Charles Darwin University

**Distributed Denial of Service Forensics using MapReduce**

Project Management Plan

**Created by:**

Mayra Dwight - s316934

Sachin Malik - s325383

Ankit Chanana - s321903

Abhijith Reddy - S323564

**PRT 581 - Principles of Software Systems**

Dr Sami Azam

17 March 2020

# Table of Content

Table of Content 2

Revision History 3

Project Charter 5

Project Background 5

Project objectives 5

Project scope 5

Project feasibility 5

Project milestones 5

Project requirements 6

Project Governance 7

Project Approach 8

Project Schedule 8

Attachment A – Initial Project Schedule 9

# Revision History

This document has been revised as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Version** | **Primary Author(s)** | **Description of Changes** |
| 17/03/2020 | Initial | Mayra Dwight  Sachin Malik  Ankit Chanana  Abhijith Reddy | First release |
|  |  |  |  |
|  |  |  |  |

Approval

|  |  |  |
| --- | --- | --- |
| **Recommended for Approval** | | |
|  | |  |
| Charles Yeo |  | Date |

Distribution

| **Name** | **Title** |
| --- | --- |
| Charles Yeo |  |
| Sami Azam |  |
| Mayra Dwight | Information Systems |
| Sachin Malik |  |
| Ankit Chanana |  |
| Abhijith Reddy |  |

# Purpose of Project Management Plan

[ADD]

# Project Charter

## Project Background

A Denial of Service (DoS) attack is used in order to make the Internet resources unavailable by overwhelming the victim with large number of requests(Karthik, S et al., 2008).In recent years with the advancement in technology, DoS attacks has shifted from single source attacks to multi-source attacks known as DDoS(Distributed Denial of Service) attacks. To perform a DDoS attack, the attacker first infects large number of devices with a malware. These infected devices are known as bots or zombies and are controlled by the attacker known as bot master. Attacker can command these infected devices to send SYN requests to the target website, server or an application at same time and use all its resources to crash it. DDoS can be broadly categorised into bandwidth attacks and application layer attacks. In bandwidth attacks the attacker floods the bandwidth of the victim with high volume of requests and as Internet is limited in its processing capability therefore it is unable to process further and fails to respond to legitimate users. Attacks such as TCP attack, UDP attack and ICMP flooding are categorised as bandwidth attacks. Application layer attacks include low-and-slow attacks, GET/POST floods and attacks that target Apache vulnerabilities. The main aim of these attacks is to crash the web server. There are various tools available online which require minimal or no technical knowledge to perform a DDoS attack. Some of the common tools that are used to launch a DDoS attack are: Tribal Flood Network (TFN), Low orbit ion cannon (LOIC), Trin00, Trinity and Mstream (Arun Raj Kumar, P et al., 2009).Based on the time of action DDoS defence mechanisms can be classified into three categories: Before the attack, During the attack and After the attack

1. Before the attack: This type of defence mechanism is used in order to prevent the attack from happening.

2. During the attack: Mechanisms in this category are used when the attack is happening. These mechanisms are used to detect the attack. Intrusion Detection and Prevention Systems (IDPS) and firewalls are used under this category.

3. After the attack: This mechanism is used once the attack is detected and is used to trace back the attacker. Forensic investigation comes under this category(Rajkumar & Nene, M, 2013).

In our project we will be focusing on the third part (After the Attack) i.e. forensic investigation.

As the amount of data produced during a DDoS attack is enormous, analysing the log files and then recovering from the attack takes a very long time. In this project we will be investigating the use of Hadoop and MapReduce to detect packets that belong to a DDoS attack. MapReduce is a software framework introduced by google for supporting the distributed computing of large data sets on cluster computers. The basic idea of our project is to divide the enormous log files into sub parts which can be then analysed parallelly to determine the packets which belong to a DDoS attack.

## Project objectives

## Project scope

## Project feasibility

## Project milestones

## Project requirements

**Business Requirements**

* Finding Source of Attacker.
* Distinguish between good traffic and bad traffic.
* Identifying the types of attacks.
* Identifying the devices that are used.
* Efficient MapReduce programs.
* Fast Recovery of the Server and its processes.
* Real-time analysis/detection (Live-Demo of the product).

**Project Requirements**

**Set-up of Infrastructure/multi-node Hadoop cluster**

* Google Cloud Platform
  + 3 servers (1 Master node, 2 Salve nodes)
  + Boot with CentOS 6.
  + Selecting CPU speed, size of RAM, Memory etc.,
    - **Master Server [CPU Core: 2, RAM: 7.5GB, CentOS 6, SSD: 50GB].**
    - **Slave Server 1 [CPU Core: 1, RAM: 3.5GB, CentOS 6, SSD: 30GB].**
    - **Slave Server 2 [CPU Core: 1, RAM: 3.5GB, CentOS 6, SSD: 30GB].**
  + Generating RSA key.
* Connecting Personal PC with the 3 servers.
  + Establishing a Secure Shell (SSH) connection between PC and the 3 servers using RSA key.
* Setting up Multi node cluster in Hadoop
  + Installing Java on each server.
  + Creating User Account.
  + Mapping the Nodes.
    - **Hosts** file should be edited in **/etc/** on all the nodes and IP address of each system followed by host names.
* Configuring Key Based Login.
  + **SSH**/telnet should be setup in each node, so they can communicate without any password.
* Installing Hadoop on Master Server.
  + Download Hadoop.
  + Configuring Hadoop.
    - **Core-site.xml** should be edited.
    - **Hdfs-site.xml** should be edited.
    - **Mapred-site.xml** should be edited.
* Installing Hadoop on Slave Servers.
  + Download Hadoop.
  + Configuration.
* Editing Additional configs and staring Hadoop services.

**Gathering Network Log files (Data Sets)**

* Based on the CPU, RAM and Memory of the servers, the log file size should be decided.
* Requesting Data sets from the providers.
* Offline analysis of data.

**MapReduce Programs that can process the log files. The code should be based on the requirements of the business/company.**

* **Mapper, Sorting, Reducer.**
* Gathering different modules from various creators/source codes and combing those modules.
* Analyzing the Log files to check the efficiency of the code before implementing in real-time analysis.

**Real-time DDoS detection**

* **DDoS** attack Stimulation (**Network Stimulators**).
* Selecting no. of Zombies.
* Selecting the rate of attack and duration.
* Implementing the Hadoop Multi node cluster in Real-time.

# Project Governance

# Project Approach

## Project Schedule

The project schedule has been included in this Project Management Plan. Refer to [Attachment A.](#_Attachment_A_–)

The project schedule is a living document that will continue to be updated by the project team as the activities are executed or scheduled.

# Attachment A – Initial Project Schedule

