FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

# Supervisor's statement of a final thesis

| | |
|---|---|
| **Student:** | Bc. Jan Vojtěšek |
| **Supervisor:** | Ing. Josef Kokeš |
| **Thesis title:** | Novel approaches to the detection of backdoors |
| **Branch of the study:** | Computer Security |

**Date:** 30. 5. 2020

| *Evaluation criterion:* | *The evaluation scale: 1 to 4.* |
|---|---|
| **1. Fulfilment of the assignment** | ***1 = assignment fulfilled,***<br>*2 = assignment fulfilled with minor objections,*<br>*3 = assignment fulfilled with major objections,*<br>*4 = assignment not fulfilled* |

*Criteria description:*
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.

*Comments:*
The student not only performed all the requirements of the assignment, but in pretty much all the parts went far beyond these requirements and created an admirable piece of work. Considering that the requirements themselves were enough for two or three theses, his product can't be rated as anything short of "amazing".

| *Evaluation criterion:* | *The evaluation scale: 0 to 100 points (grade A to F).* |
|---|---|
| **2. Main written part** | *99 (A)* |

*Criteria description:*
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. 26/2017, 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.

*Comments:*
The textual content of the thesis is very nearly perfect. It presents a very in-depth, very detailed study of the backdoor techniques which is still quite clear and easy enough to understand. It is also written in some of the best English I have seen since I started reviewing theses. My only minor complaint is that I would have preferred fewer commas in the sentences, but I understand that this is mostly a subjective issue.

| *Evaluation criterion:* | *The evaluation scale: 0 to 100 points (grade A to F).* |
|---|---|
| **3. Non-written part, attachments** | *100 (A)* |

*Criteria description:*
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.

*Comments:*
The student created two major pieces of software for this thesis: A tool for extraction of relevant information from the executable files and a set of heuristic detectors of backdoors. The extractor implementation was somewhat simplified by the fact that it is not a self-sufficient solution but rather an add-on to IDA Pro, but that is perfectly acceptable - any other approach would basically require writing a brand new decompiler, which would take high years or low decades, with no clear benefit from this approach. The code for the heuristic modules is rather simple, but that's actually good - it shows that the student did select reasonable statistics from the executable files and extracted them into such a format that it is easy to process. That gives a high probability that more heuristics could be created based on the same once-extracted data, saving a lot of effort in the process (as the extraction is by far the most resource-consuming part of the process).

That the student in addition to all that also performed reverse engineering of about 30 (!) malicious samples, is beyond words. Analyzing *one* such sample is sufficient for a master's thesis!

| *Evaluation criterion:* | *The evaluation scale: 0 to 100 points (grade A to F).* |
|---|---|

| | |
|---|---|
| **4. Evaluation of results, publication outputs and awards** | *100 (A)* |

*Criteria description:*
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.

*Comments:*

The presented work provides a framework for an important and heretofore rather neglected part of a fight against malware - detecting new, previously unknown backdoors. This is a highly specialized topic, which prevents the created tools from being used be "an average user", but I am confident they will be very useful to the experts in this field: certainly to anti-malware professionals, and in a simplified form to the maintainers of software repositories (who can use the tools to verify that the submitted packages seem to be free of hidden backdoors) and software vendors (who can use the tools to verify that their build process did not change unexpectedly). The highly developer-friendly license helps here, too. The dependency on the commercial IDA Pro package may seem a bit detrimental, but IDA Pro is a de-facto standard tool in the field and all professionals can be expected to have access to it. Also, since the software is split into two parts, the extraction of the metadata can be provided as a separate service.

| *Evaluation criterion:* | *The evaluation scale: 1 to 5.* |
|---|---|
| **5. Activity and self-reliance of the student** | *5a:*<br>***1 = excellent activity,***<br>*2 = very good activity,*<br>*3 = average activity,*<br>*4 = weaker, but still sufficient activity,*<br>*5 = insufficient activity*<br>*5b:*<br>***1 = excellent self-reliance,***<br>*2 = very good self-reliance,*<br>*3 = average self-reliance,*<br>*4 = weaker, but still sufficient self-reliance,*<br>*5 = insufficient self-reliance.* |

*Criteria description:*
From your experience with the course of the work on the thesis and its outcome, review the student's activity while working on the thesis, his/her punctuality when meeting the deadlines and whether he/she consulted you as he/she went along and also, whether he/she was well prepared for these consultations (5a). Assess the student's ability to develop independent creative work (5b).

*Comments:*

No complaints in this department, either. The student excelled in his chosen field two years ago and apparently the passage of time only improved his skills.

| *Evaluation criterion:* | *The evaluation scale: 0 to 100 points (grade A to F).* |
|---|---|
| **6. The overall evaluation** | *100 (A)* |

*Criteria description:*
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.

*Comments:*

Pretty much everything about this thesis is perfect so it's difficult to pick one specific aspect, but I would like to bring a particular attention to the very clear research of all backdoor-related topics in the textual part. It's a highly specialized topic but I am convinced the student managed to convey all the key ideas exceptionally well, so that even beginners in the field should be able to follow the work easily enough. And we should not forget the sheer scope of the thesis - the work presented here would be considered admirable even if performed by a whole research team! There's no doubt whatsoever in my mind that this thesis deserves and A and a recognition as one of the best theses created on this faculty, ever.

Signature of the supervisor: