FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

Review report of a final thesis

Reviewer:	Ing. Tomáš Čejka, Ph.D.
Student:	Anton Aheyeu
Thesis title:	Automated Creation of TLS Fingerprinting Database
Branch / specialization:	Computer Security and Information technology
Created on:	29 May 2022

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 submitted bachelor's thesis deals with development of tools to automatically generate annotated datasets of TLS traffic for fingerprinting. The presented outcomes of the thesis are a plugin for the ipfixprobe flow exporter and its integration with the OSQuery system that collects information about communication by running processes in OS.
Using the developed ipfixprobe extension, the thesis demonstrated creation of the TLS fingerprinting database.

2. Main written part 85 /100 (B)

The text is well written and well organized, but it contains some incaccuracies.

Dividing network monitoring into just 2 types ("DPI" and "Flow-based monitoring") is not accurate. Also, the author claims it is disadvantage to use DPI because of: "the fact that careful processing of packages can slow down a network." (page 4)
The author probably meant "packets" and a specific case when a monitoring process with DPI runs actively inside a network device, such as router. This is, however, not the case when passive monitoring is deployed.

Page 5: the author claims NetFlow v5 is the most used version of NetFlow and cites a URL to Cisco Systems about NetFlow v5. This page is no longer available, however, based on the URL I assume it was rather about format description than statics of use.

Page 8: Explanation of TLS should be improved.

There are some minor languague issues.

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

The implementation of the ipfixprobe plugin works and is already merged into publicly available open source repository.

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

Outcomes of the thesis are useful for network traffic monitoring on end devices. The aim is to annotate IP flow records with process identification.

The overall evaluation 95 /100 (A)

The quality of the submitted thesis is high. The text part as well as practical implementation part are well elaborated. The outcomes of the thesis are useful for practical deployment and can be used for creation of annotated datasets for network traffic classification and detection of security threats.

Questions for the defense

Is there any chance to annotate even short network connections?
Have You investigated any way to override functionality of system calls to log established connections?

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.