Project Proposal

Benjamin Lee Xin Lone

registration: 100292465

1 Description

This project aims to identify the impact that machine learning can have on the fight against ransomware while also looking into alternatives. The COVID-19 pandemic has significantly increased the amount of cyberattacks around the world due to reasons such as, working from home where the internet is less protected, and an increase in unemployment which encouraging those who had been relieved to commit cybercrimes (Beaman et al., 2021). This will continue as long as the pandemic and movement restriction are still in place. To combat this, cyber security of those working from home must be improved for the foreseeable future, with one this solutions being an Anti-Ransomware software that makes use of machine learning. Ransomware can be very difficult to detect as it is able to change its behaviour alongside the ability to camouflage into any program, email attachment, and files alike. Machine learning will be able to adapt to how existing ransomware behave, making it easier to detect currently known threats alongside ransomware with similar traits. This project will aim to solve some of the major controversy surrounding the use of machine learning to detect ransomware.

2 Research

Machine learning was created to use a set of data and algorithms that resembles human intelligence through the act of analysis (El Naqa and Murphy, 2015). When given a set of data on currently known ransomware, the machine will be able to analyse each an every known variant and how each of them behaves. The idea result would be a machine that would be able to predict and contain new variants of ransomware through the act of learning behavioural patterns in new and older variants. This however will not be sustainable in the long term as ransomware become more complicated and unpredictable. A way to get around this would be to train a machine using an One-Class Support Vector Machine that will teach it to identify positive data results, while data that is different compared to the positive data set would return as negative (Sahs and Khan, 2012). In order for One-Class SVM to work, an ample amount of data must be provided to the machine in order to get ethical results. This could lead to a case where by the machine would be trained in an unsupervised manner, furthermore, high-dimensional datasets may make it more difficult for the machine to detect anomalies (Erfani et al., 2016).

3 Approach

The project will be done using the traditional waterfall method as a majority of the deliverables would depend on each other to proceed. The first half of the project will consist of researching how ransomware is spread, current ransomware threats, reviewing recent and pass ransomware attacks, analysing existing anti malware software's, as well as researching how machine learning has already been used to combat against ransomware. The research task should be given a timeslot from the beginning of week 6 to at least the end of the autumn semester. The second half of the project would be building a simple One-Class SVM Anti-Malware software that would be able to detect malware and ransomware in the selected file or directory, another software without machine learning would be created as a controlled variable so that a comparison can conclude the effectiveness of machine learning. This half of the project will take up the rest of the academic year to ensure that the program works as intended.

4 Risks

There are a number of risk in the potential project proposal, some of them are time based risks that depends on how the other parts of the projects, and some are risks related that relate to the testing of the program itself. The list of Risks includes:

- Testing of potentially harmful malware
- Time taken to search and teach data to Machine
- Time taken to bug test machine learning program
- Program not being able to work as expected

Countermeasures would need to be placed, especially in regards to testing with malware. Testing can be done in a virtual machines such as Oracle's VM Virtual Box which simulates operating systems in a safe environment making it perfect to test the program. The use of unharmful test malware can be another alternative, provided by EICAR. In regards to time based risks, it would be better to allow for a longer testing period of the program, allowing for more time to fix bugs as well as fix any potential problems in regards to machine learning.

References

Beaman, C., Barkworth, A., Akande, T. D., Hakak, S., and Khan, M. K. (2021). Ransomware: Recent advances, analysis, challenges and future research directions. *Computers like& Security*, page 102490.

El Naqa, I. and Murphy, M. J. (2015). What is machine learning? *Machine Learning in Radiation Oncology*, page 3–11.

Erfani, S. M., Rajasegarar, S., Karunasekera, S., and Leckie, C. (2016). High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning. *Pattern Recognition*, 58:121–134.

Sahs, J. and Khan, L. (2012). A machine learning approach to android malware detection. In *2012 European Intelligence and Security Informatics Conference*, pages 141–147.

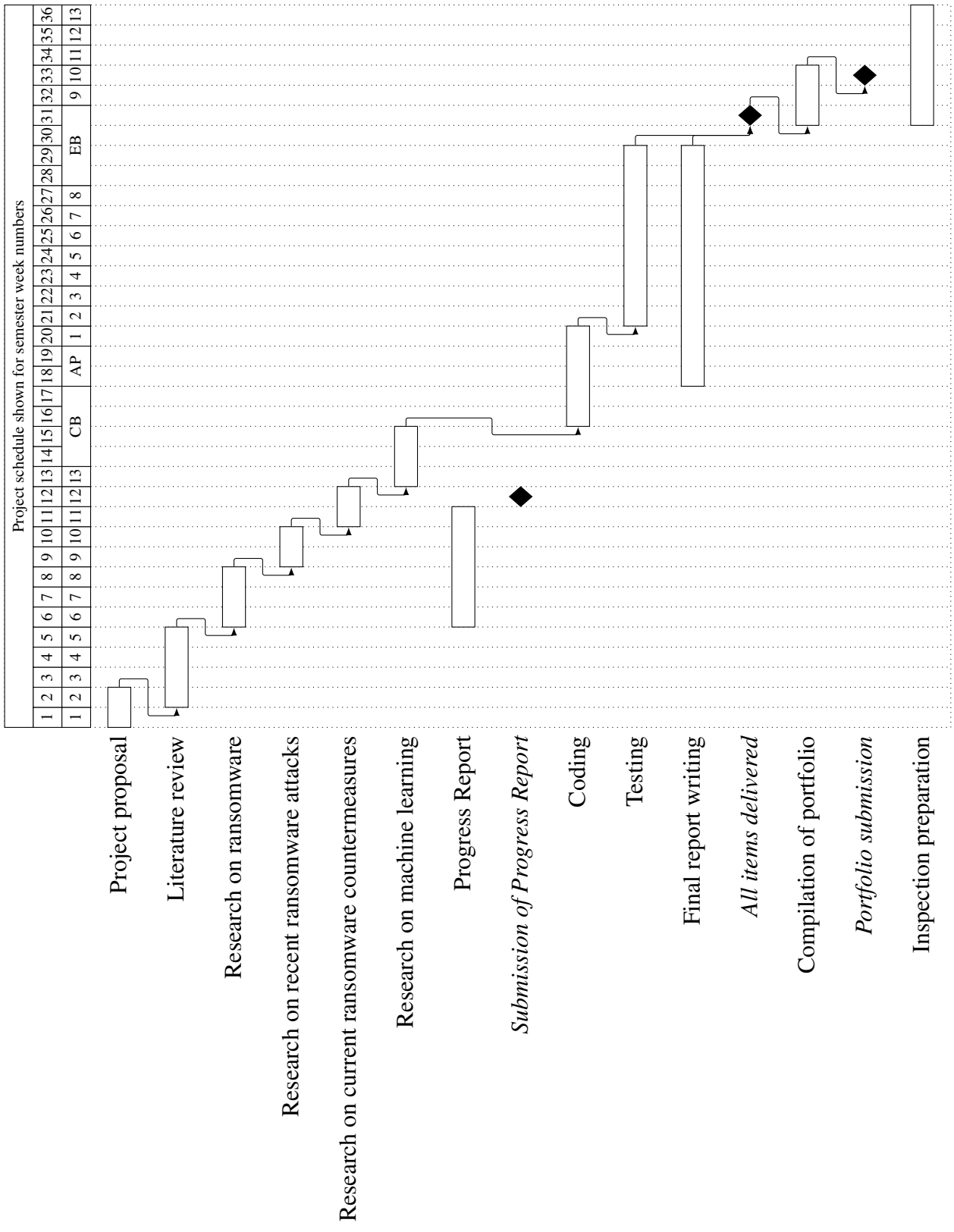

Figure 1: Project Gantt chart