

Supervisors' Opinion on the Ph.D. Thesis of

Lenka Turoňová

Lenka's dissertation focuses on regular pattern matching in performance critical applications such as scanning of large amounts of logs, emails, of network traffic, etc. The problem at hand is that, despite long development and extensive optimisation, all pattern matchers occasionally suffer from extreme spikes in the run-time. This may lead to failures of provided services. Especially painful consequence is a thread of ReDoS, regular expression denial of service attacks, where an attacker crafts a text to expose a weakness of the matcher, and consequently cause a failure of the service or an opening of a security hole.

Lenka focused on a feature of regular expressions called bounded repetition or counting which is particularly problematic. The run-time of matching with counting is in the worst case multiplied by the repetition bounds. Regexes where the worst case is possible are easy to write and hard to recognise. Counting regexes hence must be designed very carefully, or sometimes better avoided altogether, which limits their practical usability.

Lenka's main contributions is twofold. Firstly, she designed a matching algorithm that works in time independent of counter bounds for a large and practically significant class of regexes. The core of the problem is determinisation of automata with simple counters. Lenka's dissertation solves the problem by a novel determinisation algorithms. The latest version produces a new type of an automaton that stores sets of numbers in registers. This solves the problem in majority of real-life cases, and seems to be practical enough to be worth implementing in industrial solutions. Considering the practical relevance of matching and also that efficient matching with counting has been an open and thoroughly researched problem for decades, I see this as an unexpectedly good result.

Secondly, Lenka conducted a study of the impact of counting on existing matchers and proposed an algorithm for generating ReDoS vectors. It is a first such study that focuses on counting, and also a first study that focuses on automata-based matchers (they are an alternative to backtracking used in performance critical applications, faster and more stable). The study confirmed that counting is problematic for all kinds of matchers and revealed that it a serious Achilles heel of automata based matchers, contrary to the air of invincibility around them. Lenka's ReDoS generator was the only one which could touch the automata-based matchers. Some potential ReDoS vectors were discovered that could threaten most performance focused implementations of matching, such as the Snort network intrusion detection system or a hardware implementation of matching in a dedicated card by Invidia.

I consider these results exceptionally good from theoretical as well as from practical perspective. They were also published at the most competitive venues, mainly OOPSLA, USENIX-security, and APLAS. The results are a team work. Lenka was a key member of the team, provided an indispensable input in designing the solutions and carried out nearly all implementation and experimental work. The focus on relevance in practice made this very demanding, we needed to deal with huge databases of regular expressions written in various dialects and with special quirks of tens of practical matchers and ReDoS generators. Showing a ReDoS vulnerability of a real industrial solution is harder than examining a theoretical algorithm since (among other things) the engineers are aware of weaknesses and try to defend against them in pragmatic ways. Implementing a matcher which can compete against long developed industrial tools is a challenge for similar reasons, no matter a superior base algorithm. Showing that the said results are indeed relevant for practice was therefore possible only due to Lenka's exceptionally diligent and focused effort.

Lenka's Ph.D. took a long time. We have started with another problem which turned out more difficult than expected. The results started coming only after we changed the direction and it requi-

red some grit to persist. I am grateful to Lenka that she did, the results are worth it. I recommend the thesis for acceptance.

Brno, May 13, 2021

doc. Mgr. Lukáš Holík, Ph.D.

Prof. Ing. Tomáš Vojnar, Ph.D.