Technische Universität Braunschweig Institut für Theoretische Informatik Prof. Dr. Roland Meyer

Mühlenpfordtstr. 23
38106 Braunschweig
Phone: 0531 - 391 9524
Email: roland.meyer@tu-bs.de

prof. Ing. Lukáš Sekanina, Ph.D.
Department of Computer Systems
Brno University of Technology

June 9, 2022

Report on the PhD thesis of Lenka Turoňová

Matching a text to a regular expression (regex) is an evergreen topic in Computer Science with thousands of essential applications. In recent years, the subject has moved into the focus of security research with so-called *regex denial of service* (ReDoS) attacks. The idea of ReDoS is that an attacker provides an instance of a matching problem that is extremely hard to solve as an input to a system, and the computational effort of matching lets the system react in undesirable ways. For example, consider a web application in which a sanitizer is supposed to check input strings for harmful commands. An attacker may provide an input string where the computational effort of sanitization considerably slows down the system, perhaps to the degree that it becomes unresponsive. The system may react to the time consumption and stop the sanitization, in unfortunate cases so early that the harmful part of the input goes unnoticed.

Contribution

The present thesis is the first systematic study on the vulnerability of regex matchers to ReDoS attacks. The starting point is an experimental comparison that reveals a difference in vulnerability depending on the matching technique. Backtracking-based matchers are susceptible to even simple ReDoS attacks, while automata-based algorithms routinely match large texts against ordinary regexes but react sensibly to the use of counting/repetition constraints. The precise problem tackled in the thesis is therefore to mitigate the vulnerability of automata-based matchers to ReDoS attacks that are based on counting constraints. The contributions are as follows.

- A novel automaton model, so-called counting automata, that forms the algorithmic backbone of new matching algorithms.
- An elegant translation of regular expressions with counting constraints to counting automata. The translation generalizes Antimirov's idea of partial derivatives to the more general regexes and to counting automata.
- A clever determinization of counting automata. The key aspect is that the result is again a counting automaton, hence a symbolic device with a good chance of remaining compact. Experiments confirm that deterministic counting automata are exponentially more succinct than the corresponding deterministic finite automata.
- Monadic regexes as an important subclass of counting regexes that makes up 95% of practical instances.
- A specialization of the determinization to counting automata resulting from monadic regexes. A theoretical result shows that the size of the deterministic automaton is only polynomial in the highest repetition value.
- An ingenious data structure that can manipulate a set of counter values in constant time, and an automaton model that makes use of it.

- A determinization of counting automata into such counting-set automata that generalizes the well-known subset construction. Importantly, the determinization runs in time independent of the counting constraints in the regular expression, which I consider the main (and a very unexpected) achievement in the thesis.
- A ReDoS attack generator constructed from counting-set automata. This is the first device that allows us to evaluate the vulnerability of automata-based matchers against ReDoS attacks using counting constraints.
- All models and algorithms have been implemented, exercised on large data sets, and the results documented in careful evaluations.
- All theoretical results have been worked out with care and proven correct.

The thesis has a total length of 119 pages and six chapters.

Judgement

I will judge the results according to different measures.

Importance of topic and contribution ReDoS attacks are more prominent than ever, and the fact that not even the vulnerability of matcher types was understood before the thesis came about should make clear the importance of the contribution. With the ReDoS attack generator, we have the first device at hand that allows us to judge the vulnerability of a system, and with counting-set automata the first matcher that has a guaranteed independence from counting constraints.

Novelty and depth The invention of counting-set automata is nothing but an ingenious move. The key idea is to increment a set of counter values by incrementing a single value that serves as an upper bound for all of them — brilliant! Also the idea of using counting-set automata the other way around, namely to generate attacks, left me stunned.

Quality The technicalities have been worked out with mastership. The seeming ease with which classics like Antimirov's translation and Rabin and Scott's powerset construction are generalized to new and highly non-trivial automata models is fantastic. Also the experiments, which are based on considerable implementation efforts, have been done with care and documented in minute detail.

Breadth and related work The thesis links regex matching to the modern field of computer security, and manages to make high-class contributions to both of them. This demonstrates the breadth of topics Lenka Turoňová manages to cover, from compilation over determinization to attack generation. Let me say that contributing to a single field already means studying a body of literature. Lenka Turoňová has carefully placed her contributions into an extraordinarily large landscape of related work.

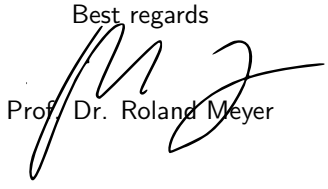
Write-up The write-up is excellent. A broad picture of what has been achieved in the context of ReDoS and what is the problem to tackle next serves as a careful motivation for the problem of interest. When it comes to the technical sections, introductory paragraphs written at precisely the right level of abstraction make clear the contribution while informal explanations accompanied by graphical illustrations help the reader grasp technical constructions.

Publications The results have been published at venues of highest rank, namely APLAS, OOPSLA, and USENIX Security.

Lenka Turoňová made a considerable step in understanding and mitigating ReDoS attacks. The results are an impressive demonstration of technical proficiency, comprehensive knowledge, creativity, and brilliance. I recommend the acceptance of the thesis and grade it

excellent (summa cum laude).

Best regards



Prof. Dr. Roland Meyer