ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

REVIEWER'S OPINION OF FINAL THESIS

I. IDENTIFICATION DATA

Thesis name:	**Graph-Based-Analysis-of-Malware-Network-Behaviors**
Author's name:	**Daniel Šmolík**
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	Ing Carlos CATANIA (PhD)
Reviewer's department:	Institute for Information Technologies and Telecommunications (ITIC) National University of Cuyo (UNCUYO) – Mendoza - Argentina

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **challenging**

Evaluation of thesis difficulty of assignment.

I personally consider this work attempts to provide a solution to a complex and significant problem using state of the art algorithms and techniques.
Despite years of research, Malware detection continues to be one of the most challenging research tasks. Its difficulty lies on the continuous development of strategies for performing new types of attacks. Normally, new attacks are developed at a higher frequency than the responses provided by security analysts. The machine learning area has been working on possible solutions for more than 20 years. In particular, the Graph based approaches are currently an active line of research, mostly perhaps the arisen of complex network algorithms and visualization tools for facilitating its application to different fields.

Satisfaction of assignment **fulfilled**

Assess that handed thesis meets assignment. Present points of assignment that fell short or were extended. Try to assess importance, impact or cause of each shortcoming.

From a general point of view, the thesis met the assignment. The four points considered were correctly developed. Perhaps, the state of the art could have included more articles that applied graph approaches for malware detection. However, those actually included in this thesis were enough to understand the scope of the work and possible strategies for solving it.

Method of conception **outstanding**

Assess that student has chosen correct approach or solution methods.

The student has follow the methodology used in Machine Learning. The student has analyzed the state of the art, and proposed a new set of features that could solve the problem. Then, the student has validated his hypothesis on real malware captures following the machine learning standards.

Technical level **A - excellent.**

Assess level of thesis specialty, use of knowledge gained by study and by expert literature, use of sources and data gained by experience.

The student has proven himself capable of dealing with a new problem and provide a valid solution using a different set of tools. He has showed expertise in several areas such as software development, machine learning and network security.

Formal and language level, scope of thesis **C - good.**

Assess correctness of usage of formal notation. Assess typographical and language arrangement of thesis.

In general the thesis was well written, the formal notation was used when needed. There were some grammatical errors in the writing but I consider they were not serious. Especially, if we consider the thesis was not written in student's mother

language.

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

Present your opinion to student's activity when obtaining and using study materials for thesis creation. Characterize selection of sources. Assess that student used all relevant sources. Verify that all used elements are correctly distinguished from own results and thoughts. Assess that citation ethics has not been breached and that all bibliographic citations are complete and in accordance with citation convention and standards.

The student has always made reference to third-party articles and software applications used for meeting the thesis assignment. All references used in the work followed the proper quality standards

Additional commentary and evaluation

Present your opinion to achieved primary goals of thesis, e.g. level of theoretical results, level and functionality of technical or software conception, publication performance, experimental dexterity etc.

III. OVERALL EVALUATION, QUESTIONS FOR DEFENSE, CLASSIFICATION SUGGESTION

Summarize thesis aspects that swayed your final evaluation. Please present apt questions which student should answer during defense.

In the present thesis the student has proposed a new method based on graph analysis for performing malware detection. A software application was developed from scratch for extracting features from network data and a machine learning approach was proposed for performing malware detection. The proposed method was validated using traffic captures considering different attacks scenarios. In addition, the developed application was opened to public for possible future improvements in the field. The student prove himself competent in several areas such as software development, machine learning, experimental design and graph analysis. Moreover, the field of security research is extremely hard and the student has been able to provide an small but significant contribution in the field.

Apt questions:

1) What are the reason behind the election of Random Forest for classification?
2) Have the student considered using other algorithms?
3) What are the ideas for improving the computational time of the application?
4) Have the student considered the application of resampling techniques such as cross validation for selecting the thresholds?
5) An Sensitivity of 92% with 0% of specificity are and outstanding results. Can we say the malware detection problem is solved? Why?
6) Why the proposed algorithm could detect the 8% of the dataset? Does the student have an idea of how to deal with those cases and achieve better results?

I evaluate handed thesis with classification grade A - excellent.

Date: **11/06/2017**

Signature: