FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

Review report of a final thesis

Reviewer:	Ing. Petr Máj, Ph.D.
Student:	Andrey Olos
Thesis title:	Java vulnerability to regular expression denial of service
Branch / specialization:	Computer Security and Information technology
Created on:	19 August 2023

Evaluation criteria

1. Fulfillment of the assignment

▸ [1] **assignment fulfilled**
[2] assignment fulfilled with minor objections
[3] assignment fulfilled with major objections
[4] assignment not fulfilled

The thesis topic is quite broad and complex, requiring the student to familiarize with a relatively diverse set of topics such as OpenJDK architecture, regular expresissions, performance and vulnerability testing and so on. While there are objections to the presented results, these are mostly omissions, that are to be expected at undergraduate thesis of this breadth.

2. Main written part

75 /100 (C)

The written part is concisely written and analyzes first the problem of ReDOS attacks, the offending regular expressions themselves and the root causes for the computational explosion. The text feels a bit fragmented and a different format, such as first introducing the problem in its entirety and then focusing on the offending patterns one by one showing their structure, node graphs and performance visualizations might be easier to read. Relatively frequent typos do not help the readability either.

In chapter 1, the definitions seem to be mostly dumped for completeness, expanding the accompanying text would be very useful. Some things, such as the OpenJDK itself should really not be defined in my opinion, but rather cited.

Chapter describing the patterns would greatly benefit for better explanation of the methodology - some patterns were not found by the author itself (this is clearly marked in the thesis), but the pattern that was contains too little information about its discovery. Was this just a random-ish permutation of the existing patterns? Was it some kind of a more exhaustive search? What is the likeliood of new patterns being identified?

The graphs in the evaluation chapter should use log scale on the vertical axis, the non-exponential parts of them are just flattened. The grid size that breaks at 0.5 input characters makes them harder to read as well.

3. Non-written part, attachments 100 /100 (A)

The code was not expected to be part of the thesis judging from the topic description. A cursory look at them reveals no problems.

4. Evaluation of results, publication outputs and awards 90 /100 (A)

I am very optimistic that if the shortcomings described will be fixed, the thesis can form a basis of a research paper.

The overall evaluation 85 /100 (B)

Overall, the thesis addresses a very broad topic with enough details and contributions. The written part would benefit from better structuring, more depth in certain parts (most notably the pattern discovery) and clearer evaluation, but with respect to its undergraduate level, presents a significant and valuable contribution.

Questions for the defense

1) Why were two different time measurements used? Woudn't nanoTime suffice for the longer times as well? What was the resolution of the nanoTime call in your setup?
2) Would limiting not the time itself, but the amount of backtracking, or state changes, make sense for preventing the DoS as well?

Instructions

Fulfillment of the assignment

Assess whether the submitted FT defines the objectives sufficiently and in line with the assignment; whether the objectives are formulated correctly and fulfilled sufficiently. In the comment, specify the points of the assignment that have not been met, assess the severity, impact, and, if appropriate, also the cause of the deficiencies. If the assignment differs substantially from the standards for the FT or if the student has developed the FT beyond the assignment, describe the way it got reflected on the quality of the assignment's fulfilment and the way it affected your final evaluation.

Main written part

Evaluate whether the extent of the FT is adequate to its content and scope: are all the parts of the FT contentful and necessary? Next, consider whether the submitted FT is actually correct – are there factual errors or inaccuracies?

Evaluate the logical structure of the FT, the thematic flow between chapters and whether the text is comprehensible to the reader. Assess whether the formal notations in the FT are used correctly. Assess the typographic and language aspects of the FT, follow the Dean's Directive No. 52/2021, Art. 3.

Evaluate whether the relevant sources are properly used, quoted and cited. Verify that all quotes are properly distinguished from the results achieved in the FT, thus, that the citation ethics has not been violated and that the citations are complete and in accordance with citation practices and standards. Finally, evaluate whether the software and other copyrighted works have been used in accordance with their license terms.

Non-written part, attachments

Depending on the nature of the FT, comment on the non-written part of the thesis. For example: SW work – the overall quality of the program. Is the technology used (from the development to deployment) suitable and adequate? HW – functional sample. Evaluate the technology and tools used. Research and experimental work – repeatability of the experiment.

Evaluation of results, publication outputs and awards

Depending on the nature of the thesis, estimate whether the thesis results could be deployed in practice; alternatively, evaluate whether the results of the FT extend the already published/known results or whether they bring in completely new findings.

The overall evaluation

Summarize which of the aspects of the FT affected your grading process the most. The overall grade does not need to be an arithmetic mean (or other value) calculated from the evaluation in the previous criteria. Generally, a well-fulfilled assignment is assessed by grade A.