HardnBot
Intelligent Server Hardening Software

Project ID: 19_20-J 01

Preliminary Progress Review Report

G.G.L Anjula IT16022416

Bachelor of Science (Hons) in Cyber Security

Sri Lanka Institute of Information Technology

August 2019

HardnBot
Intelligent Server Hardening Software

Project ID: 19_20-J 01

Preliminary Progress Review (PPR) Report

G.G.L Anjula IT16022416

Supervisor Mr. Amila Nuwan Senarathne

Bachelor of Science (Hons) in Cyber Security

Sri Lanka Institute of Information Technology

Sri Lanka

August 2019

DECLARATION

I declare that this is my own work and this Preliminary Progress Review (PPR) report does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

……

G.G.L Anjula (IT16022416)

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.

Supervisor

……

Mr. Amila Nuwan Senarathne

Table of Contents

List of Figures

1. Introduction

1.1.Purpose

This Preliminary Progress Review (PPR) report is created to give a clear idea about progress and the process of development of the research project "Intelligent Server Hardening Software (HardnBot)" to the interest parties. Through this document we will explain the purpose of the project, requirements (functional and non-functional), clear explanation about the technologies we are going to use, what are the innovative features of this project, what are the hardware/software requirements for best performance and for availability, what are the major areas we hope to cover in this research project, how are we going produce and fulfil the requirements.

By utilizing this Preliminary Progress Review anybody can survey the procedure of work flow in the various periods of the project and check whether the undertaking group meets requirement particulars which distinguished previously. This archive permits to distinguish deviations of venture results and furthermore acknowledge changes in requirements.

This PPR report is mainly created for the supervisor and the research team to make a good understand about requirements (functional and non-functional). When we are considering the next phase or another paths of this research, this PPR report will be helpful for researchers who take this research's outcome for another research.

1.2. Scope

Intelligent Server Hardening Software (HardnBot) is a software that has the capability to automatically detect poor/non-compliant configurations in Linux servers which has CentOS 7 as an operating system and applying industry recommended fixes for them.

Under this PPR document, following research components are described:

- Implement backup function
- Implement intelligence rollback function
- Setup SSH connection to servers through secure VPN tunnel

Software is divide to fourteen main parts and all of them are important to final outcome of this project. This report cover three parts of that. Through this document these components will be described with validation.

1.3 Overview

Server Hardening is the one of the most important tasks to be handled on servers. Server hardening, which is also referred to as operating system hardening, is the process of making the server stronger and more resistant to security issues. Server hardening is an inexpensive and simple task to improve the overall operating system for maximum performance and to reduce expensive

failures. Hardening is a Process requires many steps, all of which are critical to the success of the hardening system. The more steps a user follows, the safer and more resilient the system will be.

Normally these hardening processes will be done by either network administrators, system administrators, outsourced professionals or server custodians by manually running scripts, commands and queries against the server and it will roughly take more than six hours to completely harden a single server in the infrastructure. Probability of a misconfiguration occurrence is higher because hardening will carry out with human interaction. Scenarios where a misconfiguration occurs, it may be hard to detect those issues since some issues cannot be identified via an error message. So, in a scenario like that, system administrators, server custodian or network administrators need to go back to the initial state of the server operating system via a backup image which will be a time-consuming task. By introducing a special feature/function for improve backup and rollback tasks, fully backup to the initial state will not be necessary. Rather that the software itself can automatically identify the misconfigured spot and rollback only to that point.

2. Statement of work

2.1. Background information and overview of previous work based literature survey

Manual Server auditing and hardening is a process that server custodian will manually audit the server for any compliance failures and correct them manually. There is no fully automated intelligent hardening platform implemented yet. Even the network administrators plan to do the Hardening processes manually, it might take more than 6 hours for the complete harden processes to complete.

Zhe Wang, Jin Zeng, Tao Lv, Bin Shi and Bo Li published a research paper in 2016 on "*A Remote Backup Approach for Virtual Machine Images".* In there they were talking about virtual backup on a cloud storage. When we are considering cloud computing, virtualization is playing a major role, because of hosting several applications and services in virtual machines (VM) which were hosted in cloud environments. Security become a prior requirement in virtualized applications. In this research, mainly focused area is high availability issue in virtual machines. LiveRB (Live remote backup) is the proposed remote backup approach. The purpose of the Live RB is to save the running state of the VM in an online manner known as "Live Migration". This backup process will happen the background of the hosted cloud applications of the VM and is transparent to them. A virtual block device will be designed and will be used to cache of I/O Operations in memory, in order to save the incremental virtual disk data.

LiveRB will be implemented on KVM virtualization platform in order to evaluate effectiveness and efficiency using a set of comprehensive experiments. These experiments are all related to Cloud Computing and the security issues that come along with this and the key points considered in order to have successful cloud computing are security, availability & fault tolerance. The commonly used solution to handle Fault Tolerance & High Availability is using snapshots or checkpoints that periodically record the states of the software for backup and rollback the cloud applications to the previously backup up state. This procedure will be carried out when encountering Failures or Errors of the original system.

Most currently existing VMs stop the VM to take snapshots. Some VMs need to be shut down too take snapshots which this affects the ability to provide the service/ result in abnormal cloud application behavior. Some VMs suspend the current process and save the current progress onto local disks to be transferred onto remote servers later which sometimes result in data loss if a hardware failure is encountered.

The above issues can be resolved using the Live RB since it works by not stopping the VM to do the backup process. Results of this process indicate that Live RB can be used on a VM to do the backup task from VM onto a Remote Server with only a slight reduction in performance [1].

In this research, it described about method that used to back up a virtual machine, but when it comes to our research area we have to consider about live server. Therefore, no need of care about any virtual machines, but when we are talking about remote backup approach used in here, that was Live Remote Back up, so we can consider about this technique when we are dealing with our problem.

L. Farinetti and P. L. Montessoro published a research paper in 1993 which named "*An Adaptive Technique for Dynamic Rollback in Concurrent Event-Driven Fault Simulation*". In here it is discussing about automatic rollback based on an adaptive mechanism which is including advanced network/system status recording system. Time can be any time, that mean before changing of a system or after changing a system this status recording can be apply. Main feature of this research is user can define the rate for maximum acceptable level for rollback. This approach takes the average time to minimum level, that means very short time of rollback process.

To come up with proposed technique, researchers were used existing methods such as incremental backups, journal files, checkpoints, rollback, roll forward which were found on different applications, different operating systems as well as different databases. Mainly the status of the network/system is record on disk and run for negative time period to analyze previous status. If needed user can run for a positive time period as well. Those time periods are for compare with current status of the network/system. To make it happens above approaches need some fine tunes as well [2].

According to rollback techniques used by those researchers we identified some techniques and requirement that should be in our system too. For this part of the software it is necessary to detect abnormal behaviors of the server after the hardening process is done. For that we need to record system status after the hardening process. Then compare with the previous status, that means system status before the hardening process, but in our approach there are pre define models for compare with current system status. Apart from that rollback mechanism is going to adopt from this research. That are the things we are going to take from this research.

Ning Lu and Yongmin Zhao published a research paper in 2018 which named "*Research and Implementation of Data Storage Backup*". In this research, researchers were tried to discuss about features of a reliable and secure backup and types of backup. With use of applications which were depend on big data, the usage of data storage backups was became more important. Therefore, the methods used to backup should be more flexible and can be able to ensure of security and reliability of backup contents and also backup and restore should be in a convenient manner. There are several backup methods such as data backup, system backup, application backup etc. The backup contents are guaranteed to be confidential, complete and effective.

There are several specific performances in a backup,

i) Backup should be upgradable, capacity expansion
ii) Management without affecting other application in the system
iii) Implement a backup storage system combining SAN (storage area networks) and NAS (network attached storage) storage networks.
iv) Provide several backup methods such as data backup, system backup, application backup,
v) Backup contents should be secure and restore operations should be done in a convenient manner.

System backup

Reefers to the backup of the end-point operating system, server operating system and other systems. In here core files and system's registry are backed up as a data. In a matter of system crash or operation mistaken the backup can be restoring to the previous state.

Virtual tape library

Virtual tape library (VTL) considered as a world's leading modern technology to create a backup system. It can rapidly backup and rapidly recover a system that we want to backup. Main feature is no manual intervention of this technology. VTL storage media is a SATA disk and its data transfer rate is 150MS/s. That means approximately it takes 10 seconds for transferring 1.5GB data to the backup storage [3].

In our research one of a main goal is to reduce the overall hardening time. For achieving that task, we should have to minimize the overall backup time to some acceptable level using speed backup mechanism. In this point we are going to use technique which is described in this research known as virtual tape library. If we can adopt this mechanism overall hardening time will reduce averagely by 8 hours to 4 hours.

Teruaki Sakata, Teppei Hirotsu, Hiromichi Yamada and Takeshi Kataoka published a research paper in 2007 which named "*A Cost-effective Dependable Microcontroller Architecture with Instruction-level Rollback for Soft Error Recovery*". This tool is developed for detect soft errors using electronic design automation (EDU) which generates optimized soft error detecting logic circuits for flip-flops. When a soft error is detected that signal goes to a developed rollback control module (RCM). That RCM will reset the CPU and restores the CPU's register file from a backup register file using a rollback program guidance. After that CPU will able to restart from the state which is before the soft error occurred. In here researchers were developed another two modules called error reset module (ERM) that can restore the RCM from soft errors and error correction module (ECM) that corrects errors in RAM after error detection with no delay overhead. In above mentioned soft error means, which are random transient errors. Those errors are the main cause of failures in microcontrollers which include reversal of a memory element's bit data due to factors such as alpha rays in a package, neutron strike and noise of the environment [4] .

D. R. Avresky and M. I. Marinov published a research paper in 2011 which named "*Machine Learning Techniques for Predicting Web Server Anomalies*". The basic idea between servers on the web is to provide requests made by the client through the web using different transmission methods such as Services. Businesses relying on these services require the web servers to have

reliability, availability and security in order to provide constant quality in the service provided. This document describes the quality ensured in these services.

The assumption made for this problem is mainly due to Resource Starvation. Resource Starvation is when a process that functions in Concurrent Computing is unendingly denied the necessary resource to continue & process the rest of its work. Resource Starvation is measured by the response time taken to cater requests under artificial workloads while collecting data on other resource parameters. The research provides proof that these recordings gathered from different artificial workloads can be applied to real world entities as well

Machine Learning is used to monitor & correlate the high response time and this is done by observing the system data. The goal of this analysis is to resolve issues of this variety in Web Servers, Operating Systems or in VM (virtual machine) Rejuvenation.

Based on the statistics provided by the Internet World Statistics, we could clearly notice a rapid rise QoS (quality of service) Internet Service Usage users and this gave several companies & industries to exist in the current world. The below listed out Companies/ Industries who gets affected by these figures since their prime business is offering Internet QoS,

- Cloud Computing
- Data Storage
- Hosting Providers
- Content Delivery
- Application Performance Management & other

Due to this high demand and dependence on network QoS, it is important for a particular service to be aware of its own deteriorating quality. Currently there are several self-monitoring network products that ensure that the QoS of services offered through the internet. The goal of this this research is to increase this area.

The benefits taken from this research can be applied to other areas as well and they have been listed down below,

- Proactive Software Rejuvenation
- Web Server Workload Balancing
- Web Server Performance Testing

  Other… [5]

In this research mainly focused about detecting anomalies on a web server using machine learning technique. Hence we are not going to use machine learning techniques for detecting anomalies in a server this research is not a to good feed for us.

2.2. Identification and significance of the problem

In modern, information systems basically consist with operating systems, databases and applications/services. Except personal computers, majority of the servers include all above components, because of that server custodian have to pay attention about complete security of these

servers. Even though there are many of vulnerability scanners to detect vulnerabilities and auditing tools available, there are no proper tool for automatically detect vulnerabilities, automatically generate risk score of the asset, automatically backup server, automatically apply industry recommended fixes and automatically rollback if any failure is detected. Most of organizations are done above mentioned procedures manually and separately. The manual auditing procedure expend much more resources, for example human resources, time, cost moreover. Apart from that most data centers include Linux servers; Ubuntu Server, Red Hat Enterprise Linux and CentOS. Datacenter includes about 200 live servers it is very difficult to do the operating system hardening manually.

When server is ready to harden it is mandatory to take a backup of the server and it is usually happening in all of the server hardening tools to reduce risk of server failure which can be occur after the hardening process. But in this tool backup is taken automatically during the hardening process. User only has to do is choose the backup path using user confirmation pop-up which is appear during the hardening process. There are two options in the pop-up box, user can either select HardnBot node machine's disk place or user can select target server's disk place. Actually backup is going to create using Virtually Tape Library technology. This is world leading and fastest technology to create a backup.

Next phase is rollback part. In modern, most of the rollback functions are done by manually, because of that HardnBot is going to provide better solution for this. After the hardening process is done HardnBot let the server to normally run its services and analyze any abnormal behaviors on that services or the server. If any, there is a script to roll back the operations which establishes the previous state of the server using previously taken backup. This procedure is done automatically which is cannot expect form other hardening tools.

2.3. Technical objectives (specify s/w and h/w requirements)

Main Programming Language: C#

Almost all operating systems support C# language. C# is a general-purpose, multi-paradigm programming language including strong typing, lexically scoped, imperative, declarative, functional, generic, object-oriented, and component-oriented programming disciplines. And, C# contains with lots of supporting third party libraries. Therefore, C# will be used as the main programming language to implement this software.

IDE: Visual Studio

Microsoft Visual Studio is an integrated development environment from Microsoft. It is used to develop computer programs, as well as websites, web apps, web services and mobile apps. It is mainly support for C# development and contains lots of features that improve programming experience. And because of the inbuilt functions and configurations, we are going to use Microsoft visual studio as our development IDE.

Virtual Environment: VMware Workstation

VMware Workstation is the industry standard for running multiple operating systems as virtual machines (VMs) on a single Linux or Windows PC. As the need we have to run several server-based operating systems to test our product, we will use VMware workstation for run multiple operating systems and observe the product outcomes.

System monitoring tool: Nagios

Nagios, also known as Nagios Core, is a free and open source computer-software application that monitors systems, networks and infrastructure. Nagios offers monitoring

and alerting services for servers, switches, applications and services. This tool will be used to monitor our servers for any misbehaviors after the hardening. Abnormal behaviors are going to detect using predefine models which are adopt from Nagios application.

Script Language: Shell

Shell script A shell script is a list of commands (a program) designed to be run by the Unix shell. We will use shell script for scanning purposes and other types of command executions in Linux servers.

Operating System: Linux

Linux is a family of open-source Unix-like operating systems based on the Linux kernel. Research's scope is limited to Unix-like operating systems based on the Linux kernel. HardnBot is developed to comply with RHEL 7,6 and CentOS 7,6.

3. Research Methodology

1. Implement backup function.

Existing system configurations should be taken into backups as a precaution if the new configurations failed. For this purpose, scripts will be designed for each OS/DB components. These scripts will generate a single backup file which can be used to restore in case of applied fixes failed.

2. Implement Intelligent Rollback function.

Rollback function will take place after the hardening process is done. This function basically depends on the abnormal behaviors of the server which are appearing after the hardening process. In here there are several pre-define behavior models to detect those kind of abnormal

behaviors. For that purpose, we create our own models as well as we can use existing models in various tools such as NAGIOS for this task. If those models detect any anomaly regarding to server services, there is a rollback script to run for establishing previous status (backup) of the server and it will automatically run.

3. Setup SSH connection to servers through secure VPN tunnel

When connecting to the server which is going to harden, it should be done in secure manner. For fulfill this requirement SSH connection through a VPN tunnel is the better way. All the processes should be done through this secure SSH session.

4. Test data & analysis

We will use VM Ware to setup several virtual machines to take several backups using several backup technologies and test the reliability and the minimal time of taking backup. Same as the backup function, HardnBot will use several virtual machines which are deployed in VM Ware to check abnormal behaviors of servers after the hardening process is done and to check the functionality of the rollback function. All of this observation will based on the statistical analysis of test data.

5. Anticipated benefits

HardnBot has a great commercial value because of the functionalities that it provides for users as well as organizations. Our primary goal is to automate an entire server hardening process and with the unique functionalities, HardnBot's user will gain following main benefits.

- **Fully in-depth server scanning specifically for OS compliances**

With this functionality, HardnBot can scan thoroughly to find any compliance failures and any configuration errors.

- **Failed compliance classification and summarize**

After collecting data from the scan, HardnBot's machine learning algorithms will provide a details list of compliances failures and misconfigurations with their criticality level. By using these, HardnBot will provides user with a detail percentage level of severity existence (For example: 10% Critical, 20% High, 30% Medium, 40% Low). With this, information server custodians can get an overview idea about the severity level of the server.

- **Risk Score**

By this functionality, an overall detailed risk score will be displayed to the custodian/user so that the organization can get a rough idea about server's possibility of compromise, likelihood, loss and probability of compromise.

- **Intelligent Hardening**

HardnBot' s hardening function is a unique function because it will harden the system for an acceptance level as required by the organization. Obviously 0% risk acceptance is not possible but in this function it's algorithms will use pre-classified compliance failure data and harden with respect to the severity levels.

- **Intelligent rollback functions**

Using these functions, HardnBot will monitor applied fixes on the run as well as the server behavior simultaneously and identify any abnormal behaviors and go back to that point where the abnormal behavior took place and rollback to the default or previously backed-up settings.

Besides those benefits, users will not require any down time to perform hardening because HardnBot's capability of performing stepwise hardening, stepwise error checking, stepwise backup and stepwise rollback. So, each fix will be monitored and having that, these functions will help to improve security operation center's (SOC) productivity and accuracy.

6. Project plan or schedule

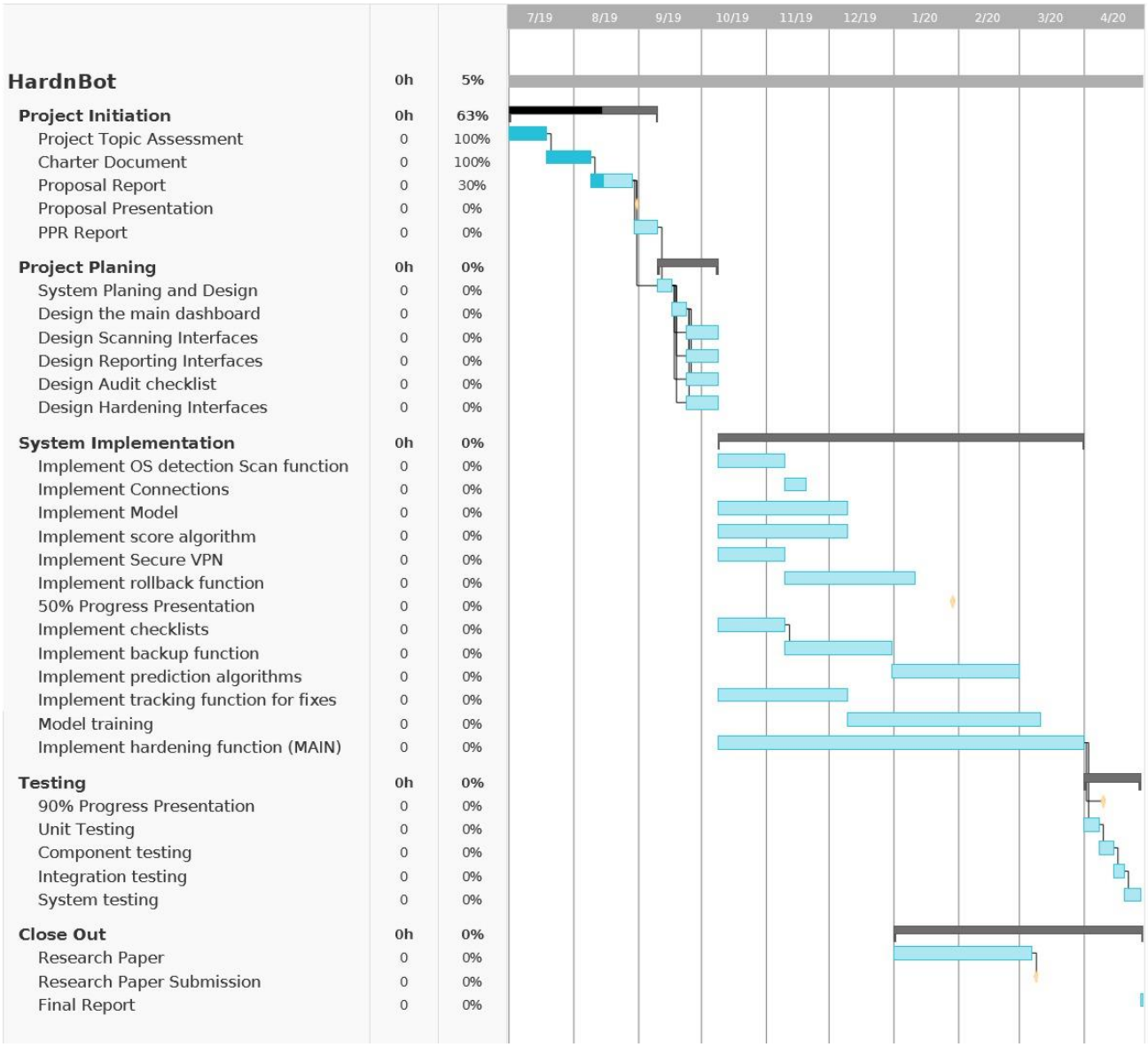

Figure 3: Project Schedule

7. Research constraints

- Limited to RHEL 7,6 and CentOS 7,6

  Because of not enough time, our research scope limited to servers which have RHEL 7,6 and CentOS 7,6.

8. Specified deliverables

Final product will contain following features:

- Automatic backup the server before it is going to harden
- Perform an intelligence role to ditect abnormal behaviours of the server
- Automatically perform rollback function after detecting any abnormal behaviour.

9. References

[1] Zhe Wang, Jin Zeng, Tao Lv Bin Shi, Bo Li, "A Remote Backup Approach for Virtual Machine Images," in *IEEE*, 2016.

[2] L. Farinetti, P. L. Montessoro, "An Adaptive Technique for Dynamic Rollback in Concurrent Event-Driven Fault Simulation," in *IEEE*, 1993.

[3] Ning Lu, Yongmin Zhao, "Research and Implementation of Data Storage Backup," in *IEEE*, 2018.

[4] Teruaki Sakata, Teppei Hirotsu, Hiromichi Yamada, Takeshi Kataoka, "A Cost-effective Dependable Microcontroller Architecture with Instruction-level Rollback for Soft Error Recovery," in *IEEE*, 2007.

[5] D. R. Avresky, M. I. Marinov, "Machine Learning Techniques for Predicting Web Server Anomalies," in *IEEE*, 2011.