**Initial Plan**

CM3203 - One Semester Individual Project – 40 Credit

# Machine Learning Model to Categorize Cyber Attacks

Author: Bedoor Bahassan

Supervisor: Amir Javed

Moderator: Dr Daniel J. Finnegan

School of Computer Science and Informatics

Cardiff University

2020

## Table of Contents

Project Description: .....2
Background Information: .....2
Brief Description: .....2
Project Aims and Objectives: .....3
Supervisor Roles: Amir Javed .....4
Ethics: .....4
Work Plan: .....4
Gant Chart: .....8
Milestones and Deliverable: .....9
Risk Plan: .....9
Conclusion: .....10
References: .....11

# Project Description:

## Background Information:

Nowadays, we live in a digital world. The Internet is defined as the worldwide interconnection and the use of internet devices are spreading between government, industry, academic, and private parties. The research shows, the rate of the internet users is growing to 11 users per second and reached to one million new users daily, they are now covering 58% of the global population (Marketing et al., 2020). However, recently, the amount of cyber security attacks increased rapidly, causing negative impact on business and individuals. New research from leading market analysts, Juniper Research, indicates that the expenses of data breach in 2020 will overreach $150 million (Juniper research, 2020). In 2012, Saudi Aramco oil company exposed to one of the biggest cyberattack that affected on 30,000 of its Windows based machines, the process of recovery from the harm took around two weeks (Taylor & Francis, 2020).

Cyber-attacks is increasing year on year, therefore it is needed to avoid attacks by increasing the security of the network base and data information which could be achieved by using protected devices; such as, antivirus services, antispam devices, firewalls, IPS (Intrusion Prevention System), IDS (Intrusion Detection System), and authentication services. These devices generate a vast number of logs per second to be controlled.

## Brief Description:

The aim of this project is to create a model to categorise cyber-attack from network traffic. The model will be trained with the given dataset, using machine learning technique. The implementation of the project and machine learning model will be using python-based programming language and all the implementation and testing phases will carry on Mac device.

All the implementation of the project will use python as programming language and will be implemented and tested on a MacOS device.

## Project Aims and Objectives:

This section will illustrate the main aims and objectives of the proposed project throughout the given time frame of the project.

**Aim:** Establish a background in several areas that need to be used in order to achieve the solution.

**Objective:**

- Understand machine learning methodologies
- Understand the different algorithm for classification
- Identify framework and tools needed to develop the engine for the project (e.g. programming language, tools, framework and techniques).

**Aim:** Identify machine learning models that can successfully detect malicious and denial Network traffic.

**Objective:**

- Determine which machine learning will be used for the project (Supervised or unsupervised)
- analysing the dataset
- Implement different algorithms for machine classifier to determine the best model to use
- train the models with the given dataset (IoT dataset).
- Explain the result
- Validate the Implementation.

**Aim:** Documentation and evaluation on the established solution.

**Objective:**

- Test the program with multiple distinct datasets to verify the solution.
- Introduce possible improvement in the code.
- Performance Analysis

## Supervisor Roles: Amir Javed

There will be a meeting each week on Wednesday for 15-20 minutes with the supervisor where advice, guidance and feedback will be provided.

## Ethics:

The project as discussed with the supervisor is likely not require an ethical approval since the dataset does not contain any sensitive or personal information and it is created for public usage. However, ethic will be taking into consideration during working on both, the project and the final report.

The given dataset is IoT-23 Dataset includes 20 captures of malware carried out from IoT devices, and 3 captures of benign IoT devices traffic. The given dataset has more than 760 million packets and 325 million labeled flows of more than 500 hours of traffic. The captures happened through 2018 and 2019 at the Stratosphere Laboratory, AIC group, FEL, CTU University, Czech Republic. This dataset and its research are sponsored by Avast Software, Prague.

## Work Plan:

This work plan is created to manage the project stages, the final report, and to deliver an efficient project. Divided by week and including regular meetings with the supervisor throughout the semester, including the Easter break.

| | Tasks | Milestone |
|---|---|---|
| Week 1<br><br>(27th Jan/2nd Feb) | • Begin writing the initial plan.<br>• Meet the supervisor for further clarification of the project objectives and aims, and feedback on the initial plan draft. | initial plan draft |

| | | |
|---|---|---|
| | • Produce and submit the initial plan, considering the supervisor's comments. | |
| Week 2<br><br>(3rd Feb/9thFeb) | • Submit initial plan<br>• Prepare final report structure<br>• Background research on:<br>  ○ Machine Learning process<br>  ○ Machine learning algorithm for cyber attacks<br>  ○ Dataset | Submit the initial plan (3rd February 2019). |
| Week 3<br><br>(10th Feb/16thFeb) | • Continue background reaching<br>• Review, understand and learn ML Python libraries and modules<br>• Determine Python libraries and modules for implementation | |
| Week 4<br><br>(17th Feb / 23rd Feb) | • Produce first draft of background section of the final report.<br>• Outline and design class | Produce first draft of background section of the final report. |

| | | |
|---|---|---|
| | diagram of ML model<br>• Initial Implementation Setup:<br>o Set up a Git repository for the project<br>o Set up a workspace environment and install required tools<br>• Progress review meeting | |
| Week 5- 7<br><br>(24th Feb/15th Mar) | • Implementation of ML model and Development<br>• Meet the supervisor to discuss progress.<br>• Start first draft of final report. | |
| Week 8<br><br>(16th Mar/22nd Mar) | • Train models with dataset<br>• Testing and debugging models<br>• Finish Implementation | Final report first draft. |
| Week 9<br><br>(23rd Mar/ 27th Mar) | • Progress review meeting<br>• Design testcases<br>• Start tool testing and evaluation<br>• Record tool testing and results | |

| | | |
|---|---|---|
| Easter Recess<br>Week 10-12<br>( 27th Mar/29thApr) | • No meetings will be scheduled<br>• Finish any falling behand work<br>• Write final report of remaining sections | |
| Week 13<br><br>(20th Apr/ 27th Apr) | • Restructure and organise all Final Report sections<br>• Add table of figures and abbreviation, glossary, appendices<br>• Arrange used references<br>• Prepare Final Report in a standard format<br>• Second Draft of the final report | Second draft of final report. |
| Week 14-15<br><br>(27th Apr / 7th May) | • Progress review meeting<br>• Adjust any errors and typos<br>• Finalise, proofread the final report. | Submitting the Final Report. (7th of May) |
| VIVA Week | • Prepare for the VIVA of the final year project. | Completion of the project |

## Gant Chart:

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Initial Plan | | | | | | | | | | | | | | | |
| Background Research | | | | | | | | | | | | | | | |
| Start Implementation | | | | | | | | | | | | | | | |
| Implementation & Development | | | | | | | | | | | | | | | |
| Final Report 1st Draft | | | | | | | | | | | | | | | |
| Complete the model & Testing | | | | | | | | | | | | | | | |
| Analyses | | | | | | | | | | | | | | | |
| Final Report 2nd Draft | | | | | | | | | | | | | | | |
| Submit the Report | | | | | | | | | | | | | | | |
| VIVA presentation | | | | | | | | | | | | | | | |

## Milestones and Deliverable:

This section will outline the most significant milestones and deliverables of the project, derived from the work plan above.

- Initial plan report submission – 3/2/2020
- First draft of some current and future work in the final report done during the implementation stages, as stated in the work plan, such as;

    - Background research.
    - Approach and implementation.
    - Results and evaluation.
    - Improvement.

- The completion of the machine learning model.
- Second Final Report Draft.

  o Completing the missing sections, after completing the implementation.

- Final report submission –7/5/2020
- Viva Presentation, to complete the project.

### Risk Plan:

The risk plan below is created for the purpose of preparing for any difficulties that could happened during the project.

| Risk | Risk Level (Low, Medium and High) | Likelihood of event (certainty, likely, somewhat Likely and Unlikely) | Solution |
|---|---|---|---|
| Data loss | High | Somewhat Likely | Insure having backups for each stage of the work. Upload files online using Google Drive. |

| | | | |
|---|---|---|---|
| Falling behind the work plan | Medium | Likely | Do the task week by week and in the event of falling behind, a time from another week could be use. |
| Unable to submit the work by the deadline in case of any unexpected issues. | High | Somewhat Likely | Enough time range is given before the deadline 7th May 2020, to resolve any unexpected issues. |
| Sickness | Low | Somewhat Likely | Distribute the work fairly with the time provided so in the case of sickness, it will be manageable to catch up later. |

## Conclusion:

To sum up, this report discussed the proposed final year project for Computer Science and Forensics Security degree. It provided a project description, project aims and objectives, and a work plan, including a Gantt chart. It also indicates the supervisor's role in the project, ethical consideration, and a risk management.

## References:


- Juniper research. (2020). *Cybercrime will Cost Businesses Over $2 Trillion by 2019*. [online] Available at: https://www.juniperresearch.com/press/press-releases/cybercrime-cost-businesses-over-2trillion-by-2019 [Accessed 1 Feb. 2020].

- Marketing, D., Community, M., learningMobileVoice, E., Hospitality, D., us, &. and Dougherty, J. (2020). *Internet growth + usage stats 2019: Time online, devices, users - ClickZ*. [online] ClickZ. Available at: https://www.clickz.com/internet-growth-usage-stats-2019-time-online-devices-users/235102/ [Accessed 1 Feb. 2020].

- Taylor & Francis. (2020). *The Cyber Attack on Saudi Aramco*. [online] Available at: https://iiss.tandfonline.com/doi/abs/10.1080/00396338.2013.784468?casa_token=-dn_g-rYS3gAAAAA:MDeTQQvJsYpqL4BWhvW6OrfmeSLZ7cZK2c8UTocQLIIYjxuE0gMYXXt8HShAkHqJ7kt7zXDr9uhN#.XjXf4S2cY0p [Accessed 1 Feb. 2020].