Doctoral Dissertation Review

Author of the dissertation: **Michael Heigl, M.Sc.**

Title of the dissertation: **Enhancing Computer Network Security through Improved Outlier Detection for Data Streams**

Ph.D. study program: **Computer Science and Engineering**

Faculty: **University of West Bohemia, Faculty of Applied Sciences**

Evaluation of the significance of the dissertation for the field

Author of this dissertation addresses the area of computer network security, which is a very important field of computer science, becoming continuously more challenging in the IoT and AI era. Author presents a generic framework that he named Anomaly-based Incident Detection and Response System (ANDERS) and formulates four specific research questions targeting different tasks of the overall network incident detection as well as response process. In all four research questions author presents some interesting research contributions. I therefore find this dissertation thesis very significant for the field of computer network security.

A statement on the problem-solving procedure, the methods used and the fulfilment of the specified goal

Author applies standard problem-solving procedure and uses suitable research methods throughout the whole dissertation in solving four different research questions (RQ). He starts with detailed analysis of the related work, following with identification of specific requirements whenever applicable (in case of RQ1 and RQ2). As next, author explains in detail original or modified algorithms he designed to address particular research questions. For each of the proposed methods author describes experimental set-up and data used, performs experiments, and discusses their results. I can state that all four research problems formulated by the author as research questions have been fulfilled with observable improvements with respect to the state-of-the-art methods.

An opinion on the results of the dissertation and on the original concrete contribution

RQ1: Author proposed an original method called Unsupervised Feature Selection for Streaming Outlier Detection (UFSSOD). The proposed solution is very well motivated and excellently grounded in the requirements analysis on one hand side and state of the art in the related areas like e.g. outlier analysis on streaming data and feature selection for outlier detection, on the other side. Operation principles as well as two alternative operation modes are very well explained and described. Core UFSSOD functionalities in terms of the proposed scoring mechanism (both, outlier, and feature) and feature clustering algorithm are carefully explained. Experiments for objective evaluation of the proposed method are very well designed, performed and discussed. It has been shown that applying UFSSOD in a streaming setting notably increases both the classification and computational performance if the data set is of high-dimensional nature and has a high number of data instances.

RQ2: I appreciate an elegant method for adaptive adjustment of the ensemble model by switching its worst performing parts whenever a concept drift is detected. It is again very well experimentally evaluated and strong as well as weaker aspects of the proposed method are correctly evaluated. Performance Counter-Based iForest framework in its core component (performance counter-based scoring) proposes a relatively simple mechanism for identification

of badly performing components of an ensemble. Is this mechanism inspired by some other work on ensemble learning for streaming data?

RQ3: Author proposed a framework with multiple configuration settings to improve the input quality for the streaming alert correlation/clustering. In addition to the typical intrinsic attributes such as IP, port or timestamp information, author uses also feature importance and the respective outlier scores. The resulting clusters can evolve over time as well as be discarded when they become irrelevant. Method proposed by the author provides better clustering of different alert situations with much shorter processing time than the state-of-the-art method (GAC). The alarms detected by means of these clusters are further processed to generate three types of signatures, based on three different sources of data. I appreciate that author very carefully and in great details experimentally analysed and nicely visualised impact of particular types of signature data sources.

RQ4: Author designed, implemented, and experimentally evaluated an improved and extended Uncoupled Message Authentication Code method as a protection-based security technique able to work also as a detection-based method. This method is based on the sampling approach having MAC Phases and Idle Phases, whereas the required security level can be adjusted with a balanced overhead on resource consumption and network utilization.

Statement on the systematics, clarity, formal arrangement, and language level of the dissertation

This dissertation is written very systematically with clear and well explained structure. The formal arrangement as well as language level of the dissertation is very high. I personally think that the extent of this dissertation is larger than it would be suitable for a dissertation and some parts of the text are repeated on more places (e.g., in background section as well as in some of the other sections). On the other side it is evident that the author performed an enormous amount of analytical and experimental research work, which deserves high recognition.

Some remarks:

- RQ1: I think that although selection of data sets from the Outlier Detection DataSets (ODDS) Library are well motivated and explained on page 87, because of the focus of this dissertation on intrusion detection, traditional datasets for IDS could also be included in the experiments. Moreover, some of the datasets do not fulfil requirements stated at the very beginning for the extent of outliers, which should be rare.
- RQ2: I appreciate clear requirements identification and explanation. However, it is not always clear, how well each of those requirements is fulfilled in your final solution. It would be nice to present a summary of them in a form analogical to the comparison provided on the existing methods in Table 4.1.
- On page 93, last sentence mentions F1 0,162 but the respective value in Table 3.5 is slightly different (0,158).
- You wrote on page 98: "... in the real-world the data stream has an infinite amount of samples." The number of samples in real-world will never be infinite.
- Page 107 as well as 108 and 126: " ... this article ..." – is it really article, or dissertation thesis?
- There are too many abbreviations used throughout the dissertation.

Comments on the student's publications

Author of this dissertation co-authored 5 journal publications (plus one submitted already with a revised version), from which one is a Journal in Q1 on WOS, two are in Q3 and one in Q4. Moreover, he co-authored 8 conference publications and provided 5 oral presentations on international or national scientific or professional conferences. Author of the dissertation has 11 publications registered in WOS with already some citations available to

them indexed in WOS (8 together, 2 without self-citations) and 38 citations indexed in GoogleScholar. I find the quality as well as the number of publications above average for a PhD-student in this research field.

Based on the statements above I fully **recommend** the dissertation of Michael Heigl, M.Sc. **for defence**.

In Košice, November 15, 2021

prof. Ing. Ján Paralič, PhD.
Dept. of Cybernetics and Artificial Intelligence
Technical University of Košice
Letná 9, 042 00 Košice, Slovakia

PhD Thesis Review

Reviewer: doc. RNDr. Petr Šaloun, Ph.D.
Title: Enhancing computer network security through improved outlier detection for data streams
Název: Zvyšování bezpečnosti počítačových sítí zesílenou detekcí odlehlých hodnot v datových tocích
Author: Michael HEIGL, M.Sc.
Institution: University of West Bohemia Faculty of Applied Sciences
Supervisor: doc. Ing. Dalibor Fiala, Ph.D.

Up-to-datedness of the dissertation

Computer networks security has been always a Holy Grail of the intersection of computer science theory and computer networks theory with the practical outcome and consequences. The described research area is the focus of the assessed dissertation. The topic of the work is current and the goals set in the work have been achieved.

Formal structure and organization of the thesis

The thesis is 247 pages long and is written in English. The thesis consists of seven chapters, two appendixes, bibliography, and list of acronyms. Some chapters dealing with publications of the author. They are written well, the chapter's sequence is easy to follow and particular problems are placed and discussed well. The introduction introduces author's motivation, gives cyber defense life cycle description and specifies the phases of a continuous incident handling process. There is given an explanation, why the outlier detection methods are (still) more efficient in terms of time and space complexity and allow feature interoperability as well as model transparency. The introduction contains research objectives, defines research questions (RQ1 – RQ4), and finally the research goal, which is *Improve Outlier Detection for Data Streams to Enhance Computer Network Security.* Mapping and discussing of particular RQ is described in section 1.5 as an overview, which is good, let me summarize, that the four core chapters (Chapter 3–6) are presented in a standalone and self-explanatory manner on pages 72 – 105. The text is enriched by 82 figures partially containing block diagrams, 29 tables, and 9 algorithms (Some of them references to the following algorithms, which is very convenient, at least for reader of the PDF version of the thesis.), where a meta-language is used.

Completion of the dissertation objectives

They are four research questions in the thesis:

RQ 1 How can unsupervised feature selection be applied on streaming data for the purpose of outlier detection?

RQ 2 How can a flexible framework for unsupervised online outlier detection be designed to provide an online scoring functionality for feature importance?

RQ 3 How can the output of online outlier detection mechanisms be exploited to characterize and compare novel attack patterns?

RQ 4 How can a cryptography scheme be leveraged to function as a detection mechanism and have its feedback be incorporated in improving performance over run-time as part of response functionality?

Assessment of the methods used in the dissertation

Methods used in the thesis are accepted in the field of research and they are selected and applied appropriate, I mean, that newest methods of deep learning cannot be used, as they work not well, besides traditional machine learning approach is suitable, as it produce excellent results.

Evaluation of the results and contributions of the dissertation

The research contributions of the thesis can be summarized as follows:

RC1 Unsupervised Feature Selection for Outlier Detection on Streaming Data to Enhance Network Security.

RC2 On the Improvement of the Isolation Forest Algorithm for Outlier Detection with Streaming Data.

RC3 Exploiting the Outcome of Outlier Detection for Novel Attack Pattern Recognition on Streaming Data.

RC4 A Resource-Preserving Self-Regulating Uncoupled MAC Algorithm to be Applied in Incident Detection.

Remarks, objections, notes, and question for the defense

They are some minor remarks about formal details of the thesis. One of them is about the way of writing the used literature does not always make it easier to find it, the presentation of the DOI would be a transparent, fast and generally accepted way of identifying the cited source. For example, I cite one of the most recently cited sources and mention its DOI. [295] A. Nazir and R. A. Khan, "A novel combinatorial optimization based feature selection method for network intrusion detection," Comput. Secur., vol. 102, no. 102164, p. 102164, 2021. `https://doi.org/10.1016/j.cose.2020.102164` I have one general question: *"What is your opinion on the future development of intrusion detection systems in the context of the steadily increasing role of the Internet-of-Things (IoT) concept? Will the standard machine learning techniques still be able to protect the vulnerable networked devices from malicious activities or will there be some need for a wholly new technology ensuring network security?"*

The overall evaluation of the dissertation

The research results were published in 6 journals and 8 conference papers, and was presented in 5 talks, all on the international level. In this year he publish results in Electonics Journal twice, this journal is evaluated in JCR between Q2 and Q3 respectively, and hist best ranked research output is dated in 2019 in Computers & Security journal, evaluated between Q1 and Q2, this journal is accepted as excellent across the research community. The author of thesis proved the ability to conduct research and achieve scientific results. In accordance with par. 47, letter (4) of the Law Nr. 111/1998 (The Higher Education Act) I do recommend the thesis for the presentation and defense with the aim of receiving the Ph.D. degree.

Ostrava, November 28, 2021

doc. RNDr. Petr Šaloun, Ph.D.