the process of digital image processing and machine learning techniques which can be described into five steps: 1) Image processing, to improve the quality of retinal images. 2) Optic disc selection to separate the optic disc from the other sections which makes it easy to analyze. 3) Features extraction of the abnormal retinal. 4) Image subtraction in order to get disorders of disease. 5) Classification, the severity scale in each level; analyze the characteristic irregularities occurrences on the retina. Artificial Neural Network Back Propagation was used to learn 100 retinal images with 11 features input of and 5 stages output. The optimum network was tested with 11:10:5 for input nodes, hidden layers and output nodes respectively. The performance testing is focused on disease diagnosis. The performance result given 96% in terms of accuracy. Therefore, the proposed method can be applied successfully to screen for retinal disease in premature infants and also reduce the amount of ophthalmologist screening mistakes of Retinopathy.

1 Introduction

Retinopathy of prematurity (ROP) previously known as retrolental fibroplasia (RLF). ROP remains a major cause of visual loss in very premature infants that first described by Terry in 1942 [1]. ROP is a disease of the retina that is found in infants, especially premature infants with low birth weight, the retinal blood vessels have not completed their development. In cases of patients with ROP, the blood vessels stop growing and new, abnormal blood vessels grow instead of the normal ones. As the blood vessels regenerate similar retinal abnormalities found in diabetes (diabetic retinopathy) or other diseases for which there is inadequate retinal ischemia. The statistical data survey from the Queen Sirikit National Institute of Child Health (QSNICH), the center of the transferred patients with abnormal retinal blood vessels in premature infants. Since the year between 2009 and 2014, premature infants with birth weight less than 2,000 grams were screened as patients equal 3,583 cases. The detection result of infants with symptoms of abnormal retinal blood vessel is 1,350 cases as ROP and 152 cases were detected as abnormal retinal blood vessel in severe called Aggressive Posterior ROP (AP-ROP) and a high risk of causing permanent blindness if not treated promptly which represented as 38 percent and 4 percent of the total screening respectively. According to the data analysis of health system research about blindness, low vision and eye disease that is a problem of Thai Child in years 2006–2007. In Thailand, the abnormal blood vessel in premature infants is a conditional disease that causes blindness in children up to 66.67 percent [2, 3].

Currently, premature infants are likely to survive more than the past year. In Thailand, still lack of ophthalmologists to screen for diseases or abnormal of the eye. The information from The Royal College of Ophthalmologists of Thailand was found in 2015, the specialists in Retina and Vitreous is equal 154 persons while the specialists in Pediatric Ophthalmology is equal 109 persons. There are only 263 persons who have the proficiency to detect and treatment for premature infants who have abnormal blood vessel as opposed to premature infants at risk for abnormal blood vessel which increased for each year.

Table 1: Number of patients for each class of disease

	2009	2010	2011	2012	2013	2014
Patients	605	679	593	569	571	566
ROP	280	261	236	180	185	208
ROP (%)	46.28	38.44	39.80	31.63	32.40	36.75
APROP	28	62	27	15	9	11
APROP (%)	4.63	9.13	4.55	2.60	1.58	1.94

Table 1 shows the number of patients for each class of disease between 2009 and 2014, collected from Queen Sirikit National Institute of Child Health. According to the problem as mentioned earlier, this research aims to classify the stage of ROP in the premature infants who have risked of an abnormal blood vessel by using digital image processing and machine learning techniques to create a model which can be used to classify the disease in the initial stage of retinal screening in order to reduce the risk that might be happened in case of lacking of ophthalmologists.

2 Related Work

Retinopathy of prematurity (ROP) is one of a few causes of childhood visual disability which is largely preventable. Many extremely preterm babies will develop some degree of ROP although in the majority never progresses beyond mild disease which resolves spontaneously without treatment. A small proportion, develop potentially severe ROP which can be detected through retinal screening. If untreated, severe disease can be resulted in serious vision impairment and consequently all babies at risk of sight-threatening ROP should be screened [4]. ROP can be divided into five stages [5, 6] as:

Stage 1 Demarcation line, signifies development of a demarcation line between the normal retinal vessels developed retina and avascular retina, where vessels have not developed.

Stage 2 Ridge, signifies development of a demarcation ridge between the normal retinal vessels developed retina and avascular retina, where vessels have not developed. The demarcation line of stage 1 gains height and width and extends above the retina surface.

Stage 3 Ridge with extra retinal fibro vascular proliferation, signifies development of a new vessels (neovascularization) in the stage 2 demarcation line and leads to extra retinal proliferation of tissue. The ridge may now bleed and cause traction on the ridge.

Stage 4 Subtotal retinal detachments, signifies that retinal detachment has started to occur due to the traction of the fibrous and vascular tissue over the ridge. It's called 4A when the retinal detachment spares the macula and stage 4B when retinal detachment involves the macula.

Stage 5 Total retinal detachment, signifies that total retinal detachment has occurred. This is the last stage of retinopathy of prematurity, and the most severe stage too. Since the total retina has detached, it becomes necessary for an eye surgeon to perform eye surgery on these eyes in an attempt to reattach retina.

In order to detect the skeletonized structure, Lassada [7] proposed the methods statistically optimized LOG edge detection filter, Otsu thresholding, Medial Axis transform skeletonization, Pruning, and edge thinning. The result from experiment was compared with ophthalmologists' hand-drawn ground truth and it can detect the blood vessel with a high specificity of 0.9879 and sensitivity of 0.8935, while [8] proposed the difference method, he use high pass filter to track the retinal vessels and the energy criterion is computed for finding the percentage of area which covered with blood vessels. The result of this study was compared with ophthalmologist's hand labels of diagnosis and it can detect the prematurity with sensitivity of 29.61% and accuracy of 59.10%. In order to detect the tortuosity, Conor [9] proposed the segmentation technique, the threshold was used to further emphasize, skeletonization. Applying a simple retrospective screening paradigm based solely on vessel width and tortuosity yields a screening test with a sensitivity and specificity of 82% and 75% respectively while [10] proposed the different method, he use a skeleton of the retinal blood vessels which extracted from the original infant retinal image using a series of morphological operators, an adaptive linear interpolation scheme and the tortuosity is calculated based on the curvature of the resulting vessel segments. The retinal images were classified into two classes using segments characterized by the highest tortuosity. In order to diagnosing Diabetic Retinopathy, [11] proposed method to improve and restore the quality of retinal image, separate elements of retinal image consisting of Optic disc, Fovea and Macula and structure of blood vessels, detect and identify the pathology of Diabetic Retinopathy by characteristic, classify the severity scale in each level which given 86% of classification in terms of accuracy.

3 Proposed Method

This paper proposed an algorithm for the diagnosis of retinal disorders in premature infants from images of the retina to the staging of the disease development. It can be divided into five stages of the process as:

1. Image preprocessing, at this stage, the retinal images derived from a database of patients to the treatment. The acquired images may have different elements from different imaging devices, the element of picture is

not clear, or other interference. This makes the quality of the picture different in terms of the clarity of the light intensity of an image size and it is necessary to adjust initially to the same norms in order to identify the composition of the image and accuracy.

   a) Image resizing, determining the size of the picture to the 512×512 pixel.

   b) Edge removing, a portion of the edge of the image to a specific analysis.

   c) Moving average filtering, this step takes the intensity of the pixels in the side of the pixel gray to the image processing average. The results of the mean is represented as a level of intensity. In this research, a number of nearby images is tamper-proof filter on average size equal $3 * 3$.

2. Optic Disc Selection, the optic disc is the area where most of the brightness image by the look. Area of a circle, which is usually the blood vessels will begin to sprout from the optic disc area and go out until the end. The purpose of optic disc area is to detect out of the picture before the split and simply to analyze the lesions.

   a) Adaptive Threshold, this step is the process of converting a color image to black and white. The image brings to this process is a green channel image through the process of filter already, the output will be in the form of a binary image, which is segmented but each pixel of the image will be calculated if the pixel is below a certain threshold that configured as background. The remaining pixels are set as the foreground.

   b) Binary Area Open, this step removes smaller objects from binary images, which in research this procedure configured P at 1,200 pixels.

3. Feature Extraction or characterization of retinal disorders in the disease, retinal disorders in premature infants with characteristics that depend on the abnormal retinal blood vessels that look different each time. Need to find lesions in order to contribute to the process of disease classification.

   a) Adaptive Threshold, this step is the process of converting a color image to black and white. The image brings to this process is a green channel image, the output will be in the form of a binary image, which is segmented. However, each pixel of the image will be calculated if the pixel is below a certain threshold that will be set as background. The remaining pixels are set as the foreground.

4. Subtraction, a process that takes a result from the Optical selection process subtract with a result from the characterization of retinal disorders in order to get a clearer disorders of disease.

   a) Convert White to Black Pixel, this step is the process after images were processed in the process, the transformed pixels. From white to black, it obtained a precise lesions leading to separate by disease.

   b) Binary Area Open, to remove all the components are connected together (Objects) with less than P pixels of an image binary and finding lesions which determined the 20 pixels.

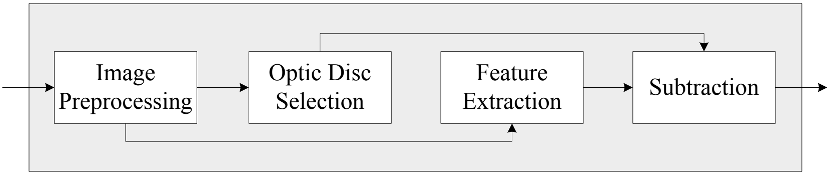


Fig. 1: Feature Extraction

5. Disease Extraction, the image is processed as showed in Fig. 1 until the image shown only disorder features of the blood vessel and brings value to the process of training and testing in order to classify a stage of ROP into 5 stages as mentioned earlier. There are two steps for learning.

   a) Feature Extraction, the separation characteristics of the image will be separated or extracted an important features from the image. The characteristics of the properties can be obtained by using image processing techniques. In this research, the basic characteristics of the retinal image is separated into shape features and color features in total of 11 features: Number of lesions, Area, Perimeter, Major Axis Length, Minor Axis Length, Compactness, Eccentricity, Standard deviation of HSV, Standard deviation of Green channel, Average of HSV, and Average of Green channel is used as showed in Table 2.

Table 2: Input details for ANNs

Input	Description
X1	Number of lesions
X2	Area
X3	Perimeter
X4	Major Axis Length
X5	Minor Axis Length
X6	Compactness
X7	Eccentricity
X8	Standard deviation of HSV
X9	Standard deviation of Green channel
X10	Average of HSV
X11	Average of Green channel

b) Artificial neural network is used in this step to create the best model of ROP classification. In this research, 100 images were collected, for learning and testing for disorders of the retina with 11 features of input and 5 stages output as shown in Fig. 2.

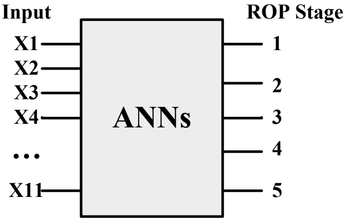


Fig. 2: ANNs for Retinopathy of Prematurity (ROP) Classification

4 Experiments and Results

The data of 100 premature infant's images was taken using a Retcam3 camera collected from the Queen Sirikit National Institute of Child Health (QSNICH) between 2009 and 2014, which taken from several ophthalmologist in QSNICH. In part of classification model, Fig. 3 shown both training and testing process. The process started with feature extraction as described earlier. The process of disease extraction generate the output of the characteristics of the premature infant retinal image of 11 features which is used as an input of ANNs to classify the stage of ROP. The best model is used to be an input of testing phases in order to generate best performance in terms of accuracy.

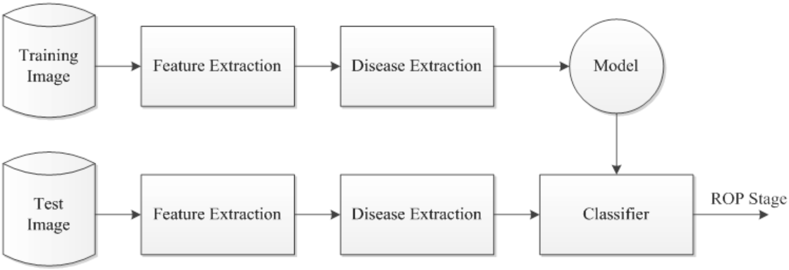
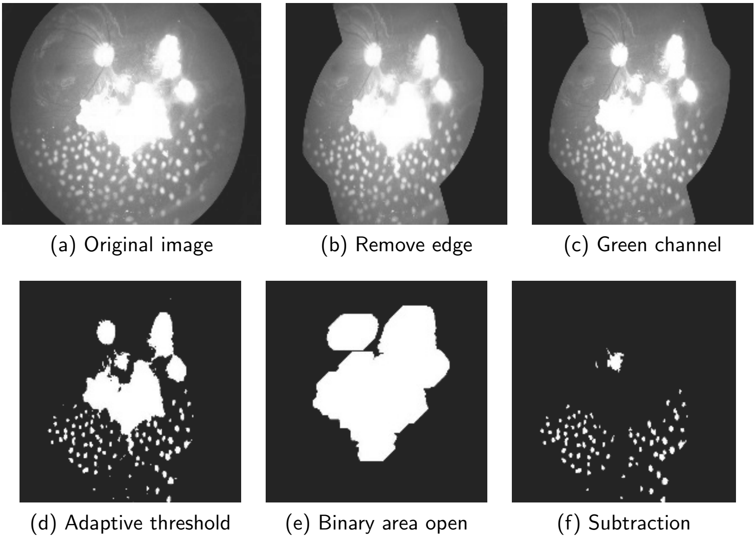


Fig. 3: ROP learning and classification


(a) Original image (b) Remove edge (c) Green channel

(d) Adaptive threshold (e) Binary area open (f) Subtraction

Fig. 4: Results of each processing step

Figure 4 shows the results from each step. Fig. 4 (a) is an original premature infant retinal image, 4 (b) is the result after the image processing process, 4 (c) is the result after green channel has selected from RGB image, 4 (d) is the output after applied Adaptive Threshold. The result of the binary area open is shown in 4 (e) and Fig. 4 (f) is the image after the subtraction process which display only the lesions of disease.

The experiment was conducted by 100 of retinal images which divided developmental stages of macular disorders in premature infants to compare with the diagnosis by the ophthalmologist specialist.

Table 3: Performance evaluation in terms of accuracy

	Predicted							
	ROP Stage	**1**	**2**	**3**	**4**	**5**	**Overall**	**Accuracy**
Actual	1	19	0	0	0	0	19	100%
	2	1	35	0	1	1	38	92%
	3	1	0	23	0	0	24	96%
	4	0	0	0	16	0	16	100%
	5	0	0	0	0	3	3	100%

In Table 3 showed the confusion matrix of classification performance. There are four images error in the prediction model which equal 8% and 4% error for stage 2 and 3 respectively. However, overall performance is good with 100% of accuracy from three of five stage ROP classification.

5 Conclusion

This research is aimed to classify the stages of Retinopathy of prematurity (ROP) which caused of childhood visual disability using image processing model and artificial neural network (ANNs). Eleven features were extracted from premature infant's images in order to use as input of ANNs. Five stages output of ANNs described the stage of ROP disease. The overall performance is equal 98% in terms of accuracy. However, in order to improve the performance of classification model, the quality of the retinal image and inputs have to be considered.

Acknowledgements

The authors would like to thank M. D. Kwanjai Wongkittirux, Head department of Pediatric Ophthalmology and M. D. Sumalin Trichaiyaporn for her very helpful support on the medical side. Also many thank the Queen Sirikit National Institute of Child Health for providing the ROP images for this research.

References

[1] Terry TL. (1942). Extreme prematurity and fibroblastic overgrowth of persistent vascular sheath behind each crystalline lens. *Am J Ophthalmol*. 25, 203–4.

[2] Kwanjai Wongkittirux. (2012). Blindness, Low Vision and Eye Diseases in Thai Children 2006–2007. *Journal of Health Systems Research*. 9(4), 501–512.

[3] Jenchitr W, Hanutsaha P, Iamsirithaworn S, Panrut U, Choosri P, Yenchitr C. (2007) The First National Survey of Visual Impairment, Blindness and Low Vision in Thailand 2006–2007 (The First TVIP 2006–2007). *Thai J Pub Hlth Ophthalmol*. 21, 1–94.

[4] Royal College of Paediatrics and Child Health. (2008). *Guideline for the Screening and Treatment of Retinopathy of Prematurity*. UK.

[5] The Committee for the Classification of Retinopathy of Prematurity. (1984). An international classification of retinopathy of prematurity. *Arch Ophthalmol*. 102, 1130–34

[6] The International Committee for the Classification of the Late Stages of Retinopathy of Prematurity. (1987). An international classification of retinopathy of prematurity II. The classification of retinal detachment. *Arch Ophthalmol*. 105, 906–12

[7] Lassada Sukkaew, et al. (2007). Automatic Extraction of the Structure of the Retinal Blood Vessel Network of Premature Infants. *J Med Assoc Thai*. Vol. 90, No. 9.

[8] Niousha Hormozi, Seyed Amirhassan Monadjemi and Gholamali Naderian. (2013). Retinal Vessel Detection in Retinopathy of Prematurity Using Butterworth High-pass Filters and SVM. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*.

[9] Conor Heneghan, John Flynn, Michael O'Keefe and Mark Cahill. (2002). Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis. *Medical Image Analysis*. 6. 407–429.

[10] Lassada Sukkaew, et al. (DECEMBER 2008). Automatic Tortuosity-Based Retinopathy of Prematurity Screening System. The Institute of Electronics, Information and Communication Engineers. *IEICE TRANS. INF. & SYST.* VOL. E91-D NO. 12

[11] Athiwath Daengphoonphol. (2013). *Digital Image Processing Algorithms for Diagnosing Diabetic Retinopathy from Fundus Photography in Diabetes Patients*. Master of Engineering Thesis in Computer Engineering, Graduate School, Khon Kaen University.

Dual Echo State Networks-based Generalized Matrix Inversion with Applications in Stochastic Time-varying Systems

Ahmad Haj Mosa[1], Kyandoghere Kyamakya[1], Mouhannad Ali[1],
Hughes Bisuta Bieto[2] and Jean Chamberlain Chedjou[1]

[1] Institutes of Smart System Technologies, Transportation Informatics
Alpen Adria University, Klagenfurt, Austria

[2] University of Kinshasa, Polytechnic Faculty, DR Congo

Abstract: This paper presents the concept of a black-box trained reservoir-neuro-computing based matrix inversion system. Hereby, the reservoir-neuro-computing processor is realized by a Dual Echo State networks. First, the importance of a real-time matrix inversion function is highlighted. Then we briefly describe a series of selected technical contexts where such a function is critical. A critical survey of related approaches involving recurrent neural networks (RNN) is conducted whereby pros and cons are presented. In contrary to these last listed concepts, to be categorized as white-box ones, we do then present our novel black-box alternative which does involve a Echo State Network as universal processor system. A comprehensive training concept is presented and validated along with a systematic extensive testing for an overall validation. For illustrative purposes, two cases of matrix inversion are studied. Case 1 contains square singular/non-singular matrix samples to be inverted of dimensions 3×3 and Case 2 contains non-square singular/non-singular matrix samples to be inverted of dimensions 5×3. The tested matrices are generated randomly with $\mathcal{N}(0,1)$ and considered to be sequential stochastic extreme time varying case. The experiments do confirm that the Dual ESNs does work and perform well.

1 Introduction

In the field of engineering and related sciences there are various contexts which are significant in different ways and manners. Matrix inversion which is real time and quick is very significant in engineering. When computer engineering is studied, matrix inversion is considered as one basic issue. This issue is studied in different scenarios. It is generally a brick of many solvers or functions or solutions in various areas like the following [1–10]: finite element solvers, preliminary steps for optimization, signal processing, electromagnetic systems, robotic control, statistics, simulations, and physics. There are many backgrounds or scenarios where matrix inversion can help in processes of solution. Some of them are more challenging than the others. To start with, the real time processing contexts are complex. Same is the case when the matrix to be inverted are stochastic and time-varying. In such cases, besides robustness one should face related computational challenges. These scenarios have complexity which is computational in nature. In these cases, matrix inversion needs to be robust as well. When neurocomputing gets involved, these problems mentioned above can only be solved by studying the previous researches done on the topic. For matrix inversion, there are many approaches already in use which are based on neural networks. The next section will however highlight that these approaches are good but not perfectly robust or efficient. Further improvement is thus needed. Section 3 and 4 will discuss the issues pertaining to neural networks which are cellular, and the training approach which is based on black box. There will be other points discussed in this research paper too. At the end, the research will be concluded and ideas for further research will be shared. Work related to matrix inversion will be studied along with various aspects which bring this study into significant contexts.

2 Related Work and Limitations of the State of the Art

There are many challenges and complexities to face when it comes to study and use of functions of matrix inversion. This is especially true in case of settings which are highly technical and need high level of safety measures. Kinematics can be taken as an example here [5, 9]. The problems that are faced in such contexts or scenario include the situation where time varying matrix is to be inverted. In other cases when real time computation is to be done, that might prove to be really challenging. Another issue in such cases is the fact that some times the outputs from the sensors are uncertain. They are said to be carrying some noise. Robust results are thus needed to overcome this issue. Also, if the whole method is to be speeded up, it might prove to be a challenge. Also, the efforts need to be scalable. The efforts made in computational sense must rise in a linear way to help the matrix size increase. Exponential growth in this case can be considered as a worst case scenario. To evaluate matrix inversion system, there are some criteria that can be followed:

1) correctness of the results or the error root mean square error.
2) length of the transient phase (resp. response time).
3) the computing time.
4) the robustness to some small/bounded data uncertainty.
5) a good scalability behavior regarding the computation effort.
6) a good speeding-up in presence of multiple processor cores.
7) in addition, the matrix inversion method is generalized, which means it can invert single/non-single and square/non-square matrix.

The weak points of the existing recurrent neural network (RNN) based matrix inversion concepts can be found in the following aspects :

(a) there is a relatively long transient phase.

(b) the accuracy is not always perfect and does strongly depends on the nonlinear activation function settings and possibly of the initial conditions of X(t=0) as well.

(c) there is no straight-forward method to optimize the settings of the nonlinear activation functions.

(d) the speeding-up potential in presence of multiple processor cores may be sub-optimal.

(e) the scalability potential regarding computation effort is questionable.

(d) there is no comprehensive evaluation regarding the stochastic-time varying system.

Regarding these limitation of RNN based concepts we do propose in the next section (Section 3) a novel black-box based solver paradigm which will be implemented and realized in Section 4 by a cellular neural network (CNN) processor system.

3 Black-box-based Matrix Inversion Concept

There are certain assumptions in case of black-box matrix inversion concept. According to this solver concept it is assumed that a robust system is available, which is based on black-box model and it has certain special qualities. The qualities include:

(1) It can reliably cope with strong nonlinearities.

(2) It can robustly cope with stochasticity in the related system behavior input-output transfer function model.

(3) It ensures a extremely high processing speeding even on one single processor core.

(4) It does offer best possible speeding-up potential in presence of multiple processor cores.

(5) It does reliably and robustly display the characteristic of universal system modeller.

(6) It is of parallel computing nature in its essence and can be implemented on various hardware and/or software platforms, even on embedded ones and it does consume minimal computing ressources.

(7) It is of a neurocomputing nature and does thereby takes profit of the analog computing emulation potential and fast computing characteristics.

(8) It is capable of handling either continuous-time or discrete-time or both types of inputs and output simultaneously.

(9) Since the training process should be kept as fast as possible, it should posess a strong extrapolation and interpolation capability w.r.t. the fact that training simples should be as small as possible but sufficient to ensure a high quality of the training result.

(10) It offers maximum flexibility in the model setting regarding the fixation and interpretation of both inputs and outputs and related range values.

(11) There exists a coherent, straight-forward and comprehensive and accurate training scheme for the black-box processor.

(12) There exists a comprehensive straight-forward, comprehensive and accurate methodology and scheme to determine the parameter settings of all involved nonlinear activation functions.

As a prising recent trending candidate, we propose to use Echo state networks (ESN) as nonlinear black box matrix inversion method.

4 Echo State Network

Echo state network is also called as ESN. When non-Gaussian sequences which are dynamic or the nonlinear sequences are to be forecasted, ESN is considered as a very useful and reliable type of RNN. ESN was initially designed in 2001. It was designed by Jaeger [13]. The basic purpose of having this ESN was to develop the ability of using this RNN to produce the desired level of sparsity. This is achieved with the help of using a big number of neurons. This helps in better extraction of information from the inputs. In other words, ESN enhances the ability to extract information from inputs in certain sequences. In comparison to other networks, especially when compared with neural concepts, ESN has many advantages over the others. The first and foremost is the fact that it shows more details from the contributions made in the past. These contributions are highlighted in a manner that the recent ones are reflected in the best possible way. This system has a short term memory as it is relying on a very heavy number of neurons. This solves a major issue which pertains to considering how much history should be considered to be included in the input. Another plus point that this concept has is that ESN is easier when it comes to training. As a result, it becomes possible to reach out to a bigger population. This proves to be a handicap for many other neural networks. First layer of ESN is reservoir which has nonlinear neurons and offers self-feedback to the system. This first layer is connected with neurons for network input. Output is also connected here. Output layer actually has the regressors which are linear (see Fig. 1). Training of network happens in the first stage. Simple least square method is used in training. ESN has its own limitations too. There are ill conducted solutions in certain cases. Input uncertainty is another problem.

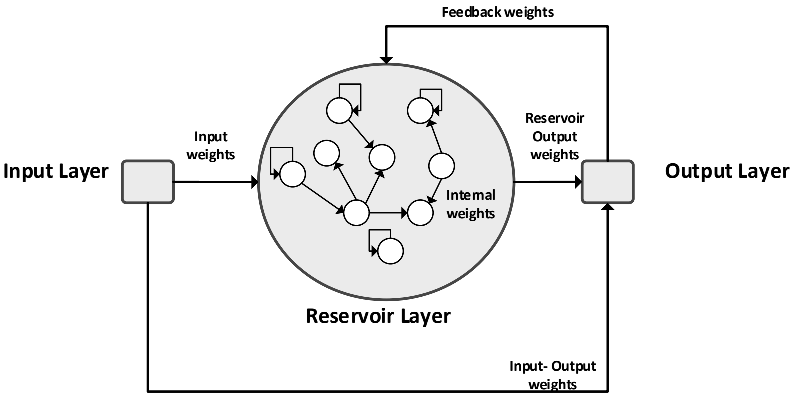


Fig. 1: The echo state network architecture. Inputs, Reservoir and Output layers are presented. The input layer feeds the reservoir which has a random internal connection between its inner cells. The reservoir feeds the output layer which also feeds back the reservoir.

5 Architecture of Dual ESN Processor System for Black-box-trained Matrix Inversion

5.1 Basics of Echo State Network

In the so called leaky ESN. The interaction between the reservoir neurons is modeled by a system of differential equations. The ordinary state equation of all reservoir neuron is given by [13]

$$\dot{X} = -X + \tanh(\mathbf{W}^{\text{in}}U + \mathbf{W}X + \mathbf{W}^{\text{fb}}Y) \tag{1}$$

where X, U and Y are respectively the current reservoir states vector, the inputs vector and the ESN output. $\mathbf{W}$ is the internal template, $\mathbf{W}^{\text{fb}}$ is the feedback template, $\mathbf{W}^{\text{in}}$ is the input/control template and

$$\tanh(r) = \frac{2}{1+e^{-2r}} - 1 \tag{2}$$

is the (Hyperbolic tangent sigmoid transfer function). The ESN output layer has the following linear model:

$$Y = \mathbf{W}^{\text{out}}[X;U] \tag{3}$$

where $[X;U]$ stands for a vertical matrix concatenation.

5.2 Echo State Training

Given an ESN with n reservoir neurons, m inputs and k outputs. The internal template, the inputs template and feedback templates are randomly generated. The random generation approach when it is associated with a high number of neurons extract valuable information coming from the inputs. The random generation process is done as the following:

a) $\mathbf{W}$ ($n \times n$ matrix) is generated as normally distributed sparse matrices with $\mathcal{N}(0, 1)$ and sparseness measure to sp . The resulted matrix is then divided on its own largest absolute eigenvalue and multiplied with a spectral radius sr. These generating constrains are important for the stability of the network as suggested by [14].

b) $\mathbf{W}^{\text{fb}}$ ($k \times n$ matrix), $\mathbf{W}^{\text{in}}$ ($m \times n$ matrix) generated randomly with a standard normal distribution $\mathcal{N}(0, 1)$ and scaled with a factor equal to f_{in} and f_{fb} respectively.

After the ESN templates are generated finally the global output linear regressors is trained using the so-called Ridge Regression as it is the most recommended method to train the regression weights [14] as described in the following:

$$\mathbf{W}^{\mathbf{out}} \quad = Y^{\mathbf{target}} V^{\mathrm{T}} (V V^{\mathrm{T}} + \beta \boldsymbol{I})^{-1} \tag{4}$$

where $Y^{\mathbf{target}}$ is the desired output;$V = [X; U]$; $\boldsymbol{I}$ is the identity matrix and β is the regularization coefficient (to avoid over fitting).

5.3 Data Generation

The black-box modeling is considered as the appropriate option when the studied problem is sophisticated enough in which, it can not be modelled using physical laws following the âœwhite-boxâ approach. The system parameters of a black box can be estimated using the corresponding inputs-outputs measurements and/or the observations that raise challenges when the studied data is highly nonlinear. Since ESN is considered to be a black-box model, we need a sufficient examples of matrix inversion to train the ESN. As a reference model we use Matlab [11] pseudo-inverse function (*pinv*). This Matlab function can be used to invert singular/non-singular square/non-square matrices. To evaluate our proposed model we consider two scenarios:

a) 3×3 square singular/non-singular matrices

b) 5×3 non-square singular/non-singular matrices

For each scenario, a total of 10000 matrices are generated using a normal random generator $\mathcal{N}(0,1)$. The 10000 samples are considered to be sequential in time, a case which represents an extreme stochastic-time-varying matrix inversion. The generated data is then splitted into 5000 training set 5000 testing set.

5.4 Dual ESN Model and Data Transformation

The fact that the inverse of an inverse is the matrix itself, in other words:

$$A = (A^{-1})^{-1}, \tag{5}$$

does inspired us to develop a closed loop matrix inversion model. Since the ESN black-box model mimics the matrix inversion, then it must have the property in Eq. 5. Accordingly, given a matrix A and its inverse B then:

$$\hat{B} = \mathbf{ESN}(A) \tag{6}$$

where $\hat{B}$ is the black-box ESN estimated inverse of A. Then, the A can be reconstructed by:

$$\hat{A} = \mathbf{ESN}(\hat{B}) = \mathbf{ESN}(\mathbf{ESN}(A)) \tag{7}$$

Consequently, the accumulated (double inverse) performance error of ESN can be measured by:

$$\mathbf{Error} = A - \hat{A} \tag{8}$$

In this paper, we introduce a dual ESNs modes that use the performance error in 8 as an additional information/input to improve the performance of the inverse system. In Figure 2 we illustrate the dual ESN model in which, two ESN models are presented, the first one is the Naive ESN Matrix Inverter, this model does a matrix inversion and trained using the samples of the pair A and B as inputs-outputs data. The second model (Robust ESN Matrix Inverter) does also a matrix inversion, however, it uses additional input to improve the performance. The second input is the estimated A such as in Eq. 7. The operation mode of the dual ESNs goes as follow:

1) Given An input A representing a matrix to be inverted

2) Given a random initial values of $\hat{A}$

3) Using The Robust-ESN-Inverter estimate $\hat{B}$

4) Using The Naive-ESN-Inverter estimate $\hat{A}$

5) Repeat steps 3 and 4 until convergence (The error of Eq.8 is stable)

6) The ordinary differential equations of both model are solved using the Matlab ode113 solver [11].

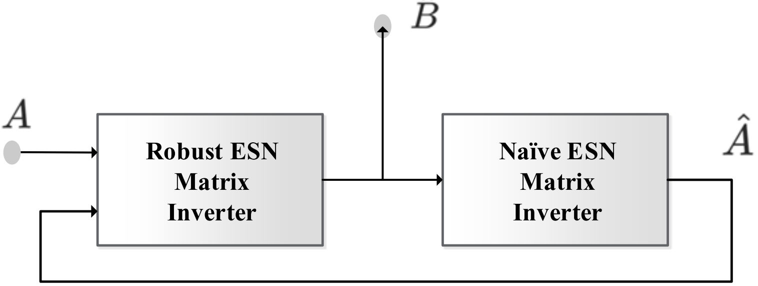


Fig. 2: The dual echo state networks architecture

6 Illustrative Examples of Validating the Novel Dual ESN-based Matrix Inversion Concept

In this section, we present the use of our novel dual echo state model for two case studies of generalized matrix inversion. The related performance is presented by the normalized root mean square error given by

$$NRMSE = \frac{1}{t}\sum_{i=1}^{t}\left(\frac{\sqrt{\frac{1}{me}\sum_{j=1}^{me}(B_{i,j}-\hat{B}_{i,j})^2}}{max(B_j)-min(B_j)}\right) \tag{9}$$

where t is the number of tested samples, n is the number of matrix elements, $B_{i,j}$ is the real expected inversion value and $\hat{B}_{i,j}$ is the predicted values.

6.1 Case Study 1

In this case study, a total of 5000 3 × 3 singular/non-singular matrices are generated and used to train and test our proposed model. The related ESNs configuration parameters are giving in Table 1 and the recored performance in Table 2. The presented ESNs parameters have been selected empirically.

Table 1: The dual ESN model configuration parameters of case 1

Parameter	Robust ESN	Naive ESN
n	2000	2000
m	9*2=18	9
k	9	9
sp	0.005	0.005
sr	0.1	0.3
f_{in}	0.1	0.1
f_{fb}	0.1	0.1
β	1e-4	1e-4

Table 2: The Evaluation of Case Study 1

ESN	NRMSE	Speed	Speed on GPU	Matlab Speed
Naive	35%	0.7s	0.005s	0.0055s
Dual	14%	1.5s	0.007s	0.0055s

6.2 Case Study 2

In this case study, a total of 5000 5 × 3 non-square singular/non-singular matrices are generated and used to train and test our proposed model. The related ESNs configuration parameters are giving in Table 3 and the recored performance in Table 4. The presented ESNs parameters have been selected empirically.

Table 3: The dual ESN model configuration parameters of case 1

Parameter	Robust ESN	Naive ESN
n	2000	2000
m	15*2=30	15
k	15	15
sp	0.005	0.005
sr	0.1	0.3
f_{in}	0.1	0.1
f_{fb}	0.1	0.1
β	1e-3	1e-4

Table 4: The Evaluation of Case Study 1

ESN	NRMSE	Speed	Speed on GPU	Matlab Speed
Naive	37%	0.9s	0.0055s	0.006s
Dual	13%	1.9s	0.0073s	0.006s

7 Conclusion

This paper has presented and validated the concept of a black-box trained reservoir-neuro-computing based matrix inversion system. The Echo State Network (ESN) system is used as the reservoir processor. We highlighted, the advantages of robust real-time matrix inversion for science and engineering. Then we briefly explored cutting edge related approaches involving recurrent neural networks (RNN) is conducted whereby pros and cons are addressed. In contrary to these last listed concepts, to be categorized as white-box ones, we do then present our novel black-box alternative which does involve Dual ESN universal processor system, which is a closed loop inverter, a feature that make the proposed model more robust for time varying system. The simple-fast training process of Echo states allowed us to explore many configuration and to obtain the best one. For illustrative purposes, some matrix samples to be inverted, of dimensions 3×3 square singular/non-singular matrices and 5×3 non-square singular/non-singular matrices are selected and the experiments do confirm that the dual esn based black-box trained concept does work and perform well. Also a benchmark is presented, which shows that the closed loop (dual) has obtained much better performance than the single ESN model.

References

[1] Zhang, Yunong, Weimu Ma, and Binghuang Cai. :From Zhang neural network to Newton iteration for matrix inversion. *Circuits and Systems I: Regular Papers, IEEE Transactions on* 56.7 (2009): 1405-1415.

[2] Zhang, Yunong, and Shuzhi Sam Ge. :Design and analysis of a general recurrent neural network model for time-varying matrix inversion. *Neural Networks, IEEE Transactions on* 16.6 (2005): 1477-1490.

[3] Steriti, Ronald J., and Michael A. Fiddy. .:Regularized image reconstruction using SVD and a neural network method for matrix inversion. *IEEE Transactions on Signal Processing* 41.10 (1993): 3074-3077.

[4] Sarkar, Tapan K., Kenneth R. Siarkiewicz, and Roy F. Stratton.: Survey of numerical methods for solution of large systems of linear equations for electromagnetic field problems. *ROCHESTER INST OF TECH NY DEPT OF ELECTRICAL ENGINEERING*, 1981.

[5] Sturges Jr, Robert H. :Analog matrix inversion [robot kinematics].: *Robotics and Automation, IEEE Journal of* 4.2 (1988): 157-162.

[6] Fa-Long, Luo, and Bao Zheng. :Neural network approach to computing matrix inversion. *Applied Mathematics and Computation* 47.2 (1992): 109-120.

[7] Cichocki, Andrzej, and Rolf Unbehauen. :Neural networks for solving systems of linear equations and related problems. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on* 39.2 (1992): 124-138.

[8] Song, Jingyan, and Yeung Yam. :Complex recurrent neural network for computing the inverse and pseudo-inverse of the complex matrix. *Applied mathematics and computation* 93.2 (1998): 195-205.APA

[9] Ge, Shuzhi Sam, and Chang Chieh Hang. :Structural network modeling and control of rigid body robots. *Robotics and Automation, IEEE Transactions on* 14.5 (1998): 823-827.

[10] Wang, Jeen-Shing, and Yen-Ping Chen. :A fully automated recurrent neural network for unknown dynamic system identification and control. *Circuits and Systems I: Regular Papers, IEEE Transactions on* 53.6 (2006): 1363-1372.

[11] MATLAB:MATLAB:2013, *The MathWorks Inc.*,

[12] G. Wu and J. Wang and J. Hootman.:A recurrent neural network for computing pseudoinverse matrices. *Math. and Computer Modelling* .13 - 21.1994.

[13] Jaeger, Herbert. :The echo state approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Tech. GMD Technical Report* .34.2001.

[14] Lukoševičius, Mantas. :A practical guide to applying echo state networks. *Neural Networks: Tricks of the Trade* .659-686.2012.

Cardinality-constrained Portfolio Optimization using an improved quick Artificial Bee Colony Algorithm

Dit Suthiwong and Maleerat Sodanil

Faculty of Information Technology
King Mongkut's University of Technology North Bangkok, Thailand

Abstract: Portfolio optimization problem is related to investment trade-off between risk and return. The Investors have to make decision to how much invest money in each company stock while in the stock market has so many stock to invest. A combination result yield return differently. This problem is one type of Non-deterministic Polynomial-time hard (NP-hard) problems which recently trend using evolutionary computation (EC) and swarm intelligence based optimization techniques. Artificial Bee Colony (ABC) algorithm is the one which can be modeled to solve Portfolio optimization as well. This paper proposed an improved quick Artificial Bee Colony (iqABC) by modified employed bees phase to choose neighborhood solution using Gbest value. Moreover, the performance of iqABC is better than normal ABC compared with the state of art algorithms performance in portfolio optimization problem.

1 Introduction

The optimizing portfolio selection has been one of the modern financial mathematics. Markowitz [1] had proposed mean-variance model with a largescale quadratic programming problem combining a covariance matrix. More specialize in portfolio optimization there are more precise objectives to ease portfolio management requirement such as limit number of company stocks, minimum transaction lots, sector capitalization constraints to limit purchased amount in batch mode. Cardinality-constrained portfolio optimization is a investor requirement to limit number of company stock which can reduce transaction cost and reduce trading complexity. Chang et al. [2] had proposed a cardinality-constrained mean-variance (CCMV) model – an extended Markowitz portfolio selection model.

As of today, evolutionary computation (EC) and swarm intelligence have become attractive to researcher to find out model solve optimizing portfolio selection included tabu search (TS) [3], simulated annealing (SA) [4], genetic algorithm (GA) [7], particle swarm optimization (PSO) [6], artificial bee colony (ABC) [5, 6], firefly algorithm (FA) [8], quick ABC [8] and so on.

Our proposed solution is a improved algorithm ABC to solve a cardinality-constrained mean-variance model in Chang et al. [2]. In addition, we have chosen Genetic Algorithm (GA), Fireflies Algorithm (FA), Artificial Bee Colony (ABC), and quick ABC (qABC) for experimental comparison. The rest of this paper is organized as follows. In Section 2, we discuss about the cardinality-constrained portfolio optimization problem definition and solution option in present day. In Section 3, we describe the ABC algorithm and introduce introduce an improved version of it. In Section 4, we discuss the computation result. In the last section, we conclude what we get from this study.

2 Cardinality-constrained Portfolio Optimization Problem – Definitions

2.1 The Markowitz Model

The basis of Modern Portfolio Theory (MPT) proposed by Harry Markowitz [1]. By focusing on the aggregate of all the investments among competing financial alternatives. MPT addresses the combining of assets to efficiently achieve a set of return and risk objectives and show it with a weight ratio for each assets. Both objectives return and risk are independent. The return had been calculated as a mean of return in definite period. The risk is calculated from the variance in price fluctuation depend on global impact from market or local impact from their asset itself.

2.2 Mean-variance Model

First MPT was modeled and solved by using quadratic programming. The key point in Mean-Variance (MV) model is to employ the expected returns of a portfolio as the investment return and the variance of returns of then portfolio as the investment risk. This model can be mathematically defined as

$$\text{minimize} \quad \sum_{i=1}^{N}\sum_{j=1}^{N} w_i\, w_j\, \sigma_{ij} \tag{1}$$

$$\text{subject to} \qquad \sum_{i=1}^{N} w_i \, \mu_i = R^* \tag{2}$$

$$\sum_{i=1}^{N} w_i = 1 \tag{3}$$

$$0 \leq w_i \leq 1, \quad i = 1, \ldots, N \tag{4}$$

In equation (1) minimizes the total variance (risk) associated with the portfolio, N represents the number of asset options, w_i and w_j denotes the investment percentage on asset i and j, σ_{ij} denotes the covariance between assets i and j. Equation (2) ensures that the portfolio has an expected return of R^*. Equation (3) ensures that the proportions add to one.

2.3 Cardinality-constrained Mean-variance Model

Chang et al. [2] extended the basic mean-variance model to a cardinality-constrained mean-variance model. The details of the model is defined as follows.

$$\text{minimize} \qquad \lambda \sum_{i=1}^{N} \sum_{j=1}^{N} w_i \, w_j \, \sigma_{ij} - (1-\lambda) \sum_{i=1}^{N} w_i \, \mu_i \tag{5}$$

$$\text{subject to} \qquad \sum_{i=1}^{N} z_i = K \tag{6}$$

$$\sum_{i=1}^{N} w_i = 1 \tag{7}$$

$$\varepsilon_i \, z_i \leq w_i \leq \delta_i z_i, \quad i = 1, \ldots, N \tag{8}$$

$$z_1 \in \{0, 1\}, \quad i = 1, \ldots, N \tag{9}$$

$$0 \leq \lambda \leq 1 \tag{10}$$

In equation (5) λ is critical parameter. It controls the relative importance of the mean return for the investor. When λ is set to 0, it means return of portfolio is being maximized without considering the risk. When λ is set to 1, it means risk of the portfolio is minimized regardless importance of the mean return. Each λ value generates different objective function value which is composed of the mean return and variance. By using λ, a continuous curve can be drawn which called an efficient frontier in the Markowitz's modern portfolio theory.

In equation (6) defines a limit K on desired number of assets in the portfolio while if an asset is selected $z_i = 1$ otherwise $z_i = 0$. Equation (8) ensures that if asset i is held (i. e., $z_i = 1$) its proportional w_i must lie between a given range $[\varepsilon_i, \delta_i]$ specified by investors.

3 Artificial Bee Colony Algorithm

Artificial bee colony (ABC) algorithm [4] is a population-based evolutionary algorithm. In real-life, Honey bees live in colonies and have a social communication. In Generally, there are three groups of bees in the colony who responsible to find and collect food to nectar. They are employed bees, onlookers and scouts.

Employed bees take charge of exploring the solution space to search food sources and when they come back to nectar they perform waggle dances to transmit the nectar information to Onlooker bees. Onlooker bees choose a number of food sources to exploit by make decision from information got from the waggle dances. When there are no improvements to the food sources after several times Scout bees will be sent out randomly to find new brand food sources. If the new food source has better nectar they will memorize the new position and forget the previous one.

3.1 Implementation

The position of a food sources represented to a possible solution to the problem space. The nectar amount of a food source denotes the quality of the associated solution. The number of employed bees is equal to the number of food sources and it is also equal to the number of solutions in the population. The pseudocode is show as below.

```
Set the parameter values
Initialization phase
  Establish solutions randomly
  Evaluate fitness of the population
  While (the stopping criterion not met)
  Employed bee phase
     Random select sites for neighborhood search
     Search the selected sites and evaluate fitness
     Update the efficiency frontier and food sources
  Onlooker bee phase
     Select sites by using the probability based
         on the food sources fitness quality
     Search the selected sites and evaluate fitness
     Update the efficiency frontier and food sources
  Scout bee phase
     Abandon food sources which not improved fitness quality
         over iteration limit
     Scout bees will be assigned to randomly find new food sources
     Update the efficiency frontier and food sources
  End While
End
```

3.2 Initialization Phase

The ABC search algorithm starts with the generation of the initial population. With no prior knowledge about the search space, Random generation is generated. The size of population denotes by NS means the number of food sources and the number of scout bees being dispatched initially.

$$x_{m,i} = l_i + \text{rand}(0,1) \cdot (u_i - l_i) \tag{11}$$

where $x_{m,i}$ is the value of the i, dimension of the m. l_i represents the lower bound and u_i represents the upper bound of the parameter $x_{m,i}$. Then evaluate fitness by using equation (12).

$$fit(x_m) = \begin{cases} 1/(1 + f(x_m)) & \text{if } f(x_m) \geq 0 \\ 1 + \text{abs}(f(x_m)) & \text{if } f(x_m) < 0 \end{cases} \tag{12}$$

3.3 Employed Bees Phase

In this phase, each food source will be further exploited by one and only one employed bee. Then it search for the neighborhood of a target food source. This study focuses on the change of the associated vm values for each asset selected by equation as follows:

$$v_{m,i} = x_{m,i} + \phi_{m,i}(x_{m,i} - x_{k,i}) \tag{13}$$

In equation (13) x_k is a food source selected from neighborhood randomly and i is also a randomly weight parameter. $\phi_{m,i}$ denotes a random number generated from a uniform distribution with the range of $[-1, 1]$. v_m is the new candidate food source which will be calculated for fitness of a solution. Then greedy selection is applied between v_m and x_m. If a food source is not be improved employed bee will increase a certain number of iteration (limit) by one.

3.4 Onlooker Bees Phase

The onlooker bees determine the food sources to search using the probability based on the quality of each food source. The probability of selection pm can be calculated as follows.

$$p_m = \frac{fit(x_m)}{\sum\limits_{m=1}^{SN} fit(x_m)} \tag{14}$$

According to the probability, Onlooker bees choose a food source v_m to exploit by using equation (13) similar to employed bees, and its fitness value is computed. Then, a greedy selection is used to determine between v_m and x_m.

3.5 Scout Bees Phase

If some of food sources cannot be improved through a certain number of iterations (limit). scout bees will be dispatched to explore new brand food sources. This make new solutions randomly generated. If new solution is better than the old solution, then the old food source will be replaced with new food source.

3.6 quick Artificial Bee Colony (qABC) Algorithm

The standard ABC algorithm is powerful. But it still need more improvement in exploitation. Karaboga [8] introduced new equation of onlooker bees phase modified. Comparing to nature of honey bees, Employed bees and Onlooker bees exploit foods in different ways. Employed bees exploit the food source that they visit before. Onlooker bees exploit food source based on communication from employed bee dancing (we called "wagged dance") which will be interpreted for which food sources will be selected. The equation for onlooker bee had been modified as follow.

$$v_{N_m,i}^{best} = x_{N_m,i}^{best} + \phi_{m,i}(x_{N_m,i}^{best} - x_{k,i}) \tag{15}$$

From equation (15), $x_{N_m,i}^{best}$ represents the best solution between the neighbors of x_m and itself N_m. The neighbourhood of individual m is determined by the Euclidean distance between X_{N_m} and the other food sources. The mean Euclidean distance between x_m and the rest food sources is calculated and then compare it with a new parameter r which refers to the ″neighborhood radius″ is added into the parameters of standard ABC algorithm. If a solution which Euclidean distance from x_m is less than the mean Euclidean distance md_m then this food sources could be accepted as a neighbour of x_m as equation (16).

$$md_m = \frac{\sum\limits_{j=1}^{SN} d(m,j)}{SN-1} \tag{16}$$

This solution is similar to nature that onlooker bees selects the region which is centered by the food source x_m. The pseudocode for determine a neighbor of x_m is given as follow.

```
if $d(m,j) \leq rxmd_m$ then $x_j$ is a neighbor of $x_m$, else not
```

3.7 improved quick Artificial Bee Colony (iqABC) Algorithm

Our proposed improved quick ABC algorithm mainly consider Employed bee phased. Inspired by the Particle Swarm Optimization algorithm, we proposed to use global best (gbest) apply as follows.

$$v_{m,i} = x_{m,i} + \phi_{m,i}(x_{m,i} - x_{k,i}) + \varphi_{i,j} \cdot (gbest_i - x_{k,i}) \tag{17}$$

In employed bee phase, we replace equation (13) with (17). Where φ represents a uniformly distributed random number in $[0, 1.5]$, $gbest$ is the current global best solution in the whole swarm, and $gbest_i$ represents the ith variable of $gbest$.

4 Experimental Analysis

Our proposed improved ABC algorithm is compared with two other methods - GA, original ABC. Coding in Matlab R2013a and run on a notebook computer with Intel Core i7-4510U CPU 2.00 GHz and 4.0 GB memory.

4.1 Test Data Sets and Unconstrained Efficient Frontier

We brought trading data from Stock Exchange Thailand from 2008 to 2014 and then select 50 stocks from SET50 listing. We get close price at the end of month and then compare previous to current to get return and calculate co-variance.

The Unconstrained Efficient Frontier (UEF) can be calculated from Quadratic Programming and we use this as benchmark solutions. From the computational results we took 2,000 return values.

4.2 Performance Measures

To compare performance we apply the multi-objective optimization performance measures concept to measure our algorithm performance. We focus on accuracy, diversity and coverage performance aspects. In term of accuracy, We apply the $D1_R$ value [9]. The $D1_R$ measure refers mainly the average minimum distance from each point in the reference Pareto front to the approximated Pareto front. The small value of the $D1_R$ measure means the better of the non-dominated solutions. In equation (18) computes the average of the minimum distance. Note that the smaller the $D1_R$ value is, the better an approximated Pareto frontier is.

$$D1_R = \frac{1}{|X_{ref}|} \sum_{x^* \in X_{ref}} min\{d_{x^*x} | x \in X_{app} \tag{18}$$

$$d_{x^*x} = \sqrt{(f_r(x^*) - f_{er}(x))^2 + (f_{er}(x^*) - f_{er}(x))^2} \tag{19}$$

where X_{ref} denotes the reference Pareto front and $|X_{ref}|$ is the number of non-dominated solutions in X_{ref}. The algorithm generates the approximated Pareto efficiency frontier, denotes X_{app}. In equation (18) and (19), x^* is a solution in X_{ref} and x is a solution in X_{app}. In equation (19) calculates the Euclidean distance between x^* and x in the space of objective functions by using $f_r(.)$ represents the corresponding function value of risk objective for a given solution and $f_{er}(.)$ denotes the associated function of the expected return objective for a non-dominated solution.

In term of diversity and convergence, the Δ measure (a spread metric) [10] are adopted in this study. The spread metric Δ evaluate the diversity property of the approximated Pareto efficiency frontier by an algorithm.

$$\Delta = \frac{d_f + d_l + \sum_{i=1}^{|X_{app}|-1} |d_i - \overline{d}|}{d_f + d_l + (|X_{app}| - 1)\,\overline{d}} \tag{20}$$

where d_f and d_l denote the Euclidean distances between the extreme solutions in the reference frontier and the boundary solutions in the approximated frontier. d_i represents the Euclidean distance between two successive solutions in the approximated efficiency frontier while $\overline{d}$ is an average of those d_i values. $|X_{app}|$ denotes the number of non-dominated solutions in an approximated efficiency frontier. The number of d_i values equal to $|X_{app}| - 1$. The larger Δ value shows that the approximated Pareto efficiency may not distribute evenly in the objective function space, and the search may stuck only in certain regions.

In term of coverage, we simply compare number of the reference Pareto front with the number of non-dominated solutions.

$$\text{Non-dominated Point Ratio } = \frac{|X_{app}|}{|X_{ref}|} \tag{21}$$

where X_{ref} denotes the reference Pareto front and $|X_{ref}|$ is the number of non-dominated solutions in X_{ref} and $|X_{app}|$ is the number of non-dominated solutions from algorithms.

4.3 Computational Results

The best parameter values were obtained from the preliminary experiments. The parameter of ABC, qABC and iqABC are set as follow: Number of food sources $(NS) = N/4$ and its threshold value is 50 (iterations). The setting mainly follows Chang et al. in order to conduct a fair comparison among competing algorithms. The number of stocks selected is 10, the upper and lower bounds of the weights are equal to 1.0 to 0.01 accordingly. The setting of the λ values start from zero, and increase at 0.02 until reaches 1. The maximum number of evaluations is set as a stopping criterion for each λ value. All algorithms are run ten times, and calculating for the average to compare performance.

The result of comparisons has been shown in table 1. The result point out that iqABC performs the best with the lowest $D1_R$ value. While comparing with the GA, FA, previous ABC and qABC is also able to achieve 85.14%, 85.66%, 60.89% and 9.09% improvement. It shows that iqABC provides a superior solution accuracy quality. For Δ result, GA is the best while iqABC give the result with most wide range of non-dominated solution when compare to other algorithms. And for the last measure, Non-dominated Point Ratio, iqABC shows that it can

Table 1: Comparison on the $D1_R$, Δ and Non-dominated Point Ratio values

	Algorithms				
Measures	GA	FA	ABC	qABC	iqABC
$D1_R$	0.0471	0.0488	0.0179	0.0077	**0.0070**
Δ	**0.9105**	1.0893	1.2742	1.3745	1.4107
Non-dominated Point Ratio	0.0092	0.1553	0.1963	0.1966	**1.1967**

deliver more non-dominated point for pareto front compare to others while GA is the worst.

5 Conclusion

The Artificial Bee Colony Algorithm had shown its powerful to solve the portfolio optimization problem. Especially for Cardinality-constrained portfolio optimization problem which is NP-hard problem. For qABC, it also shows improvement for accuracy and convergence. And we have proposed iqABC which developed based on qABC. By put concept Gbest direction in employed bee phase, we get the more accuracy result and more non-dominated points compare to others.

For future works, in algorithm development topic about convergence should be developed more. In term of portfolio management, more type of risk models [3] should be applied to find out more better return of investment during maintain low risk.

References

[1] Markowitz H.: Portfolio selection, *Journal of Finance*, 7, 77-91, 1952

[2] Chang, T. J., Meade, N., Beasley, J. E., Sharaiha, Y. M.: Heuristics for cardinality constrained portfolio optimization, *Computers and Operations Research*, 27, 13, 1271-1302, 2000

[3] Chang, Tun-Jen, Sang-Chin Yang, and Kuang-Jung Chang,: Portfolio Optimization Problems in Different Risk Measures Using Genetic Algorithm, *Expert Systems with Applications*, 36, 1052937, 2009

[4] D. Karaboga,: An Idea Based on Honey Bee Swarm for Numerical Optimization, *Technical Report-TR06*, 1-10, 2005

[5] Chen A.H.L., Yun-Chia Liang, and Chia-Chien Liu: An Artificial Bee Colony Algorithm for the Cardinality-Constrained Portfolio Optimization Problems, In: *IEEE Congress on Evolutionary Computation (CEC)*, 1-8, 2012

[6] Chen, A.H.L., Yun-Chia Liang, and Chia-Chien Liu,: Portfolio Optimization Using Improved Artificial Bee Colony Approach, In: *IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr)*, 60–67, 2013

[7] Soam, V., L. Palafox, and H. Iba,: Multi-Objective Portfolio Optimization and Rebalancing Using Genetic Algorithms with Local Search, In: *IEEE Congress on Evolutionary Computation (CEC)*, 1–7, 2012

[8] Karaboga, Dervis, and Beyza Gorkemli: A Quick Artificial Bee Colony (qABC) Algorithm and Its Performance on Optimization Problems, *Applied Soft Computing*, 23, 227-38, 2014

[9] Ishibuchi, H., T. Yoshida, and T. Murata,: Balance between Genetic Search and Local Search in Memetic Algorithms for Multiobjective Permutation Flowshop Scheduling, In: *IEEE Transactions on Evolutionary Computation*, 204-23, 200

[10] Deb, K., A. Pratap, S. Agarwal, and T. Meyarivan,: A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II, In: *IEEE Transactions on Evolutionary Computation*, 182-97, 2002

Centroid Terms and their Use in Natural Language Processing

Mario M. Kubek and Herwig Unger

Chair of Communication Networks
FernUniversität in Hagen, Germany

Abstract: The calculation of semantic similarities between text documents plays an important role in automatic text processing. For example, algorithms to topically cluster and classify texts heavily rely on this information. Standard methods for doing so are usually based on the bag-of-words model and thus return only rough estimations regarding the relatedness of texts. Moreover, they are unable to find generalising terms or abstractions describing the textual contents. Therefore, a new graph-based method to determine centroid terms as text representatives will be introduced. It is shown, that – among further application scenarios – this method is able to compute the similarity of texts even if they have no terms in common. In first experiments, its results and advantages will be discussed in detail.

1 Motivation

After only a few lines of reading, a human reader is able to determine which category of texts and which abstract topic category a given document belongs to. This is a strong demonstration of how well and fast the human brain, especially the human cortex, can process and interpret data. It is able to not only understand the meaning of single words – as representations of real-world entities – but a certain composition of them [1], too.

In many text mining applications, the topical grouping of texts and terms/words contained in them are common tasks. In order to group semantically related terms, unsupervised topic modeling techniques such as LDA [2] have been successfully applied. This technique tries to infer word clusters from a set of documents based on the assumption that words from the same topic are likely to appear next to each other and therefore share a related meaning. Here, deep and computationally expensive (hyper)parameter estimations are carried out and for each word, the probability to belong to specific topic is computed in order to create those constructions. The graph-based Chinese Whispers algorithm [3] is another interesting clustering technique that can be used in the field of natural language problems, especially to semantically group terms. It is usually applied on undirected semantic graphs that contain statistically significant term relations found in texts.

Also, it is usual to apply the k-means algorithm [4] to group terms. For this purpose, it necessary to determine their semantic distance. Here, several methods can be applied. The frequency of the co-occurrence of two terms in close proximity (in a window of n words or on sentence level) is a first indication for their semantic distance. Terms that frequently co-occur together are usually semantically related. Several graph-based distance measures [7, 8] consult manually created semantic networks such as WordNet [5], a large lexical database containing semantic relationships for the English language that covers relations like polysemy, synonymy, antonymy, hypernymy and hyponymy (i.e. more general and more specific concepts), as well as part-of-relationships. These measures apply shortest path algorithms or take into account the depth of the least common subsumer concept (LCS) to determine the closest semantic relationship between two given input terms or concepts. It is also common to measure the similarity of term contexts [6] that contain terms that often co-occur with the ones in question. Technically, these contexts are realised as term vectors following the bag-of-words model.

The same approach is applied when the semantic similarity or distance of any two documents should be determined. Here, the term vectors to be compared contain the texts' characterising terms and their score (typically, a TF-IDF-based statistic [9] is used) as a measure for their importance. The similarity of two term vectors can be determined using the cosine similarity measure or by calculating the overlap of term vectors, e.g. using the Dice coefficient [10]. The commonly used Euclidean distance and the Manhattan distance are further examples to measure the closeness of term vectors at low computational costs.

However, in some cases, these measures do not work correctly (with respect to human judgement), mostly if different people write about the same topic but are using a completely different vocabulary for doing so. The reason for this circumstance can be seen in the isolated view of the words found in documents to be compared without including any relation to the vocabulary of other, context-related documents. Moreover, short texts as often found in posts in online social networks or short (web) search queries with a low number of decriptive terms can therefore often not be correctly classified or disambiguated. Another disadvantage is that these measures cannot find abstractions or generalising terms by just analysing the textual data provided. For this purpose, static lexical databases such as WordNet [5] must be consulted as a reference. Despite their usefulness, these resources are – in contrast to the human brain – not able to learn about new concepts and their relationships.

In order to address these problems, this article presents a new graph-based approach to determine centroid terms of text documents. It is shown that those terms can actually represent text documents in automatic text processing, e.g. to determine their semantic distances. In the next section, the fundamentals of this method are presented. Afterwards, section 3 describes its mathematical and technical details. Section 4 proves the validity of this approach by explaining the results of first experiments. In section 5, the method's working principles and advantages are discussed. Section 6 presents numerous application scenarios for it in the fields of text mining and information retrieval while also elaborating on technological aspects of its practical implementation. Section 7 summarises the article and suggests further application fields of the introduced method.

2 Fundamentals

For the approach presented herein, co-occurrences and co-occurrence graphs are the basic means to obtain more detailed information about text documents than term frequency vectors etc. could ever offer. The reason for this decision is that co-occurrence graphs are able to accumulate a certain knowledge obtained from a few selected or all documents of a text corpus while (at least to some extent) maintaining the semantic connection of terms found in them.

Two words w_i and w_j are called *co-occurrents*, if they appear together in close proximity in a document D. The most prominent kinds of such co-occurrences are word pairs that appear as immediate neighbours or together in a sentence. A *co-occurrence graph* $G = (W, E)$ may be obtained, if all words of a document or set of documents W are used to build its set of nodes which are then connected by an edge $(w_a, w_b) \in E$ if $w_a \in W$ and $w_b \in W$ are co-occurrents. A weight function $g((w_a, w_b))$ indicates, how significant the respective co-occurrence is in a document. If the significance value is greater than a pre-set threshold, the co-occurrence can be regarded as significant and a semantic relation between the words involved can often be derived from it. Commonly used significance measures are the Dice coefficient [10], the mutual information measure [11], the Poisson collocation measure [12] and the log-likelihood ratio [13].

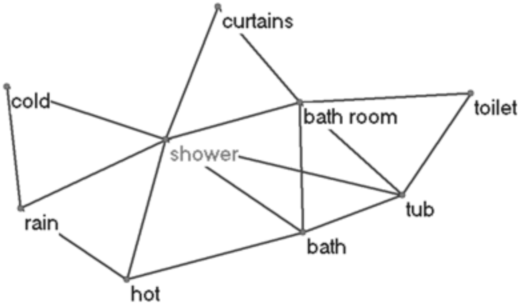


Fig. 1: A co-occurrence graph for the word "shower"

A co-occurrence graph – similarly to the knowledge in the human brain – may be built step by step over a long time taking one document after another into consideration. From the literature [6] and own experiments (see Figure 2) it is known that the out-degrees of nodes in co-occurrence graphs follow a power-law distribution and the whole graph exhibit small-world properties with a high clustering coefficient as well as a short average path length between any two nodes. This way, a co-occurence graph's structure also reflects the organisation of human lexical knowledge.

The use of the immediate neighbourhood of nodes in a co-occurrence graph has been widely considered in literature, e.g. to cluster terms [3] and to determine the global context (vector) of terms in order to evaluate their similarity [6] or to derive paradigmatic relations between them [14]. In the authors' view, indirect neighbourhoods of terms in co-occurrence graphs (nodes that can be reached only using two or more edges from a node of interest) and the respective paths with a length ≥ 2 should be considered as well as indirectly reachable nodes may still be of topical relevance, especially when the co-occurence graph is large. The benefit of using such nodes/terms in co-occurrence graphs has already been shown by the authors for the expansion of web search queries using a spreading activation technique applied on local and user-defined corpora [15]. The precision of web search results can be noticably improved when taking those terms into account, too.

The field of application of indirect term neighbourhoods in co-occurrence graphs shall be extended in the next section by introducing an approach to determine centroid terms of text documents that can act as their representatives in further text processing tasks. These centroid terms can be regarded as the texts'

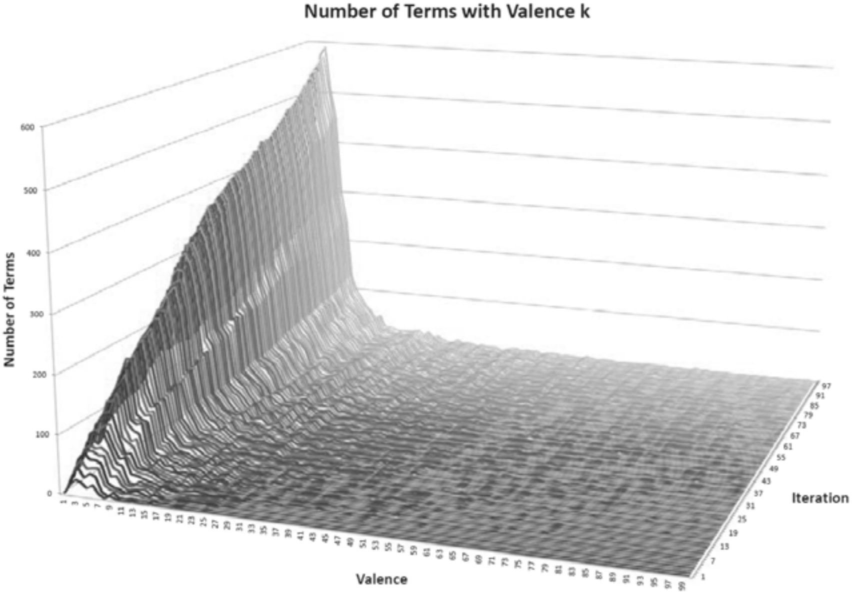


Fig. 2: Distribution of out-degrees in a co-occurrence graph over time

topical centers of interest (a notion normally used to describe the part of a picture that attracts the eye and mind) that the authors' thoughts revolve around.

3 Finding Centroid Terms

In physics, complex bodies consisting of several single mass points are usually represented and considered by their so-called center of mass, as seen in Figure 3. The distribution of mass is balanced around this center and the average of the weighted coordinates of the distributed mass defines its coordinates and therefore its position.

For discrete systems, i.e. systems consisting of n single mass points $m_1, m_2, .., m_i$ in a $3D-$space at positions $\vec{r}_1, \vec{r}_2, .., \vec{r}_i$, the center of mass $\vec{r}_s$ can be found by

$$\vec{r}_s = \frac{1}{M} \sum_{i=1}^{n} m_i \vec{r}_i, \tag{1}$$

whereby

$$M = \sum_{i=1}^{n} m_i. \tag{2}$$

Usually, this model simplifies calculations with complex bodies in mechanics by representing the whole system by a single mass at the position of the center of mass. Exactly the same problem exists in automatic text processing: a whole text shall be represented or classified by one or a few single, descriptive terms which must be found.

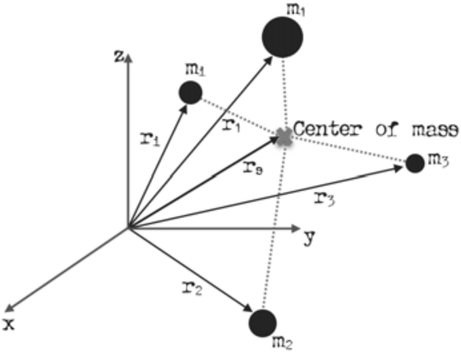


Fig. 3: The physical center of mass

To adapt the situation for this application field, first of all, a *distance d* shall be introduced in a co-occurrence graph G. From literature it is known that two words are semantically close, if $g((w_a, w_b))$ is high, i.e. they often appear together in a sentence or in another predefined window of n words. Consequently, a distance $d(w_a, w_b)$ of two words in G can be defined by

$$d(w_a, w_b) = \frac{1}{g((w_a, w_b))}, \tag{3}$$

if w_a and w_b are co-occurrents. In all other cases (assuming that the co-occurrence graph is connected[1]) there is a shortest path $p = (w_1, w_2), (w_2, w_3), .., (w_k, w_k + 1)$ with $w_1 = w_a$, $w_{k+1} = w_b$ and $w_i, w_{i+1} \in E$ for all $i = 1(1)k$ such that

$$d(w_a, w_b) = \sum_{i=1}^{k} d((w_i, w_{i+1})) = MIN, \tag{4}$$

[1] This can be achieved by adding a sufficiently high number of documents to it during its building process.

whereby in case of a partially connected co-occurrence graph $d(w_a, w_b) = \infty$ must be set. Note, that differing from the physical model, there is a distance between any two words but no direction vector, since there is no embedding of the co-occurrence graph in the $2-$ or $3-$dimensional space. Consequently, the impact of a word depends only on its scalar distance.

In continuation of the previous idea, the distance between a given term t and a document D containing N words $w_1, w_2, .., w_N \in D$ that are reachable from t in G can be defined by

$$d(D,t) = \frac{\sum_{i=1}^{N} d(w_i, t)}{N}, \tag{5}$$

i.e. the average sum of the lengths of the shortest paths between t and all words $w_i \in D$ that can be reached from it. Note that -differing from many methods found in literature- it is not assumed that $t \in D$ holds! Also, it might happen in some cases that the minimal distance is not uniquely defined, consequently a text may have more than one centroid term (as long as no other methods decide which one is to use). In order to define the centroid-based distance ζ between any two documents D_1 and D_2, let t_1 be the center term or *centroid term of* D_1 with $d(D_1, t_1) = MIN$. If at the same time t_2 is the centroid term of D_2,

$$\zeta(D_1, D_2) = d(t_1, t_2) \tag{6}$$

can be understood as the semantic distance ζ of the two documents D_1 and D_2. In order to obtain a similarity value instead,

$$\zeta_{sim}(D_1, D_2) = \frac{1}{1 + \zeta(D_1, D_2)} \tag{7}$$

can be applied.

It is another important property of the described distance calculation that documents regardless of their length as well as single words can be assigned a centroid term by one and the same method in a unique manner. The presented approach relies on the preferably large co-occurrence graph G as its reference. It may be constructed from any text corpus in any language available or directly from the sets of documents whose semantic distance shall be determined. The usage of external resources such as lexical databases or reference corpora is common in text mining: as an example, the so-called difference analysis [6, 16] which measures the deviation of word frequencies in single texts from their frequencies in general usage (a large topically well-balanced reference corpus is needed for this purpose) is an example for it. The larger the deviation is, the more likely it is that a term or keyword of a single text has been found. Furthermore, the presented distance measure is not only based on a physical analogon and bears (at least to a certain extent) resemblance to the well-known difference analysis as discussed, the measure's approach is brain-inspired, too. Further considerations in this respect will be discussed in section 5.

In the following section, the quality and properties of the centroid terms and the new centroid-based distance measure shall be investigated and discussed.

4 First Experiments

For all of the exemplary experiments (many more have been conducted) discussed herein, linguistic preprocessing has been applied on the documents to be analysed whereby stop words have been removed and only nouns (in their base form), proper nouns and names have been extracted. In order to build the undirected co-occurrence graph G (as the reference for the centroid distance measure), co-occurrences on sentence level have been extracted. Their significance values have been determined using the Dice coefficient [10]. The particularly used sets of documents will be described in the respective subsections[2].

4.1 Centroids of Wikipedia Articles

As the centroid terms are the basic components for the centroid-based distance measure, it is useful to get a first impression of their quality in terms of whether they are actual useful representatives of documents. Table 1 therefore presents the centroid terms of 30 English Wikipedia articles. The corpus used to create the reference co-occurrence graph G consisted of 100 randomly selected articles (including the mentioned 30 ones) from an offline English Wikipedia corpus from `http://www.kiwix.org`. It can be seen that almost all centroids properly represent their respective articles.

4.2 Comparing Similarity Measures

In order to evaluate the effectiveness of the new centroid-based distance measure, its results will be presented and compared to those of the cosine similarity measure while the same 100 online news articles from the German newspaper "Süddeutsche Zeitung" from the months September, October and November of 2015 have been selected (25 articles from each of the four topical categories 'car',

[2] Interested researchers can download these sets (1,3 MB) from `http://www.docanalyser.de/cd-corpora.zip`

Table 1: Centroids of 30 Wikipedia articles

Title of Wikipedia Article	Centroid Term
Art competitions at the Olympic Games	sculpture
Tay-Sachs disease	mutation
Pythagoras	Pythagoras
Canberra	Canberra
Eye (cyclone)	storm
Blade Runner	Ridley Scott
CPU cache	cache miss
Rembrandt	Louvre
Common Unix Printing System	filter
Psychology	psychology
Religion	religion
Universe	shape
Mass media	database
Rio de Janeiro	sport
Stroke	blood
Mark Twain	tale
Ludwig van Beethoven	violin
Oxyrhynchus	papyrus
Fermi paradox	civilization
Milk	dairy
Corinthian War	Sparta
Health	fitness
Tourette syndrome	tic
Agriculture	crop
Finland	tourism
Malaria	disease
Fiberglass	fiber
Continent	continent
United States Congress	Senate
Turquoise	turquoise

'travel', 'finance' and 'sports' have been randomly chosen) for this purpose. As the cosine similarity measure operates on term vectors, the articles' most important terms along with their scores have been determined using the extended PageRank [17] algorithm which has been applied on their own separate (local) co-occurrence graphs (here, another term weighting scheme such as a TF-IDF variant [9] could have been used as well). The cosine similarity measure has then been applied on all pairs of the term vectors. For each article A, a list of the names of the remaining 99 articles has been generated and arranged in descending order according to their cosine similarity to A. A most similar article can therefore be found at the top of this list.

In order to apply the new centroid distance measure to determine the articles' semantic distance, for each article, its centroid term has been determined with the help of the co-occurrence graph G using formula 5. The pairwise distance between all centroid terms of all articles in G has then been calculated. Additionally, to make the results of the cosine similarity measure and the centroid distance measure comparable, the centroid distance values have been converted into similarity values using formula 7.

The exemplary diagram in Figure 4 shows for the reference article ("Abgas-Skandal – Schummel-Motor steckt auch in Audi A4 und A6") its similarity to the 50 most similar articles. The cosine similarity measure was used as the reference measure. Therefore, the most similar article received rank 1 using this measure (lower bars). Although the similarity values of the two measures seem uncorrelated, it is recognisable that especially the articles with a low rank (high similarity) according to the cosine similarity measure are generally regarded as similar by the centroid distance measure, too. In case of Figure 4, the reference article dealt with the car emissions scandal (a heavily dicussed topic in late 2015). The articles at the ranks 3 ("Abgas-Affäre – Volkswagen holt fünf Millionen VWs in die Werkstätten"), 7 ("Diesel von Volkswagen – Was VW-Kunden jetzt wissen müssen") and 12 ("Abgas-Skandal – Was auf VW- und Audi-Kunden zukommt") according to the cosine similarity measure have been considered most similar by the centroid distance measure, all of which were indeed related to the reference article. The strongly related articles at the ranks 1, 4, 6 and 9 have been regarded as similar by the centroid distance measure, too. In many experiments, however, the centroid distance measure considered articles as similar although the cosine similarity measure did not.

Here, another implicit yet important advantage of the new centroid distance measure becomes obvious: two documents can be regarded as similar although their wording differs (the overlap of their term vectors would be small or even empty and the cosine similarity value would be very low or 0). The article at rank 49 ("Jaguar XF im Fahrbericht – Krallen statt Samtpfoten") is an example

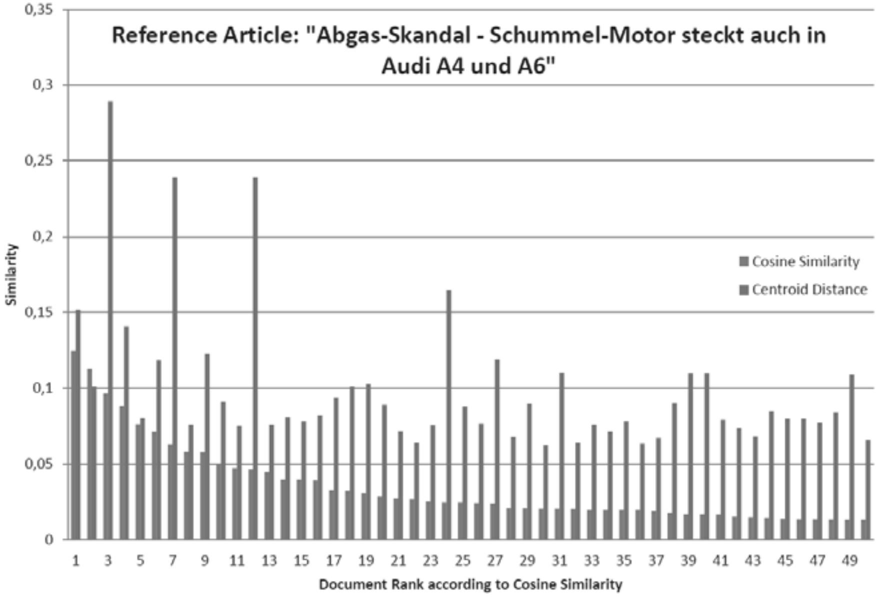


Fig. 4: Cosine similarity vs. centroid distance (topic: car emissions scandal)

for such a case. The centroid distance measure uncovered a topical relationship to the reference article, as both texts are car-related and deal with engine types.

Figure 5 depicts another case of this kind: the article with rank 29 received the highest similarity score from the centroid distance measure. A close examination of this case revealed that the centroids of the reference article ("Deutschland – Ausgebucht") and the article in question ("Briefporto – Post lässt schon mal 70-Cent-Marken drucken") are located close to each other in the reference co-occurrence graph. The reference article's main topic was on financial investments in the German hotel business and the article at rank 29 dealt with postage prices of Deutsche Post AG. Both articles also provided short reports on business-related statistics and strategies.

4.3 Searching for Text Documents

The previous experiments suggest that the centroid distance measure might be applicable to search for text documents, too. In this sense, one might consider a query as a short text document whose centroid term is determined as described

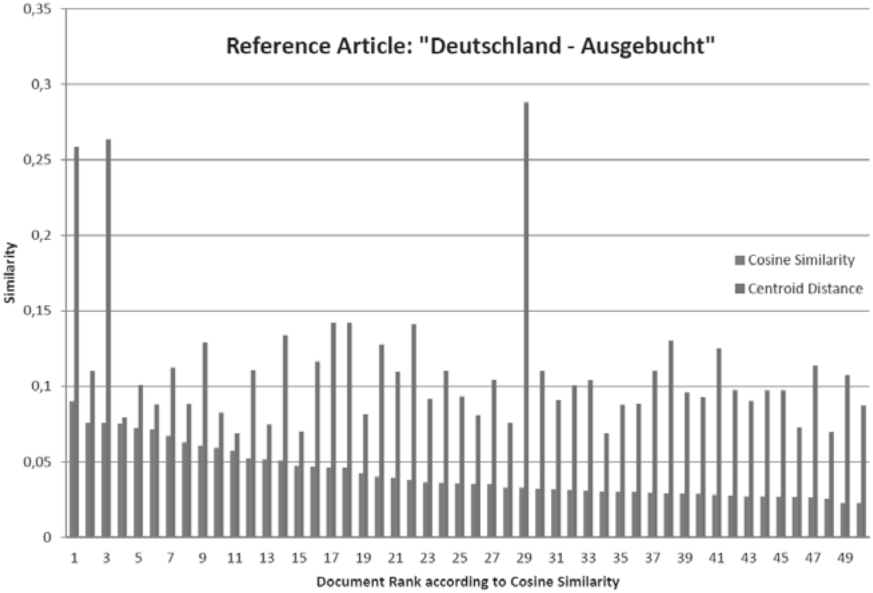


Fig. 5: Cosine similarity vs. centroid distance measure (topic: business-related statistics and strategies)

before and the *k* documents whose centroid terms are closest to the query's centroid term are returned as matches. These *k* nearest neighbours are implicitly ranked by the centroid distance measure, too. The best matching document's centroid term has the lowest distance to the query's centroid term.

The following two tables show for two exemplary queries ”VW Audi Abgas” (centroid term: ”Seat”) and ”Fußball Geld Fifa” (centroid term: ”Affäre”) their respective top 10 articles from the German newspaper ”Süddeutsche Zeitung” along with their own centroid terms whereby the distances from the queries' centroid terms to all 100 mentioned articles' centroid terms in the co-occurrence graph *G* have been calculated.

It can be seen that most of the documents can actually satisfy the information need expressed by the queries. This kind of search will, however, not return exact matches as known from the popular keyword-based web search. Instead, documents will be returned that are in general topically related to the query. As the query and the documents to be searched for are both represented by just one centroid term, an exact match is not possible when applying this approach.

Table 2: Top 10 documents for the query "VW Audi Abgas" (Seat)

Filename of News Article	Centroid Term
auto_abgas-skandal-vw-richtet...	Audi
geld_aktien-oeko-fonds-schmeissen-volkswagen-raus	Ethik
auto_bmw-siebener-im-fahrbericht-luxus-laeuft	S-Klasse
auto_abgas-affaere-volkswagen-ruft...	Schadstoffausstoß
auto_abgas-skandal-schummel-motor...	Schadstoffausstoß
geld_briefporto-post-laesst-schon...	Marktanteil
auto_abgas-skandal-was-auf-vw-und-audi...	EA189
auto_abgas-skandal-acht-millionen-vw-autos...	Software
auto_diesel-von-volkswagen-was-vw-kunden...	Motor
auto_abgas-affaere-schmutzige-tricks	Motor

Table 3: Top 10 documents for the query "Fußball Geld Fifa" (Affäre)

Filename of News Article	Centroid Term
sport_affaere-um-wm-mehr-als-nur-ein-fehler	Fifa
sport_angreifer-von-real-madrid-karim-benzema...	Videoaufnahme
sport_affaere-um-wm-vergabe-zwanziger-schiesst...	Zwanziger
sport_affaere-um-fussball-wm-wie-beckenbauers...	Organisationskomitee
sport_affaere-um-wm-zwanziger-es-gab-eine...	Organisationskomitee
sport_affaere-um-wm-vergabe-zwanziger-legt...	Gerichtsverfahren
sport_affaeren-um-wm-vergaben-die-fifa...	Zahlung
sport_affaere-um-wm-netzer-wirft-zwanziger...	Fifa-Funktionär
geld_ehrenamt-fluechtlingshilfe-die-sich...	Sonderausgabe
sport_affaere-um-wm-wie-zwanziger-niersbach...	Präsident

However, this method can still be of use when a preferably large set of topically matching documents is needed. This kind of recall-oriented search is of interest e.g. for people that want to get an overview of a topic or during patent searches when exact query matches might lower the chance of finding possibly relevant documents that nevertheless do not contain all query terms but related terms instead. A typical precision-oriented search would then be harmful. In these cases, a search system would first determine documents that contain the input query terms using its inverted index and then rank these documents by computing e.g. their term vectors' cosine similarity with the query. That means that a highly relevant document will contain (almost) all query terms.

In order to optimise both recall using the centroid distance measure and also the precision for the k top documents (precision@k) using a variant of the aforementioned procedure, it might be sensible to calculate a combined rank that factors in the rankings of both approaches. Also, it is imaginable to use the centroid distance measure (as a substitute for the Boolean model) to pre-select those documents that are in a second step ranked according to the cosine similarity measure. Still, other well-known techniques such as expanding queries using highly related and synonymous terms [15] are suitable options to increase recall as well. More experiments in this regard taking all these approaches into account will be conducted.

Also, in the experiments presented herein, mostly topically homogeneous texts (except for the book analyses) have been used in order to demonstrate the validity of the centroid distance measure and the role of centroid terms as text representatives. In future experiments, it will be interesting to evaluate the effectiveness of this approach when it is applied on more topically heterogeneous documents.

4.4 Analysing Full Books

Additionally, full books (not in combination with other texts) have been analysed to determine their centroid terms. In these cases, the books' own co-occurrence graphs G have been used to determine their important terms and to find their respective centroid terms (one for each book). In case of the English King James version of the Holy Bible, the centroid term determined that has the shortest average distance in the book's (almost fully connected) graph G to all other 7211 reachable terms is 'Horeb'. This experiment has been repeated while only using the k ($k = 25, 50, 75 \ldots$) most frequent terms for this purpose. Here, besides 'Horeb' and others, the terms 'God' and 'gladness' have been determined as the centroid terms. It is to be pointed out that all of these terms have a low distance to each other in the co-occurrence graph G, meaning they are all good representations of the text no matter what actual centroid term is used for further considerations and applications. This also shows, that it is sufficient to take into account only a few prominent terms of a text in order to determine its centroid term in the co-occurrence graph G while at the same time the algorithm's execution time is drastically lowered.

5 Discussion

The presented approach of using a reference co-occurrence graph to determine the semantic distance of texts is brain-inspired, too. Humans naturally, unconsciously and constantly learn about the entities/objects and their relationships surrounding them and build representations of these perceptions in form of concept maps as well as their terminologies in their minds. New experiences are automatically and in a fraction of a second matched with those previously learned. The same principle is applied when using the centroid distance measure. An incoming text A – regardless of whether it was previously used to construct the co-occurrence graph G or not – whose centroid term shall be found, must at least partially be matched against G. In this sense, G takes on the role of the brain and acts as a global and semantic knowledge base. The only prerequisite is that the graph G must contain enough terms that the incoming text's terms can be matched with. However, it is not necessary to find all of A's terms in G for its at least rough topical classification. The human brain does the same. A non-expert reading an online article about biotechnology may not fully understand its terminology, but can at least roughly grasp its content. However, in doing so, this person will gradually learn about the new concepts, a process that is not yet carried out in the herein presented approach. In later publications, the inclusion of this process will be examined.

In order to find proper centroid terms for documents whose topical orientation is unknown, it is important to construct the co-occurrence graph G from a preferably large amount of texts covering a wide range of topics. That is why, in the previous section, the 100 documents to build the respective corpora have been randomly chosen to create G as a topically well-balanced reference. However, the authors assume that topically oriented corpora can be used as a reference when dealing with documents whose terminology and topical orientation is known in advance, too. This way, the quality of the determined centroid terms should increase as they are expected to be better representations for the individual texts' special topical characteristics. Therefore, a more fine-grained automatic classification of a text should be possible. Further experiments are planned to investigate this assumption.

The bag-of-words model that e.g. the cosine similarity measure solely relies on is used by the centroid-based measure as well, but only to the extend that the entries in the term vectors of documents are used as anchor points in the reference co-occurrence graph G (to 'position' the documents in G) in order to determine their centroid terms. Also, it needs to be pointed out once again that a document's centroid term does not have to occur even once in it. In other words, a centroid term can represent a document, even when it is not mentioned in it.

However, as seen in the experiments, while the cosine similarity measure and the centroid distance measure both often regard especially those documents as similar that actually contain the same terms (their term vectors have a significantly large overlap), one still might argue that both measures can complement each other. The reason for this can be seen in their totally different working principles. While the cosine similarity measure will return a high similarity value for those documents that contain the same terms, the centroid distance measure can uncover a topical relationship between documents even if their wording differs. This is why it might be sensible to combine both approaches in a new measure that factors in the results of both methods. Additional experiments in this regard will be conducted.

Additionally, the herein presented experiments have shown another advantage of the centroid distance measure: its language-independence. It relies on the term relations and term distances in the reference co-occurrence graph G that has been naturally created using text documents of any language.

6 Application Scenarios

The presented centroid distance measure can naturally be applied by text mining algorithms that topically cluster or classify documents. These algorithms make heavy use of similarity and distance measures in order to group semantically similar documents or terms. Here, the new measure can be perfectly applied as an alternative to the well-known measures mentioned above. It will be especially useful, when it comes to grouping topically documents that – despite their topical relatedness – have only a limited amount of terms in common.

However, as shown in the experiments, search applications can make use of this measure, too. Also in this case, documents can be found that do not even share a single query term, yet are highly relevant to the query. Even so, as users are often interested in documents that actually contain the entered query terms but make mistakes in finding the right terms for their information needs, it might be sensible to expand the original query terms with the determined centroid term along with some of its neighbouring terms in the co-occurrence graph G. Matching documents containing these terms could be ranked in reverse order of their similarity to the expanded query. By using this approach, the search results' recall and precision are both expected to increase as common terms in a topical field (the included centroid term and/or its immediate neighbours) as well as the original query terms are used to find matching documents. This approach will be examined and discussed in further publications.

Interactive search applications such as "DocAnalyser" [18] that aim at helping users to find topically similar and related documents in the World Wide Web could benefit from employing the centroid distance measure, too. Starting with a document of the user's interest, the application could determine the document's centroid term as described before and send this term (to increase the search results' recall) as well as some characteristic terms of the document as an automatically formulated query to a web search engine which will (hopefully) return relevant documents.

From the technological point of view, it becomes obvious that it is necessary to be able to manage large graph structures efficiently and effectively. Graph databases such as Neo4j [19] are specifically designed for this purpose. They are also well-suited to support graph-based text mining algorithms [20]. This kind of databases is not only useful to solely store and query the herein discussed co-occurrence graphs, with the help of the property graph model of these databases, nodes (terms) in co-occurrence graphs can be enriched with additional attributes such as the names of the documents they occur in as well as the number of their occurrences in them, too. Also, the co-occurrence significances can be persistently saved as edge attributes. Graph databases are therefore an urgently necessary tool as a basis for future and scalable text mining solutions.

7 Conclusion

A new physics-inspired method has been introduced to determine centroid terms of particular text documents which are strongly related to them and yet do not need to occur in them. As text representatives, these terms are useful to determine the semantic distance and similarity of text documents. Especially, texts with similar topics yet different descriptive terms, may be classified more precisely than by commonly used measures. As the text length's influence does not play a role in doing so, even short texts or (search) queries may be matched with other texts using the same approach. It may therefore be applied in future (decentralised) search engines and text clustering solutions.

References

[1] Hawkins, J., Blakeslee, S.: *On Intelligence*, Times Books, New York, NY, USA, 2004

[2] Blei, D. M., Ng, A. Y., Jordan, M. I.: Latent Dirichlet Allocation, In: *The Journal of Machine Learning Research*, Vol. 3, pp. 993–1022, 2003

[3] Biemann, C.: Chinese Whispers: An Efficient Graph Clustering Algorithm and its Application to Natural Language Processing Problems, In: *Proceedings of the HLT-NAACL-06 Workshop on Textgraphs-06*, pp. 73–80, ACL, New York City, 2006

[4] MacQueen, J. B.: Some Methods for Classification and Analysis of Multivariate Observations, In: *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 281–297, University of California Press, 1967

[5] Miller, G. A.: WordNet: A Lexical Database for English, In: *Communications of the ACM*, Vol. 38, Issue 11, pp. 39–41, Nov. 1995

[6] Heyer, G., Quasthoff, U., Wittig, T.: *Text Mining - Wissensrohstoff Text*, W3L Verlag Bochum, 2006

[7] Budanitsky, A., Hirst, G.: Evaluating WordNet-based measures of semantic distance, In: *Computational Linguistics*, Vol. 32, Issue 1, pp. 13–47, 2006

[8] Resnik, P.: Using information content to evaluate semantic similarity, In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pp. 448–453, Montreal, Canada, 1995

[9] Baeza-Yates, R. A., Ribeiro-Neto, B.: *Modern Information Retrieval*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999

[10] Dice, L. R.: Measures of the amount of ecologic association between species, In: *Ecology*, Vol. 26, No. 3, pp. 297–302, 1945

[11] Church, K. W., Hanks, P.: Word association norms, mutual information, and lexicography, In: *Computational Linguistics*, Vol. 16, Issue 1, pp. 22–29, Mar. 1990

[12] Quasthoff, U., Wolff, C.: The poisson collocations measure and its application, In: *Workshop on Computational Approaches to Collocations*, Wien, Austria, 2002

[13] Dunning, T.: Accurate methods for the statistics of surprise and coincidence, In: *Computational Linguistics*, Vol. 19, Issue 1, pp. 61–74, MIT Press, Cambridge, 1993

[14] Biemann, C., Bordag, S., Quasthoff, U.: Automatic acquisition of paradigmatic relations using iterated co-occurrences, In: *Proceedings of LREC2004*, Lisboa, Portugal, 2004

[15] Kubek, M., Witschel, H. F.: Searching the Web by Using the Knowledge in Local Text Documents, In: *Proceedings of Mallorca Workshop 2010 Autonomous Systems*, Shaker Verlag Aachen, 2010

[16] Witschel, H. F.: *Terminologie-Extraktion: Möglichkeiten der Kombination statistischer und musterbasierter Verfahren*, Ergon-Verlag, Würzburg, 2004

[17] Kubek, M., Unger, H.: Search Word Extraction Using Extended PageRank Calculations, In: *Autonomous Systems: Developments and Trends*, Studies in Computational Intelligence, Vol. 391, pp. 325–337, Springer Berlin Heidelberg, 2011

[18] Kubek, M.: DocAnalyser - Searching with Web Documents. In: *Autonomous Systems 2014, Fortschritt-Berichte VDI*, Vol. 10, Nr. 835, pp. 221–234, VDI-Verlag Düsseldorf, 2014

[19] Website of Neo4j, `https://neo4j.com/`, 2016, Last retrieved on 07/22/2016

[20] Efer, T.: Text Mining with Graph Databases: Traversal of Persisted Token-level Representations for Flexible On-demand Processing, In: *Autonomous Systems 2015, Fortschritt-Berichte VDI*, Vol. 10, Nr. 842, pp. 157–167, VDI-Verlag Düsseldorf, 2015

Thai Language Segmentation by Automatic Ranking Trie

Chalermpol Tapsai[1], Phayung Meesad[1] and Choochart Haruechaiyasak[2]

[1]Faculty of Information Technology
King Mongkut's University of Technology North Bangkok, Thailand

[2]National Electronics and Computer Technology Center, Thailand

Abstract: Thai language is a non-segmentation Natural Language (NL) in which all words continuously present in sentences without any delimiters. It is difficult for Word Segmentation (WS) to process. Thai WS programs were developed and continuously improved by many researchers. The most widely used is Thai Lexeme Tokenizer (LexTo) using trie structure and longest match technique. LexTo works well but has 2 main disadvantages; (1) dictionary size is too big and (2) too many excessive matching dispensable words before the correct word is found. In this research, Thai Language Segmentation using Automatic Ranking Trie (TLS-ART) is proposed. TLS-ART uses Word Usage Frequency (WUF) to exclude unused words from the dictionary and reorganize words in trie structure to reduce matching task which significantly improves efficiency. The experimental results showed that accuracy, precision, recall, and f-measure values are comparable to LexTo; however, the dictionary size is 86.07% smaller and matching task 12.73% decrease.

1 Introduction

In the modern era like the present, computers are the devices that play an important role in human daily life, being used widely and clearly to claim that majority of human's works today are inevitably related with computer processing. To command a computer, we needs to understand "Computer Language" that is the special language used for creating programs, a set of instructions which instruct computers to read data as well as to process and display the results according to user needs. Although many computer languages have been developed to mimic the syntax more closely to human language, these computer languages are still hard to understand by non-technician users. In addition, unexperienced programmers may take long time to learn how to develop an efficient computer program. These problems inspire to instead of making human to understand computer language; a better way is to make computers to understand human Natural Language, which is used in our everyday life. This concept will help users to command computer with their own language and express their requirements correctly without extra training.

On the contrary, many Natural Language Processing (NLP) techniques that have been developed in numerous ways are not significantly progressive or widely used. This is due to various problems, i. e. diversity of natural language which is different for each race, country or region where people live. Moreover, natural languages are complex, some words may have multiple types and meanings. One sentence can be interpreted to more than one meaning. Moreover, one answer can be a result of different sentences because of the rhetoric or familiarity of each user. This is the main cause of inaccurate word segmentation in NLP and still a major obstacle of successful research in this field.

In general case, NLP techniques include 4 major steps: 1) Lexical Analysis, 2) Syntactic Analysis, 3) Semantic Analysis, and 4) Output Transformation. Lexical Analysis is an important process which analyses natural language sentences that is split into small units called Token together with type and essential information used by next step. Faulty analysis and segmented wrong words will contribute to wrong interpretation, incorrect meaning, and erroneous output results, consequently. Especially in non-segmentation languages such as Thai, Laos, Burmese, Chinese, Japanese, Korean, etc., all words in sentences are written in a cursive without spaces or special characters distinguish between words. They are complex and highly possible to segment wrong words.

In case of Thai, many researchers have developed word segmentation algorithms by using various techniques. For example, Chaloenpomsawat *et al.* [3] used a feature based approach with RIPPER and Winnow learning algorithms. Besides, Henasanangul *et al.* [7] used string matching and word identification in dictionary to identify unknown-word boundary of partially hidden words and explicit unknown words. In addition, Tepdang *et al.* [14] improved Thai word segmentation with Named Entity Recognition by using the Conditional Random Fields (CRFs) algorithm for training and recognizing Thai named entities. Moreover, Suwannawach *et al.* [13] used Maximum Matching and Tri-gram Technique. Haruechaiyasak *et. al.* [6] conducted experiments comparing performance between the DCB method with Trie algorithm technique and MLB method with 4 techniques: Naive Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM) and Conditional Random Field (CRF). The result showed that the DCB with Trie algorithm and MLB with CRF technique gave the best results in Precision and Recall measures.

Since 2003, many TLS programs were developed and distributed to public usage and one of the most illustrious TLS programs is LexTo. By using DCB method with Trie algorithm and the longest-word matching technique, LexTo can analyze sentences and split Thai words with high accuracy, but there are two main problems. Firstly, size of dictionary is too big especially when more than 40,000 words are included with a large number of unused words while the necessary and frequently used words are not stored. This results in many unknown words and forces users to add more words to dictionary. Secondly, the word organization is not efficient in Trie to reduce matching task, causing excessive matching with dispensable words and taking a long time to find a word. According to the mentioned problems, Thai Language Segmentation by Automatic Ranking Trie (TLS-ART) is proposed herein to improve Trie by using actual Word Usage Frequency (WUF) of Thai words to exclude unused words from dictionary and reorganized words in Trie, seeking more frequently used words before less used words to reduce matching task.

The remainder of this article is organized as follows. Section 2 presents overview researches and problems related to NLP. Section 3 concerns with natural language processing concepts. Section 4 illustrates detailed presentation of TLS-ART. Section 5 shows the researchs results. Finally, Section 6 concludes remarks as well as future research direction.

2 Related Work and Existing Problems

2.1 Research in NLP

Recently there are many researchers have conducted a study in many different ways of NLP. For example, in [4] the researchers studied on NLP to conclude requirements specification of software handbook written in unfix-pattern natural language and translate to a format defined language. In addition, automatic indexing document by the number of occurrences of each substring in the document and tree structure were studied by [15]; [11] proposed a technique to classify legal contract agreement documents by using NLP. Moreover, NLP interface with computer systems to retrieve information from databases have been studied by using various techniques and algorithms, such as, LUNAR [16], FREyA

[8], and NLKBIDB [10]. These technique cove natural language in both simple sentences and negation sentences including words, ″outside″, ″exclude″, ″does not″, ″not″, ″no″ and so on.

2.2 Existing Problems

Androutsopoulos *et al.* [1] mentioned that key factors in development of NLP is the expertise in linguistics and specialisation of researchs work. The lack of expertise would hinder progression of NLP researches and developments. This is consistent with Rodolfo *et al.* [9] that mentioned four major problems often occurring from the use of Natural Language to Interface Database as follows: 1) Various grammatical forms of Natural Language; 2) Missing of some important words to convey meaning of sentence; 3) Querying for information which relates to many tables and using of aggregate function; and 4) Problems caused by human errors.

3 Methods and Techniques of NLP

For more than 40 years, NLP have been conducted in many researches to facilitate computer utilization by using numerous methods and techniques. Main processes in NLP can be divided into four steps [10] as shown in Figure 1.

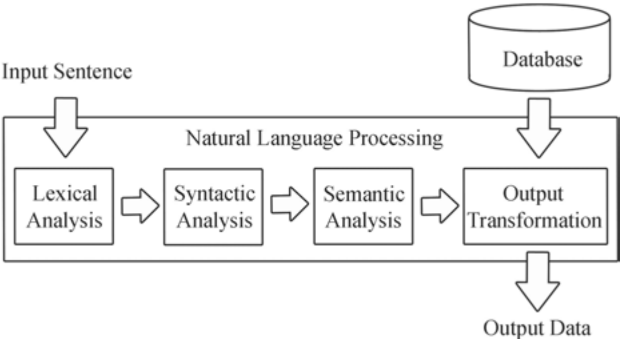


Fig. 1: Natural Language Processing steps

1. *Lexical Analysis*: This step analyses natural language sentences by splitting into small items each called Token. In addition, the Tokens is identified types and some essential information will be used in the next step.

2. *Syntactic Analysis*: In this step, all tokens are parsed with predefined sentence structure (Syntax) for validity checking and provided some information to be used in the meaning analysis process.

3. *Semantic Analysis*: The semantic analysis process interprets the meaning of a sentence by parsing information, which derives from the previous step with a semantic structure such as an ontology or a semantic web structure to provide some data that represent the meaning of a sentence.

4. *Output Transformation Process*: This step transforms outputs derived from Semantic Analysis into the results that meet the objectives of targets work, such as SQL commands for information retrieval from databases.

As mentioned before, Thai language which is a non-segmentation Natural Language, Word Segmentation in lexical analysis is a very important process due to the difficulty when splitting words from sentences. If the analysis is not effective enough, it will produce wrong results. The consequent process, i.e., syntax analysis and semantic analysis will inevitably produce wrong output too. At present, some lexical analysis systems are available for public usage. For example, WordNet is a system that can analyze English words with a large online database. For Thai, LexTo is a program developed by National Electronics and Computer Technology Center (NECTEC) widely used for Thai Word Segmentation.

4 The Proposed TLS-ART

The main idea of this research is to improve the efficiency of dictionary based method by exclude excessive words from dictionary and organized words in Trie. The proposed technique, Thai Language Segmentation by Automatic Ranking Trie or TLS-ART, employs actual usage frequency to reduce the dictionary size and number of matching tasks for splitting and identifying words in sentences. There are three steps in the proposed technique as shown in Figure 2.

1. *Dataset Preparation*: In this step data sets are prepared to build dictionary and Trie. The dataset used for this research is a set of Thai Language sample text files, collected from actual daily life usage of Thai people. The sam-

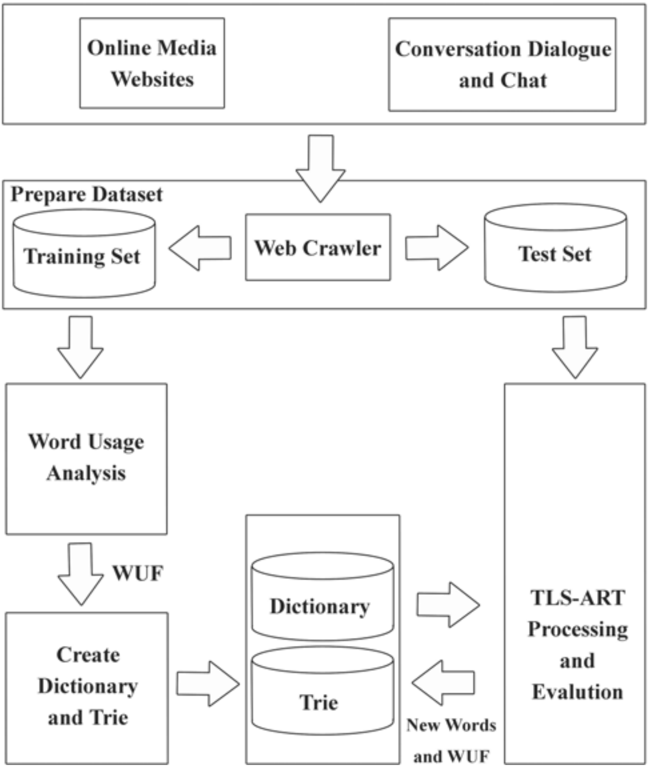


Fig. 2: TLS-ART Research Steps

ple text files were created from sentences, randomly collected from popular websites and conversation dialog chats covering all major fields including economics, social, political, entertainment, and others. There were 48 websites and 1,320 files collected as shown in Table 1. From Table 1, the dataset was divided into two sets: Training Set used for Word Usage Analysis process and Test Set used for TLS-ART Evaluation process.

2. *Word Usage Analysis*: This process analyses texts in Training Dataset and count for number of appearance of each word to provide Word Usage Frequency (WUF).

3. *Create dictionary and Trie*: This process saves words and WUF of each word to dictionary and creates Trie by placing ordered words based on usage frequency from high to low.

4. *TLS-ART Processing and Evaluation*: The main task of TLS-ART processing is to parse input text files from Test Dataset with Trie, comparing character by character to find a word with longest match and count the number of words appearance used for Improving Trie and dictionary. In case of no matched word is found, the unknown string will be shown to user for verification and add as a new word to the dictionary and Trie. To prove the effectiveness, TLS-ART is compared with LexTo by using Test Dataset which are 100 text files randomly collected from popular websites in eight categories as shown in Table 1.

Table 1: Number of sample websites and text files in each category of Dataset

Category	No.Websites	No.Training Files	No.Test Files
Economics	8	200	20
Social	8	200	20
Political	8	200	20
Entertainment	8	200	20
Chat room	8	200	20
Others	8	200	20
Total	48	1200	120

5 Experimental Results

Experiments for the performance comparisons between word segmentation using TLS-ART and LexTo shows that TLS-ART can reduce the size of dictionary from 42,222 words to 5,881 words or 86.07%. In addition, the number of link search used by TLS-ART is 925,467 or 12.73% less than LexTo while accuracy, precision, recall and f-measure values are nearly equal as shown in Tables 2 and 3.

6 Conclusion, Discussion and Future Work

To reduce matching task and improve efficiency of Thai segmentation, in this research, Thai Language Segmentation using Automatic Ranking Trie or TLS-ART is proposed. Word Usage Frequency (WUF) is used to exclude unused

Table 2: Dictionary size and number of link search using by TLS-ART compared with Lexto

Dictionary	No. words	No.link search
LexTo	42,222	1,060,514
TLS-ART	5,881	925,467
Decrease	86.07%	12.73%

Table 3: Performance Evaluation

Techniques	Accuracy	Precision	Recall	F-measure
LexTo	0.935	0.957	0.976	0.967
TLS-ART	0.936	0.958	0.976	0.967

words from the dictionary and reorganize words in trie structure. Experimental results showed that TLS-ART can significantly reduce the size of dictionary. Besides, reorganization of words in Trie using WUF reduces the number of Link Search explicitly.

From this study, it is observed that almost of unknown words are specific name of people, places, things and some other rarely used vocabulary. Moreover, some specific names may include other words as their parts. It is a challenge issue for future research to segment and identify these unknown words in a correct way.

References

[1] Androutsopoulos, G., Ritchie, D., Thanisch, P.: Natural Language Interfaces to Databases-An Introduction, *Natural Language Engineering*, 1, 1, pp. 29–81, 1995

[2] Al-Suwaiyel, M., Horowitz, E.: Algorithms for trie compaction, *ACM Trans. Database Syst*, 9, 2, pp. 243–263, 1984

[3] Chaloenpomsawat, P.: *Feature-Based Thai Word Segmentation*, Master Thesis, Chulalongkorn University, 1998

[4] Fatwanto, A.: Software Requirements Specification Analysis Using Natural Language Processing Technique, *IEEE Quality in Research*, 2013

[5] Fellbaum, C.: WordNet and wordnets, In: *Brown, K. et al. (eds.), Encyclopedia of Language and Linguistics*, 2nd ed., Oxford: Elsevier, pp. 665–670, 2005 [`http://wordnetweb.princeton.edu/perl/webwn`]

[6] Haruechaiyasak, C., Kongyoung, S. and Dailey, M.: A Comparative Study on Thai Word Segmentation Approaches, In: *IEEE Proceedings of 5th International Conference on ECTI-CON 2008*, pp. 125–128, 2008

[7] Henasanangul, T., Seresangtakul, P.: Thai Text with Unknown Word Segmentation Using the Word Identification, *Khonkaen University Research Journal*, 6, 2, pp. 48–57, 2006

[8] Miguel Liopis, A. f.: Computer Standards and interfaces How to make a natural language interface to query datasets accessible to everyone: An Example, 2012

[9] Rodolfo, A., Pazos, R., Juan, J., Gonzalez, B., Marco, A., Aguirre, L.: Semantic Model for Improving the Performance of Natural Language Interfaces to Databases, In: *Advances in Artificial Intelligence: 10th Mexican International Conference on Artificial Intelligence, MICAI 2011*, Vol. 7094, pp. 277-290, 2011

[10] Shah, A., Pareek, J., Patel, H., Panchal, N.: NLKBIDB - Natural language and keyword based interface to database, In: *IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI 2013)*, pp. 1569–1576, 2013

[11] Slankas, J., Williams, L.: Classifying Natural Language Sentences for Policy. *IEEE International Symposium the Policies for Distributed Systems and Networks (POLICY)*, pp. 34–36, 2012

[12] Song, P., Shu, A., Phipps, D.: Language Without Words: A Pointillist Model for Natural Language Processing, *SCIS-ISIS 2012*, Kobe, Japan, 2012

[13] Suwannawach, P. *Thai Word Segmentation Improvement using Maximum Matching and Tri-gram Technique*, Master Thesis, King Monkut's Institute of Technology Ladkrabang, 2012

[14] Tepdang, S., Haruechaiyasak, C., Kongkachandra, R.: Improving Thai word segmentation with Named Entity Recognition. In: *International Symposium on Communications and Information Technologies (ISCIT)*, 2010

[15] Todsanai, C.: An automatic indexing technique for Thai texts using frequent max substring, In: *IEEE Eighth International Symposium on Natural Language Processing*, 2009

[16] Woods, W.A., Kaplan, R.M., Webber, B.L.: *The Lunar Sciences Natural Language Information System*, Final Report, BBN Report 2378, Cambridge, Massachusetts: Bolt Beranek and Newman Inc., 1972

Didactics in Computer Science Education at Universities of Applied Sciences

A Personal Review of Modern Teaching Methods

Theodor Tempelmeier

University of Applied Sciences, Rosenheim, Germany

Abstract: Modern teaching methods as advocated by educationalists are briefly covered and contrasted with more traditional approaches. The focus hereby is on education in computer science or similar disciplines. Especially at universities of applied sciences with their emphasis on teaching, didactics is of great importance. It is shown that traditional teaching methods contain the same elements as modern methods, without using fancy new names for these methods, though. In summary Interactive-Style Teaching (IST) as usually employed at universities of applied sciences is considered equivalent – if not superior – to many of the “modern” teaching approaches.

1 Introduction

> ***What Impertinence!***
> *Since the days of Plato teaching means that the master speaks to his students. Western civilization rests on this. Today we are told that teachers must learn to teach in a different way. But this dis-intellectualizes the teachers. What impertinence!*
> *Alain Finkielkraut*[1]

The author was struck by Finkielkraut’s statement, although it was meant against school reform, not directed at university teaching, and although “the master speaking to his students” is not considered appropriate. But teachers at universities of applied sciences[2] in Germany are constantly confronted with

[1] French philosopher, on school reform in France. Originally in German (translated by the author): „Welche Unverschämtheit! Seit den Tagen Platons bedeutet unterrichten, dass der Meister zu seinen Schülern spricht. Die westliche Zivilisation beruht darauf. Heute erzählt man uns, dass Lehrer lernen müssen, anders zu unterrichten. Aber das entintellektualisiert die Lehrer. Welche Unverschämtheit!“ [2]

[2] Fachhochschulen in German

new and “modern” teaching methods they should use (e.g. as in [1]), so they might feel like Finkielkraut.

This contribution examines some of the advertised modern teaching methods and compares them to traditional teaching as it is (or was?) common at universities of applied sciences. Only education in computer science or in similar disciplines (e.g. in electrical engineering) is considered. The following may probably not apply to other disciplines, such as educational, social, or psychological sciences.

Note that this contribution does not contain an extensive literature study or a statistical evaluation of various teaching methods on a large sample size. Rather, it constitutes a very personal review of almost 40 years of teaching computer science (and more than 45 years of learning it), based on the very small sample of the author′s former students and a few scientists in his personal environment. Further it is biased by the author′s special field of expertise, embedded systems, safety-critical systems, and the like.

This contribution may also be seen as a dispute about or a counterpoint against demands to use more “modern” teaching methods at universities of applied sciences. Hopefully it will spur discussion.

2 Teaching Computer Science at Universities

The Parties Involved

Parties involved in education in universities are of course the teachers (the professors), and the learners (the students). However, for this contribution these will be distinguished between teachers and learners of computer science or similar disciplines on the one hand, and teachers and learners of educational, social, or psychological sciences on the other hand. These groups will be termed “engineers”[3] and “educationalists” for short in the following. See figure 1 for an overview. Teachers at secondary or primary schools are not considered in this contribution and are only shown for completeness.

The Students

Students in computer science (or in engineering in general) learn differently! As one student of electrical engineering put it after having finished his master′s degree: “In my studies, once I had understood things, I did not have to memorize

[3] Engineers are kindly asked to tolerate this slightly incorrect usage of their job title.

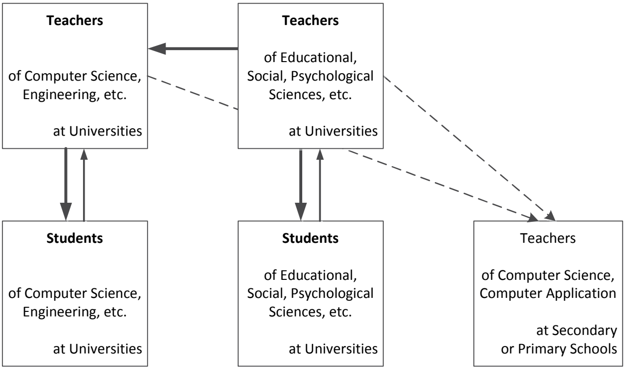

Teachers
of Computer Science, Engineering, etc.
at Universities
Teachers
of Educational, Social, Psychological Sciences, etc.
at Universities
Students
of Computer Science, Engineering, etc.
at Universities
Students
of Educational, Social, Psychological Sciences, etc.
at Universities
Teachers
of Computer Science, Computer Application
at Secondary or Primary Schools

Fig. 1: Teaching at Universities: The Parties Involved and their Influence on each other

them (only perhaps a little bit)". In the view of the author this is different in other disciplines, where "learning" has much more to do with memorizing (e.g. in medicine), whereas in engineering "learning" mostly means understanding with comparatively little memorizing.

This difference in learning is occasionally acknowledged by the educationalists (e.g. in [3], [4]), but too often (in the view of the author) educationalists ignore this different mind-set of engineers.

The Subject

The reason for the differences in learning described in the last section probably lies in the very abstract nature of the subject to be learned. Computer science excessively uses abstraction over abstraction. Thus one gets rid of many complicating details (at the highest abstraction levels) and arrives at a situation where the creation of new computer applications is more at ease. However, this enables more complex applications (with respect to software and hardware), which entail more abstractions. It poses indeed a huge challenge to understand these abstractions and their interaction with each other. And there are almost no tangible or palpable objects to ease understanding.

Especially for readers from the educational sciences an appendix is enclosed which shows *just one* of these abstractions, in fact one of the more concrete ones. The idea behind this is that the educationalists gain some comprehension of the difficulty of the subject to be taught, eventually resulting in a mutual understanding of engineers and educationalists.

The Teachers

Teachers at universities tend to teach in the same way as they encountered it during their own studies. The vertical arrows in fig. 1 show this interrelationship between teachers and their former experience as students. This is a good thing, and one may well assume that this is the best way to teach. Given the distinction of teachers and students into engineers and educationalists from above, there is no problem at all, when teachers of the educational sciences teach students of the educational sciences.

However, when educationalists give didactic advice or rules to engineers (the horizontal arrow in fig. 1), a clash between the differing conceptions of learning (as described in the two preceding chapters) arises. Educationalists tend to base all didactics on their own way of studying – that is extracting from hard to understand scientific texts into an understandable form, "learn" the latter, rephrase it into scientific (hard to understand) form again, and reproduce the latter in form of presentations[4]. This way of studying is not criticized here; however, it is not suited for computer science and the engineering disciplines. Thus the following requirement arises.

The demand is that university teachers of the educational sciences must have profound knowledge in engineering (i.e. a university degree with highest marks) in order to advise university teachers of computer science or similar disciplines. Otherwise their advice is rejected as inadequate here.

One might argue that general advice concerning rhetorics, usage of media, communication with the students, and so on is irrespective of the subject of studies, but this is only true in part, because engineers have a different mind-set. As an example, the author encountered situations where educationalists taught about "analogue" and "digital" communication (between humans and in the meaning of Watzlawick [5]), without any thought about the usage of these terms (let alone about the different meaning of these terms) in the engineering disciplines. Such egocentric behaviour is not acceptable and might as well be considered an offence to the audience.

[4] Admittedly, this is only a personal impression from observing cases in the the author's personal environment.

3 Teaching Methods – Traditional vs. Modern

3.1 Traditional Teaching at Universities of Applied Sciences: Interactive-Style Teaching (IST)

Teaching at universities varies depending on the subject taught and on the type of university (applied sciences vs. more research oriented). Here the focus is on universities of applied sciences and on teaching in the engineering disciplines.

Within this context education is traditionally organized as follows. (But note that due to "freedom of teaching" as guaranteed by the constitution in Germany, teachers at universities may deviate from this style.)

"Seminaristischer Unterricht". In contrast to lecture-style teaching (the professor literally "reading" the contents of learning units to the students), also called frontal teaching [6] or "chalk and talk" [7], universities of applied sciences employ what is called "seminaristischer Unterricht" in German, perhaps best described by interactive lecturing or by seminar-like tuition. A small group of listeners (max. 50) is a prerequisite. The professor will do lecture-style teaching in front of the audience. But in addition the professor may ask questions to the audience, short quizzes may be interspersed with lecturing, or short discussions in small groups among the participants may be initiated by the professor. The audience is encouraged to ask questions to the professor immediately during the interactive lecture.

Exercise courses and practical courses. In addition to these interactive lectures exercise courses or practical courses are held on a weekly basis. In these courses small problems have to be solved by the students, sometimes similar to the contents of the interactive lecture, sometimes going a bit beyond that. The students usually work in groups discussing with each other. The professor or a teaching assistant is available to answer questions, to ask additional questions and to supervise the discussions. The number of participants must be smaller than in interactive lecturing, 20 to 25 participants is a maximum.

Projects. Some larger projects, lasting for a full or a half semester, are also typically included in the curriculum. Besides consolidating or widening technical competence during these projects, organizational and social skills are also trained (and accompanied by respective courses on project management, group communication, and so on.) These project assignments simulate real life conditions as encountered in industry, though on a smaller scale.

Teaching as described in this paragraph (typically found at but not restricted to universities of applied sciences) will be termed *"interactive-style teaching" (IST)* for short in the following.

3.2 Peer Instruction (PI)

Traditional teaching at research oriented universities, i.e. lecturing in front of a large audience with no or little interaction with the audience, has its limitations: it is deemed inefficient and brings little benefit to the audience. The inventor of peer instruction (PI), E. Mazur excellently elaborates on this (e.g. in [8]). The proposed solution is to let students independently prepare the lecture and to use the time in the classroom for fostering understanding. This is done by asking the students questions, letting the students discuss these questions with each other, and again checking the students' understanding after the discussions.

Obviously, judging from a theoretical viewpoint, one will immediately assume that this method is superior to simple lecturing as described above. Stringent comparative tests of this teaching method clearly support this assumption [8]. So the superiority of peer instruction over simple lecturing in front of the students is evident.

However, one should compare PI not only to "chalk and talk" lecturing, but also to IST. Peer instruction, i.e. discussion of small problems in small groups of students, is done in every exercise course and in every practical course. Due to the small groups the professor can engage with the groups during their discussions on an individual basis. The professor is usually walking around, only listening to the discussions of the student groups which are on the right way, sometimes intervening and bringing the discussion back right on track, sometimes answering questions and helping those who cannot even start to ask a question or to discuss. So this style of teaching is even more intense, more interactive, more individual than peer instruction in its original setting.

3.3 Just-in-time Teaching (JiTT)

In just-in-time teaching (JiTT) the students must prepare for their next meeting in class. They receive pre-class assignments which have to be done before class meetings and usually have to be submitted online. The professor can adapt to the results just before the class meeting (hence the name of this method), and classroom time is more efficiently spent on working on the students' problem areas.

Clearly, JiTT seems to be a very good method to encourage the students to come to the classroom *prepared*. And the professor can identify what he or she should focus on during classroom time.

However, this can also be achieved the other way round. In IST, after a weekly cycle of interactive lecture combined with exercise courses, the next interactive lecture should start with a repetition of last week's exercise problems. This should be done interactively, with similar questions as in the exercise assignment, or with simpler or slightly more difficult ones. The professor thus *afterwards adjusts to the students' skills*, remembering the observations and problematic topics from the previous exercise lesson or lecture. If there is a need for a nice name for this teaching method, one could call it later-on teaching (LOT), analogous to the naming of just-in-time teaching. However, the author prefers to just call it "recapitulation".

3.4 Learning by Teaching ("Lernen durch Lehren" (LdL) in German)

The basic idea of learning by teaching is that students prepare and teach lessons (facilitating and supervising classroom discussions, bringing in new ideas, but not just giving presentations).

In short: Learning by teaching does not work, at least when difficult subjects are involved. Some experiments in the master's degree courses have severely discouraged the author, perhaps because the teaching content was too difficult. But, almost all teaching content is difficult in a master's degree program, even for the professor, especially if it is new and fast evolving content as is the case in computer science.

At best, the best 30 percent of students will manage to fully understand *for themselves* the difficult subjects and all their ramifications as they are typical during master's degree courses. Yes, these students will learn a lot during the preparation of their lessons! But usually they will not be able to convey their understanding to the next 30 percent of less capable students, let alone to the weakest students who would need maximum tuition. The lower half of the students will mostly only acquire a moderate understanding for themselves, and naturally be less successful in teaching the other students.

The method may be well suited for subjects which are less abstract as compared to computer science in a master's degree program, for instance in the context of teaching one's own language as a foreign language, where it originates from.

3.5 Problem-based Learning, Project-based Learning (PBL)

Taking a naive view as a start, problem-based learning (PBL) might be interpreted as learning along an example. For instance, a professor might explain various concepts and solutions along the example of an MP3-Player.

However, this is not what is meant with problem-based learning in the field of pedagogy. In this concept the students themselves should solve the problem in groups following a series of steps (clarify terminology, define problem, analyse problem, define open points and search for information, consolidate all information in the group). And they would learn during these activities.

There are many problems in computer science which are clearly unsuited for problem-based learning, because they are by far too difficult to be solved by students. Say, for instance, the problem given to the students would be "How do we deal with the fact that real numbers in computers have finite precision, i.e. they are not really 'real' numbers in the sense of mathematics?"[5] It is hardly conceivable that the students could solve this problem and come up with the concept and the laws of numerical mathematics. Breaking down the problem into smaller pieces would still overstrain most students.

Project-based learning is seen similar here to problem-based learning, though with a broader range, a task closer to real life, and a finished product as "learning" outcome. See [6] for a more precise differentiation of these two teaching methods.

Judging the usefulness of PBL with common sense will immediately show a great benefit of this method. In curricula where a high percentage of "learning" necessarily means memorizing – as in medicine, where PBL originates – PBL is surely a huge improvement. And generally, PBL is surely by far superior to "chalk and talk" frontal teaching.

However, looking at computer science and at universities of applied sciences, almost all teaching *contains in fact* already elements of problem-based or project-based learning. As shown in chapter 3.1 all exercise courses or practical courses supplement the interactive lectures by problem-based learning. And projects are a significant part of the curriculum, too. However, in addition to problems and projects interactive lecturing is still indispensable in the view of the author.

[5] Hint to unaware readers: "Normal" laws of computing do no longer apply. For example, given that $b = 1$ the computation $a + b - a$ may well yield the result 0 instead of 1, if a is sufficiently large.

And recently there are also cautiously questioning doubts about the usability of PBL in the engineering disciplines from the educationalists, as in [9], last paragraph.

It should be noted that the first steps in problem-based learning (clarify terminology, define problem, etc.) are so important in applied computer science that there is usually a special one-semester course on this alone: it is called requirements engineering. And systematic procedures for doing projects are taught in accompanying courses such as software engineering, project management, etc. As a consequence problem-based learning in applied computer science education does not follow the original seven-step proceedings of PBL ("seven-jump"), but is problem-based nevertheless. The same holds for project-based learning.

In summary it is astonishing (not to say ridiculous) that problem-based or project-based learning is recommended as a teaching method to the engineering disciplines or to computer science. They already do it and have been doing it for decades, though in their own way and adapted to their own needs.

3.6 Further Methods

Other modern teaching methods – characterized by the buzzwords "self-directed learning", "inverted classroom", and others – are not included here due to space limitations. The author has a similar opinion on these methods as on those described above.

A final comment has to be given on "getting to know each other" and icebreaker approaches. Methods such as "ball bearing", "throw the ball", etc. are occasionally recommended even for the universities (e.g. in [1]). They may be appropriate in some situations, but in general are considered inappropriate for teaching at universities by the author. Students do indeed have to learn to correctly introduce themselves, but in a technical meeting of engineers and in an industrial setting this cannot be done by means of kindergarten-like games.

3.7 Evaluating the Modern Teaching Methods

All new teaching methods should be evaluated whether they result in an improvement in student learning. There are serious and stringent studies conducting such a comparison, e.g. [8] just to name one.

The considerations presented here do not contain such a study, because there were no resources available to conduct such a study and because these considerations are only a personal review and a personal estimation of the (very high) capabilities of the author's students after having finished their studies.

Another larger study of applying these new methods was done, comparing them with traditional teaching methods [9], [10], [11]. Surprisingly there was no objective and fair comparison of the outcome, i.e. the students' advancement in competency, skills, or knowledge. *Rather, the students rated their capabilities by themselves ("self-estimation" of their competency, etc.)*[6]; this means it has been measured whether they "feel" better about their studies. This can in no way be taken seriously, and is in no way more scientific than the personal estimation of the author described in the paragraph above.

The author has requested long ago to obtain

- "reproducible, empirical, scientific studies
- for a comparison of traditional teaching [that is IST] with modern teaching methods
- with regard to **learning outcome**
- with **identical teaching content** and with **identical teaching time**
- for the engineering disciplines." [13]

In the study mentioned above *learning outcome* has not been investigated objectively.

In other success reports *much teaching time* is spent on *considerably smaller teaching content*, e.g. in [14] where the authors use PBL and call the course a "valuable supplement" to the usual courses in mathematics. Yes, one can see that this course can be done successfully. But how should the "usual courses" in mathematics be given with their much larger teaching content or (in comparison) much shorter teaching time? One has to think economically in terms of teaching time versus lessons learned!

One former student gave a nice example: Pupils would probably find Pythagoras' theorem if one hands out the right problem, a few triangles, a few squares, and a few strings with knots in distance 3, 4, and 5. But from an economical view, with respect to teaching time, it is much better to just tell them the theorem. And, professors are *paid* for teaching, the former student noticed. Why should I myself laboriously develop the theorem, when the professor can just simply tell it to me?

[6] "Correct, no diagnostic tests" upon inquiry during [11]. The authors of the study have probably been misguided by [12], according to a hint in [9].

Another way to handle large teaching content is described in [15]: the “80:20” rule (with 20% of effort, one will get 80% of the envisaged benefit). The implied suggestion to reduce teaching content (which would make modern teaching methods more feasible) is clearly totally unacceptable in the engineering disciplines. Is anybody in favour of “with 20% of effort, we will bring 80% of all aircraft to a safe landing”?

4 Recommendations for (Prospective) Teachers at Universities

All teaching methods mentioned so far have their merits and contain good ideas for better teaching. However, the author does not see that focusing on *just one* of the modern methods would be desirable. Some of the modern methods may even be totally unsuitable for the engineering disciplines. As a consequence the traditional style, that is interactive-style teaching (IST) as described above, is to be preferred. But some additional recommendations for prospective teachers at universities are due.

Explain! Explain! Explain! It is most important to patiently explain the subject to the students, re-phrase it and give the explanation again in different words, with different examples, and again once more, with another wording, and so on. This is best done during the exercise lessons or practical lessons, where the professor should talk with (not “speak to”, cf. the introduction) the small student groups, which form during these lessons (two to five students). As a side effect, students also improve their expressive power in their own mother language, a deficit often complained about, too [2]. The conversation may also switch languages, if both teacher and learners are bilingual. And even the local dialect may be used instead of the standard language, if it helps.

Visualize! All engineering disciplines have their own formalism for depicting objects, (intermediate) results, etc. within their domain. In addition to this, animations, simulations etc. should be used whenever possible. However, the author has made the experience that students tend to quickly “click through” such simulations and animations, without really and deeply understanding what is going on. More fruitful is the next point.

Set up a Play! Employ Play-Assisted Teaching (PAT)! Setting up a role play to simulate some course of events in computer hardware or software has shown to be very helpful for the students.

As an example for PAT, the MESI cache coherency protocol has often been conveyed to the students in this way by the author. One has to set up five groups of, say, four persons (this fits nice with the number of participants in an exercise course). Each group takes one of the following roles: processor 1, processor 2, cache 1, cache 2, main memory. Now the processor groups generate read or write accesses to memory, and all other groups have to act according to the MESI protocol. If groups fail, they are usually corrected and helped by the others ("Hey, you have to ...!").

Of course, this simulation in form of a role play proceeds very slowly. Other subjects could also be taught this way, e.g. pipelining, where the students would probably detect for themselves the idea of register forwarding.

Remember! Think about your own learning during your own studies! Think about what was helpful and what was bad! Avoid the latter and employ the former!

Be self-confident! Reject advice from educationalists which do not have any knowledge of your field of expertise, if you feel that this advice does not fit with your subject of teaching!

And don't get confused by all those "new" teaching methods. You are probably already doing it quite well, when you use interactive-style teaching (IST) and follow the previous recommendations!

5 Conclusion

As a summary, the traditional teaching style in the engineering disciplines might be termed "peer-instruction-/ problem-and-project-based-/ later-on-/ interactive-lecturing- teaching". However, there is no need for a new name, because traditional education in computer science and engineering at universities of applied sciences is well-established and proven in practice. Speaking technically in ISO 26262 jargon one might say it is "proven-in-use" or (in German) "betriebsbewährt".

Appendix: An Example of Abstraction in Computer Science

This appendix is provided especially for the educationalists (if they happen to read this paper). It contains *just one* example of abstraction in computer science, in fact one of the more concrete ones. The idea behind this appendix is that the educationalists gain comprehension of the difficulty of the subject to be taught, eventually resulting in a better mutual understanding of engineers and educationalists. Readers familiar with computer science may well skip the following (or enjoy reading it).

Assume that a computer program accesses a variable, say x. This variable is assigned by the compiler to a particular memory location, say 4660 [7]. However, the memory cell with this address is only *virtual*, existing only in imagination. The computer translates (during execution of the program) this virtual address to a real address of memory, say 4547124. The memory cell with that address is indeed real, because it is a tangible and visible thing, although a microscope would be needed to see it.

But, some milliseconds later, the original mapping of the virtual address 4660 may have changed and the virtual address is now translated to another real address, say, 7901748. Some milliseconds later this mapping may have changed again, and so on. This address translation is a four-stage process, i.e. it requires four more memory accesses to be achieved. For performance reasons, if one is lucky, an additional, but identical translation is provided as a short-cut, to do the translation faster.

And above all, eventually it may not even be necessary to access the memory cell, because a copy is available in faster so-called cache memory, if one is lucky. (Cache memory is typically organized in a hierarchy of three levels, but this additional complexity is omitted here).

To sum up, virtual addresses are translated to real addresses (of memory cells existing in reality) by some complicated mechanisms. The mapping of virtual to real addresses may change within short periods of time.

Now there is a concept, called virtual machines. A virtual machine is a computer program, which simulates another computer, e.g. some older computer hardware, or the same computer hardware but with different software installed, etc. Now, within this virtual machine basically the same address translation process happens. But the outcome, the “real” address of the virtual machine, is not really real. It is only a virtual address to the computer running the virtual machine software. So, that “real” address is only virtual in this context and has to be translated once more.

[7] In hexadecimal notation this would be 1234_{16}

In summary, apart from some tiny memory cells (which might be seen in a microscope) nothing is real, almost all is virtual in this example although the subject is computer hardware, which is wrongly assumed to be tangible by many.

Note that this is just one example of abstraction in computer science. There are many more. And by the nature of software, almost all software is virtual in a way.

References

[1] Waldherr, F., Walter, C.: didaktisch und praktisch. Ideen und Methoden für die Hochschullehre. (didactical and practical. Ideas and methods for teaching at universities. In German) 2. Auflage. Schäffer-Poeschel Verlag Stuttgart, 2014.

[2] Finkielkraut, A.: »Welche Unverschämtheit!« Interview in: Die Zeit, 21. Mai 2015, Seite 48.

[3] Eschner, A.: Brauchen Ingenieure eine spezielle Didaktik? Ingenieure ticken anders. (Do engineers need special didactics? There is a difference in what makes engineers tick. In German). In: DiNa 05/2009, Zentrum für Hochschuldidaktik der bayerischen Fachhochschulen (DiZ), Ingolstadt. https://www.diz-bayern.de/images/documents/82/DiNa05-2009web.pdf. Accessed 2016 April, 30.

[4] Pace, D., Middendorf, J. (eds.): Decoding the Disciplines: Helping Students Learn Disciplinary Ways of Thinking: New Directions for Teaching and Learning, Number 98. Wiley (Jossey-Bass) 2004.

[5] Watzlawick, P., Beavin, J., Jackson, D. D.: Pragmatics of Human Communication. A study of interactional patterns, pathologies, and paradoxes. New York, Norton 1967.

[6] de Graaff, E.: Problem- versus Project-Based Learning in Engineering: Antagonist or Complementary Pedagogical Approaches. 1. VDI-Workshop Projektorientiertes und Problem-Basiertes Lernen (PBL) in der Ingenieurausbildung. 22./23. November 2012. Darmstadt. https://www.vdi.de/fileadmin/vdi_de/redakteur/bg-bilder/Qualitaetsdialog/Workshop/3_EdeGraaff-PBL-Project.pdf. Accessed 2016 April, 30.

[7] Mills, J.E., Treagust, D.F.: Engineering Education – Is Problem-Based or Project-Based Learning the Answer? Australasian Journal of Engineering Education, January 2003. https://www.researchgate.net/publication/

246069451_Engineering_Education_Is_Problem-Based_or_Project-Based_Learning_the_Answer. Accessed 2016 May, 17.

[8] Mazur, E.: Memorization or understanding: are we teaching the right thing?, presented by Eric Mazur at the University of Waterloo in Waterloo, ON, Canada on 1 December 2010. http://mazur.harvard.edu/talks.php or https://www.youtube.com/watch?v=tn1DLFnbGOo. Accessed 2016 April, 30. Many other publications on peer instruction also available at http://mazur.harvard.edu/.

[9] Keller, U., Köhler, T.: Vergleich der Anwendbarkeit von PBL in verschiedenen MINT-Fächern. (Comparing the usability of PBL in various MINT disciplines. In German) Zeitschrift für Hochschulentwicklung, Jg.11 / Nr. 3 (Mai 2016) S. 153-172.

[10] Zentrum für Hochschuldidaktik (Hrsg.): DiNa Sonderausgabe. Das Projekt HD-MINT 2012-2016. https://www.diz-bayern.de/publikationen/dina. To appear.

[11] Advance information to [10]: Köhler, T.: Abschlusspräsentation des HD-MINT Projekts. (Final presentation of project HD-MINT. In German.) 2016, May 30. University of Applied Sciences, Rosenheim.

[12] Braun, E.: Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). (The Berlin evaluation instrument for self-evaluated student competences. In German.) Göttingen Vandenhoeck & Ruprecht, Unipress, 2007. Zugl.: Berlin, Univ., Diss., 2006. See also: Braun, E., Gusy, B., Leidner, B., Hannover, B.: BEvaKomp - Berliner Evaluationsinstrument für selbsteingeschätzte studentische Kompetenzen. Diagnostica 54, pp. 30-42, 2008.

[13] Tempelmeier, T.: Private communication with an institution in charge of didactic training of university teachers. 2006.

[14] Eich-Soellner, E., Fischer, R., Wolf, K.: Ein Praxisbeispiel: Problembasiertes Lernen in der Veranstaltung “Angewandte Mathematik”. (A practical example: Problem-based learning in a course on applied mathematics. In German.) In J. Roth & J. Ames (Hrsg.), Beiträge zum Mathematikunterricht. S. 1345-1346, WTM-Verlag, Münster, 2014. See also: PBL in der Mathematik – ein Umsetzungsbeispiel. (PBL in mathematics – An example of implementing it. In German.) In: Zentrum für Hochschuldidaktik (Hrsg.): DiNa 10/2014. pp. 12-17. https://www.diz-bayern.de/publikationen/dina. Accessed 2016 June, 27.

[15] Course number 19 at an institution in charge of didactic training of university teachers. 2002. Name of the institution withheld, because it is not the intent of this contribution to blame a particular institution.

Towards Live Feedback in Online Exercise Systems

Hauke Coltzau and Marco Badalus

Chair of Communication Networks
FernUniversität in Hagen, Germany

Abstract: We give an overview of a system for online exercises in STEM education that allows students to enter their solutions with much higher degree of freedom than in existing systems. The approach allows for detailed automatically generated feedback without forcing teachers to invest a large amount of time for preparation of assignments. We describe feedback triggers as well as appropriate reactions of the system on these triggers.

1 Introduction

In STEM education, online interactive studying systems like Maple T.A. [7] are used to let students train methods and concept taught in lectures and seminars. Students can take advantage of possible direct feedback regarding the correctness of their results and, therefore, increase the efficiency of their self-studies. Teachers benefit from reduced need for perpetual manual correction of student's solutions as well as from automatically generated information about their student's abilities, skills and possible understanding problems both on a group level as well as individually.

Still, the feedback given towards students within the training situation is rather limited. The systems usually evaluates the correctness of the student's *results*, which is only usefull, if the student is able to find these result at all. To handle cases, in which student's need help on their way to the solution, teachers can organize the exercises into smaller, predetermined sections, where each section depends on intermediate results gathered from previous sections. This way, feedback can be given by the system up to the point on where a student does not know, how to continue. Teachers can even implement static feedback for predefined failures in the intermediate results and let the system give hints on how to continue.

This approach, however, does not only require a high amount of time and effort to create the exercises and the predefined feedback, it also drastically reduces the student's decision horizon and freedom of expressing a solution, resulting in a system that tells the students exactly, how to solve the task instead of letting them find an appropriate way on their own.

In this article, we discuss an approach that gives students and teachers a high degree of freedom in expressing solutions but still remains being analyzable automatically due to its component-oriented architecture. Additionally, we discuss some triggers for feedback that can be given to the students *during* their attempt to solve an assignment.

The article is structured as followed: In section 2, an overview over related works and existing systems is given. Section 3 discusses the system model of our approach. In section 4, triggers for synchronous feedback and possible feedback contents are described. The article closes with an outlook on future works.

2 Related Work

The Online Exercise System (Online Übungssystem) of FernUniversität in Hagen is a teaching and learning platform based on the WebAssign instructional system [1] adapted to the specific requirements of asynchronous learning scenarios inherent to distance universities. Teachers can easily create simple assignments with automatic evaluation based on value comparison and selection elements. In contrast to similar systems, the system can also transfer the student's inputs to arbitrary backends for evaluation and processing. This gives high flexibility on what kind of assignments can be created and how the student's inputs are evaluated. In programming courses, for example, students can enter source code as an answer to an assignment. The code is compiled and executed in the backend and has to undertake a blackbox testing procedure. The results of the test are transferred into a rating of the student's solution.

The main disadvantage of the system in the context of this article is the high implementation effort required to create the evaluation logic in these cases. Direct feedback during input of a solution is not explicitly provided, but can be implemented manually.

A more general learning platform is Moodle, which integrates collaboration and discussion possibilities [2]. Individual exercises and assessments are possible, but only have a low level of freedom in their inputs. More complex answers given by students (e. g. as free text) need to be corrected manually.

A commercial solution for testing and assessment is Maple T.A. by Maplesoft [7], backed by the Maple computer algebra system. Besides being able to provide simple evaluation similar to Moodle, Maple T.A. can be combined with MapleSim, their system-level modelling and simulation tool. This way, highly interactive assignments with direct feedback given in form of system behaviour can be created, especially for physical systems but also for function analysis and similar tasks in mathematics. An online community exists that provides a large amount of different tasks and assignments. Still, the main disadvantages are the same as already described for the previous systems. Feedback and freedom of expressing a solution must be implemented with high effort on the teacher's side.

AutoTutor [3] is "an intelligent tutoring system that holds conversations with the human in natural language. AutoTutor has produced learning gains across multiple domains (e.g., computer literacy, physics, critical thinking)." [4]. A first implementation was already presented in 1999 [5]. The Software uses algorithms of language processing, semantic analysis as well als classic techniques like regular expressions. It aims to be a platform for teaching in tutorial dialogues and therefore lies outside the focus of this article.

The transition from AutoTutor to cognitive systems is blurred. Cognitive systems structure and analyse knowledge and try to simulate human problem solving behaviour [6]. They deal with objects, their properties and relations between them. They attempt to derive knowledge from these relations as well as possible (re-)actions on specific situations. Cognitive systems can interact with users and also learn from their input by adapting objects, relations or other system properties [8]. Although their complexity is too high to be maintained in a dynamic practical learning environment, these systems may be of interest in future learning scenarios.

As for today, no system or approach exists that allows for an appropriate degree of freedom to enter an individual solution for a task into an online system without forcing teachers to invest a high amount of effort into preparation, evaluation and possible synchronous feedback.

3 Model

In our approach, students enter their solution step-by-step, where each step contains of a set of input elements, the actual step itself and a set of output

Table 1: Symbols to describe a solution

Name	Symbol	Description
Steps	S	Set of all valid steps
Solution elements	L	Set of all valid Elements, on which steps can be performed or which can be a result of a step.
Dependencies	$D \subseteq L \times S \cup S \times L$	Valid aassignments of steps to solution elements and vice versa
Step	$s \in S : L^{\|{}^{\cdot}s\|} \rightarrow L^{\|s^{\cdot}\|}$	Analogue to a transistion in Petri nets. From C/S perspective, a step is a function that assigns input solution elements to output solution elements.
Input elements	${}^{\cdot}s = \{l \in L \| (l,s) \in D\}$	Analogue to a transistions's *preset* in Petri net graphs. These are all solution elements $\in L$ that are an input for a given step $s \in S$.
Output elements	$s^{\cdot} = \{l \in L \| (s,l) \in D\}$	Analogue to a transistions's *postset* in Petri net graphs. These are all solution elements $\in L$ that are an output of a given step $s \in S$.

elements (results). Students can select each step from a set of valid step templates, containing e. g. elementary row operations for linear equation systems or equivalent transformations. The available step templates have to be created beforehand and are part of the exercise system. For each step template, parameters including their types and value ranges need to be declared as well as the number of results and their type. Additionally, each step template contains a function that is able to execute the step (i. e. calculate the results) automatically for any given valid set of parameters. This allows to crosscheck each of the steps, a student enters while expressing their solution and give automatic live feedback.

A student's solution is represented by a linear execution of steps, where input elements for any step can either be chosen from output elements of previous steps or can be created on demand, e.g. by using information from the assignment or from other knowledge bases. From a computer-science perspective, a single step is a function that assigns input elements to output elements and a solution is a function that assigns preliminary knowledge to the solution element(s) of the assignment.

When both solution elements and steps are represented by nodes and their relations as directed edges, the resulting graph is directed and acyclic. Each path that starts from any node in the graph will end in a node that represents a solution element. Steps, input and output elements (*solution elements*) can be interpreted as nodes in a *Petri net graph*, where steps represent transitions, while solution elements are places. We therefore denote these graphs as *Petri-net-alike solution graphs*.

A solution I is given by $I = \{L_I, S_I, D_I\}$ consisting of the solution elements L_I, the steps S_I and the connections between them D_I (see table 1). Any solution element $l \in L_I$ either exists

- only in presets of steps, i.e. $(\exists s \in S_I \ (l,s) \in D) \wedge (\neg\exists s \in S_I \ (s,l) \in D)$. If so, l is *preliminary knowledge* (e.g. derived from the assignment)
- only in postsets, i.e. $(\neg\exists s \in S_I \ (l,s) \in D) \wedge (\exists s \in S_I \ (s,l) \in D)$. If so, l is a *result*
- both in presets and in postsets, i.e. $(\exists s \in S_I \ (l,s) \in D) \wedge (\exists s \in S_I \ (s,l) \in D)$. If so, l is an *intermediate result*

For each assignment, there exists a *reference solution* $M = \{L_M, S_M, D_M\}$ that can be used to (help) evaluate the correctness of a student's solution. The minimal reference solution only consists of a single generic step, transforming the necessary input elements derived from the assignment into the correct output elements representing the solution. More fine-granular reference solutions can either be created by teachers, manually, or can be gradually learned from correct solutions provided by students. Reference solutions differ from student's solutions in such a way that a reference solution may have alternative solutions paths, especially, when the reference solution "learned" from student's solutions. The alternative paths may either be a result of different step execution orders or of alternative solution approaches for the given assignment.

To be able to compare a student's solution to a reference solution, it is therefore often necessary to break both the student's solution and the reference solution into comparable partial solutions.

Formally, to check, if a subgraph $P_I = L_P \subseteq L_I, S_P \subseteq S_I, D_P \subseteq D_I$ of a student's solution I can be a partial solution of I, the set of solution elements L_P must be dividable into three disjunct sets $L_{P,I}$, $L_{P,H}$ and $L_{P,O}$:

- The preset $L_{P,I} \subset L_P$ of the partial solution, such that $\forall \in L_{P,I}, \forall s \in S_P : l \notin s^\cdot \wedge \forall l \in L_(P,I) : \exists s \in S_P : l \in {}^\cdot s$ contains only solution elements that do not occur in one of the postsets of any of the partial solutions steps and at least in one preset. This set contains the preliminary knowledge for the partial solution.
- The postset $L_{P,O} \subset L_P$, such that $\forall \in L_{P,I}, \forall s \in S_P : l \notin {}^\cdot s \wedge \forall l \in L_(P,I) : \exists s \in S_P : l \in s^\cdot$ contains only solution elements that do not occur in one of the presets of any of the partial solutions steps and at least in one postset. This set contains the results of the partial solution.
- The intermediate (or hidden) results $L_{P,H} = L_P \setminus (L_{P,I} \cup L_{P,O})$ contains $\forall l \in L_{P,H} : (\exists s_i \in S_P : l \in {}^\cdot s_i \vee \exists s_o \in S_P : l \in s_a^\cdot)$.

If these three sets can be found, then P_I is a partial solution of I, if $\forall s \in S_P : {}^\cdot s \subset L_P \vee s^\cdot \subset L_P$, i.e. all solution elements of all presets and postsets of all steps within P_I are part of the partial solution. In other words: A partial solution is a cut of I with a closed surface, such that the surface only cuts edges between steps outside of the partial sultion and solution elements inside of it.

The comparison between student's solution and reference solution mainly comes down to the task of finding partial solutions of the same granularity for both of them and compare these partial solutions. The exact process on how to find and compare these solutions is not part of this article.

4 Triggers for Live Feedback

Students enter their solution for any given assignment step-by-step. Every new step starts with the selection of the step from the set of available step templates. This action can be seen as a declaration of the intention (*annotation*), what the student wants to do. After selecting a step template, the student will execute the step by selecting input elements and determining the appropriate output elements (results) for that step. With this information, the system can evaluate the student's actions and give feedback (see Table 2) on two different (sub-)levels:

1. on the *approach level*, the system can crosscheck with the reference solution, if the currently selected step is eligible to lead towards the solution of the assignment.
2. on the *execution level*, it can be checked, if the step itself was executed correctly.

Table 2: System feedback on steps

Trigger	Interpretation	Feedback
type error	Student cannot correctly identify step inputs or outputs	• give step template description, • give example for step execution, • refer to lecture material, • · · ·
Range error	either as above or student misses important information in assignment	• give hint to recheck assignment description, • further feedback as above
step execution error	Student either has simply made an error or has not yet mastered the specific step	• show correct execution, • further feedback same as with type error
no step selection	Student does not know how to continue	• show excerpt of reference solution, • show comments from other students

Formally, to be able to check, if a step leads toward the solution of the assignment, the step itself must already be part of the reference solution. If so, it will usually lead towards the solution. Additionally, when the reference solution is enhanced with solution attempts from students, there may be alternative paths to the solutions having different lengths. In this case, the shortest path could be preferred or given as feedback by the system.

On the execution level, for each step $s \in S_M$, it can be checked, if

- all input and output elements $l \in {}^{\cdot}s \cup s^{\cdot}$
  - are of the correct (expected) type for the current step.
  - are within valid value ranges (if applicable),
- all input elements $l \in {}^{\cdot}s$ are consistent, i.e. allow a valid execution of the step.
- the output elements $l \in {}^{\cdot}s$ are the correct results for the execution of the step.

While the approach level checks need to be done in a self-implemented logic, the execution level checks can at least partially be performed by a computer algebra system (CAS) like MapleNet.

Table 3: System feedback on solutions

Trigger	Possible feedback
incomplete solution	• identify missing partial solution and show first step of it, • refer to lecture material
incorrect solution	• find and show consistency error (if exists), • show reference solution, • refer to lecture material
inconsistent solution	• show consistency error, • provide alternative (partial) solution to bridge the inconsistency
no solution	• show first step(s) of reference solution, • refer to lecture material, • show comments from other students

The student's solution as a whole is of course also subject to feedback and assistance from the system (see Table 3). Even without consulting the reference solution at all, the system can check the consistency of the student's solution. An inconsistent solution either contains inconsistent steps (see above) or has missing connections from solution elements to existing or missing (but necessary) steps.

When comparing the student's solution to the reference solution, the postset of the student's solution should contain all elements of the postset of the reference solution to be correct. If the postset of the student's solution is larger than that of the reference solution, the student's solution is assumed to be correct. If it is smaller, it is assumed to be incomplete and the system can try to identify the missing partial solution and give appropriate feedback e.g. by naming the first step of the missing partial solution.

It is important to point out that a consistent solution does not necessarily lead to a correct solution in terms of the assignment. Vice versa, an inconsistent solution still may be correct in terms of their results. Consistency and correctness checks may overlap but need to be interpreted separately and may lead to different feedback to the student.

5 Summary and Outlook

In this article, we propose a system that allows students to enter solutions in an online exercise system with a much higher degree of freedom than in existing systems. The underlying model allows for a generic and fine-granular automatic analysis of the students input and is able to provide feedback both on step level and on solution level. It can especially cover situations, in which a student does not know, how to continue, without having to prescribe the solution approach as part of the assignment.

Future works will focus on the implementation of a system for real-life education scenarios as part of our online exercise system. Additionally, more theoretic work needs to be done concerning consistency analysis as well as enhancing the reference solution based on student's solutions.

References

[1] Brunsmann, J., Homrighausen, A., Six, H. W., Voss, J.: Assignments in a virtual university – the WebAssign-System. In: *Proceedings of the 19th World Conference on Open Learning and Distance Education*, 1999

[2] Cole, J., Foster, H.: *Using Moodle: Teaching with the Popular Open Source Course Management System*, O'Reilly Media, 2008

[3] Graesser, A., Chipman, P., Haynes, B., Olney, A.: AutoTutor: an intelligent tutoring system with mixed-initiative dialogue, In: *IEEE Transactions on Education*, 48,8, 612–618, 2005

[4] Graesser, A., Hu, X.: AutoTutor, online at `http://www.autotutor.org`, July 2016

[5] Graesser, A., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R.: AutoTutor: A simulation of a human tutor., In: *Cognitive Systems Research*, 1,1, 35–51, 1999

[6] Kurbel, K.: *Entwicklung und Einsatz von Expertensystemen: Eine anwendungsorientierte Einführung in wissensbasierte Systeme*, Springer-Verlag, Berlin Heidelberg, 2013

[7] Maple T.A. – Online Testing and Assessment Software for STEM education, online at `http://www.maplesoft.com/products/mapleta/`, July 2016

[8] Strohner, H.: *Kognitive Systeme: Eine Einführung in die Kognitionswissenschaft*, Springer Fachmedien, Wiesbaden, 1995

Ubiquitous Communication Channels and E-Learning

Djamshid Tavangarian and Ingolf Waßmann

Faculty of Computer Science and Electrical Engineering
University of Rostock, Germany

Abstract: The new web technologies, ubiquitous communication and social networks gave the initial impetus for new forms of online tasks and processes, which bring together the comprehensive networking of objects, spaces, services and users on the Internet sustainably. However, the diversity of different systems makes it difficult to use for a comprehensive holistic learning scenario that meets the requirements of targeted learning in a modern information society. In network technology, a hub is seen as a central instance in a communication environment, with which individual network nodes, in a central instance, connects and enables the exchange of information between the nodes by forwarding communication data.

The concept and the functionality of a hub are still used in many areas. For example, the conception of a ″social media hub″, then one speaks of social functions on websites as a focal point with a central view for users, being integrated from social networks in the various elements. These include aggregation of information from various social networks (such as plug-ins and sharing features) in a central application or a centralized portal. Users get a quick overview on relevant topics and discussions in social networks. An example application is Flipboard, a news aggregator that also involves social networks such as Facebook and Twitter as sources. The user can directly comment, edit content and share from the app.

Based on the idea of Connectivism theory, the hub concept provides an instance of Web-based learning as a ″learning hub″. So a Learning Hub is a quasi-central instance that allows communication between students, teachers, content, tools, services and institutions for obtaining qualifications.

In this presentation the term ″hub″ is considered as being especially closer to a ubiquitous online learning with the involved communication channels. It is developed as a formal representation, the core entities recognized users, content, communication channels and both human and machine services and providing ubiquitous learning in a central environment for the user. Further integrating the hub is shown in the web application wiki-Learnia for finding, creating and distributing educational content, which are individually tailored to each user. To support a semantic metasearch engine is used to search different learning repositories to select matching content. After filtering and preparation of the searched and selected contents, the contents, in a structured form, will be delivered to the learners.

Coordinate-based and Service-centred Addressing for Routing Design in MANETs

Adrian Dilo

FernUniversität in Hagen, Germany

Abstract: Apart from well established MAC and IP based addressing and routing techniques, an IP-addressless, coordinate based, ISO low level framework is a real competitive alternative, which is presented here. IP v4 and v6 based designs for mobile environments are complex and cumbersome. The presented design is characterised by overcoming IP routing difficulties in MANETs due to provider network address mixings in user habitations and reflects geografic singularities to optimize routing and service support. A new routing algorithm based on user density and user movements in habitations is part of the framework. The integrated service awareness of addressing and routing accomplishes this new design presented, which is suitable to be embedded in existing IP environments.

On Event-triggered Control with Application to Multi-agent Systems

Gang Feng

Mechanical and Biomedical Engineering
City University of Hong Kong, China

Abstract: In this talk event-triggered control will be first overviewed. The motivation and major event-triggering mechanisms will be discussed. The challenging issue on exclusion of Zeno behavior will be highlighted. Then event-triggered control will be considered for heterogeneous multi-agent systems. A distributed even-triggered control algorithm will be presented for output consensus of such multi-agent systems. It is shown that the output consensus problem can be solved by the proposed event-triggered control algorithms if a necessary and sufficient condition is satisfied. Then a self-triggered control scheme is also developed, where continuous monitoring of measurement errors can be avoided. The feasibility of both proposed control schemes is discussed by excluding Zeno behavior. A numerical example is given to illustrate the effectiveness of the proposed control schemes.

All about Chaos

Zhong Li

Chair of Computer Engineering
FernUniversität in Hagen, Germany

Abstract: This talk is dedicated to Prof. Halang's 65th birthday, which is to summarize our research work related to chaos in the past 15 years under the leadership of Prof. Halang. It involves chaos-based cryptography, electromagnetic compatibility (EMC) with chaos control, which is to use the pseudo-random and continuous power spectrum feature of chaos to spread peaks over a wide frequency band, thus to suppress electromagnetic interference (EMI), high quality converters, which is based on a novel impedance source to overcome the drawbacks of traditional converters, as well something about complex networks, including some fundamental concepts and our research work on smart grins.

Aperiodic Pulse Width Modulation in Power Converters for EMI Suppression

Hong Li

Electrical Engineering School
Beijing Jiaotong University, China

Abstract: Pulse width modulation (PWM) technique has been widely used in power converters owing to its simplicity and ability to generate the output waveform with good quality, however, PWM results in serious electromagnetic interference (EMI) problems. In order to suppress EMI under traditional PWM with fixed switching frequency, aperiodic PWM which includes random PWM and chaotic PWM has been proposed and applied to power converters. Since the real random signals for random PWM are difficult to generate in practical applications, chaotic signals generated by certain chaotic mappings are always used to replace the random signals due to their pseudo-random characteristics. This presentation is to introduce the EMI suppression mechanism of aperiodic PWM for power converters, then introduced the realization method of chaotic PWM. The chaotic PWM spectrum analysis approach based on double Fourier series, and the thermal impact of chaotic PWM will be introduced. Finally, new benefits and challenges for aperiodic PWM in power converters for EMI suppression will be discussed.

Trainable COSFIRE Filters for Pattern Recognition

Nicolai Petkov

Intelligent Systems, University of Groningen, Netherlands

Abstract: A trainable filter is a filter that is configured by the automatic analysis of a pattern specified by a user. Subsequently, such a filter can detect similar patterns. This approach is illustrated by the design of filters that can detect bifurcations in retinal fundus images. The user presents a vascular bifurcation as a local pattern of interest. The automatic analysis system applies a bank of Gabor filters to this pattern and identifies which of them respond most strongly and in which positions. The response of the composite trainable filter is then computed as a combination (e.g. a geometric mean) of the responses of the selected Gabor filters, shifted by certain off-set vectors determined in the analysis phase. We call this method Combination of Shifted Filter Responses (COSFIRE). An advantage of this approach is its ease of use, as it requires no programming effort the parameters of a filter are derived automatically from a single training pattern. This approach is further illustrated by the segmentation of blood vessels and the localization and segmentation of the optic nerve head in retinal fundus images.

Counting Efficiently Large Subgraphs Approximately

Hanno Lefmann

Chair of Theoretical Computer Science and Information Security
Technische Universität Chemnitz, Germany

Abstract: There has been interest in estimating the value of a graph parameter, i.e., of a function defined on the set of finite graphs G, by sampling a randomly chosen substructure whose size is independent of the size of the input. Graph parameters that may be successfully estimated in this way are said to be testable or estimable, and the sample complexity $q_f = q_f(e)$ of an estimable parameter f is the size of the random sample required to guarantee that the value of $f(G)$ may be estimated within an error of at most e with probability at least $2/3$. We consider the sample complexity of estimating two graph parameters associated with a monotone graph property. To obtain the results, we prove that the vertex set of any graph that satisfies a monotone property P may be partitioned equitably into a constant number of classes in such a way that the cluster graph induced by the partition is not far from satisfying a natural weighted graph generalization of property P.

Index of Authors

Fahrzeugkonzepte für das 2. Jahrhundert Automobiltechnik
Vehicle Concepts for the 2nd Century of Automotive Technology
VDI-Berichte 1653

Fortschritt-Berichte VDI
Weiterentwicklung der Endfunktionsprüfung von Verbrennungskraftmaschinen in der Großserienproduktion

Die Reihen der Fortschritt-Berichte VDI:

1 Konstruktionstechnik/Maschinenelemente
2 Fertigungstechnik
3 Verfahrenstechnik
4 Bauingenieurwesen
5 Grund- und Werkstoffe/Kunststoffe
6 Energietechnik
7 Strömungstechnik
8 Mess-, Steuerungs- und Regelungstechnik
9 Elektronik/Mikro- und Nanotechnik
10 Informatik/Kommunikation
11 Schwingungstechnik
12 Verkehrstechnik/Fahrzeugtechnik
13 Fördertechnik/Logistik
14 Landtechnik/Lebensmitteltechnik
15 Umwelttechnik
16 Technik und Wirtschaft
17 Biotechnik/Medizintechnik
18 Mechanik/Bruchmechanik
19 Wärmetechnik/Kältetechnik
20 Rechnerunterstützte Verfahren (CAD, CAM, CAE CAQ, CIM ...)
21 Elektrotechnik
22 Mensch-Maschine-Systeme
23 Technische Gebäudeausrüstung

ISBN 978-3-18-3**84810**-2