THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Malware detection based on call graph similarities**
Author's name:	**Štěpán Dvořák**
Type of thesis :	Master
Faculty/Institute:	FEL
Department:	katedra počítačů
Thesis reviewer:	Tomáš Pevný
Reviewer's department:	katedra počítačů

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **A**

How demanding was the assigned project?

The assignment is of standard difficulty in Open Informatics. The point two and three seems to be very vague, as "propose a representation" means anything and it will be always fulfilled.

Fulfilment of assignment **B**

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.

I am not entirely satisfied with fulfilling the last bulet, namely the comparison to the state of the art. The method was compared to unpublished industrial solution that uses different information about binaries than those used by the proposed algorithm. This information is interesting, but I am missing the comparison with the state of the art that uses call graphs. This means that we cannot compare the proposed Graph Neural Networks (GNN) to the SOTA. I had the impression that the student fall into the trap of playing with GNN framework rather than focusing on the solved problem.

Methodology **A**

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

Methodology is sound.

Technical level **A**

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?

Technical level is good.

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

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?

I have found some parts of the thesis, mainly sections 2 and 3 very superficial with a lots of structure in the text, which further decreases the content. A reader is frequently referred to the references for further citations.

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?

Sources are sufficient.

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.

The findings of the thesis are interesting, as the author says that the structure of the (call) graph is less informative than the content of the function (description of vertices). This contradicts referenced SOTA, which has shown that the call graph is indeed informative. Why this is the case we cannot conclude, because (a) proper comparison to SOTA is missing and (b) the proposed approach mixed information about the call graph with informations about functions itself. It is therefore possible (likely) that informations about functions outweighed informations from the call-graph. Alternative explanation might be that GNN failed to explain the structure. And yet another explanation might be that the malware distribution has changed in such a way that SOTA method does not work anymore. I think that the investigation would not difficult and would greatly improve the quality and impact of the work.

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 **B**

Date: 12.6.2020 Signature: