

**SCHOOL OF COMPUTER SCIENCE, ENGINEERING AND TECHNOLOGY**

**DEPARTMENT OF COMPUTER SCIENCE AND IT**

**PROJECT PROPOSAL**

**A Project Documentation Submitted in The Department of Computer Science and IT in partial fulfillment of the degree of Computer Science**

**SUBMITTED ON,** November 11, 2020**.**

**TIKIVUN, A VULNERABLE VIRTUAL MACHINE**

**SUBMITTED BY: EVANS** KIPKOGEI KIPCHUMBA

**REG NO:** BMIT/MG/0538/05/17

**UNIT:** PROJECT PROPOSAL 1

**COURSE CODE:** INTE 414

# DECLARATION

This is to certify that to the best of my knowledge, the content of this project is my own work except where stated otherwise. This work has not been submitted for any degree or any other awards and all assistance received and /or sources has been acknowledged.

**APPROVAL**

**Student**

……………………….. Date: ……………….

………………………..

Evans Kipkogei Kipchumba

BMIT/MG/0538/05/17.

Supervisor:

……………………………… Date: ……………....

……………………………….

Joseph Mengo Orori

Department of Computer Science and

Information Technology

Kabarak University

# ACKNOWLEDGEMENTS

I want to thank My Supervisor for his support, guidance, and devotion throughout this semester for the amazing guidance in the documentation and of this project proposal, and above all to appreciate The Lord for granting this opportunity in the pursuit of success

Table of Contents

[DECLARATION 1](#_Toc55995055)

[ACKNOWLEDGEMENTS 2](#_Toc55995056)

[DEDICATION 4](#_Toc55995057)

[ABSTRACT 4](#_Toc55995058)

[CHAPTER ONE 5](#_Toc55995059)

[INTRODUCTION 5](#_Toc55995060)

[1.1 Background of the study 5](#_Toc55995061)

[1.2 Problem statement 6](#_Toc55995062)

[1.3 Objectives 6](#_Toc55995063)

[1.3.1 Main Objective 6](#_Toc55995064)

[1.3.2 Specific objectives 6](#_Toc55995065)

[1.4 Justification 6](#_Toc55995066)

[1.5 Significance of the study 6](#_Toc55995067)

[1.6 Scope of the study 7](#_Toc55995068)

[CHAPTER TWO 7](#_Toc55995069)

[LITERATURE REVIEW 7](#_Toc55995070)

[2.1 INTRODUCTION 7](#_Toc55995071)

[2.2 REVIEW OF RELATED LITERATURE 8](#_Toc55995072)

[2.2.1 WEBGOAT 8](#_Toc55995073)

[2.2.2 Metasploitable 9](#_Toc55995074)

[2.2.3 Tagua VM 9](#_Toc55995075)

[2.3 THEORATICAL FRAMEWORK 9](#_Toc55995076)

[2.4 IDENTIFICATION OF KNOWLEDGE GAP 9](#_Toc55995077)

[CHAPTER THREE 11](#_Toc55995078)

[RESEARCH METHODOLOGY 11](#_Toc55995079)

[3.1 INTRODUCTION 11](#_Toc55995080)

[3.2 DATA COLLECTION METHODS 11](#_Toc55995081)

[3.2.1 PRIMARY COLLECTION METHODS 11](#_Toc55995082)

[3.2.1.1 INTERVIEWS 11](#_Toc55995083)

[3.2.2.2 QUESTIONAIRES 12](#_Toc55995084)

[3.2.2.3 OBSERVATION 12](#_Toc55995085)

[3.2.2 SECONDARY COLLECTION METHODS 12](#_Toc55995086)

[3.3 VALIDITY 13](#_Toc55995087)

[3.4 RELIABILITY 13](#_Toc55995088)

[3.5 ETHICAL CONSIDERATIONS 14](#_Toc55995089)

[3.6 DATA ANALYSIS 14](#_Toc55995090)

[3.7 TESTING 14](#_Toc55995091)

[3.8 IMPLEMANTATION 14](#_Toc55995092)

[3.8.1 Software Design Life Cycle. 14](#_Toc55995093)

[3.8.2 Phases 15](#_Toc55995094)

[CONTEXT DIAGRAM 16](#_Toc55995095)

[DATA FLOW DIAGRAMS 16](#_Toc55995096)

[REFERENCES 17](#_Toc55995097)

[APPENDICES 18](#_Toc55995098)

# DEDICATION

I dedicate this project to my loving parents and lecturers for the tireless effort they have offered and shown towards my academic progression. They made sure I had access to every resource I needed in relation to my academics.

# ABSTRACT

TikiVun is a vulnerable virtual machine with more holes than a Swiss cheese! From my experience as an Ethical Hacking student, I got stuck several times because I did not have a platform to test my theoretical skills learnt in class. Considering the fact that it is illegal to perform a hack on an unauthorized system-even just basic information gathering or vulnerability analysis- many Ethical hacking students find it hard to put their skills learnt in class into practice. Considering the reality facing them which is presented by the curriculum they are using, there are no government institutions offering tertiary education which has Ethical Hacking as a course on offer. This forces those who want to do the same to choose between two available options, either enroll in a private institution of learning (and this is costly) or do self-study on the same. With this option, the resources needed are not available as open source which presents another challenge. But with TikiVun, these challenges will be addressed as it will give the students a platform to practice their skills free. The virtual machine is also set to cover majority if not all of the attacks, exploits and vulnerabilities that can be performed on any system.

# CHAPTER ONE

# INTRODUCTION

# Background of the study

TikiVun is a vulnerable virtual machine whose aim is to provide a platform where student doing Ethical Hacking and Security courses can expound on their skills. It is on the backgrounds that it is illegal to perform these tests on a live host or an unauthorized machine that this machine is built. It emulates the functionality of a normal machine.

Since multiple objectives rarely have points that maximize all the objectives simultaneously, the proposal aims to maximize the objectives as much as it can. The main focus is on those students who are just beginning their journey on the Cyber Security world and wish to gain skills that emulate the real world problems.

# 1.2 Problem statement

With the growth in technology in the world, many systems are created. Many of these systems perform very critical tasks and other systems also store and handle critical information which cannot be allowed to land in the wrong hands. To ensure this, there should be enough system security personnel who have vast skills in system security and cybersecurity. TikiVun ensures that those aspiring to major in the system security and cybersecurity gain these skills by presenting a platform where these skills can be tested and perfected. The virtual machine is to be designed in a way that it covers all majority vulnerabilities and weaknesses a system can have. It also will cover all exploits that can be performed on any given system. The machine is to be designed to replicate all the components that a system can have i.e. Database functionality, User authorization, User-data encryption amongst other functions.

# 1.3 Objectives

# 1.3.1 Main Objective

To create a platform where students doing Ethical Hacking courses can test their skills and also the various attacks learnt in class which models all the real world security flaws present in those systems in use.

# 1.3.2 Specific objectives

1. To enable people in the security sector to gain enough practical skills which enables them to help protect user information and also data from prying eyes and unauthorized access.
2. To avail an open-source platform for people in the computer security sector to test their skills and also expound on them.
3. To develop a system that ensures skill integrity of those using it by ensuring its components replicate the real world components of any given system.
4. To design a system that works perfectly on any given environment, be it a real host or a virtualized environment.

# 1.4 Justification

This will be the first system of its kind in the market that will eliminate the problem faced by those trying to enhance their skills in the Ethical Hacking field. The system will reduce the total costs incurred and will provide improved efficiency and availability. It will provide a user-friendly interface that will enabled quick adaptability to users, therefore, limiting their resistance to this technology. The system will accomplish the credibility in its general functionality.

# 1.5 Significance of the study

The study is significant with the fact that it presents a platform for students to test their skills learnt in class free. It will help create a well conversant community of security specialists. This is because it will cover majority important ideas and stuff concerning cybersecurity and Ethical hacking fields.

# 1.6 Scope of the study

**Time Scope-** The estimated time this project will take to be completed is 3 months.

**Geographical Scope-** This project will be implemented in any urban/semi-urban, or rural location. This is because it will be designed to work a local computer even one without internet connection.

**Content Scope-** This system will provide an interface and a link up aiding in communication between the student and the developer, giving a unique form of interaction that will provide a fast and responsive system. This will be for the sole purpose of fixing any kind of bugs that the system might have.

# CHAPTER TWO

# LITERATURE REVIEW

# 2.1 INTRODUCTION

A virtual machine is an emulation of a particular computer system. They operate based on the computer architecture and functions of a real or hypothetical computer, and their implementations may involve specialized hardware, software or a combination of both.

In this chapter we will analyze the evolution of virtual machines since the invention of virtualized environments. The problem has been under research for a very long time and now days it has grown to have a lot of types and classifications.

In the creation of this machine, some critical aspects have to be considered which include, Efficiency of the system, ease of use and it should also be easy to understand. I Intend to use a language which is easy to understand and also easy to implement. Considering the fact that this machine majorly targets ‘newbies’ in the Ethical Hacking and Cybersecurity world, the language to be used to code the system should be an easy one and one which is widely used. Then language should also be portable across various platforms.

# 2.2 REVIEW OF RELATED LITERATURE

The concept of virtualization is believed to have its origin in the mainframe days in the late 1960s and early 1970s, when IBM invested a lot of time and resources in developing robust time-sharing solutions. Time sharing refers to the shared usage of resources among a large group of users, aiming at increasing the efficiency of both users and the expensive computer resources they share. This model represented a major breakthrough in computer technology: the cost of providing computing capability dropped considerably and it became possible for organizations and even individuals to use a computer without owning one.

Back to the year 1999 when VMware introduced the first x86 virtualization product, VMware platform, based on earlier research by its founders at Stanford University. This is considered now the breakthrough in the virtualization world.

TikiVun has to be implemented in a way that is portable and will run in any host be it a real host or a virtualized one. The fact that both VMware has made it easy to implement both Unix and Windows hosts on the same platform, I won’t have to code the system for a specific operating system

# 2.2.1 WEBGOAT

WebGoat is a deliberately insecure web application maintained by [OWASP](http://www.owasp.org/) designed to teach web application security lessons. This program is a demonstration of common server-side application flaws. The exercises are intended to be used by people to learn about application security and penetration testing techniques.

The major limitation with WebGoat is that it is not customizable in that it does not allow the user to add or remove some of its features.

# 2.2.2 Metasploitable

The Metasploitable virtual machine is an intentionally vulnerable version of Ubuntu Linux designed for testing security tools and demonstrating common vulnerabilities. Currently missing is documentation on the web server and web application flaws as well as vulnerabilities that allow a local user to escalate to root privileges.

# 2.2.3 Tagua VM

This is a php virtual machine created as an experimental virtual machine. Its biggest limitation is that it has a limited set of vulnerable services. From the developer’s description, “Tagua vm

# 2.3 THEORATICAL FRAMEWORK

The above literature references two examples of widely used virtual machines which are intentionally vulnerable, Metasplotable and Owasp’s WebGoat.

Metasploitable is considered one of the most used, most studied as it is open source and is available for download free. Created by Rapid7.

# 2.4 IDENTIFICATION OF KNOWLEDGE GAP

I managed to identify that the current systems in use have a lot of gaps which I intend to breach. These gaps include,

* Complicated deployment

For instance, Owasp and Metasploitable are somewhat hard to deploy in your system or even set them up. This is caused mainly by its platform dependence, i.e. it can only be run in a certain version of virtual machines

* Hard to understand user interface

The Metasploitable vulnerable virtual machines has a really difficult to use interface. It only comes with a command line interface (CLI) which many beginner users are not well conversant with. It has not present a user interface (UI) which is considered a better option for beginners.

* Needs a lot of prerequisites to get started

As mentioned above, some of these vulnerable virtual machines such as Metasploitable requires one to understand some concepts beforehand. These concepts include a good understanding of using the command line interface (CLI) and understanding of the Linux operating system or other Ubuntu operating system flavors.

* Are not customizable

It would be fine for one to have a virtual machine whose difficulty levels he/she can control. This would allow one to improve his/her skills day by day. Owasp provides this feature by allowing the user to set his/her own level of difficulty. Metasploitable on the other hand does not provide this feature thus making it hard to use by many people.

Some online platforms which provide online labs for users to use have tried their best to create a knowledgeable online cybersecurity community with enough and good skills. However, this has not fixed most of the issues being faced by those who are starting up in this career path. They base their labs and resources on the basis that those who acquire and use them have some levels of understanding and knowledge of how these labs work.

Also the issues being tackled by these sites are more advances and would require one to have some good knowledge on a lot of concepts including networking, scripting languages and advances server side languages. For example, the Portswigger labs would require one to have a critical knowledge of how the entire world wide web works and how requests are sent be it using the GET method or POST method.

Taking for instance the Portswigger HTTP header vulnerability, it requires one to know firsthand how to use Burpsuite to strip off parts of the HTTP/HTTPS header which is something one would not expect a beginner to know very well.

# CHAPTER THREE

# RESEARCH METHODOLOGY

# 3.1 INTRODUCTION

Research methodology represents the framework that the researcher used to illustrate the procedures for collecting data for this study. The methodology for this study involves research design, examining the types of data required and their sources, methods of data collection, data analysis and presentation techniques.

# 3.2 DATA COLLECTION METHODS

The data collection methods targeted various institutions of learning offering Information Technology courses within Nakuru County. I will administer interviews, questionnaires and also administer some observation skills in order to get their reaction and feedback on the existing and the proposed system.

# 3.2.1 PRIMARY COLLECTION METHODS

# 3.2.1.1 INTERVIEWS

. In an interview, the investigator and the respondent are set together in a face to face conversation between the two for a specific purpose. The interviews will target the students and tutors within the various institutions targeted.

Advantages of an interview are;

Less costly – Very simple, prompt and low cost.

Time-saving – Within a very short time, the conversation can be done.

Flexible – An interview is feasible therefore information obtained is abundant.

Sufficient – Question and Answer captures adequate information.

Development of Relationship—Increases mutual and beneficial understanding.

# 3.2.2.2 QUESTIONAIRES

An instrument of data collection that involves asking a given subject to respond to a set of written questions. The questionnaires will target the students, the lecturers who will in turn give credible information which we will use in the analysis of data.

Advantages of questionnaires are;

* Covers every aspect of the topic.
* Answers are easy to analyze and get the right information.
* A large number of people can be easily reached.
* Less costly.
* Maintains the privacy of the respondent.

# 3.2.2.3 OBSERVATION

Also known as field research as it involves direct and actual observation of the operations as it unfolds. The observation will be applied in the monitoring of the processes that take place in the lecture halls and the computer labs where the students do their practical classes.

Advantages of Direct observation are;

* It is the simplest Method—doesn’t require much technical knowledge.
* Useful in forming a hypothesis—Knowledge obtained is first hand.
* It is cheap as no costs are incurred in the observation process.

# 3.2.2 SECONDARY COLLECTION METHODS

A researcher can obtain secondary data from various sources. Secondary data may be published data or unpublished data.

Published data are available in: Publications, Journals, Reports, Public Records, etc.

Unpublished data may be found in: Letters, Diaries, etc.

Advantages

* The primary advantage of secondary data is that it is cheaper and faster to access.
* Secondly, it provides a way to access the work of the best scholars all over the world.
* Thirdly, secondary data gives a frame of mind to the researcher that in which direction he/she should go for the specific research.
* Fourthly secondary data save time, efforts and money and add to the value of the research study.

Disadvantages

* The data collected by the third party may not be a reliable party so the reliability and accuracy of data go down.
* Data collected in one location may not be suitable for the other one due variable environmental factor.
* With the passage of time the data becomes obsolete and very old
* Secondary data collected can distort the results of the research. For using secondary data, a special care is required to amend or modify for use.
* Secondary data can also raise issues of authenticity and copyright.

This study entailed identifying relevant documents in budgeting in various financial departments. This was mainly done using online review of the documents.

# 3.3 VALIDITY

Validity is the degree to which an instrument measures what it purports measure. The research concentrated on content validity by performing a pre-test so as to adjust the research tool to meet the required standards. The results of the study were validated by reviewing it with other similar researches done.

# 3.4 RELIABILITY

This is a measure of the degree to which a research instrument would yield the same results after repeated trials. The reliability of the questionnaire will be established through split half techniques where the pretest data set will be split into 2 equal data sets and the Cronbach Alpha evaluated. The results indicated a Cronbach Alpha score of 0.75 greater than 0.7 which shows that the study results were reliable.

# 3.5 ETHICAL CONSIDERATIONS

Prior to data collection exercise, the respondents were informed of what the research was about, its research goals and objectives. Additionally, all respondents were guaranteed confidentiality during this process. Data collected was kept confidential and be used only purposes of this research.

# 3.6 DATA ANALYSIS

Statistical Package of Social Science (SPSS) is used to analyze data from the questionnaire. The analysis that is conducted by using the tool is the descriptive statistics, reliability analysis, and regression analysis.

# 3.7 TESTING

In this stage the system will be tested to check on its functionality. This will be done in two main platforms which are The Linux Operating System and The Windows Operating system. This will be done with the aim of establishing the best platform where the virtual machine runs better.

# 3.8 IMPLEMANTATION

# 3.8.1 Software Design Life Cycle.

This is a process that describes the steps involved in the creation of software. It covers a detailed plan for building, deploying and maintaining the software. These steps will be modified into phases where each phase will contain a specific action. The phases include; Requirements gathering, Analysis, and Design, Implementation, Testing, Deployment, and Maintenance.

The Waterfall Methodwill be used in the implementation the system. In this model, the output of one phase (a previous phase) is the input of the next phase. Development of the next phase will only begin when a previous phase is completed. Below are some advantages of waterfall model;

It is a simple method which is easily understood.

Phases are done step by step.

It is easily manageable; not complex.

# 3.8.2 Phases

1. **Requirement Gathering**

At this phase, all information collected from the clients will be used to come up with the final product. It will contain the basis and foundation of the system that will be constructed. This methodology involved the collection of all system requirements at this stage. Here the SRS- Software Requirement Specification document is created. The output of the requirement phase will be the input of the analysis and design phase.

1. **Analysis and Design**

The requirements of the system will be analyzed until the functionality of the system is understood. This phase is usually accompanied by documentation for each requirement. If the requirements are not well understood, it will allow jumping back to the requirement phase. Documents which are inputs of the next phase will be derived from this phase.

1. **Implementation**

Once the analysis and documentation are done, coding begins in this phase and all the software requirements will be implemented in this phase. The software developed will be the input of the next phase.

1. **Testing**

This phase begins once the coding is complete and each module has been released. The modules will be run through a number of tests to identify any defects that may make the system fail and these will be corrected.

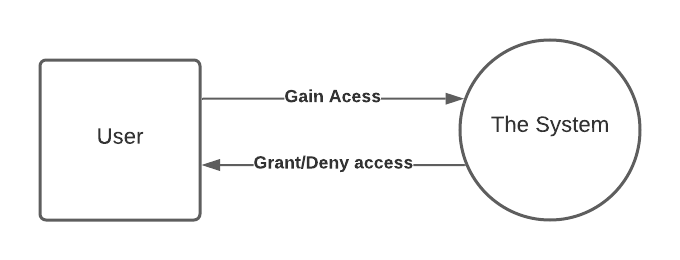
1. **Deployment**

Once the system has been tested and is free from errors, it will be deployed.

1. **Maintenance**

After deployment of the system, maintenance will be necessary because the system could later develop issues that need immediate attention or the system may need improvement in functionality.

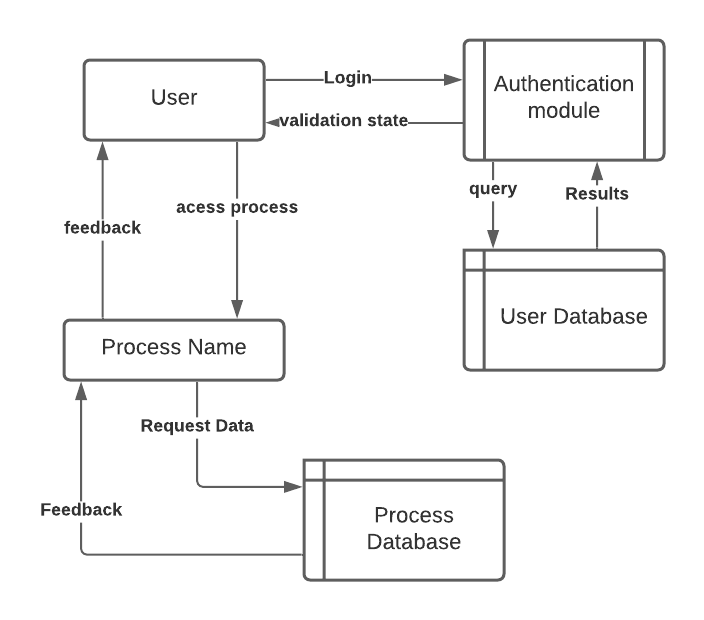
# CONTEXT DIAGRAM

****

# DATA FLOW DIAGRAMS

A data flow diagram is a traditional way to visualize the information flows within a system. A neat and clear DFD can depict a good amount of the system requirements graphically. It can be manual, automated or a combination of both.

It shows how information enters and leaves a system, what changes the information and where the information is stored. The purpose of a DFD is to show the scope and boundaries of a system as a whole. It may be used as a communication tool between the system analyst and any person who plays a part that acts as a starting point for redesigning a system



# REFERENCES

1. **The Owasp Open Web Application Project and Owasp Foundation**

**[https://owasp.org/Owasp Foundation](https://owasp.org/Owasp%20Foundation)**

1. Rapid7 The Metasploitable2 Documentation

[**https://docs.rapid7.com/metasploit/metasploitable-2/**](https://docs.rapid7.com/metasploit/metasploitable-2/)

1. **Rapid3 Metasploitable3 Documentation by Wei Chen**

[**https://blog.rapid7.com/2016/11/15/test-your-might-with-the-shiny-new-metasploitable3/**](https://blog.rapid7.com/2016/11/15/test-your-might-with-the-shiny-new-metasploitable3/)

1. **The Portswigger HTTP Host Header Attacks**

[**https://portswigger.net/web-security/host-header**](https://portswigger.net/web-security/host-header)

1. **The Core of IT, The history of Virtual Machines**

**Authors**

**Eric Kohl Brenner**

**Dana Morris**

**Brett Morris**

[**http://denninginstitute.com/itcore/virtualmachine/history.htm**](http://denninginstitute.com/itcore/virtualmachine/history.htm)

# APPENDICES

APPENDIX 1: SCHEDULE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ACTIVITY | 1ST MONTH | 2ND MONTH | 3RD MONTH | 4TH MONTH |
| Software requirement |  |  |  |  |
| Logical design |  |  |  |  |
| Physical design |  |  |  |  |
| Documentation |  |  |  |  |
| Presentation |  |  |  |  |

The above schedule represents the timeline which will take me to complete my project. The 1st month represents my first month of my semester and so on.

APPENDIX 2: BUDGET

|  |  |  |  |
| --- | --- | --- | --- |
| ITEM | QUANTITY | UNIT PRICE IN KSHS | COST IN KSHS |
| LAPTOP/DESKTOP | 1 | 50,000 | 40,000 |
| DISK STORAGE | 1 | 5,000 | 5,000 |
| OTHER COSTS | - | 10,000 | 10,000 |
| TOTAL |  |  | 55,000 |