School of Computer Science and Informatics

CM3203 - One Semester Individual Project - 40 credits

INITIAL PLAN

Machine learning model to identify attacks on IoT devices.

Author: *Shaikha Alshehhi*
Supervisor: *Amir Javed*
Moderator: *Hiroyuki Kido*

TABLE OF CONTENT

1. Project Description

1.1 Brief Description

The project aims to build a machine learning model to identify an ongoing cyberattack on IoT devices. The Machine learning model will be built by analysing a publicly available labelled dataset. IoT-23 is a new network traffic dataset derived from the Internet of Things (IoT) devices. It has 20 malware and three benign IoT device traffic captures. It was initially released in January 2020, featuring captures spanning the years 2018 to 2019. The reason for choosing this IoT dataset is due to the fact that it provides a huge set of real and labelled IoT malware infections as well as IoT benign traffic which is essential for training the Machine Learning model. The Machine Learning model will be implemented using Python.

1.2 Background information

The Internet of Things (IoT) is a network of physical objects embedded with sensors, software, and other technologies for connecting and sharing data with other devices and systems through the internet without the need for human intervention. These devices range from ordinary household objects to sophisticated industrial tools (Malathi and Padmaja, 2021). The Internet of Things (IoT) has become one of the most popular technologies and a successful topic of study in the commercial sector. The demand for and use of IoT is quickly growing. Experts predict that there will be more than 7 billion linked IoT devices by 2020, and 22 billion by 2025 (Oracle 2020). However, the essential built-in protection to address security threats is frequently missing from IoT devices. Cyber attackers can compromise the device and use it as a launchpad for sophisticated cyberattacks due to common vulnerabilities and exposures. According to Kaspersky(2021), there were 1.51 billion breaches of Internet of Things (IoT) devices from January to June this year, up from 639 million in 2020. As a result, in recent years, researchers have begun to investigate more complex security solutions to address this problem. The use of machine learning to identify and categorise attacks is one of the new approaches. IoT systems can offer the vast amounts of data that machine learning algorithms use to create their detection models. Furthermore, the sheer number of different sorts of attacks and their manifestations makes it nearly hard for human operators to recognise and categorise them (Zeadally and Tsikerdekis, 2019).

---

2. Ethics

Since the dataset does not contain any sensitive or personal information and is published for public use, it is unlikely that ethical approval will be required for this project. The goal of publishing the dataset publicly is to provide researchers with a huge dataset of real and labelled IoT malware infections as well as IoT benign traffic to train machine learning algorithms on, which is the main aim of this project. Nevertheless, ethics will be considered while working on the project and writing the report.

---

3. Aims and Objectives

- **Aim:** Establish an understanding of the problem and possible solutions:
  - **Objectives:**
    A. Carry out research into security issues associated with IoT systems and devices.
    B. Conduct literature review on existing solutions and machine learning models.
    C. Background research on machine learning processes, models, algorithms.
    D. Explore different tools and frameworks that are required for implementation.

- **Aim:** Build a Machine Learning model that can identify an ongoing cyberattack on IoT devices and categorise them into different kinds of attacks
  - **Objectives:**
    1. Explore and establish an understanding of Python's machine learning libraries and modules
    2. Download and clean the dataset of captured PCAP files of malicious traffic.
    3. Transform the dataset into a machine-readable format.
    4. Build a deep learning model and optimise the model by changing different parameters of the model using python.
    5. Test the model using k-fold validation techniques and on an unseen dataset.

- **Desirable aim:** Build a GUI to show how a network packet can be classified into malicious or benign, thus, demonstrating how an ongoing attack can be detected.
  - **Objectives:**
    1. Establish an understanding of Python's standard GUI library, Tkinter.
    2. Create Graphical Interface.
    3. Deploy the Machine learning model

---

4. Work plan

4.1 Supervision roles

I decided with my supervisor, Amir Javed, to have a meeting each week on Tuesday for a duration of 30 minutes. These meetings will be an opportunity to share progress, raise issues, seek help and receive feedback.

4.2 Project Timeline

This project will run from January 31st through May 13th, with a three-week break after week 9. The project will be completed and delivered by May 13th, with a time frame of 12 weeks. Below, section 4.3, is a breakdown of the project's weekly schedule, including tasks to be completed and major milestones. This process is more likely to be iterative than linear. In addition, it's conceivable that some unexpected problems might occur that can not be predicted beforehand. Hence, the project timeline will be altered accordingly, and significant effort will be made to overcome any problems as quickly as possible in order to keep the project on track. Section 4.5 provide a risk plan for the purpose of preparing for any challenges that might be faced during the project

4.3 Weekly Plan

Week	**Tasks**	**Milestones and Deliverables**
Week 1: 31/1/2022 - 4/2/2022	• Write the 1st draft of the Initial Plan. • Meeting with my supervisor for project clarifications on aims and objectives, and get suggestions on project planning and useful resources. • Finish Initial plan after receiving feedback from my supervisor.	I. Initial Plan 1st draft II. Send 1st draft of the Initial Plan by the 3rd of February to my supervisor. III. Final Initial Plan
Week 2: 7/2/2022 - 11/2/2022	- Literature review/background research on: • Machine Learning process. • Machine Learning models. • Machine Learning algorithms. • Python Machine Learning libraries and modules.	Submit Initial Plan by the 7th of February

Week 3: 14/2/2022 - 18/2/2022	- Continue literature review/ background research - Have a clear idea on the implementation of ML model - Start writing the first draft of the Background section of the report - Make implementation decisions: • Determine what models, algorithms, libraries and modules that would be used based on the background research.	Produce the first draft of the Background section of the report
Week 4: 21/2/2022 - 25/2/2022	- Initial implementation: download software, tools, libraries and modules. - Download the dataset of captured PCAP files of malicious traffic.	Write up the Approach section of the Final report
Week 5-7: 28/2/2022 - 18/3/2022	- Understand and clean the dataset - Transform the dataset into a machine-readable format. - Implementation and development of the machine learning model. - Document everything for the implementation section of the Final Report.	
Week 8: 21/3/2022 - 25/3/2022	- Testing and debugging the ML model. - Finish implementation.	The machine learning model works and functions correctly.

Week 9: 28/3/2022 - 1/4/2022	- Produce test cases - Test the model using k-fold validation techniques and on an unseen dataset - Document everything for the Results and Evaluation sections of the Final Report.	I. Write up the Results and Evaluation sections of the Final Report. II. Produce the first draft of the Final Report.
Easter break: 2/4/2022 - 24/4/2022	- No scheduled meetings - During the break, I have no plans to work. I am well aware that projects might overrun owing to unforeseen issues, particularly during the implementation phase. As a result, if the implementation stage takes longer than anticipated, I will use the Easter break to complete the implementation and produce the final report.	
Week 10: 25/4/2022 - 29/4/2022	- Structure and organise the final report. - Add table of content, figures, appendices and references. - Construct the final report in the standard format	I. Write up the remaining sections of the Final Report. II. Produce the 2nd draft of the Final Report.

Week 11-12: 2/5/2022 - 13/5/2022	- Adjust any errors and typos - Finalise and proofread the Final report.	I. Final Report with all sections written up. II. Submit the Final Report by the 13th of May.

4.4 Gantt Chart

Week/milestone	1	2	3	4	5	6	7	8	9	Easter break	10	11	12
Initils plan draft	▨												
Literature review		▨											
Background research on machine lerning model implementation		▨											
First draft of the Background section of the report			▨										
Implementation decisions				▨									
Implementation of ML model					▨	▨	▨						
Finish implementation and start testing					▨	▨	▨	▨					
First draft of the Final Report									▨				
Documentation and evaluation of the implementation and testing process									▨				
Finish any falling behind work										▨			
Second draft of the Final Report											▨		
Finalise and submit the Final Report												▨	▨

4.5 Risk Plan

Risks Identified	**Likelihood** (High, Medium, Low)	**Impact** (Large, Medium, Small)	**Treatment** Strategy to minimise disruption
Possible loss of data.	Low	Large	Make sure that online and offline backups are kept for all stages of the project
Delay in implementation resulting from difficulties in the coding stage.	Low	Large	I. Sufficiently research the tools, libraries and algorithms that will be used. II. Seek help and advise from supervisor. III. Use the Easter break to complete any uncompleted work and go back on track.
Sickness / issues associated with mental health .	Medium	Medium	I. Fair distribution of tasks throughout the project period. II. Take necessary breaks to avoid over exhaustion and stress.

Falling behind on work	Medium	High	I. Complete all tasks assigned for each week. II. Use the weekend to complete any task that was not completed during the weekdays. III. Use the Easter break to complete any uncompleted work and go back on track.

---

5. Conclusion

In conclusion, the proposed final year project for the Computer Science and Forensics Security degree was discussed in this report. It included the project’s description, ethical considerations, the project's aims and objectives, and a work plan with a weekly plan, a Gantt chart, a risk plan and the supervisor's role in the project.

6. References

[1] Kaspersky. 2022. *Kaspersky Cyber Security Solutions for Home & Business*. [online] Available at: <https://www.kaspersky.co.uk> [Accessed 9 December 2021].

[2] Malathi, C. and Padmaja, I., 2021. Identification of cyber attacks using machine learning in smart IoT networks. *Materials Today: Proceedings*.

[3] Oracle. 2022. *What is the Internet of Things (IoT)?*. [online] Available at: <https://www.oracle.com/uk/internet-of-things/what-is-iot/> [Accessed 9 December 2021].

[4] Sebastian Garcia, Agustin Parmisano, & Maria Jose Erquiaga. (2020). IoT-23: A labeled dataset with malicious and benign IoT network traffic (Version 1.0.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4743746

[5] Zeadally, S. and Tsikerdekis, M., 2019. Securing Internet of Things (IoT) with machine learning. *International Journal of Communication Systems*, 33(1).