Initial Plan
Simulating the Effects of Releasing Malware into the Internet of Things

CM3203 - One Semester Individual Project (40 Credits)
Author: Jamie Knowles
Supervisor: Eirini Anthi

February 5, 2018

Project Description

By the end of this year, there will be more than eight billion internet of things (IoT) devices connected worldwide [1]. Many of these devices are left vulnerable and are therefore constantly exploited by hackers across the globe. In October 2016, a company that controls a large majority of the internet's domain name system (DNS) infrastructure called Dyn was hit by a distributed denial of service (DDoS) attack that let to numerous popular sites suffering downtime including Twitter, Netflix and Reddit. The attack was orchestrated by a strain of malware known as Mirai, the strain infects vulnerable internet-connected devices to form a botnet. In this instance, the botnet was used to coordinate a DDoS attack on Dyn. The attack is known as one of the largest disruptions the internet has suffered. [2] It is clear that, with the amount of internet-connected devices rapidly rising, more needs to be done to raise awareness of how vulnerable the Internet of Things is and what the consequences of these vulnerabilities could be.

Therefore, I aim to create an application that demonstrates how vulnerable the Internet of Things is by simulating the effects of releasing specific strains of malware such as Mirai and Brickerbot as well as provide information on the most common potential attack vectors. To achieve this I will require real device data that will be used to simulate internet-connected devices, which I will acquire from Shodan (`shodan.io`). In order to simulate the effects of releasing a strain of malware it is necessary to know the potential attack vectors for each device and so, using maching learning methodologies, I will classify these attack vectors. Using the knowledge of these attack vectors and other information supplied by Shodan the effects of releasing a specific strain of malware can be simulated.

Ethical Considerations

Although the data that I will be using is all publicly available on the Internet it is still necessary to consider that the devices in my dataset may belong to individuals and could disclose information that individually identifies users. After conversing with the COMSC Ethics Committee they concluded that ethical approval was not needed.

Aims and Objectives

The principal goal for this project is to create a system that simulates and visualises the effects of releasing certain strains of malware. This can be achieved by completing the following objectives:

- Gathering Requirements -
  This phase will involve gathering the applications functional and non-functional requirements of the system.
- System Design -
  After I have gathered requirements I will then need to design the overall architecture of the system. The system will be designed with modularity in mind so that each service is lightweight and has it's own unique, well-defined goal. Further to the design of the overall architecture I will also need to design the user-interface.
- Researching IoT malware -
  I will need to research different strains of malware to get a better understanding of them, the attack vectors they try to exploit and the types of devices they try to target. This research will aid me when I reach the stage of creating a knowledge base of the different types of malware the application will be able to simulate.
- Gathering a Dataset -
  1. Retrieving the data -
     The majority, if not all, of the data I retrieve on internet-connected devices will be from Shodan.
  2. Pre-processing the data -
     Pre-processing the dataset will allow me to prepare the dataset in the way I see fit.
  3. Classifying the data -
     Once I have pre-processed my data I will then train a classifier to predict potentially exploitable attack vectors on each device.
  4. Storing the data -
     After I have trained a model and used the model to classify my dataset I will then store the classified data in a database hosted either on my personal computer or on an AWS EC2 instance.
- Developing the API -
  During this phase I will create the back-end API, this will allow the user interface to retrieve the data from the database. The back-end API will not only contain a means of communication between the user interface and the database but also a knowledge base on the various types of malware the system can 'release'. Once I have finished implementing the back-end I will then perform automated testing to ensure quality.
- Developing the User Interface -
  This phase will involve implementing the user interface. The main goal of the user interface is to visualise the effects of releasing the different strains of malware. I will visualise this with a map showing devices getting infected 1-by-1 and a command-line interface displaying the more-detailed actions taken by the malware in it's simulation. In addition to selecting the strain of malware to release, I will add additional options that will allow the user to perform a more specific attack such as by only performing an attack within a specific area. After the implementation has been completed I will then perform manual and automated testing.

Workplan

Week	Objective	Deliverables	Milestones
29.01-04.02	Research malware/technologies and plan project	Initial plan	Initial plan created and submitted. Research conducted on malware and potential technologies to use.
05.02-11.02	Start final report and research malware		Draft of introduction, background and approach written in initial report. Further research conducted on malware.
12.02-18.02	Research malware and gather dataset		Large dataset retrieved and pre-processed. Further research conducted on malware.
19.02-25.02	Research machine learning techniques and start developing classification model.		Knowledge on the best way to implement a classifier with my dataset will have been obtained.
26.02-04.03	Continue developing model and test model.		Model trained and tested. An accurately classified dataset of considerable size will have been obtained. A project review meeting with my supervisor will have taken place.
05.03-11.03	Develop API		
12.03-18.03	Develop API		
19.03-25.03	Test API		Back-end API implemented and tested thoroughly.
26.03-01.04	Develop user interface		
02.04-08.04	Develop and test user interface		User interface implemented and tested thoroughly.
09.04-15.04	System review and integration testing		System implemented and tested as a whole unit. A user is able to simulate the effects of releasing a specific strain of malware. A project review meeting with my supervisor will have taken place.
16.04-22.04	Write report		
23.04-29.04	Write report		Draft of final report completed.
30.04-06.04	Finalize report		Final report completed.
07.04-11.04	Contingencies	Final report, source code and datasets	Final report finished and submitted along with all source code and necessary datasets.

Further to my weekly objectives, I expect to have weekly meetings with my supervisor. I also expect to continuously be working on my final report throughout each week with weeks 12-14 solely focused on report writing.

References

[1] Gartner. “Gartner Says 8.4 Billion Connected ’Things’ Will Be in Use in 2017, Up 31 Percent From 2016.” In: (2017). URL: https://www.gartner.com/newsroom/id/3598917. (Accessed on 2018-01-29).

[2] N Woolf. “DDoS attack that disrupted internet was largest of its kind in history, experts say.” In: (2016). URL: https://www.theguardian.com/technology/2016/oct/26/ddos-attack-dyn-mirai-botnet. (Accessed on 2018-02-02).