ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

REVIEWER'S OPINION OF FINAL THESIS

I. IDENTIFICATION DATA

Thesis name:	**Anomaly detection of host roles in computer networks**
Author's name:	**Bc. Yury Kasimov**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Computer Science
Thesis reviewer:	Ing. Martin Grill, Ph.D.
Reviewer's department:	External – Cisco Systems

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **ordinarily challenging**

Evaluation of thesis difficulty of assignment.

The research assignment is ambitious as well as relevant. The detection of malware infections in computer networks using User or Entity Behavior Analysis (UEBA) is currently very popular research topic in the field of network security. One of the main concepts of UEBA is the process of building behavioral profiles that can capture normal user behavior to be later used for malicious or abusive network traffic detection. Therefore, introduction of better profiles together with their evaluation is important research that will resonate within the security community.

Satisfaction of assignment **fulfilled with minor objections**

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.

Thesis fulfills all the objectives defined in the assignment. However, the possibility of application of the proposed solution in practice is questionable.

Method of conception **partially applicable**

Assess that student has chosen correct approach or solution methods.

The proposed features seem to be suited to capture the normal behavior of network users, however the normalization and quantization could be improved. The selected anomaly detection approaches are some of the most commonly applied methods in the field which justifies their selection. There is clearly a lot of work invested in the creation of reliable datasets needed for the evaluation of the proposed method. Even so, an experimental evaluation on real network users is missing, which leaves unanswered questions about the practical application of the proposed solution.

Technical level **C - good.**

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

The presented approach uses standard algorithms from anomaly detection field that are already implemented in the scikit python library. Student proved his full understanding of these method by describing them in the thesis and by their proper usage.

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

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

The thesis is well-structured with number of figures and tables that allow for easy understanding of the presented concepts. The thesis is written in English containing several typos that do not affect its overall intelligibility. Additionally, there are typographical issues in some equations (e.g. Equation 3.3, Table 5.17), itemizations (missing dots or commas) and non-capitalized references (figures, equations and tables). There are two paragraphs that are identical in Sections 5.1, 5.2, 5.3, 5.4. Finally, some terms are not defined when first used in the text (NetFlow, Argus, etc.).

Selection of sources, citation correctness **C - good.**

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.
All resources are well cited throughout the thesis. Student reviews some of the relevant existing network anomaly detection methods in the related work chapter. However, these are mostly outdated not reflecting the current state-of-the-art in the UEBA field.

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.

Student proposes an anomaly detection method that identifies changes in users' behavior to detect malware. The method uses a set of already existing features that capture the network behavior aggregated over five-minute intervals. The features are quantized to reduce the dimensionality and standardized to be later used in one of the three anomaly detection methods available in the scikit library: One-class SVM and LOF. The final decision about the anomalousness is made using majority voting that in some cases reduces efficacy of the solution. Experimental evaluation shows that the detector is able to learn baseline of simulated, simple and stable behavior to identify change of behavior caused by a malware infection. However, an evaluation on a real network user data is missing. This evaluation is vital to assess its ability to effectively learn baseline of real network users.

Questions:

- *What is the reason for the large amount of communication to port 443 over UDP shown in the Figure 3.1b?*
- *Why is the quantization of the destination ports (described in Section 3.4.1) not applied to first 1-1000? Wouldn't it be better to skip the system ports (0-1024)?*
- *How would you design an experiment to show that the proposed method is able to effectively learn baseline that does not generate excessive number of false positives for a real network user that is typically using number of various applications in a chaotic way?*

I evaluate handed thesis with classification grade **C - good.**

Date: **4.6.2018** Signature: