THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Hierarchical density-based clustering and interpretation for network measurements**
Author's name:	**Ing. Pavol Mulinka**
Type of thesis :	doctoral
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Dept. of Telecommunication Engineering
Thesis reviewer:	Ing. Alexandru Mihnea Moucha, PhD.
Reviewer's department:	Dept. of Computer Systems, FIT, CVUT

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **extraordinarily challenging**

How demanding was the assigned project?

I consider the work as very challenging because the student had to assemble together - in a consistent idea (followed, of course, by implementation and testing) - pieces from lots of areas of information technology: networking, services, data analysis and processing, protocols, statistics, artificial intelligence, machine learning and math. In order to deliver such a work in a consistent way it is obvious that you need deep understanding of all these fields.

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 goals of the thesis (traffic classification and anomaly detection without prior knowledge about the monitored system, using machine-learning) are fulfilled, my feeling is that in fact the methodology can be applied beyond networking – in fact to any system with a behavior which can be described and interpreted.

Methodology **outstanding**

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

The involved methodology is explained in a clear manner, with reproducible experiments and verifiable results. Due to the combination of so many areas of expertise, I quantise the methodology as at an outstanding level.

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?

This – in my opinion – was already explained in my previous comments. I am very fond of the fact that the student could orient his work into so many areas of research in which it is very easy to get lost (due to their complexity and ways they are interconnected).

Formal and language level, scope of thesis **A - excellent.**

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 thesis is very well written, a pleasure to read and understand. The language is appropriate, the structure is clear, the content is well presented from both technical and explanatory perspectives. I found some minor English mistakes, with absolutely no impact on the clarity and quality of the text.

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?

This comment here is only a formality. It is obvious that the student worked with citations – this is part of his research career and published papers.

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.

I shall here talk freely about the thesis content.

Being myself a networking engineer with hardware and software basis, I had the privilege of seeing and understanding the development of large, scalable networked infrastructures from classic routing and switching to added services (voice, video, integrated data, etc.), to the development of segmentation and virtualization. We are now in the middle of migrating all these complex infrastructures to seeing them as pieces of software, governed by DevOps, SecDevOps, automatization, self-healing by proactive, detective and reactive systems, maintenance by repositories and – only to conclude – offering a complete glass-pane overview of the entire state of the infrastructure. And all of these offered in a more open and flexible way than ever before.

As I work with Cisco technologies and systems, I am fully aware of the revolution Software Defined Networks (Cisco DNA Center et alii) bring to datacentres, enterprise and open infrastructures. For me – as a networking engineer and professional – these proprietary solutions (which, without doubt, I trust as part of my professional career) are more towards "magic", as in the networked systems the analysis is performed by a proprietary solution which provides an analysis (which is correct in the overwhelming majority of the cases), however the details of how the solution got to that result are hidden (being a proprietary, commercially available, integrated solution).

It was thus a privilege for me to be able to see the insights of how artificial intelligence and machine learning can be involved in the processes which – when using proprietary solutions – are hidden. It was also an "expected surprise" to see that (with appropriate adaptions, of course) these methods (as the ones used for traffic classification and anomaly detection) can be in fact applied to any system which has a trackable and describable behaviour, being thus generally appliable principles.

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.

In my humble opinion I consider the thesis (and the work) as excellent and I would like to address only two related questions:

It is clear that the methodology here described works, I am curious how fast it can be. As modern networks rely on rapid reaction to prevent and limit the damage of increasingly sophisticated attacks, can the analysis be performed in real-time and with acceptable computational resources to be integrated as a live subsystem? Is it easily scalable or elastic in resource demand?

I am fully confident that the student is perfectly capable of performing independent research and presenting his findings and I kindly recommend the thesis for the defense.

Date: **2.8.2021**

Signature: Alex Moucha