**Data Leak Detection and Prevention Application for**

**Android Smart Phones**

**Project ID - 22\_23-J 28**

## Project Proposal Report

**Silva K.R.G.T.**

Student ID – IT19969688

Supervisor – Mr. Amila Senarathne

**Bachelor of Science (Hons) in Information Technology Specializing in Cyber Security**

**Department of Computer Systems and Engineering, Faculty of Computing, SLIIT**

September 2022

**Data Leak Detection and Prevention Application for**

**Android Smart Phones**

**Project ID - 22\_23-J 28**

## Project Proposal Report

**Bachelor of Science (Hons) in Information Technology Specializing in Cyber Security**

**Department of Computer Systems and Engineering, Faculty of Computing, SLIIT**

September 2022

# Declaration

We declare that this is our own work, and this proposal 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 our 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.

|  |  |  |
| --- | --- | --- |
| Name | Student ID  Number | Signature |
| Silva K.R.G.T. | IT19969688 |  |

Signature of the Supervisor Date

------------------------------- -----------------------------------

(Mr. Amila Senarathne)

Signature of the Co - Supervisor Date

------------------------------- -----------------------------------

(Mr. Binura Ganegoda)

# Abstract

There has been a recent increase in the attention dedicated to assaults on mobile devices. The reason for this is that more and more people are switching to smartphones from traditional nonsmartphones. As a result, fraudsters and other sorts of attackers are jumping on the bandwagon in order to increase their chances of acquiring information that is stored on mobile devices. Because of attacks like the Acoustic Side-Channel Attack, Clipboard Jacking, Permission Misuse, and Malicious Apps, we explain an Android mobile application that may assist reduce the amount of data that is exfiltrated from the device. Data exfiltration was made possible through exploits such as side-channel attacks and clipboard jacking, both of which are discussed in detail in the introduction to this article. The course will officially begin with this introduction. There are several possible solutions that have previously been created to solve these challenges. In the next section, we'll talk about the method through which we came up with a solution and the procedures we followed to secure the smartphone. To wrap up this post, we will review the project's outcomes and conclude on how to enhance this project in the future. It is our mission to guarantee that our mobile application will continue to keep the user's data safe from the hands of criminals.

## Table of Contents

[Declaration 3](#_Toc18016)

[Abstract 4](#_Toc18017)

[Introduction 7](#_Toc18018)

[1.1 Background and Literature Survey 7](#_Toc18019)

[1.2 Research Gap 8](#_Toc18020)

[1.3 Research Problem 9](#_Toc18021)

[Objectives 10](#_Toc18022)

[2.1 Main Objective 10](#_Toc18023)

[2.2 Specific Objectives 10](#_Toc18024)

[Methodology 11](#_Toc18025)

[3.1 System Overview 11](#_Toc18026)

[3.2 Flow of the Project 14](#_Toc18027)

[Description of Personal and Facilities 16](#_Toc18028)

[Budget and Budget Justification 16](#_Toc18029)

[References 17](#_Toc18030)

Apendices………………………………………………………………………………………...18

## LIST OF FIGURES

Figure 1 – Main System High Level Diagram

Figure 2 – Malware Detection Process

Figure 3 – Gantt Chart

Figure 4 – Work Breakdown Structure

## LIST OF TABLES

Table 1 - Comparing functional requirements of existing and proposed system Table 2 - Comparing non-functional requirements of existing and proposed system Table 3 - Description of personal & facilities.

Table 4 - Budget Justification

## LIST OF ABBREVIATIONS

|  |  |
| --- | --- |
| Abbreviation | Description |
| IoC | Indicators of Compromise |
| IoA | Indicators of Attack |
| TTP | Tactics Techniques and Procedures |
| ATT&CK | The Adversarial Tactics, Techniques, and Common Knowledge |
| ML | Machine Learning |

# 1. Introduction

# 1.1 Background and Literature Survey

In today’s world, from large enterprises to startups, IT infrastructure is used as a tool to make the life easier. Computers , servers , networking devices are mainly part of this IT infrastructure.

Attackers have a great interest in this IT infrastructure, finding loopholes every time in order to exploit it, which results in a data breach of the company. Most of the time, malware comes in handy for an attacker to exploit an entire organization. An attacker will be able to install malware on systems using various ways such as exploiting a vulnerability in the system or using phishing attacks.

Due to the the increase in cost and frequency of cybersecurity breaches, enterprise have taken many countermeasures. Running a security operations center which monitors the endpoints and network for cybersecurity related incidents is one example. However, running a security operations center will be not a viable option for a startup company because of budgetary constraints. Nowadays, security teams should focus on protect-the-data instead of protect-the-endpoint, because data is more valuable than the endpoint for a company.

Many enterprises have comprehensive monitoring procedures for servers, desktop and laptop computers, but when it comes to the mobile devices such as Android and iOS, there is a lack of same telemetry for monitoring those devices. Attacker only needs a single point of failure like that, in the security posture of the company, which will lead to the compromise of the entire IT systems of the organization. In each and every organization, employees are allowed to bring mobile devices and employees have increased their use of mobile devices for work because of the ease of use. So, cyberattacks on these endpoints have increased.

Mobile attack surface is not something we can neglect when we consider overall security posture of a company. According to a recent survey, 67% of organizations accept it was certain or likely they had a data breach because of employees using their mobile devices to access the confidential information of the company.[1]

Traditional Anti-Virus solutions is not viable in modern mobile threat landscape. Traditional, AV detection is signature-based, which is no longer effective at identifying latest malware due to the rapid evolution of malware and the use of unique malware and infrastructure for cyberattack campaigns. Furthermore, malware developers are using different AV evasion techniques such as fileless malware to evade detection by antivirus solutions.

Detection of modern threats to mobile endpoints requires more complex technologies and procedure than which available in AV systems, in order to protect organizations from data breaches due to the failure of securing mobile endpoints of employees.

# 1.2 Research Gap

In Sri Lankan context, most of the organizations do not use a separate application to secure the mobile endpoints of the employees. This is mainly because of the huge cost of the currently available advanced AV protection systems and the negligence of the mobile threat landscape. Research for a cost-effective data breach protection system for mobile devices are very few. Current cost-effective AV solutions are signature based and they cannot identify advanced methodologies which the threat actors are now using.[2]

Following is a comparison between proposed system and existing systems.

|  |  |  |
| --- | --- | --- |
| Features | Existing  Systems | Proposed  System |
| Signature based malware detection |  |  |
| Detect threat actor tactics and techniques with the MITRE ATT&CK® for Mobile framework |  |  |
| Application behavior monitoring |  |  |
| Vulnerable mobile endpoint detection |  |  |

*Table 1-Comparing functional requirements of existing and proposed system*

|  |  |  |
| --- | --- | --- |
| Features | Existing  Systems | Proposed  System |
| Protects user privacy |  |  |
| Preserves device resources |  |  |
| Cost effectiveness |  |  |
| Simplifies the process of mobile threat hunting, triage and response |  |  |
| Data Confidentiality |  |  |
| Data Integrity |  |  |
| Availability |  |  |
| Authentication |  |  |
| Authorization |  |  |
| Cross-platform availability |  |  |

*Table 2-Comparing non-functional requirements of existing and proposed system*

# 1.3 Research Problem

Implementing an effective data leak detection and prevention platform for mobile is totally different than implementing the same for legacy devices such as servers, desktops and laptops.

When it comes to developing such an app, the following factors should be considered thoroughly,

* Draining of the battery.
* Invasion of privacy.
* Requiring privileged access to the operating system and applications.

The threat detection is to be leveraged using machine learning. So, there is the problem of gathering machine learning data sets. This platform is to be implemented so that it automatically detects threats and take an appropriate action without human interaction. So, it will arise the problem of false positives.

A major security gap is created by mobile devices for the entire security architecture of an organization. Many organizations still consider mobile endpoints an afterthought in their overall security strategy. There is an incorrect assumption of the organizations where they think mobile endpoints do not pose a risk to the organization or they rely on mobile device management solutions which do not provide adequate security protection.[3]

The reality of the situation is that each employee of the organization uses at least one mobile endpoint at the work. It is true that modern mobile operating systems are capable of defending against traditional cyber attacks such as buffer overflows. It does not imply that mobile devices cannot be exploited by attackers. Mobile endpoints are massively vulnerable to malware and social engineering attacks because of its small form factor and personal nature. So, there should be proper policies and procedures for an organization to ensure the security of mobile devices.

# 2. Objectives

# 2.1 Main Objective

The objective of this research is to address the problems of battery draining , privacy issues and requiring privileged access to the system, while developing an effective data leak detection and prevention application for mobile endpoints, which detects known threats and vulnerabilities as well as new threats based on machine learning in order to stop data breaches of organizations which happens because of the use of mobile devices.

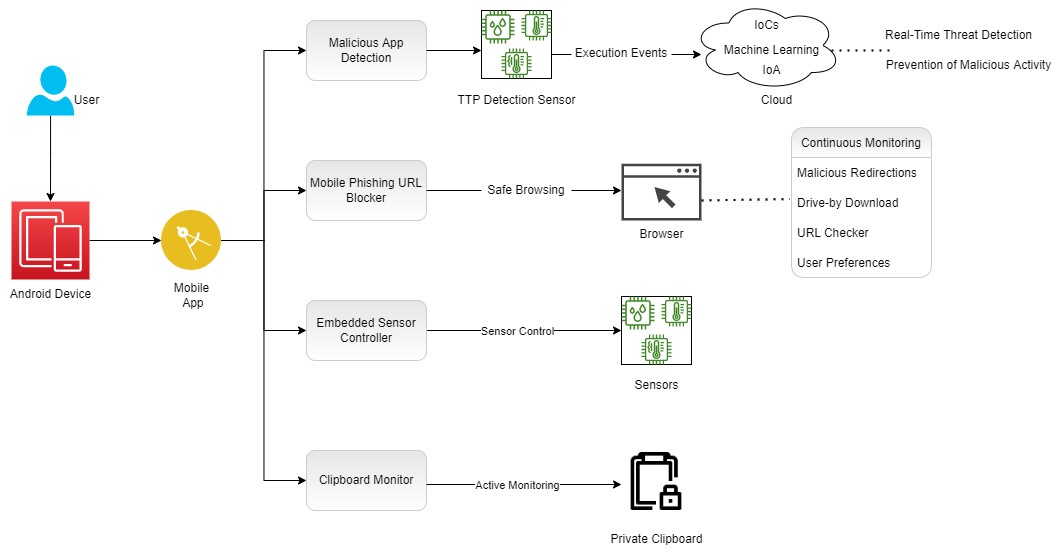
# 2.2 Specific Objectives

* Machine learning algorithms should be able to identify tactics , techniques and procedures (TTPs) of threat actors who are trying to exploit the mobile endpoints.
* Mobile malware should be statically analyzed.
* Necessary actions should be taken by the application such as whether to block, quarantine the file or to keep the file for further analysis by the user.
* Applications in the mobile endpoint should be monitored continuously to identify anomaly behaviors.
* Privacy of the user data stored in the mobile endpoint and other apps should be confidentiality , integrity and availability protected.
* Resources usage of the mobile endpoint should be minimum.
* Application should be tamper-protected. Other applications or data or user should not be able to tamper the application or its processes.
* Gathered data which is used to analyze malicious TTPs should be confidentiality , integrity and availability protected.
* Application should restrict access to other users or applications.

# 3. Methodology

# 3.1 System Overview

### Main System



*Figure 1-Main System High Level Diagram*

Main system comprises of four main components which are clipboard monitoring , embedded sensor controller, mobile phishing URL blocker and Malicious App detection. Those components ensure the total protection over an mobile endpoint which ensures the protection against malicious threat actors.

Among these components this report covers the implementing of malicious app detection component.

### Malicious App Detection

When it comes to preventing malicious apps and threat actors, there are two main things to consider, which are protection and visibility. The proposed system provides both of those things without the need of on-premises equipment, with a unique cloud-based architecture.

A lightweight sensor will be integrated into the main mobile application. This sensor will continuously monitor the execution events at the kernel level and the data will be transmitted to the cloud in real-time.

Real-time threat detection and prevention of malicious activity can be achieved because of this architecture. There are many different ways of detection and prevention of mobile threats,

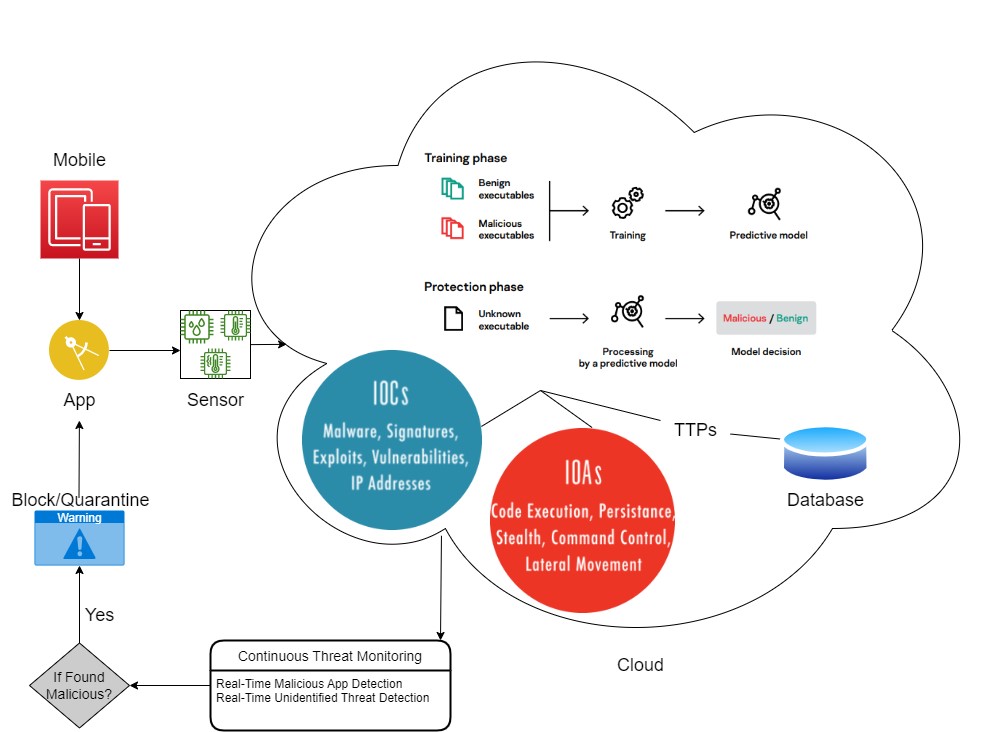
* Leveraging Indicators of Compromise or known signatures such as hashes , IPs , domains of threat intelligence sources which have been identified before.
* Using machine learning algorithms to identify tactics , techniques and procedures of unknown threat actors.
* Using Indicators of Attack to identify dynamic threat behavior, regardless of the exploit or malware used in the attack.

With the use of above-mentioned methods, the proposed system will be intelligent enough to identify not only the known threat actors, but also novel threat actors which have never been identified before.

After the identification of malicious behavior or malicious application, the application will take a necessary action such as quarantining the file or blocking the malicious process.

Furthermore, if the organization has a security operation center, SOC team will be able to hunt for threat activities in the mobile endpoints based on past threat detections data.

Machine learning algorithms should be trained beforehand for malware detection. Following diagrams illustrates the process.



*Figure 2-Malware Detection Process*

# 3.2. Flow of the Project

Agile development framework will be used throughout the project in order to develop the project as well as designing and testing to implement an effective data leak prevention application for mobile endpoints.

#### 3.2.1. Feasibility study

This section deeply focuses on the background and literature survey performed on identifying and analyzing the present statistics highlighting the existing systems to prevent the threats which occur due to mobile endpoints in an organization.

#### 3.2.2. Requirement Gathering and Analysis

Once the problems and research gap that need to be solved have been recognized, the best solutions must be clarified and identified. Research papers, threat hunting reports, and law enforcement agency websites, all of which may be found online, were used to acquire the essential data. Then I was able to talk with security teams of some large organizations in Sri Lanka and Getting in touch with these resources allowed us to learn about the internal security operation strategies of organizations, present threats and problems and potential gaps in security services.

#### 3.2.3. Design

After the requirement gathering and analyzing phase, the design phase was started to identify the technologies and design the overall system architecture. After researching for the best solution for this kind of project, we decided to go with machine learning, which is the best solution for our purposed system considering various key functionalities and drawbacks to back the best solution.

#### 3.2.4. Implementation

In this phase, the objective is to implement the components of the purposed system according to the requirements using tools and technologies that previously identified during the design phase. The proposed system is a fully functional mobile application which is capable of identifying current and unidentified threats to mobile endpoints which will stop data breaches in an organization due to the lack of security in mobile endpoints of the employees. The final product will be intelligent than a next generation anti-virus solution.

#### 3.2.5. Integration and testing

During this integration and testing phase, three testing methods will be used. They are unit testing, integration testing, and system testing.

* Unit testing is used to test individual components in isolation mode to check that every component is free from coding errors and configurations are set up properly.
* Integration testing is used to check each component is working as intended when connected.
* System testing is used to test the entire system as a whole. Attack simulations will be run against a mobile endpoint and we will check whether they are detected by the proposed system.

After completing all the testing phases, an effective product can be delivered into the target audience.

#### 3.2.6. Maintenance

In the maintenance process, the project will be continuously monitored, and it will be hosted in a cloud platform after the development phase. We will be able to identify the errors which we could not find in the development phase and the testing phase because of that.

# 4. DESCRIPTION OF PERSONAL AND FACILITIES

|  |  |  |
| --- | --- | --- |
| Registration Number | Name | Task Description |
| IT19969688 | Silva K. R. G. T | Analyzing known mobile malware using static methods leveraging IoCs.  Identifying unidentified mobile threat actors using dynamic methods leveraging machine learning and IoAs. |

*Table 3-Description of personal & facilities.*

# 5. Budget and Budget Justification

|  |  |
| --- | --- |
| Resource Type | Amount (LKR) |
| Preparation of Reports and Printing | 750 |
| Internet Usage for research and meetings | 10000 |
| Storage Charges | 2000 |
| AWS Servers | 2000 |
| Total | 14750 |

*Table 4-Budget Justification*

# References

1. S. Vanjire and M. Lakshmi, "Behavior-Based Malware Detection System Approach For Mobile Security Using Machine Learning," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), 2021, pp. 1-4, doi:

10.1109/AIMV53313.2021.9671009.

1. P. R. K. Varma, K. P. Raj and K. V. S. Raju, "Android mobile security by detecting and classification of malware based on permissions using machine learning algorithms," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017, pp. 294-299, doi: 10.1109/I-SMAC.2017.8058358.

1. A. K. Pathak, S. Saguna, K. Mitra and C. Åhlund, "Anomaly Detection using Machine Learning to Discover Sensor Tampering in IoT Systems," ICC 2021 - IEEE International Conference on Communications, 2021, pp. 1-6, doi: 10.1109/ICC42927.2021.9500825.

1. X. Miao, "Research and Practical Exploration of Mobile Application Software Security Detection Technology," 2021 International Conference on Management Science and Software Engineering (ICMSSE), 2021, pp. 9-12, doi: 10.1109/ICMSSE53595.2021.00009.

1. W. van der Lee and S. Verwer, "Vulnerability Detection on Mobile Applications Using State Machine Inference," 2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), 2018, pp. 1-10, doi: 10.1109/EuroSPW.2018.00008.

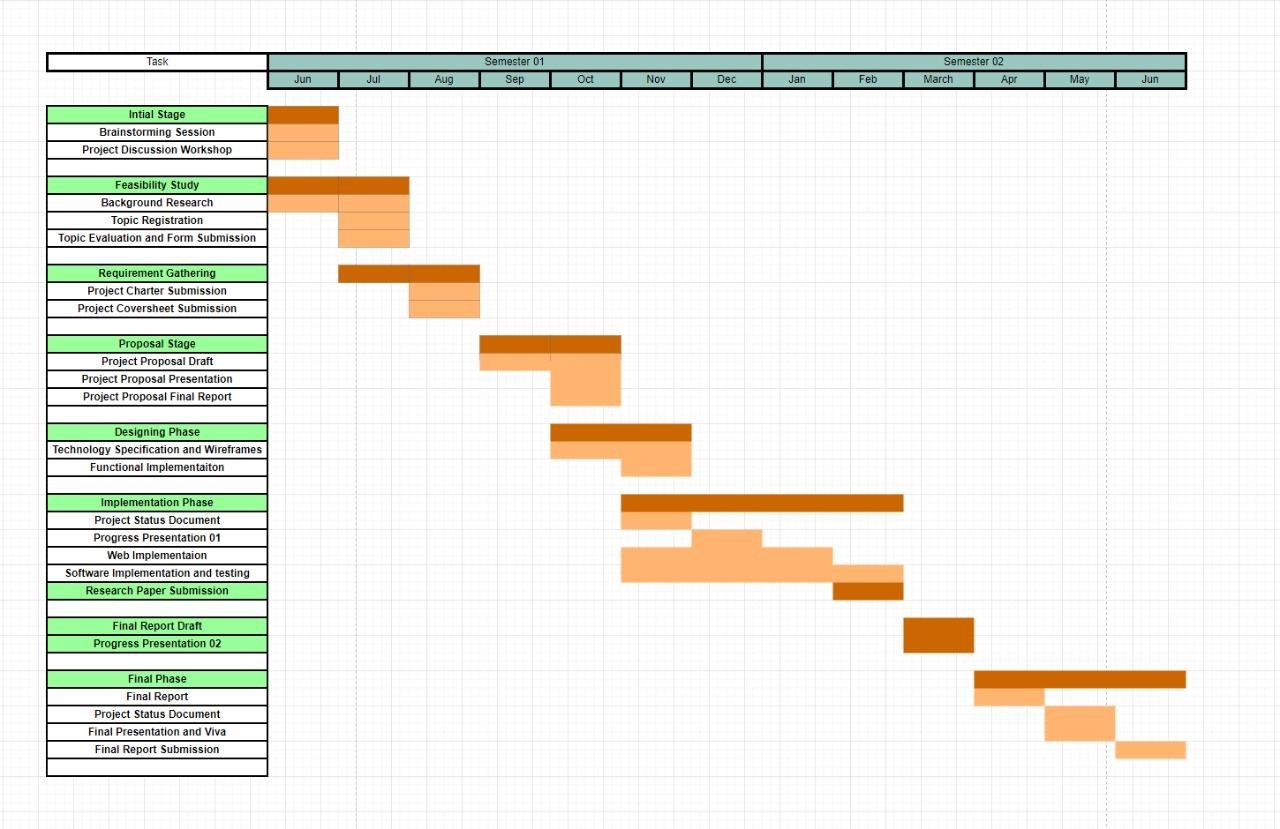
1. Wei-Ling Chang, Hung-Min Sun and Wei Wu, "An Android Behavior-Based Malware Detection Method using Machine Learning," 2016 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2016, pp. 1-4, doi:

10.1109/ICSPCC.2016.7753624.

1. S. Kumar, A. Viinikainen and T. Hamalainen, "Evaluation of ensemble machine learning methods in mobile threat detection," 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST), 2017, pp. 261-268, doi: 10.23919/ICITST.2017.8356396.

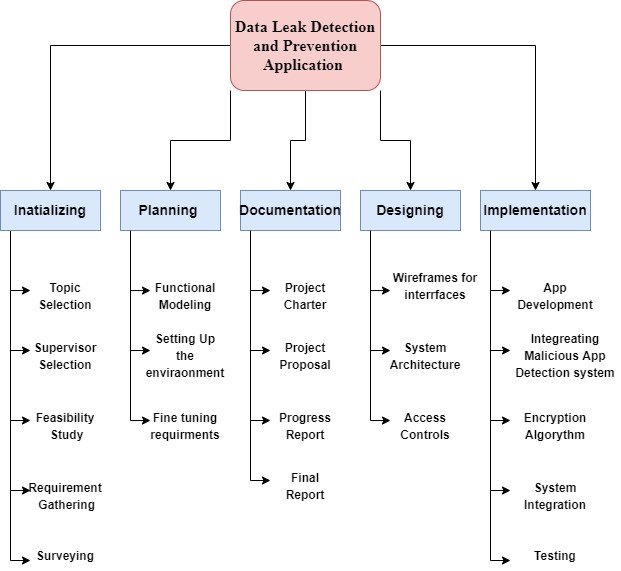
## Appendices

Appendix A – Gantt Chart



*Figure 3-Gantt Chart*

Appendix B – Work Breakdown Structure



*Figure 4-Work Breakdown Structure*