Doctoral Thesis Report

Name of student: *Ing. Tomáš Čejka*
Thesis title: *Stream-wise Parallel Anomaly Detection in Computer Networks*

Reviewer: *prof. Ing. Václav Přenosil, CSc.*
Institution: *Faculty of Informatics, Masaryk University*

Up-to-dateness of the dissertation

The topic of dissertation is very up-to-date. Securing computer and communication networks plays a vital role in the modern world of digital communication. At present, the Internet of Things is a reality. It is the research area of many researchers and security companies around the world. However, there are still many open research questions. The presented dissertation deals with several problems related to network security and parallel processing of anomalies hidden in the large volume of network traffic data.

Formal structures and organization of the dissertation

The presented work has 120 pages including the references and appendices. The thesis document has the following structure:
1. Introduction
2. State-of-the Art
3. Topics and Contribution in Details
4. Conclusions
5. Includes Papers

General structure and organization of the dissertation is adequate, correct and logical, although I would prefer to see the subchapter 1.3 "*Goals and Contribution of the Dissertation Thesis*" as a separate chapter.

The author articles in Appendix A describe in detail the individual aspects of detection and parallel processing of detected anomalies in computer networks, which are discussed in the dissertation text.

The contents of the thesis document are technically sound and deep, covering all aspects of the described work. In addition, the dissertation provides an excellent overview of the current state of the art of computer network traffic monitoring, detection of the traffic anomalies and big data processing. The bibliography section contains majority of relevant work and is fully adequate to the presented thesis.

Completion of the dissertation objectives

The aims of the dissertation are presented somewhat unclearly in the first paragraph of chapter 1.3. From my point of view, the aim of this dissertation is *processing the flow data in large backbone networks*. In this text, I have found three partial goals:

- to design mechanisms to process huge volume of flow data as soon as possible especially for network security purposes.
- to extend the monitoring and analysis infrastructure to support additional information.
- to design an algorithm for splitting, which ensures that the machine receives sufficient data to identify security threats, allowing for parallel processing based on splitting a flow of data into independent subsets.

I can say that both the main goal and all three sub-goals are met in the chapters 3.2, 3.3, 3,4 a 3,5 and in appendix A.

Assessment of the methods used in the dissertation

The methods used for the dissertation work are in accordance with the methodology of the scientific work. The dissertation contains description of the problem, problem formulation, hypothesis determination, deduction from the hypothesis, experiment and evaluation of the experiment. I have no comments on the methodology used.

Evaluation of the results and contributions of the dissertation

Thera are five main results:

- description of a new stream-wise approach to network traffic analysis and anomaly detection using flow data without long-term storage,
- design and development of stream-wise and application-aware detection methods discovering threats on application layer through a framework that allows rapid prototyping of the detection methods,
- design and development of the scalable architecture for parallel processing the flow data with respect to witnesses.
- realization and evaluation experiments with real backbone traffic using the developed detection algorithms,
- implementation the formal symbolic description of the algorithms as the modules for the NEMEA system (open source framework and a set of modules for a stream-wise analysis of flow data).

All results and contributions are convincingly documented in this dissertation.

Remarks, objections, notes, and questions for the defense

The author managed to design the system that can run in a parallel distributed environment to handle big data. So he solved the crucial problem the stream-wise concept, especially for large networks. He has managed to find a way to minimize delays of detection and keeps the storage requirements low. For increasing reliability and precision of detection he uses some additional information from the L7 layer.

Questions for oral defense:

1) In the Chapter 1.1, Motivation, you mentioned *Distributed on-the-fly anomaly detection*. Please indicate the difficulty of this type of detection and how it should be addressed?

2) Extended flow records increase the accuracy and reliability of detection of the anomalies, but this tool works with unencrypted traffic only. Do you have any idea of how to detect anomalies in encrypted communication?

The overall evaluation of the dissertation

The use of English language is adequate, although a number of minor formal issues can be found. However, the overall quality of text is good and the thesis as a whole document provides clear description of the subject matter and the contributions made by the student. Quality of all graphics and illustrations provided in the thesis is very good, further contributing to the overall merit of the thesis.

The whole work is abundantly accompanied by the author's publication work. Citations and publishing activities are in line with the usual standards. I have not found indication of plagiarism.

Statement on the recommendation of the dissertation for the defense

Author in this thesis **has demonstrated** the ability to work independently in the specified field, and the **thesis meets** the standard requirements on a dissertation in the field. In accordance with par. 47 letter (4) of the Law Nr. 111/1998 (The Higher Education Act) I **recommend** that the dissertation is **accepted** and the candidate proceeds to oral defense with the aim of receiving the Ph.D. degree.

Brno September 11, 2018

prof. Ing. Václav Přenosil, CSc.
Faculty of Informatics Masaryk University