Doctoral Research Days at FIT 2019

November 1 and November 8, 2019

FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

Abstract Proceedings.

Supported by the Grant Agency of the Czech Technical University in Prague, grant No. SVK 51/19/F8.

Doctoral Research Days at FIT 2019: Program and contents

Friday 1 November 2019

Session 1

Session chair
Pavel Tvrdík

Doctoral Research Days at FIT 2019

Processing, Checking, and Modelling of Textual Requirements Specifications

David Šenkýř

The quality of Requirements Engineering plays an essential role in the whole development life cycle of every software project – because the other phases depend on it. Writing requirements specifications in natural language is a common practice. The natural language is, unfortunately, prone to a number of inaccuracies like ambiguity, inconsistency, and incompleteness. We investigate methods of grammatical inspection to identify patterns in requirements specification written in the textual form. Based on them, we can extract the information from the text, and also eliminate some of the mentioned problems [I] . As a result, we present the CASE tool called *TEMOS* that is able to generate fragments of the UML class model from textual requirements specification and also to help the user with the detection of some inaccuracies in the text. A similar approach should be used to generate fragments of various models, such as SHACL Shapes [IV] , too.

With regard to ambiguity, we focus on the structural ambiguity. We show that the standard methods of linguistics are not enough in some cases, and we describe a class of ambiguity caused by coreference that needs an underlying domain model or a knowledge base to be solved [II] . Part of our implemented solution is a cooperation of our tool *TEMOS* with the Prolog inference machine working with facts and rules acquired from OCL conditions of the domain model.

We also investigate the incompleteness problem. Incompleteness is a typical problem that arises when stakeholders (e.g., domain experts) hold some information for generally known, and they do not mention it to the analyst. A model based on the incomplete requirements suffers from missing objects, properties, or relationships. Our methods are based on grammatical inspection, semantic networks (*ConceptNet* and *BabelNet*), and pre-configured data from on-line dictionaries [III] . Additionally, we show how a domain model has to be used to reveal some missing parts of it.

List of publications:

Published:

[I] **David Šenkýř**, Petr Kroha; *Patterns in Textual Requirements Specification* In: Proceedings of the 13th International Conference on Software Technologies, pp. 197–204, Porto, Portugal, 2018. SCITEPRESS – Science and Technology Publications. ISBN 978-989-758-320-9.

[II] **David Šenkýř**, Petr Kroha; *Patterns of Ambiguity in Textual Requirements Specification*; In: New Knowledge in Information Systems and Technologies, volume 1, pp. 886–895, Cham, 2019. Springer International Publishing. ISBN 978-3-030-16181-1.

[III] **David Šenkýř**, Petr Kroha; *Problem of Incompleteness in Textual Requirements Specification*; In: Proceedings of the 14th International Conference on Software Technologies, pp. 323–330, Porto, Portugal, 2019. SCITEPRESS – Science and Technology Publications. ISBN 978-989-758-379-7.

Accepted:

[IV] **David Šenkýř**; *SHACL Shapes Generation from Textual Documents* In: EOMAS 2019, 15th International Workshop on Enterprise & Organizational Modeling and Simulation.

Doctoral Research Days at FIT 2019

Generating UML Models with Inferred Types from Smalltalk Code

Jan Blizničenko

Generating structural UML models from code of dynamically typed languages poses several problems that need to be addressed. A structure of classes with variables and methods has to be gathered, types of instance variables have to be found for associations and the model has to be importable into common modeling tools. We bring a review of current ideas and solutions to these and similar problems and it presents our ongoing effort towards this goal – our current solution we developed so far. For now, we focus on Pharo Smalltalk code.

List of publications:

1. **Jan Blizničenko**, Robert Pergl; *Generating UML Models from Pharo Code* In: Proceedings of the 14th edition of the International Workshop on Smalltalk Technologies, 2019. (not yet published).

Evolvable Architecture of Client Applications with the use of Normalised Systems Theory

Jan Slifka

Client applications are an essential part of the most software because they provide interfaces with which the users interact. We can think of different dimensions that the client application consist of – a platform, business logic and design system. These dimensions could be independent and reusable to a large extent. However, they usually are tightly coupled, which impede the evolvability.
The problem is also a rapid development of technologies used for client applications (e.g., new frameworks for web applications). Thus these applications become obsolete every few years and have to be rewritten, usually from scratch. Nowadays, with the increasing number of mobile devices, one client application is usually not enough. Each platform needs a specific client application. Therefore, more applications have to be rewritten, which tremendously increases the costs and time demand for software development.
The research of the Normalised Systems at the University of Antwerp focuses on evolvable systems, where the domain and technical parts are separated. So-called expansion is used to generate the code from the domain specification. Systems can be regenerated to the new technologies without the loss of the domain layer.
The goal of this research is to use existing and design new ontologies and modelling languages to describe different dimensions of client applications and use them for generating the actual software using expanders following the theory of Normalised Systems. The main focus is on business applications which usually have a limited set of features with only minor customisations, so they have the highest potential to save time and the costs.

List of publications:

Accepted:

[I] Marek Suchánek, **Jan Slifka**, Robert Pergl; *Evolvable and Machine-Actionable Modular Reports for Service-Oriented Architecture* In: EOMAS 2019, 15th International Workshop on Enterprise & Organizational Modeling and Simulation.

Design of Systems Supporting Compliance Management

Marek Skotnica

Compliance management is a process which ensures that a set of people follows a given set of rules. These rules are usually a mix of procedures, documentation, policies, industry standards, and legal obligations. A primary goal of this research is to describe how to design systems which support this domain.

Our approach reviews all steps in building compliance management systems. These usually involve 1) Domain Modeling - a creation of high-quality machine-readable domain models, 2) System Specification - a machine-readable system formalization, 3) System Execution - execution of a formalized system specification, 4) System Optimization - measuring and suggestion of domain model improvements. We do introduce novel approaches in system specification and execution.

We do showcase our methods and techniques on the domain of procedural law and legal contracts. Such systems can be executed as software artefacts or as a Blockchain smart contracts (SC). The SC has enormous potential in automating traditional paper contracts and encoding contract logic into program code. Thus, replacing the role of a notary and a central authority. It may dramatically reduce an effort with administration workload and the enforcement of such contracts.

List of publications:

1. Marek Skotnica, Steven J. H. van Kervel, and Robert Pergl. “Towards the Ontological Foundations for the Software Executable DEMO Action and Fact Models”. en. In: Advances in Enterprise Engineering X. Funchal, Madeira: Springer International Publishing, May 2016, pp. 151–165.

2. Marek Skotnica, Steven J. H. van Kervel, and Robert Pergl. “A DEMO Machine - A Formal Foundation for Executionof DEMO Models”. In: Advances in Enterprise Engineering XI. Ed. by David Aveiro et al. Cham: Springer International Publishing, 2017, pp. 18–32. ISBN: 978-3-319-57955-9.

3. Ondřej Mráz, Robert Pergl, Pavel Náplava, and Marek Skotnica. “Converting DEMO PSI Transaction Pattern into BPMN: A Complete Method”. In: Advancesin Enterprise Engineering XI: 7th Enterprise Engineering Working Conference, EEWC 2017, Antwerp, Belgium, May 8-12, 2017, Proceedings. Ed. by David Aveiro et al. Cham: Springer International Publishing, 2017, pp. 85–98. ISBN: 978-3-319-57955-9

4. Barbora Hornáčková, Marek Skotnica, and Robert Pergl. “Exploring a Role of Blockchain Smart Contracts in Enterprise Engineering”. In: Advances in Enterprise Engineering XII. Ed. by David Aveiro et al. Cham: Springer International Publishing, 2019, pp. 113–127. ISBN: 978-3-030-06097-8.

Integrating Conceptual Models and Implementations Using Ontologies

Marek Suchánek

Conceptual modelling as the activity of creating a description of a problem domain in order to promote understanding and communication between people is widely used in software engineering as part of analysis during traditional software development. Based on a conceptual model, solution in the form of software (or other) is devised to solve or mitigate the identified problems. Although there are multiple methods related to Model-Driven Development, the transformation from conceptual models to software suffer by the significant loss of semantic information and thus breaking the consistency between model and software. Another disadvantage of current methods is that it is limited to specific modelling languages and can be used to generate basic skeletons of enterprise information systems that then need to be finished by people. Our approach is different in this manner. We focus on integration on the conceptual level and be able to combine various conceptual models made in different languages to capture more aspects. Such integration can be done using ontology and technologies related to the Web Ontology Language [V] . This results in need of identifying overlaps and similarities between various modelling languages, for example, in process modelling languages [VI] . The implementation-related part of our work is oriented to Normalized Systems (NS), that is proven by both theory and practice to be a way for building sustainable and evolvable enterprise information systems but in other domains as well [1] [3] . The metamodel used in Normalized System consisting of so-called *Elements* and be also integrated using ontologies with specified transformations from different or combined conceptual models. Because the NS metamodel is very technology-oriented, we also work on enhancing to capture more semantics from conceptual level and also to reduce the need of manual programming (which is already low, less than 10 %) [4] [VII] . We designed the architecture of the *Normalized Systems Gateway Ontology for Conceptual Models* and set up requirements for the related tooling that will be developed. The future research will be focused on the use of developed architecture in practice, mapping various modelling languages, and enhancing the architecture if needed.

List of publications:

[1] **Marek Suchánek**, Robert Pergl; *Evolvable Documents – an Initial Conceptualization* In: PATTERNS 2018, The Tenth International Conference on Pervasive Patterns and Applications. Wilmington: IARIA, 2018. p. 39-44. ISSN 2308-3557. ISBN 978-1-61208-612-5.

[2] **Marek Suchánek**, Robert Pergl; *Data Stewardship Wizard for Open Science* In: Data a znalosti & WIKT. Brno: Vysoké učení technické v Brně. Fakulta informačních technologií, 2018. p. 121-125. 1. ISBN 978-80-214-5679-2.

[3] **Marek Suchánek**, Robert Pergl; *Towards Evolvable Documents with a Conceptualization-Based Case Study* International Journal on Advances in Intelligent Systems. 2018, 11(3&4), 212-223. ISSN 1942-2679.

[4] **Marek Suchánek**, Robert Pergl; *Evolvability Evaluation of Conceptual-Level Inheritance Implementation Patterns* In: PATTERNS 2019, The Eleventh International Conference on Pervasive Patterns and Applications. Wilmington: IARIA, 2019. p. 1-6. ISSN 2308-3557. ISBN 978-1-61208-612-5.

Accepted:

[V] **Marek Suchánek**, Robert Pergl; *Designing an Ontology for Semantic Integration of Various Conceptual Models* In: EOMAS 2019, 15th International Workshop on Enterprise & Organizational Modeling and Simulation.

[VI] **Marek Suchánek**, Robert Pergl; *Mapping UFO-B to BPMN, BORM, and UML Activity Diagram* In: EOMAS 2019, 15th International Workshop on Enterprise & Organizational Modeling and Simulation.

[VII] **Marek Suchánek**, Jan Slifka, Robert Pergl; *Evolvable and Machine-Actionable Modular Reports for Service-Oriented Architecture* In: EOMAS 2019, 15th International Workshop on Enterprise & Organizational Modeling and Simulation.

The problem of evolvability in Data Stewardship Planning

Ing. Vojtěch Knaisl

My topic is to focus on the problem of evolvability in Data Stewardship Planning. In general, the problem of evolvability and long-term sustainability is a big challenge nowadays. My focus is to improve the evolvability in the Data Stewardship Planning process.
In terms of the increasing amount of data in research experiments, the importance of right planning, storing data, and long-term sustainability increase too. Data Stewardship (DS) contains all these phases. In CCMi [1], we are developing a unique tool Data Stewardship Wizard [2]. In this area, we face a problem with evolvability due to changing requirements and standards on DS in every phase of the projects, further, the changing requirements on output and format of desired plans.
During my first year, I focused on the general evolvability of structured and well-formatted documents. I wrote one article and made some proposals. But because the topic was too wide, we narrow it to the Data Stewardship domain where I continue on my research. My goal to apply Normalized Systems Theory [3] and other formal methods for creating a methodical framework and proposal of a technical solution for Data Stewardship Planning. The Normalized Systems Theory developed in the University of Antwerp brings the principles for the construction of evolvable systems. This general theory was applied, among other things, on the evolvability of documents [4].

List of publications:

Accepted:

[I] **Vojtěch Knaisl**;*Proposing an Architecture of an Intelligent Evolvable Document Generation System based on the Normalized Systems Theory*: Proceedings of the 15th International Workshop on Enterprise & Organizational Modeling and Simulation (EOMAS)

References

[1] *Centre for Conceptual Modelling [online], [cit. 2019-10-14].* URL: https://ccmi.fit.cvut.cz.

[2] *Data Stewardship Wizard [online], [cit. 2019-10-14].* URL: https://ds-wizard.org.

[3] Herwig Mannaert, Jan Verelst, and Peter De Bruyn. *Normalized Systems Theory: From Foundations for Evolvable Software toward a General Theory for Evolvable Design*. 2016. ISBN: 978-90-77160-09-1.

[4] Gilles Oorts, Herwig Mannaert, and Peter De Bruyn. "Exploring Design Aspects of Modular and Evolvable Document Management". In: *Advances in Enterprise Engineering XI*. Ed. by David Aveiro et al. Cham: Springer International Publishing, 2017, pp. 126–140. ISBN: 978-3-319-57955-9.

Session 2

Session chair
Petr Fišer

Doctoral Research Days at FIT 2019

Faster FPT Algorithm for 5-Path Vertex Cover

Radovan Červený

The problem of d-Path Vertex Cover, d-PVC lies in determining a subset F of vertices of a given graph $G = (V, E)$ such that $G \setminus F$ does not contain a path on d vertices. The paths we aim to cover need not to be induced. It is known that the d-PVC problem is NP-complete for any $d \geq 2$. When parameterized by the size of the solution k, 5-PVC has direct trivial algorithm with $\mathcal{O}(5^k n^{\mathcal{O}(1)})$ running time and, since d-PVC is a special case of d-Hitting Set, an algorithm running in $\mathcal{O}(4.0755^k n^{\mathcal{O}(1)})$ time is known. In this paper we present an iterative compression algorithm that solves the 5-PVC problem in $\mathcal{O}(4^k n^{\mathcal{O}(1)})$ time.
Presented results are part of [I] .

List of publications:

Published:

[I] **Radovan Červený**, Ondřej Suchý; *Faster FPT Algorithm for* 5-Path Vertex Cover

In: Proceedings of the 44th International Symposium on Mathematical Foundations of Computer Science, MFCS 2019, August 26-30, 2019, Aachen Germany. ISBN 978-3-95977-117-7, LIPICS Vol. 138 http://www.dagstuhl.de/dagpub/978-3-95977-117-7

Full version available: http://arxiv.org/abs/1906.09213

Regular Tree Expressions, Finite Tree Automata and Pushdown Automata

Tomáš Pecka

Regular tree expressions are a formalism for describing regular tree languages, which can be accepted by a finite tree automaton as a standard model of computation. It was proved in [3] that the class of regular tree languages is a proper subclass of tree languages whose linear notations can be accepted by deterministic string pushdown automata. In this talk, we firstly present regular tree expressions and then we present an algorithm for transforming regular tree expressions to equivalent real-time height-deterministic pushdown automata that accept the trees in their postfix notation. This algorithm is a modification of well-known Glushkov's algorithm [2, 1] for regular (string) expressions. This was published in [I] . This approach can also be modified so its output is not a pushdown automaton but a bottom up finite tree automaton. Lastly, we show the idea behind the opposite conversion, i.e., that every finite tree automaton can be transformed into equivalent regular tree expression.

List of publications:
Published:

[I] **Tomáš Pecka**, Jan Trávníček, Radomír Polách, Jan Janoušek; *Construction of a Pushdown Automaton Accepting a Postfix Notation of a Tree Language Given by a Regular Tree Expression* In: 7th Symposium on Languages, Applications and Technologies, SLATE 2018, June 21-22, 2018, Guimaraes, Portugal.

Submitted:

[II] Jan Trávníček, **Tomáš Pecka**, Robin Obůrka, Jan Janoušek; *Forward Linearised Tree Pattern Matching Using Tree Border Array* In: Language and Automata Theory and Applications - 14th International Conference, LATA 2020, Milan, Italy, March 2-6, 2020.

References

[1] Gérard Berry and Ravi Sethi. "From Regular Expressions to Deterministic Automata". In: *Theor. Comput. Sci.* 48.3 (1986), pp. 117–126. DOI: 10.1016/0304-3975(86)90088-5. URL: https://doi.org/10.1016/0304-3975(86)90088-5.

[2] V. M. Glushkov. "THE ABSTRACT THEORY OF AUTOMATA". In: *Russian Mathematical Surveys* 16.5 (1961). URL: http://stacks.iop.org/0036-0279/16/i=5/a=A01.

[3] Jan Janoušek and Bořivoj Melichar. "On Regular Tree Languages and Deterministic Pushdown Automata". In: *Acta Inf.* 46.7 (2009), pp. 533–547. DOI: 10.1007/s00236-009-0104-9. URL: http://dx.doi.org/10.1007/s00236-009-0104-9.

Tuning generative models for medical imaging data augmentation

Jakub Žitný

Generative Adversarial Networks (GANs) have been recently successful in several machine learning challenges, especially in niche computer vision problems. Datasets enlarged by synthetic data generated by GANs perform better in classification and segmentation tasks in comparison with traditional data augmentation methods. One of the areas where GANs helped significantly is medical imaging research where the size of datasets is among the biggest challenges. GANs have been used to improve medical datasets by translating labels to segmented data, segmented data to labels, generating alternative modalities (such as creating CT scans from MRI) or in-painting segmented parts into different regions of images. We review current state-of-the-art usage of GAN framework in the medical imaging domain and compare GAN data synthesis with traditional data augmentation methods.

GANs are widely used for mentioned use-cases and often lack proper evaluation mechanisms, e.g. with respect to adversarial examples. Comparing them with other generative models will be essential before we can safely use them in medical imaging practice. Exploring these problems will also be useful for neural networks interpretation and architecture optimisation.

List of publications:

Published:

[I] Cristina V. Lopes, Petr Maj, Pedro Martins, Vaibhav Saini, Di Yang, **Jakub Zitny**, Hitesh Sajnani, Jan Vitek; *DéjàVu: A Map of Code Duplicates on GitHub* In: Proceedings of the ACM on Programming Languages, Volume 1 Issue OOPSLA. ACM New York, NY, USA, October 2017. EISSN: 2475-1421.

Side-Channel Analysis of Cryptographic Hardware Implementations

Petr Socha

Cryptography has been evolving for a hundreds of years now, as a way to secure confident information against third party. Nowaday cryptographic systems include many diverse embedded devices, such as smartcards used e.g. for identification or for prepaid services, various IoT applications or even smart cars. While many ciphers currently in use (such as AES) are considered mathematically secure, their implementations may be vulnerable to side channel attacks, such as differential power analysis or its enhanced variant, correlation power analysis. This kind of attack exploits the fact that an intermediate value is processed in the implementation, that correlates with power consumption of the device, and with some other known information (e.g. plaintext or ciphertext). We have examined different approaches to the statistical computations necessary for the first-order [1] and arbitrary-order [4] side-channel analysis in order to perform these in a numerically stable, parallel and robust fashion, and we have evaluated the memory and time performance of these approaches, as well as their properties regarding practical usage. In order to sucessfully attack implementations in a noisy environment [2], we have proposed and evaluated a novel method for attack evaluation based on a correlation trace derivative [3 , 5], which significantly reduces a number of measurements required to mount an attack and in some cases makes the attack even feasible.

List of publications:

1. **Socha, P.**; Miškovský, V.; Kubátová, H.; Novotný, M. *Optimization of Pearson correlation coefficient calculation for DPA and comparison of different approaches* In: Proceedings of the 2017 IEEE 20th International Symposium on Design and Diagnotics of Electronic Circuit & Systems. Piscataway, NJ: IEEE, 2017. p. 184-189. ISSN 2473-2117. ISBN 978-1-5386-0472-4.

2. **Socha, P.**; Brejník, J.; Bartík, M. *Attacking AES Implementations Using Correlation Power Analysis on ZYBO Zynq-7000 SoC Board* In: 2018 7th Mediterranean Conference on Embedded Computing (MECO). Piscataway: IEEE, 2018. p. 29-32. ISBN 978-1-5386-5683-9.

3. **Socha, P.**; Miškovský, V.; Kubátová, H.; Novotný, M. *Correlation Power Analysis Distinguisher Based on the Correlation Trace Derivative* In: Proceedings of the 21st Euromicro Conference on Digital System Design. Piscataway: IEEE, 2018. p. 565-568. ISBN 978-1-5386-7376-8.

4. **Socha, P.**; Miškovský, V.; Novotný, M. *First-Order and Higher-Order Power Analysis: Computational Approaches and Aspects* In: Proceedings of the 8th Mediterranean Conference on Embedded Computing - MECO'2019. Institute of Electrical and Electronics Engineers, Inc., 2019. p. 83-87. ISSN 2377-5475. ISBN 978-1-7281-1739-3.

5. **Socha, P.**; Miškovský, V.; Kubátová, H.; Novotný, M. *Efficient algorithmic evaluation of correlation power analysis: Key distinguisher based on the correlation trace derivative* Microprocessors and Microsystems. 2019, 2019(71), 1-8. ISSN 0141-9331.

Dummy Rounds Method as Countermeasure against Side Channel Attacks

Stanislav Jeřábek

The Dummy Rounds protection scheme is intended to offer resistance against Side Channel Attacks (SCA) such as Differential Power Analysis (DPA) [4] to Feistel [3] and Substitution-Permutation [8] ciphers. Its principle is inspired by several well-known countermeasures used in hardware as Hiding [5] and Dynamic Logic Reconfiguration [6] as well as countermeasures used in software implementations as Dummy Cycles [2] or Random Order Execution [9]. The Dummy Rounds method, as we propose, combines software hiding in time with common hardware hiding of the circuitry power consumption. There are more parts of hardware design which are executed, but their outputs are randomly used or not used for computation in every single clock cycle. So, the structure of the design is the same for every clock cycle and also the power consumption stays the same. The final result stays correct due to round scheduling. Experimental evaluation of Dummy Rounds proposed above revealed weaknesses, most notably in the first and last round. The situation can be greatly improved by controlling the transition probabilities in the state space of the algorithm. We derived necessary and sufficient conditions for the round execution probabilities to be uniform and hence the minimum possible. The optimum trajectories over the state space are regular and easy to implement. For the dummy rounds scheme, there is always an optimum set of transition probabilities which makes the round execution probabilities uniform for a particular round now. This ensures maximum resistance against an SCA targeted to a particular round. A trajectory in the optimum set executes a random number of redundant rounds first, then all the active rounds, and then redundant rounds again. Now we are going to use better evaluation method. It seems, that Leakage Assessment at Register-Transfer Level [1] will be less time-consuming and more conclusive than T-Test [7] used till now.

List of publications:

1. **Stanislav Jeřábek**, Jan Schmidt, Martin Novotný; *Dynamic Reconfiguration as Countermeasure against DPA* In: Proceedings of the Work in Progress Session SEAA/DSD 2017. Linz: Johannes Kepler University, 2017. ISBN 978-3-902457-48-6.

2. **Stanislav Jeřábek**, Jan Schmidt, Martin Novotný, Vojtěch Miškovský; *Dummy Rounds as a DPA countermeasure in hardware* In: Proceedings of the 21st Euromicro Conference on Digital System Design. Piscataway: IEEE, 2018. p. 523-528. ISBN 978-1-5386-7376-8.

3. **Stanislav Jeřábek**, Jan Schmidt; *Analyzing and Optimizing the Dummy Rounds Scheme* In: Proceedings of the 22nd International Symposium on Design and Diagnostics of Electronic Circuits and Systems. Piscataway, NJ: IEEE, 2019. ISBN 978-1-7281-0072-2.

4. Petr Socha, Jan Brejník, **Stanislav Jeřábek**, Martin Novotný, Nele Mentens; *Dynamic Logic Reconfiguration Based Side-Channel Protection of AES and Serpent* In: Proceedings of the 22nd Euromicro Conference on Digital Systems Design. Los Alamitos, CA: IEEE Computer Soc., 2019. p. 277-282. ISBN 978-1-7281-2862-7.

References

[1] Miao (Tony)He et al. “RTL-PSC: Automated Power Side-Channel Leakage Assessment at Register-Transfer Level”. In: *37th IEEE VLSI Test Symposium (VTS’19)*. Apr. 2019. DOI: 10.1109/VTS.2019.8758600.

[2] Christophe Clavier, Jean-Sébastien Coron, and Nora Dabbous. “Differential Power Analysis in the Presence of Hardware Countermeasures”. In: *Cryptographic Hardware and Embedded Systems — CHES 2000*. Ed. by Çetin K. Koç and Christof Paar. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 252–263. ISBN: 978-3-540-44499-2.

[3] Horst Feistel. “Cryptography and computer privacy”. In: *Scientific american* 228.5 (1973), pp. 15–23.

[4] Paul Kocher, Joshua Jaffe, and Benjamin Jun. “Differential Power Analysis”. In: *Advances in Cryptology — CRYPTO’ 99*. Ed. by Michael Wiener. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 388–397. ISBN: 978-3-540-48405-9.

[5] S. Mangard, E. Oswald, and T. Popp. *Power Analysis Attacks*. 2007, p. 272.

[6] Pascal Sasdrich et al. “Achieving Side-Channel Protection with Dynamic Logic Reconfiguration on Modern FPGAs”. In: *Journal of Cryptographic Engineering* 2 (June 2014), pp. 107–121. ISSN: 2190-8508. DOI: 10.1007/s13389-013-0067-1.

[7] Tobias Schneider and Amir Moradi. “Leakage assessment methodology”. In: *Journal of Cryptographic Engineering* 6.2 (June 2016), pp. 85–99. ISSN: 2190-8516. DOI: 10.1007/s13389-016-0120-y. URL: https://doi.org/10.1007/s13389-016-0120-y.

[8] C. E. Shannon. “Communication theory of secrecy systems”. In: *The Bell System Technical Journal* 28.4 (Oct. 1949), pp. 656–715. ISSN: 0005-8580. DOI: 10.1002/j.1538-7305.1949.tb00928.x.

[9] Stefan Tillich, Christoph Herbst, and Stefan Mangard. “Protecting AES Software Implementations on 32-Bit Processors Against Power Analysis”. In: *Applied Cryptography and Network Security*. Ed. by Jonathan Katz and Moti Yung. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 141–157. ISBN: 978-3-540-72738-5.

Augmenting ATPG to Achieve Zero Aliasing in Output Response Compaction

Robert Hülle

Introduction

One of the long-standing problems in digital circuit testing is fault aliasing in the response compaction. Fault aliasing is an important source of coverage loss, especially if we strive to achieve a high compaction ratio. Existing methods to lower or eliminate aliasing mostly require changes to the compactor design [6, 4], while some have partial control over a test sequence, reordering test vectors but not influencing test patterns further, [2, 3, 1].

Method

We have proposed a method to lower or eliminate aliasing without the need to modify the compactor design. The basic idea is to constrain a test pattern generator (ATPG) itself to produce a test with zero aliasing. We are focusing on aliasing in temporal compaction; aliasing in spatial compaction is easier to prevent and can be solved independently of our method. In the rest of this abstract, we assume a spatial compactor that does not introduce new redundant faults.

Constraining the ATPG

Our method, *zero-aliasing test patterns generator* (ZATPG), is based on a Boolean-satisfiability ATPG (SAT-ATPG) [I] . An SAT-ATPG works by modeling the fault detection problem as a conceptual circuit, *miter*, with a fault inserted in a replica of the circuit under test (CUT). The miter is described as a Boolean formula in conjunctive normal form (CNF-SAT). The resulting SAT instance is solved by an SAT solver [5].
We expand the miter with anti-aliasing constraints in the following way. First, we construct classical miter for a tested fault f_i and generate a test pattern. All faults are simulated using the test pattern; note that this is a sequential simulation. For aliased faults $(f_{s1}, f_{s2}, \cdots f_{sm})$, we then append constraints that would prevent aliasing for these faults and generate different test pattern for the fault f_i. We repeat this process until we find a test pattern that causes no aliasing or we prove that no such test pattern exists. In practice, we have relaxed the requirement of zero aliasing to allow aliasing to occur if the overall fault coverage is increased.
Aliasing happens *after* the application of the test pattern that is being generated. To know the future state of the compactor, we need to unroll the combinational part of the compactor. The previous state of the compactor is also encoded in the SAT instance. Computation of the compactor state is then part of the SAT solving.

Results

In our experiments, we have used a simplification for linear compactors, where the future state can be partially precomputed. The experiments were performed on benchmark circuits from the ISCAS'85 and selected ITC'99 benchmarks.
For tested circuits, we have achieved zero aliasing and full fault coverage for smaller compactors, while keeping the same design (LFSR). With a fixed size of compactors, we have achieved lower aliasing and higher fault coverage, with an exception for very small compactors. For full coverage, observed gain in the size of compactors was between 2 and 5 bits of LFSR.

Presented results are part of [III] , [IV] , [V] , [VI] .

List of publications:

Published:

[I] **Hülle, R.**; Fišer, P.; Schmidt, J.; Borecký, J. *SAT-ATPG for Application-Oriented FPGA Testing* In: Proceedings of the 15th Biennial Baltic Electronics Conference. Tallin: Tallin University of Technology, 2016. p. 83-86. ISSN 1736-3705. ISBN 978-1-5090-1393-7.

[II] **Hülle, R.**; Fišer, P.; Schmidt, J. *Generování testu pro prostředky vestavěné diagnostiky* In: Počítačové Architektury & Diagnostika PAD 2016 - Sborník příspěvků. Brno: Vysoké učení technické v Brně, 2016, ISBN 978-80-214-5376-0.

[III] **Hülle, R.**; Fišer, P.; Schmidt, J. *SAT-based ATPG for Zero-Aliasing Compaction* In: Proc. of the 20th Euromicro Conference on Digital System Design. Piscataway, NJ: IEEE, 2017. p. 307-314. ISBN 978-1-5386-2146-2.

[IV] **Hülle, R.**; Fišer, P.; Schmidt, J. *Generování testu s nulovým maskováním poruch* In: Počítačové architektúry & diagnostika PAD 2017 - Zborník príspevkov. Bratislava: STU Scientific, 2017. pp. 35-38. ISBN 978-80-972784-0-3.

[V] **Hülle, R.**; Fišer, P.; Schmidt, J. *ZATPG: SAT-based Test Patterns Generator with Zero-Aliasing in Temporal Compaction* Microprocessors and Microsystems. 2018, 2018(61), 43-57. ISSN 0141-9331.

[VI] **Hülle, R.**; Fišer, P.; Schmidt, J. *ZATPG: SAT-based ATPG for Zero-Aliasing Compaction* In: Proceedings of the 6th Prague Embedded Systems Workshop. ČVUT v Praze, Fakulta informačních technologií, 2018. p. 8-11. ISBN 978-80-01-06456-6.

[VII] **Hülle, R.**; Fišer, P.; Schmidt, J. *PBO-Based Fault Selection for Compact Test Generation* In: Proceedings of the 7th Prague Embedded Systems Workshop. Praha: ČVUT FIT, Katedra číslicového návrhu, 2019. ISBN 978-80-01-06607-2.

Other Results:

[VIII] **Hülle, R.**; Fišer, P.; Schmidt, J. *ZATPG: Zero-Aliasing Test Pattern Generation* The Biannual European – Latin American Summer School on Design, Test and Reliability, Rotterdam, Netherlands, 2017 (unpublished poster presentation).

[IX] **Hülle, R.**; Fišer, P.; Schmidt, J. *Constrainig ATPG to Achieve Zero Aliasing* Vienna-Bratislava-Brno-Prague Joint Research Seminar for PhD Students, Brno, Czech Republic, 2017 (unpublished presentation).

Not Accepted:

[X] **Hülle, R.**; Fišer, P.; Schmidt, J.; Borecký J. *On Properties of ATPG SAT Instances* Euromicro Conference on Digital System Design, Architectures, Methods and Tools, 2016, Cyprus

[XI] **Hülle, R.**; Fišer, P.; Schmidt, J. *Compact Test Generation by Optimized Fault Selection* The 32nd IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems, 2019, Delft, Netherlands

[XII] Borecký J.; **Hülle, R.**; Fišer, P. *An Extended Fault Model for Application-Oriented FPGA Testing* The 32nd IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems, 2019, Delft, Netherlands

References

[1] T. Bogue et al. “Built-in self-test with an alternating output”. In: *Proceedings Design, Automation and Test in Europe*. Feb. 1998, pp. 180–184. DOI: 10.1109/DATE.1998.655854.

[2] Geetani Edirisooriya and P. Robinson John. “Test Generation to Minimize Error Masking”. In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 12.4 (Apr. 1993), pp. 540–549.

[3] Geetani Edirisooriya, P. Robinson John, and Samantha Edirisooriya. “On the performance of augmented signature testing”. In: *IEEE International Symposium on Circuits and Systems*. May 1993, pp. 1607–1610.

[4] M. Kopec. “Can nonlinear compactors be better than linear ones?” In: *IEEE Transactions on Computers* 44.11 (Nov. 1995), pp. 1275–1282. ISSN: 0018-9340.

[5] T. Larrabee. “Test pattern generation using Boolean satisfiability”. In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 11.1 (Jan. 1992), pp. 4–15.

[6] K. Pradhan D. and K. Gupta Sandeep. “A new framework for designing and analyzing BIST techniques and zero aliasing compression”. In: *IEEE Transactions on Computers* 40.6 (1991), pp. 743–763.

Friday 8 November 2019

Session 3

Session chair
Štěpán Starosta

Doctoral Research Days at FIT 2019

SAT modulo Differential Equations

Tomáš Kolárik

Many systems, namely embedded systems or cyber-physical systems, are nowadays insisted to satisfy high requirements which, in addition, often depend on physical phenomena of the real world. Formal verification is a convenient method to guarantee fulfillment of specifications of such complex systems, as it proves whether a mathematical model of a system fits given properties exactly. A widely used approach is SAT (Boolean satisfiability) [4], which, however, is unsufficient for modelling continuous behaviour of mentioned systems. ODEs, on the other hand, describe such phenomena natively, which is why it is a good idea to include them into the verification process.
We take advantage of efficient SAT solvers, which provide fast solutions of complex Boolean formulas. Current state-of-the-art solvers, however, which deal with ODEs and also embed a SAT solver, do not meet the needs of industry tasks too well: although they decide satisfiability of a formula exactly with arbitrary maximum error, based on interval arithmetic [3], their solution turns to be very time-consuming in case of complex models. Our ultimate goal is to use classic numerical methods [2], which are generally less precise and robust, but usually solve the differential equations much faster. We have already tested our current approach on several interesting experiments which deal with ODEs. Our tool [5] usually performed much faster than a state-of-the-art solver [1], which was not unexpected though. Still, we seek to further improve our methods to allow verification of real world systems with rich Boolean state space, which is still an unsolved problem so far. It will be the hard tasks which will actually show whether the precision of the solution being relaxed is limiting significantly, or not.

List of publications:

Submitted:

[I] Tomáš Kolárik, Stefan Ratschan; *SAT Modulo Differential Equation Simulations* To: 26th International Conference on Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2020, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2020, Dublin, Ireland, April 25–30, 2020.

References

[1] Sicun Gao, Soonho Kong, and Edmund M. Clarke. "dReal: An SMT Solver for Nonlinear Theories over the Reals". In: *Proceedings of the 24th International Conference on Automated Deduction*. CADE'13. Lake Placid, NY: Springer-Verlag, 2013, pp. 208–214. ISBN: 978-3-642-38573-5. DOI: 10.1007/978-3-642-38574-2_14.

[2] Kendall Atkinson et al. *Numerical Solution of Ordinary Differential Equations*. John Wiley, Feb. 2009, p. 272. ISBN: 978-0-470-04294-6.

[3] Julien Alexandre dit Sandretto and Alexandre Chapoutot. "Validated Explicit and Implicit Runge-Kutta Methods". In: *Reliable Computing electronic edition*. Special issue devoted to material presented at SWIM 2015 22 (July 2016). URL: https://hal.archives-ouvertes.fr/hal-01243053.

[4] Armin Biere et al. *Handbook of Satisfiability*. IOS Press, 2009.

[5] Tomáš Kolárik. *UN/SOT (UN/SAT modulo ODES Not SOT)*. 2019. URL: %7Bhttps://gitlab.com/Tomaqa/unsot%7D.

Missing Features Reconstruction in the Context of Asymmetric Heterogeneous Transfer Learning

Magda Friedjungová

The success of machine learning algorithms depends on data representation and the amount of available data for the training of a model. To gain sufficient information we sometimes have to combine datasets. The datasets can come from different domains represented by different feature spaces. These different datasets can be mapped one dataset to the other. This approach is related to asymmetric heterogeneous transfer learning methods. One of the very first tasks of transfer learning is the reconstruction of missing features - a scenario where entire features are missing while solving a machine learning task. We make use of traditional imputation methods such as k-NN, linear regression, and MICE within a not so traditional scenario where the reconstructed dataset is used in a predictive model trained on complete data. Furthermore, up-to-date methods for feature reconstruction such as multi-layer perceptron, gradient boosting trees, autoencoders, and variational autoencoders are presented. Our aim is to experimentally research the influence of various imputation methods on the performance of several predictive models. The experiments are performed on both real world and artificial datasets with continuous features. Within each experiment a varying number of features (ranging from 1 to 50%) is missing. The results show that MICE and denoising autoencoders can be considered as the most applicable feature reconstruction methods regardless of the amount of missing features or the classification model used. Moreover, the benefits of autoencoders lie in the fact that they represent universal imputers, which do not need to store training data or know possible missing feature combinations in advance. This is a clear advantage when the imputation has to be performed in an online production environment. Some results are presented as part of [1].

List of publications:

1. **Magda Friedjungová**, Daniel Vašata, Marcel Jiřina; *Missing Features Reconstruction and Its Impact on Classification Accuracy.* Computational Science - ICCS 2019, Springer International Publishing, pp. 207–220, 2019. ISBN 978-3-030-22744-9.

2. **Magda Friedjungová**, Marcel Jiřina; *An Overview of Transfer Learning Focused on Asymmetric Heterogeneous Approaches.* Data Management Technologies and Applications, Springer International Publishing, pp. 3–26, 2018. ISBN: 978-3-319-94809-6.

3. **Magda Friedjungová**, Marcel Jiřina; *Asymmetric Heterogeneous Transfer Learning: A Survey.* In: Proceedings of the 6th International Conference on Data Science, Technology and Applications, SciTePress, pp. 17–27, 2017. ISBN: 978-989-758-255-4.

Active semi-supervised learning in sequence labelling

Tomáš Šabata

Sequence labelling is a type of machine learning problem that involves assigning a label to each sequence member. This type of task can be found in many fields, such as object or activity recognition in video, speech recognition or natural language processing. In some of the areas, collecting of unlabelled data is cheap but labelling expensive. In such a situation, active learning can bring considerable improvement. In the standard active learning framework for sequence labelling, the most informative sequence is found and given to an annotator to be labelled. However, labelling the entire sequence may be inefficient as for some of its parts, the labels can be predicted using a model. Moreover labelling of these parts usually brigs only a little new information. To tackle this issue combination of active learning and semi-supervised learning, self-training, can be used.
In recent years, deep learning has shown supreme results in many sequence labelling tasks, especially in natural language processing. However, the standard deep learning tools do not capture model uncertainty well, as the neural nets are usually overconfident and can be uncertain despite high values of the soft-max function. Model uncertainty is indispensable for active learning as well as for semi-supervised learning. Therefore we aim to use Bayesian neural networks to overcome this issue. Our approach presented in 1 utilizes an approximation of Bayesian inference for neural nets using Monte Carlo dropout. Two proposed active learning query strategies outperform other existing techniques for deep neural nets in the field of natural language processing. Moreover, semi-supervised learning reduced the labelling effort by almost 80%.

List of publications:

1. **Tomáš Šabata**, Juraj Eduard Páll, Martin Holeňa; *Deep Bayesian Semi-Supervised Active Learning for Sequence Labelling* In: Proceedings of the Workshop on Interactive Adaptive Learning co-located with European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2019). CEUR Workshop Proceedings, 2019. p. 80-95. vol. 2444. ISSN 1613-0073.

2. **Tomáš Šabata**, Petr Pulc, Martin Holeňa; *Semi-supervised and Active Learning in Video Scene Classification from Statistical Features* In: Proceedings of the Workshop on Interactive Adaptive Learning (IAL 2018) co-located with European Conference on Machine Learning (ECML 2018) and Principles and Practice of Knowledge Discovery in Databases (PKDD 2018). Aachen: CEUR Workshop Proceedings, 2018. p. 24-35. ISSN 1613-0073.

3. Petr Pulc, Oliver Keruĺ-Kmec, **Tomáš Šabata**, Martin Holeňa; *Motion Segmentation by Semi-Supervised Classification in Dynamic Scenery* In: Proceedings of Poster Session of 3rd ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data (AALTD 2018). Dublin: University College Dublin, 2018. p. 65-72.

4. **Tomáš Šabata**, Tomáš Borovička, Martin Holeňa; *K-best Viterbi Semi-supervized Active Learning in Sequence Labelling* CEUR workshop proceedings. 2017, 2017 144-152. ISSN 1613-0073.

5. **Tomáš Šabata**, Tomáš Borovička, Martin Holeňa; *Modeling and Clustering the Behavior of Animals Using Hidden Markov Models* CEUR workshop proceedings. 2016, 2016(1649), 172-178. ISSN 1613-0073.

Physical Unclonable Functions on FPGAs

Filip Kodýtek

PUFs (Physical Unclonable Function) are increasingly used in proposals of security architectures for device identification and cryptographic key generation. Many PUF designs for FPGAs proposed up to this day are based on ring oscillators (RO). The classical approach is to compare frequencies of ROs and produce a single output bit from each pair of ROs based on the result of comparison of their frequencies. Such ROPUF design requires all ROs to be mutually symmetric and also the number of pairs of ROs is limited in order to preserve the independence of bits in the PUF response. This led us to design a new ROPUF on FPGA which is capable of generating multiple output bits from each pair of ROs and is also allowing to create higher number of pairs of ROs, thereby making the use of ROs more efficient than the classical approach. Our PUF design is based on selecting a particular part of a counter value and using it for the PUF output. In principle, this PUF design does not need the ROs to be mutually symmetric, however, it is shown that this ROPUF design has significantly better properties with varying supply voltage and temperature when symmetric ROs are used.

Moreover, we observed that we can use the same design as a TRNG (True Random Number Generator). This enables us to use the same design for various applications – PUF can be used for generating and storing cryptographic keys, TRNG for generating session and ephemeral keys, nonces and salts. The proposed TRNG design exhibited satisfactory behaviour as it passed NIST statistical test suite. However, in order to evaluate the TRNG thoroughly, a statistical model of its source of randomness is needed.

List of publications:

1. **Kodýtek, F.**; Lórencz, R.: A design of ring oscillator based PUF on FPGA. In *18th IEEE Symposium on Design and Diagnostics of Electronic Circuits and Systems*. April 22–24, 2015 – Belgrade, Serbia.

2. **Kodýtek, F.**; Lórencz, R.: Proposal and Properties of Ring Oscillator Based PUF on FPGA, 2016. In *Journal of Circuits, Systems and Computers*. March 2016, Vol. 25, No. 03. ISSN 0218-1266.

3. **Kodýtek, F.**; Lórencz, R.; Buček, J.: Improved ring oscillator PUF on FPGA and its properties. In *Microprocessors and Microsystems*. 2016, ISSN 0141-9331, http://dx.doi.org/10.1016/j.micpro.2016.02.005.

4. **Kodýtek, F.**; Lórencz, R.; Buček, J.; Buchovecká, S.: Temperature dependence of ROPUF on FPGA. In *Euromicro Conference on Digital System Design* (Poster). August 31 – September 2, 2016 – Limassol, Cyprus.

5. Buchovecká, S.; Lórencz, R.; **Kodýtek, F.**; Buček, J.: True Random Number Generator based on ROPUF circuit. In *Euromicro Conference on Digital System Design*. August 31 – September 2, 2016 – Limassol, Cyprus

6. Buchovecká, S.; Lórencz, R.; **Kodýtek, F.**; Buček, J.: True random number generator based on ring oscillator PUF circuit. In *Microprocessors and Microsystems*. 2017, ISSN 0141-9331, https://doi.org/10.1016/j.micpro.2017.06.021.

R Melts Brains: An IR for First-Class Environments and Lazy Effectful Arguments

Jan Ječmen

The R programming language combines a number of features considered hard to analyze and implement efficiently: dynamic typing, reflection, lazy evaluation, vectorized primitive types, first-class closures, and extensive use of native code. Additionally, variable scopes are reified at runtime as first-class environments. The combination of these features renders most static program analysis techniques impractical, and thus, compiler optimizations based on them ineffective. We present our work on PIR, an intermediate representation with explicit support for first-class environments and effectful lazy evaluation. We describe two dataflow analyses on PIR: the first enables reasoning about variables and their environments, and the second infers where arguments are evaluated. Leveraging their results, we show how to elide environment creation and inline functions.

List of publications:

1. Flückiger, Olivier and Chari, Guido and Ječmen, Jan and Yee, Ming-Ho and Hain, Jakob and Vitek, Jan *R Melts Brains: An IR for First-class Environments and Lazy Effectful Arguments* In: Proceedings of the 15th ACM SIGPLAN International Symposium on Dynamic Languages, pp. 55–66, 2019. ISBN 978-1-4503-6996-1.

Approximate string matching and k-mer analysis

Ondřej Cvacho, Jan Holub

Approximate string matching is an essential operation in many bioinformatics applications. For example, mapping reads from high-throughput sequencing onto a reference genome, or in the microarray probe design process. Approximate string matching is defined as a task to find all occurrences of given pattern P in a text T with some maximum number k of mismatches allowed. The number of mismatches is measured using distance metric called Hamming distance that expresses the minimum number of replace operations required to transform one sequence into another. Our solution is based on filtering technique for reduction of strings in Hamming neighbourhood of P. The idea is to generate only perspective strings in the neighbourhood of P and use a self-index for answering the exact pattern matching queries. Generation of strings in pattern neighbourhood is done by traversing de Bruijn Graph constructed from k-mers occurring in the indexed text, where k is fixed. Presented results are part of publication 1 and extended in collaboration with L. Hrbek in publication 2 .

Another task we work on now is a practical k-mer counting application. K-mer counters return the number of unique k-mers for any given k occurring in the given text. Our goal is to create a solution that counts the number of unique k-mers for arbitrarily large k up to the length of the text itself. Our proposed solution is based on Suffix array (SA) and Longest common prefix (LCP) array. SA contains starting positions of lexicographically sorted suffixes of the input text. LCP array contains the length of the longest common prefix of two consecutive suffixes in SA. The counting algorithm uses LCP values to count the number of unique k-mers. Extensions for counting canonical k-mers, counting in a collection of sequences, or ability to show a histogram of numbers of occurrences is required for any practical k-mer counter application and its usage by bioinformatics. K-mer counting in a collection of sequences is done by constructing SA from a sequence that is a concatenation of all sequences with a unique separator between every two sequences.

List of publications:

1. **Ondřej Cvacho**, Jan Holub; *Filtering Invalid Off-Targets in CRISPR/Cas9 Design Tools* In: Proceedings of Data Compression Conference 2018. New York: IEEE Computer Society Press, 2018. p. 403. ISSN 2375-0359.

2. **Ondřej Cvacho**, Lukáš Hrbek, Jan Holub; *Approximate string matching approaches for genomic data* In: ENBIK2018 Conference proceedings. Praha: Vysoká škola chemicko-technologická, 2018. pp. 48. ISBN 978-80-7592-017-1.

On Induced Online Ramsey Number of Paths, Cycles, and Trees

Václav Blažej

An online Ramsey game is a game between Builder and Painter, alternating in turns. They are given a fixed graph H and a an infinite set of independent vertices G. In each round Builder draws a new edge in G and Painter colors it either red or blue. Builder wins if after some finite round there is a monochromatic copy of the graph H, otherwise Painter wins. The online Ramsey number $\widetilde{r}(H)$ is the minimum number of rounds such that Builder can force a monochromatic copy of H in G. This is an analogy to the size-Ramsey number $\overline{r}(H)$ defined as the minimum number such that there exists graph G with $\overline{r}(H)$ edges where for any edge two-coloring G contains a monochromatic copy of H.

In this paper, we introduce the concept of induced online Ramsey numbers: the induced online Ramsey number $\widetilde{r}_{ind}(H)$ is the minimum number of rounds Builder can force an induced monochromatic copy of H in G. We prove asymptotically tight bounds on the induced online Ramsey numbers of paths, cycles and two families of trees. Moreover, we provide a result analogous to Conlon [On-line Ramsey Numbers, SIAM J. Discr. Math. 2009], showing that there is an infinite family of trees $T_1, T_2, \ldots$, $|T_i| < |T_{i+1}|$ for $i \geq 1$, such that

$$\lim_{i \to \infty} \frac{\widetilde{r}(T_i)}{\overline{r}(T_i)} = 0.$$

List of publications:

Parameterized Algorithms for Hard Problems

Josef Malík

Parameterized analysis provides a framework to closely describe complexity of NP-hard problems in specific circumstances. Precisely, it allows us to measure the efficiency of algorithms not only with respect to the size of the input instance, but also with respect to a given parameter. As a result, we are able to pulverize classical complexity results into finer classes.

First, we show an example of a parameterized algorithm within the description of result [III] . In this result, we consider the NP-hard subgraph isomorphism problem where, given two graphs G (source graph) and F (pattern graph), one is to decide whether there is a (not necessarily induced) subgraph of G isomorphic to F . While many practical heuristic algorithms have been developed for the problem, as pointed out by McCreesh et al. [3], for each of them there are rather small instances which they cannot cope. Therefore, an alternative approach that could possibly cope with these hard instances is of interest. We tackle this problem by parameterizing it by the treewidth of the pattern graph. Treewidth of a graph, roughly speaking, resembles a similarity of the graph to a tree. More precisely, we follow the approach introduced in a seminal paper by Alon, Yuster and Zwick [1], which uses the color coding principle to solve the problem. The main part of this approach is a dynamic programming over color subsets and partial mappings. As with many exponential-time dynamic programming algorithms, the memory requirements constitute the main limiting factor for its usage. Because these requirements grow exponentially with the treewidth of the pattern graph, all existing implementations based on the color coding principle restrict themselves to specific pattern graphs, e.g., paths or trees. In contrast, we provide an efficient implementation of the algorithm significantly reducing its memory requirements so that it can be used for pattern graphs of larger treewidth. Moreover, our implementation not only decides the existence of an isomorphic subgraph, but it also enumerates all such subgraphs (or given number of them).

Secondly, we present author's current research direction in the area of parameterization. The research direction is in the study of kernelization algorithms, which focuses on effective preprocessing of problem instances. The goal of kernelization is to design a polynomial-time subroutine, which solves parts of a problem instance that are easy (or in other words non-interesting) and reduces the instance to its kernel – an interesting part of the instance. As the kernel is still a computationally difficult instance of some problem, we naturally want the kernel sizes to be as small as possible. In particular, we are often interested in knowing, whether there exists a polynomial-sized kernel for some problem. However, there are problems that do not admit a polynomial kernel.

For this problems, we try to employ a stronger, generalized version of kernelization, called Turing kernelization. This type of kernelization introduces an oracle, which can be used in the preprocessing subroutine to decide the problem for possibly multiple reduced instances of the problem. There are several recent studies ([2, 4]) which focus on the topic of Turing kernelization and which closely describe hierarchies of Turing kernels.

List of publications:

Published:

[I] Matthias Bentert, **Josef Malík**, and Mathias Weller; *Tree Containment With Soft Polytomies*; In proceedings of the 16th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2018), pp. 9:1–9:14, ISBN 978-3-95977-068-2

[II] Tereza Klimošová, **Josef Malík**, Tomáš Masařík, Jana Novotná, Daniël Paulusma, and Veronika Slívová; *Colouring* $(P_r + P_s)$*-Free Graphs*; In proceedings of the 29th International Symposium on Algorithms and Computation (ISAAC 2018), pp. 5:1–5:13, ISBN 978-3-95977-094-1

Accepted:

[III] **Josef Malík**, Ondřej Suchý, and Tomáš Valla; *Efficient Implementation of Color Coding Algorithm for Subgraph Isomorphism Problem*; To appear in proceedings of the Special Event on Analysis of Experimental Algorithms (SEA^2 2019)

References

[1] Noga Alon, Raphael Yuster, and Uri Zwick. “Color-coding”. In: *J. ACM* 42.4 (1995), pp. 844–856.

[2] Danny Hermelin et al. “A Completeness Theory for Polynomial (Turing) Kernelization”. In: *Parameterized and Exact Computation*. 2013, pp. 202–215.

[3] Ciaran McCreesh et al. “When Subgraph Isomorphism is Really Hard, and Why This Matters for Graph Databases”. In: *J. Artif. Intell. Res.* 61 (2018), pp. 723–759.

[4] Jouke Witteveen, Ralph Bottesch, and Leen Torenvliet. “A Hierarchy of Polynomial Kernels”. In: *SOFSEM 2019: Theory and Practice of Computer Science*. 2019, pp. 504–518.

On Synchronizing Tree Automata and Their Work–Optimal Parallel Run, Usable for Parallel Tree Pattern Matching

Štěpán Plachý

Finite tree automaton is a standard model of computation for the class of regular tree languages. One of well studied principles in the theory of string languages is automata synchronization. A synchronizing word from any configuration sets a deterministic finite string automaton in a well-defined state. A stronger property of k-locality occurs when all words of length at least k are synchronizing. Such property is useful for parallelization of the run of the automaton. We present a way of synchronizing finite tree automata: We define a synchronizing tree pattern and a k-local deterministic finite bottom–up tree automaton. Furthermore, we present a work–optimal parallel algorithm for parallel run of the deterministic k-local tree automaton in $O(\log n)$ time with $\lceil \frac{n}{\log n} \rceil$ processors, for $k \leq \log n$, or in $O(k)$ time with $\lceil \frac{n}{k} \rceil$ processors, for $k \geq \log n$, where n is the number of nodes of an input tree, on EREW PRAM. Finally, we prove that the deterministic finite bottom–up tree automaton that is used as a standard tree pattern matcher is k-local with respect to the height of a tree pattern.

List of publications:

Accepted:

[I] **Štěpán Plachý**, Jan Janoušek; *On Synchronizing Tree Automata and Their Work–Optimal Parallel Run, Usable for Parallel Tree Pattern Matching* In: SOFSEM 2020.

Index