FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

Supervisor's statement of a final thesis

Supervisor:	Dawid Machala, Ph.D.
Student:	Bc. Mikuláš Formánek
Thesis title:	Automatic detection of malicious activity in the internal network
Branch / specialization:	Computer Security
Created on:	30 May 2021

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 has been dedicated to the automatic detection of malicious activity within the internal network of Showmax. The topic has been sufficiently investigated, and all the objectives have been achieved:
- A literature survey has been conducted and led to selection of several algorithms for unsupervised anomaly detection.
- Showmax's logs have been analyzed and transformed into a dataset on which the selected algorithms have been tested.
- A valid proof-of-concept of detection mechanism has been implemented in Python
- Sensitivity analysis has been performed on the solution; its results suggest that the algorithms have been correctly implemented.

2. Main written part 80 / 100 (B)

The report is written in a mostly clear manner, with minor punctuation issues (such as: occasional lack of spaces before opening brackets, inconsistent use of commas and colons, duplicated dots, etc.) and occasionally confusing grammar in some of the sections. The structure is mostly correct: the work starts with a general theoretical introduction, then it describes a detailed survey of modern research literature, followed by analysis of Showmax's infrastructure, implementation of the proposed solution and testing its performance. There are occasional sections that seem a bit out of place (for example, section 3.1.4 describes the general notions of precession, recall, and accuracy; these notions would fit better to the first chapter, where the general theory is described) and some of the conclusions are not clearly justified (e.g., in 2.1.7. DeepAnt is suggested as an

optimal choice, but afterwards in 3.1.4. it is considered merely as one of several approaches. Had it been optimal, why the need for more algorithms? The decisions made are clear, but their description is not always).

On the factual level, the thesis is developed correctly, without any visible mistakes and problems. Some of the decisions taken in the thesis could benefit from a more in-depth commentary. For example, reasons for choosing independent ensembles over sequential ensembles are suggested in the first chapter, but are never again referred to, even in the third chapter, where the actual selection takes place. Especially the first chapter is written in a very broad way, with sometimes not enough depth. It could be better highlighted which aspects of it will be used in the future chapters. However, once again, the thesis seems correct on the factual level.

Novel contributions are clearly distinguished from these found in the subject literature. All the methods are linked to the relevant research papers, and quotes are visibly distinguishable using italics. Some of the introductory descriptions in the first chapter could benefit from denser citations (e.g., the taxonomy of detection systems in section 1.1.), and description of individual algorithms would benefit from citing more than one source per algorithm, but overall, the thesis follows the principles of academic integrity.

License, confidentiality and privacy laws have been observed by the student.

3. Non-written part, attachments 95 /100 (A)

The code has been written in Python in Jupyter Notebooks, which seems to be the optimal choice for the type of analysis. Several algorithms have been implemented, and their performance has been tested using sensitivity analysis. Overall, the tools and approach has been appropriate for developing the Proof of Concept. The clarity of the code could be improved, but is on the correct level.

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

The Proof of Concept has been successfully developed, and its implications will be used in the future works of the Analytics team at Showmax. The literature survey, being one of the thesis outcomes, could be extended, but meets the expectations.

5. Activity of the student

▸ [1] **excellent activity**
[2] very good activity
[3] average activity
[4] weaker, but still sufficient activity
[5] insufficient activity

The student has been punctual, responsive, and reported the progress on a weekly basis. Suggestions made during these meetings were then usually implemented.

6. Self-reliance of the student

▸ [1] **excellent self-reliance**
[2] very good self-reliance

[3] average self-reliance
[4] weaker, but still sufficient self-reliance
[5] insufficient self-reliance

The student has shown good initiative, has been consulting his problems with several people within the company, and overall has proven well-capable of conducting independent work.

The overall evaluation 85 /100 (B)

The thesis fulfills the suggested objectives. Several improvements to the structure of the thesis, as well to the justification of decisions and conclusions could be made. However, overall, the thesis meets the necessary expectations.

I do, therefore, recommend the student to obtaining the master's degree on the basis of the presented thesis.

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. 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.

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.

Activity of the student

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.

Self-reliance of the student

From your experience with the course of the work on the thesis and its outcome, assess the student's ability to develop independent creative work.

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.