THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**A Network Dataset of Normal, Malware, Attack, and Background Traffic on a Real Network.**
Author's name:	**Štěpán Bendl**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Computer Science
Thesis reviewer:	Ing. Karel Hynek
Reviewer's department:	Department of Digital Design, Faculty of Information Technology

Text

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **challenging**

How demanding was the assigned project?

The proper creation of a rich and heterogeneous dataset is always a challenging task. The student had to study the properties of current cybersecurity datasets and identify their flaws. He had to conduct survey research among cybersecurity experts to properly design the creation methodology, design the methodology and conduct all the data collection and malware execution. Thus, I consider the assignment extremely challenging.

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.

All points from the thesis assignment have been fulfilled. The published dataset is publicly accessible and has already been downloaded several times.

Methodology **correct**

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

I do not have any objections to the designed and used methodology. At first, student learned about existing datasets and their flaws. The conducted survey among cybersecurity professionals brought a novel perspective that was then applied to the creation methodology. The selection of malware families used hardware and whole network topology was made properly based on the data obtained in the survey and existing dataset research. Apart from that, he identified most of the artifacts and outlined possible use.

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?

The technical level of the thesis is excellent. The student clearly explained his approach to dataset creation. The student learned about the existing dataset, researched the "data wishes" of the cybersecurity community, and applied it in the creation of the dataset.

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

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 well structured and covers all topics required by the assignment. However, I noticed the usage of some previously unexplained abbreviations that are still not standard due to the novelty of technologies (such as DoT or DoH). Some parts of the thesis should contain a more precise description. A survey among the cybersecurity professionals was performed, but no additional details were provided. The number of respondents or the types of questions is not specified. Also, the design of the exfiltration via DNS would deserve more space and explanatory figures. Besides, I found some typographical errors and text inconsistencies. Nevertheless, the thesis is easily understandable and is written in clear English.

Selection of sources, citation correctness	**B - very good.**
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?	
Some parts of the thesis and some statements would benefit from additional references. Besides that, I do not have any problem with the bibliography and the used citation practice. Moreover, I always found the referenced information in the provided source.	

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 really enjoyed reading the thesis, which deals with the creation of the dataset. The created dataset contains unique and high-quality data samples that will allow research progress in various cybersecurity tasks. Moreover, I personally look forward to using the dataset in my future research.

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.

The grade that I award for the thesis is **A - excellent.**

Štěpán Bendl created a novel dataset containing benign and malicious communication. The community desperately needs high-quality cybersecurity datasets, and I consider the resulting dataset as one of the few high-quality datasets that are up-to-date and usable by researchers worldwide.

The thesis describes the dataset creation methodology based on previous cybersecurity datasets and interviews among professionals from the field. Despite the minor changes in the final methodology due to software bugs and non-functioning packet captures, I consider the creation methodology correct without any significant errors. More importantly, all potential data-artifact reasons and the data limitations are correctly described in the thesis. I also found extremely useful the recommendation of specific dataset parts for particular research tasks, such as malware classification or anomaly detection.

The resulting dataset is already available to the broad public on the Zenodo platform and has already been downloaded multiple times. I suggest linking the text of the thesis directly to the Zenodo readme website since the thesis documents the methodology. Nevertheless, some parts of the methodology could be described further; thus, I have some additional **questions:**

Can you provide some details about the interviews among cybersecurity professionals? How many respondents did you have? How many questions did you have? What were the questions?

Can you be more specific about the DNS exfiltration attack design? Did you use a recursive DNS (such as dns.google) to add real-world timing and packet loss properties to the exfiltration attack?

Even thought, the thesis has some weak spots—mainly in the missing description of the performed survey and some attack design, I found them marginal compared to the amount of work and the extremely useful outcome. Therefore, I consider the thesis excellent—grade A.

Date: **16.6.2023**

Signature: