Supervisors' Opinion on the Ph.D. Thesis of

Vojtěch Havlena

Finite automata (FAs) are used pervasively in many areas of computer science, so efficient algorithms for manipulating them are in high demand. The uses of FAs are not only as acceptors of regular languages in programming languages' parsers or tools like `grep` but also as means for encoding (potentially infinite) sets of configurations of (infinite-)state systems. Although automata theory has been alive already for over half a century, state-of-the-art techniques for manipulation with automata are in many cases still too inefficient, which hinders their use in applications. The goal of the work of Vojtěch was therefore to significantly improve this situation by means of developing new techniques for *efficient manipulation with finite automata* (over finite and infinite words as well as finite trees) and their use in applications such as detecting malicious network traffic or in decision procedures of various logics.

The research of Vojtěch was supervised by me, co-supervised by dr. Ondřej Lengál, and conducted within the VeriFIT research group at the Faculty of Information Technology of the Brno University of Technology (FIT BUT). The research was an important part of multiple research projects including projects of the Czech Science Foundation (projects 16-17538S, 16-24707Y, 17-12465S, 19-24397S, 20-07487S, and 20-02328Y), the Czech Ministry of Education (the ERC.CZ project LL1908 and the NPUII project LQ1602), as well as two projects of the internal grant agency of the BUT (projects FIT-S-17-4014 and FIT-S-20-6427). The results achieved by Vojtěch were an important contribution to all these projects.

The main contributions of the research of Vojtěch Havlena presented in his thesis include the following:

- Development of two novel techniques for *reducing the size of finite word automata* used in detection of malicious network traffic by regular expressions (regexes). The first technique uses reduction with theoretical guarantees on the error given with respect to a probabilistic automaton modeling the network traffic (obtained via the technique of learning probabilistic automata). The other technique does not give theoretical guarantees but evaluates the error on a representative sample of network traffic. Using the techniques, together with a novel multi-stage design of a regex matching pipeline, enabled to significantly decrease HW resources needed to match a set of regexes, which allows one to use more matching units in parallel and achieve unprecedented speeds of regex-matching in a single-box network probe.

- Optimization of a decision procedure for the *weak monadic second order logic of k successors* (WSkS). The optimization develops a new framework for working with tree automata that allows one to use *lazy algorithms*. These lazy algorithms allow one, for instance, to test language emptiness of a tree automaton constructed using the operations of intersection, union, complement, and projection, without constructing it in the first place (and only exploring its state-space while using subsumption and other techniques to prune some parts). The proposed approach led to significantly improved performance of deciding certain classes of WSkS formulae over other state-of-the-art approaches. The paper presenting this approach received the **Best Paper Award** of CADE-27.

- *Antiprenexing for WSkS*. This contribution builds on the previous one and develops new ways for preprocessing WSkS formulae that helps to obtain formulae whose structure is good for the underlying solver. The high-level goal of the preprocessing is to move quantifiers as deep as possible (an approach sometimes called *antiprenexing*). Aggressive antiprenexing can sometimes make the job of the underlying solver harder, for instance, when

distributive laws are applied so that antiprenexing can be performed. The distributive laws can sometimes increase the size of the formula too much, negating the saving from the antiprenexing. This work develops a machine-learning-based algorithm that decides when to apply some rules. There is a strong experimental evidence that this preprocessing can significantly extend the limits of WSkS solvers.

- *Symbolic encoding* of Nielsen transformation for *solving string constraints*. The contribution encodes word equations into finite word automata and Nielsen transformation rules into finite word transducers. Solving word equations then amounts to solving an instance of the *regular model checking* problem, i.e., trying to decide whether the transitive closure of the relation induced by the transducers can lead to a configuration establishing equality of the word equations. The basic algorithm for word equations was extended in a non-trivial way to richer classes of strings constraints, such as Boolean combinations of word equations with length constraints and regular expression membership constraints. The approach was able to solve some hard benchmarks where state-of-the-art solvers could not provide the answer.

- A series of optimizations of rank-based *Büchi automata complementation*. The optimizations are based on (i) using *simulations* to prune redundant states of the complement (in a way that cannot be done by standard simulation-based reduction techniques) and (ii) defining the notion of *super-tight* runs, which are the only runs that need to stay in the complement, and pruning states for which it is known that they cannot be a part of any super-tight run. The optimizations significantly improved rank-based Büchi automata complementation, in many cases obtaining exponentially smaller automata. Also, when compared to other than rank-based techniques, in many cases obtained *the smallest* complement known.

The above mentioned works have been published in six papers at highly ranked international conferences (TACAS'18, FCCM'19, CADE-27, APLAS'19, LPAR'20, APLAS'20, CONCUR'21) and in two journal papers (International Journal on Software Tools for Technology Transfer and Journal of Automated Reasoning). Moreover, Vojtěch was the main developer of the tools that were used for evaluating the developped techniques. All the mentioned works have several co-authors, but we can acknowledge that Vojtěch contributed by key ideas as well as by a very sophisticated implementation and experiments to all of them.

During his Ph.D. studies, Vojtěch Havlena has proved to have creative abilities, independence, and to be able to work hard. He has also proved to be capable of a tight international cooperation with researchers from leading international teams. In our opinion, the thesis of Vojtěch Havlena satisfies all requirements usually associated with Ph.D. theses in the area of computer science, and we therefore recommend it to be accepted.

Brno, June 29, 2021

Prof. Ing. Tomáš Vojnar, Ph.D.

Ing. Ondřej Lengál, Ph.D.