CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS SUPERVISOR'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Analyzing the execution of malware in a sandbox using hierarchical multiple instance learning**
Author's name:	**Vojtěch Kozel**
Type of thesis :	Bachelor Thesis
Faculty/Institute:	Faculty of Electrical Engineering
Department:	Department of Control Engineering
Thesis reviewer:	Tomáš Pevný
Reviewer's department:	Department of Computer Science

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **A**

How demanding was the assigned project?

The assignment was demanding, since the student had to master two very different fields: the computer (network) security and machine learning (statistic), where he has also studied different models.

Fulfilment of assignment **A**

How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.

The assignment was fulfilled completely.

Activity and independence when creating final thesis **A**

Assess whether the student had a positive approach, whether the time limits were met, whether the conception was regularly consulted and whether the student was well prepared for the consultations. Assess the student's ability to work independently.

The student was very active. For discussions of his progress, which has been held every week, he had came with a list of problems (and questions) he needs to discuss.

Technical level **A**

Is the thesis technically sound? How well did the student employ expertise in his/her field of study? Does the student explain clearly what he/she has done?

The level of the thesis is good. The student has demonstrated he can master theoretical and also the practical aspect of the problem.

Formal level and language level, scope of thesis **A**

Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS SUPERVISOR'S REPORT

The language and organisation of the thesis is acceptable. While there is still room for improvement, I consider it sufficient for.

Selection of sources, citation correctness **A**

Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?

Forty citations correlates with the volume of information the student went through.

Additional commentary and evaluation (optional)

Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.

Please insert your comments here.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

As mentioned on the beginning, the complexity of the thesis is in the fact that the student had to study two topics: computer security and machine learning. In the latter field, he studied three very different approaches to apply neural networks to the structured data in a time series. The experimental comparison has met goals of initial study, yet to make it statistically valid it is needed to repeat the experiments multiple times, which has been skipped due to the high computational complexity. Nevertheless the study is interesting, as it clearly demonstrates that exploiting structure of the data simplifies the training of classifiers, leading to better accuracy of the prediction. I also consider this work to be a starting point, as I believe that are interesting unanswered question, for example if a time dependency between packets is important for classification or not. The thesis therefore builds a solid foundation for a future work.

The grade that I award for the thesis is **A**

Date: 4.6.2021 Signature:

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Hierarchical models of network traffic**
Author's name:	**Vojtěch Kozel**
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Control Engineering
Thesis reviewer:	Dr Fabio Pierazzi
Reviewer's department:	Department of Informatics, King's College London, UK

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment — **challenging**

How demanding was the assigned project?

The complexity of the project is more than adequate, and especially challenging for a bachelor student.

Fulfilment of assignment — **fulfilled**

How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.

The thesis fulfills the assigned task fully and thoroughly. It is very appreciated that the student compared many alternative methods, and also managed to implement a few variants of the Multi-Instance learning method.

Methodology — **outstanding**

Comment on the correctness of the approach and/or the solution methods.

The proposed methodology and evaluation is correct and very good. The student followed systematical evaluation and comparison with other approaches, and defined many possible variants from the state of the art. The actual dataset used for the evaluation is relatively small, but it is more than adequate for the scope of a BSc thesis.

Technical level — **A - excellent.**

Is the thesis technically sound? How well did the student employ expertise in the field of his/her field of study? Does the student explain clearly what he/she has done?

The technical level of the thesis is very high and well done. There are some cases there is some confusion in presentation of the high-level methods, especially for the HMill variants and its results – for example it is unclear why using the world map to visualize HMill results. Nevertheless, the project's technical level is very high.

Formal and language level, scope of thesis — **B - very good.**

Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?

The language is generally very good, despite some occasional areas in which the explanation of the results could have been clearer, and better compare-and-contrasted between different approaches.

Selection of sources, citation correctness — **A - excellent.**

Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?

Related work and sources are relevant and appropriately cited. This is excellent also when considering that the candidate is only at the Bachelor level.

Additional commentary and evaluation (optional)

Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS REVIEWER'S REPORT

Please insert your comments here.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

Summarize your opinion on the thesis and explain your final grading. Pose questions that should be answered during the presentation and defense of the student's work.

The grade that I award for the thesis is **A - excellent.**

The thesis tackles a state-of-the-art challenge with an innovative solution based on multi-instance learning. There is a comparison also with other possible methods that treated network traffic communication as images and sequences, and there is an attempt at explaining the identified results in all scenarios. A thorough comparison of the proposed approaches is presented in a formal and convincing way. Despite some minor lack of clarities here and there when presenting the multi-instance learning approach, the technical level of the thesis and the soundness of the findings is very strong, as well as the citations and sources reported by the student. I am hence recommending to award the thesis the mark of A – "excellent".

Date: **25.5.2021**

Signature: