CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Large Language Models as Defensive Honeypots**
Author's name:	**Muris Sladic**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Computer Science
Thesis reviewer:	Chang Ee-Chien
Reviewer's department:	School of Computing, National University of Singapore

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **challenging**

How demanding was the assigned project?

This project venture into a new area with many uncertainties. The project requires strong practical computer system hand-on skill, and there are many tough design decisions to be made in the user studies.

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.

Please insert your comments here.

Methodology **outstanding**

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

Please insert your comments here.

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?

Please insert your comments here.

Formal and language level, scope of thesis **B - very 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?

Please insert your comments here.

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?

Please insert your comments here.

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.

Please insert your comments here.

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS REVIEWER'S REPORT

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.**

The thesis proposes using pre-trained LLM (with fine tuning) to generate events that are difficult for humans to distinguish from actual events. Two event types are considered: Unix Shell's output, and a set of selected applications API (MySQL, POP3, HTTP). A main challenge is to ensure consistencies and diversity of the generated events. The thesis proposes a few methods to handle the challenge, include fine tuning, chain-of-thought and prompt engineering. User studies are conducted to evaluate the proposed methods.

Applying LLM to generate "normal" events for the purpose honeynet is a new topic. Although intuitively LLM could help, there are many challenging questions such as the type of events and method of evaluations. This thesis focuses on a few carefully selected types and provides a well-thought-out evaluation approach.

I think consistency of the generated events would be a huge challenge, and this is particularly so for shelLM since the OS's state is huge and complex. For instance, a user could change the filename and verify later, or monitoring the processes state. I guess a reason that many participants in the experiment didn't detect was because they were not told that events could be generated in the beginning of the experiments and was told to focus on another task (Figure 7.1). In practice, intruders would attempt to detect whether they are in sandboxes or honeynets and thus would actively distinguish the events. Hence, it could be an interesting future work to evaluate the generation effectiveness with such knowledgeable intruders.

Since the topic is new, certainly, there are still many gaps which could be addressed in future works. As far as this thesis is concerned, the works done is more than sufficient and thus I suggest a grade of A.

Date: **11.6.2024** Signature: Chang-En Chien