Dissertation Report on

SCORPION SHIELD

Submitted in partial fulfillment of the requirements of the degree of

Bachelor of Engineering

`

by

HIMANSHU DHANDE (52)
DIVYA KARWANDE (53)
HIMANSHU DHANDE (54)

Supervisor:

PROF. AKSHATA RAUT

Computer Engineering Department

VIVA Institute of Technology

University of Mumbai

2021-2022

CERTIFICATE

This is to certify that the project entitled **"Scorpion Shield"** is a bonafide work of **"Himanshu Dhande, Divya Karwande, Atharv Kadam"** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of "**Bachelor of Engineering in Computer Engineering".**

Prof. Akshata Raut
Guide

Prof. Ashwini Save
Head of Department

Dr. Arun Kumar
Principal

Project Dissertation Approval for B.E.

This project report entitled **Scorpion Shield** by **Himanshu Dhande, Divya Karwande, Atharv Kadam** is approved for the degree of **Bachelor of Engineering in Computer Engineering**.

Examiners

1. _____

2. _____

Date:

Place

Abstract

People use computers for all kinds of activities: online gaming, shopping, entertainment, emails, facebook, study, research, etc. At the same time, the risk of infection by malicious programs in these computers is rising. The main issue is that general users don't understand what a virus is and how computers get infected. On the other hand, many vendors produce antivirus software with different features to prevent or remove these viruses from people's computers. General users don't understand the concept of each feature in these programs, nor is there a tool to advise users about what the features mean and help them select the right software for personal or business needs. The topic "Antivirus Software " deals with software which is used to prevent or detect malware. Antivirus software is used to prevent, detect and remove all sorts of malware such as computer viruses, hijackers, worms, Trojan horses, etc. The App will begin by checking your os programs and comparing them to known types of malware. It will also scan your mobile operating system as well as installed applications for behaviors that may signal the presence of a new and unknown malware.

Table of Contents

List of Figures

List of Tables

Declaration

We declare that this written submission represents our ideas in my own words and where others ideas or words have been included. We have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will result in disciplinary action by the Institute and can also evoke penal action from the sourceswhich have thus not been properly cited or from whom proper permission has not beentaken when needed.

sign

Himanshu Dhande

sign

Divya Karwande

sign

Atharv Kadam

Date:

Chapter 1

Introduction

Antivirus is a kind of software used to prevent, scan, detect and delete viruses from a computer[1]. Once installed, most antivirus software runs automatically in the background to

provide real-time protection against virus attacks.Comprehensive virus protection programs help protect your files and hardware from malware such as worms, Trojan horses and spyware, and may also offer additional protection such as customizable firewalls and website blocking. Antivirus programs and computer protection software are designed to evaluate data such as web pages, files, software and applications to help find and eradicate malware as quickly as possible. Most provide real-time protection, which can protect your devices from incoming threats; scan your entire computer regularly for known threats and provide automatic updates; and identify, block and delete malicious codes and software[2]. Because so many activities are now conducted online and new threats emerge continuously, it's more important than ever to install a protective antivirus program. Fortunately, there are a number of excellent products on the market today to choose from. Antivirus software begins operatingby checking your computer programs and files against a database of known types of malware[9]. Since new viruses are constantly created and distributed by hackers, it will also scan computers for the possibility of new or unknown types of malware threats. With so manyinternet-connected devices in the home today, technology has made everyday living more convenient but also riskier[8]. To help protect your devices, Verizon offers TechSure, which includes a combination of anti-virus software, 24/7 tech support for Verizon services and hardware; identity theft protection, password management; and repair insurance for damaged and broken devices[7]. Antivirus software is a type of program designed and developed to protect computers from malware like viruses, computer worms, spyware, botnets, rootkits, keyloggers and such[6]. Antivirus programs function to scan, detect and remove viruses from your computer. There are many versions and types of anti-virus programs that are on the market. However, the prime objective of any antivirus program is to protect computers and remove viruses once detected[2].

Most antivirus programs incorporate both automated and manual filtering abilities. The instant scanning option may check files - downloaded from the Internet, discs that are embedded into the PC, and files that are made by software installers. The programmed scanning process may likewise check the entire hard drive on a day-to-day basis[12]. The manual scanning system

enables you to check single documents or even to scan the complete network at whatever point you feel it is necessary. Since new infections are always being made by PC programmers, antivirus programs must keep an updated database of the most recent malware codes[10]. This database incorporates a list of "malware definitions" that the antivirus software implements when filtering records. Since new infections evolve every day , it is essential to keep your product's infection database up and coming. Luckily, most antivirus programs naturally refresh the infection database all the time[14].

While antivirus software is basically intended to ensure complete protection for PCs against virus infections, numerous antivirus programs now secure against different sorts of malware for example, spyware, adware, and rootkits as well[12]. Antivirus software may likewise include firewall features, which anticipate unapproved access to your PC[5]. Utilities that incorporate both antivirus and firewall abilities are commonly called "Internet Security Suite"[9].

While antivirus programs are accessible for Windows, Macintosh, and Unix platforms, most antivirus software is compatible with Windows operating systems[8].This is on account of the fact that most infections are focused towards Windows PCs and subsequently virus protection is particularly imperative for Windows clients[4]. If you are a Windows user, it is important to install a third party , feature-packed, robust antivirus program on your PC[6]. Comodo Antivirus has been the best and compelling solution to outsmart even the zero-day andunknown threats with efficient features and technologies like default-deny protection, Host Intrusion Prevention, Auto sandboxing solutions and Containment technology[5].

Typically, most programs will use three different detection devices: specific detection, which identifies known malware; generic detection, which looks for known parts or types of malware or patterns that are related by a common codebase; and heuristic detection, which scans for unknown viruses by identifying known suspicious file structures[8]. When the

program finds a file that contains a virus, it will usually quarantine it and/or mark it for deletion, making it inaccessible and removing the risk to your device[9]. Antivirus software isdesigned to prevent computer infections by detecting malicious software, commonly called malware, on our computer and, when appropriate, removing the malware and disinfecting the computer. Malware can be classified into various kinds, namely, Trojans, viruses (infectors), rootkits, droppers, worms, and so on. Antivirus software is special security software that aims to give better protection than that offered by the underlying operating system (such as Windows or Mac OS X)[5]. In most cases, it is used as a preventive solution. However, when that fails, the AV software is used to disinfect the infected programs or to completely clean malicious software from the operating system[4]. AV software uses various techniques to identify malicious software, which often self-protects and hides deep in an operating system. Advanced malware may use undocumented operating system functionality and obscure techniques in order to persist and avoid being detected[12]. Because of the large attack surface these days, AV software is designed to deal with all kinds of malicious payloads coming from both trusted and untrusted sources[6].

Chapter 2

Literature Survey

The following chapter is a literature survey of the previous research papers and research which gives detailed information about the previous system along with its advantages and disadvantages to make the system.

2.1 Survey of existing system

Y.-Z. Li et.al [1], malwares are being produced at an unprecedented scale with hundreds of new entities targeting users across all of technology as malware developers explore new ways or exploit old ones to evade detection and defeat analysis. The process of obfuscation was originally used to protect benign applications from code alterations, manipulations and reverse engineering, but this mechanic is also a tool malware developers could manipulate to mask the malicious applications they create.

B. Fechner et.al [2], in the cloud environment there are many unique and different on demand services available for users. Those services allocate workable and easy accessibility to use various web applications. In this environment malicious attacks and threats can occur anytime and can destroy useful files and applications. So the cloud needs to be well secure and highly maintained. In this paper we will discuss various malicious attacks which can dismantle our cloud environment and how we can make our cloud more powerful by using strong and robust cloud antivirus.

X. Zha and S. Sahni et.al [3], Banking Malware, has become a popular and ever more prevalent mechanism to monetise malware development. Since the development of the Zeus malware kit in 2007, the frequency and complexity of banking malware has been in- creasing. Developing a good understanding of the operation of a malware family is a first step in the reverse engineering required to create tools to extract the malware configuration, which is used in the remediation of malware infrastructure. This reverse engineering process in recent years has become increasingly challenging. This manuscript provides a brief summary of the reverse engineering of banking malware families over a two year period and emphasises the anti-analysis techniques employed by the authors of six families of banking malware. The manuscript presents this analysis, and examines trends in the development of these anti-analysis techniques.

K. Nakano et.al [4], In the last several decades, the arms race between malware writers and antivirus programmers has become more and more severe. The simplest way for a computer

user to secure his computer is to install antivirus software on his computer. As antivirus software becomes more sophisticated and powerful, evading the detection of antivirus software becomes an important part of malware. As a result, malware writers have developed various approaches to increase the survivability and concealment of their malware. One of these technologies is to terminate antivirus software right after the execution of the malware.

D. Man, K. Nakano, and Y. Ito el.al [5], malware Analysis is the top trend in the security industry. The number of new malware samples and toolkits for automated malware generation are growing exponentially, whereas the analysis capacity and knowledge are going down. In this paper we are going to discuss the infrastructure we created for malware analysis, with network dissection of traffic, execution of samples on multiple virtual machines or in real ones if required. The architecture performs fast analysis, comparing the results of multiple different anti-viruses and uses customized kernel-drivers, loaders and a clustered environment. New machines can be easily added to increase performance. Dispatchers, memory dumpers and dissectors are going to be discussed, as well as results we got in our livelab.

A. K. Sahoo, K. S. Sahoo, and M. Tiwary el.al [6], Antivirus is most widely used to detect and stop malware and other unwanted files. Cloud antivirus is a malware detector architecture where virus definitions and other behaviors of suspicious files are analyzed on cloud and controlled by a lightweight Agent on the client system. We suggest using a two-way caching scheme where local-cache is stored on the client system and cloud-cache is present onnetwork cloud, where we store virus definitions and behaviors according to collective intelligence techniques. Local-cache is used to detect the virus and other malware files while offline and cloud cache uses the Artificial Intelligence Techniques for the whole client base toget the most susceptible and prone virus and malware definitions thus increasing the optimality of virus definition search and hence the speed of the whole process gets increased.

M. Vincent el.al [7], Hackers use malware to gain access to target computers. Malicious payloads are usually generated using tools such as Metasploit. As a means of defense, the target computers deploy anti-virus solutions to detect these malicious payloads and protect thevictim machines. In a reaction to this, the hackers created anti-virus evasion tools to evade detection by these antivirus solutions. But how effective are these antivirus evasion tools? This paper seeks to evaluate the effectiveness of some selected antivirus evasion tools: Avet, Veil 3.0, The Fat Rat, PeCloak.py, Phantom-Evasion, Shellter, Unicorn and Hercules against current best Antivirus Solutions on Windows and Android platforms.

B. Rajesh et.al [8], countering the proliferation of malware has been for recent years one of the top priorities for governments, businesses, critical infrastructure, and end users. Despite the apparent evolution of anti-virus (AV) systems, malicious authors have managed to create asense of insecurity amongst computer users. Security controls do not appear to be sufficiently strong to stop malware proliferating. There seems to be a disconnect between public reportson AV tests and what people are experiencing on a daily basis. In this research, we are testing the efficiency of AV products and their ability to detect malicious files commonly known as malware

A. P. Namanya et.al [9], there are different types of embedded portable iDevices which can be used in criminal activities. The most commonly used gadget in the field of embedded portable iDevices is the iPad. Techniques used to acquire data from iPad include jail breaking, using inbuilt operating system utilities and using forensic tools (open source, freeware or commercial). Data integrity is a vital element of digital forensics which must be ensured for acceptability of findings (retrieved forensic artifacts) in a court of law. In order to establish data integrity in iDevice forensics, investigations were performed using different techniques, specifically an operating system inbuilt utility, a freeware tool and a commercial tool. The forensic artifacts acquired from these tools and techniques were then compared to ascertain their data integrity. The results have shown that on one hand the freeware tools, under certain circumstances, also preserve data integrity as their commercial counterparts but on the other

hand the commercial tools, under certain circumstances, also make data integrity doubtful as generally believed for freeware tools. Based on the results, the research has also recommended various data acquisition tools that the forensic examiner can select depending on the requirement.

M. Zakeri et.al [10], In this paper are given mean and types of audit, probabilities of transitions between the functioning states of information of protection tools with discrete work time. Also, in this work are given four types of antivirus errors with possible situationsin the performance of means of discrete work time. Performance cases of antivirus software are described in the form of directed Graph. By four types of antivirus errors is calculated average time of operation without breaking, intensity of appearing of error, average time of backup and intensity of backup of error.

T. Mithal et.al [11], International Conference on, vol. 1, 2009. Signature-based detection is a key process of any virus, intrusion, or malware detection system. This process works by looking for certain patterns (i.e., signatures) of viruses in a large database of signatures. Current antivirus products heavily rely on signatures to identify malware. However, there has been a debate about their effectiveness against the emergence of a tremendous number of malware types every day. Nevertheless, signature-based techniques are still in use by various antivirus products hoping that there would be an up-to-date database in the future that would contain allvirus signatures. If no comprehensive database is available, signature-based approaches wouldfail and become ineffective. Consequently, antivirus products would needto perform another type of virus detection that does not rely on the virus signature (e.g.,behavior-based detection).

W. Fleshman et.al [12], android is currently the most used smart-mobile de- vice platform in the world, occupying 82.8% of market share . As of now, there are nearly 2 million apps available for downloading from Google Play, and more than 50 billion downloads to date.

Unfortunately, the popularity of Android also creates interest from cyber-criminals who create malicious apps that can steal sensitive information and compromise systems.

R. Agrawal el.al [13], the high frequency of crime and fraud occurring during the SMS delivery process has become a problem for the users of the telecommunication service. It requires the existence of a mechanism to ensure the authenticity of SMS. The method often used to overcome the problem is by adding (to embed) signatures on the text. The signature onSMS is called digital signature. By using digital signatures, the integrity of the data can be guaranteed, and it is also used to prove the originality of the message (for instance the validityof the sender and anti- denial).

Bauer el.al [14], to provide Android device users with deeper assurance that their applications are secure, we developed AndroTotal, which is based on the model that VirusTotal successfully implemented in the desktop world. Like VirusTotal desktop users, any Android device user can submit an application to a website to check how it is classified by commercial mobile AV products. Unlike existing methods, AndroTotal uses a completely automatic approach to scan hundreds of suspicious applications per day against all major AV applicationversions. Unlike VirusTotal, it creates reproducible, self-contained testing environments for each AV-malware pair, while ensuring a high throughput because of its inherent scalability.

W. Wang el.al [15], independent developers have created apps to manage root privilege, but the root-management model underlying these apps remains vulnerable. Security Enhanced Android3 and its extensions might offer some protection, but they do not give users root access and complicated policies make them difficult to manage. Malware variants can
steal confidential data, initialize distributed denial of service (DDoS) attacks, and perform disruptive damage to the computer systems. New malware variants use concealing techniques such as encryption and packing to remain invisible in the victim's system. Those new variants spread by exploiting human trust as an infection vector. For instance, opening email

attachments, downloading fake applications, visiting and downloading files from phony websites are well-known methods of malware spreading vectors.

R. Lyer el.al [16], Recent technological developments in computer systems transfer human life from real to virtual environments. Covid-19 disease has accelerated this process. Cyber criminals' interest has shifted in real to virtual life as well. This is because it is easier to commit a crime in cyberspace rather than regular life. Malicious software (malware) isunwanted software which is frequently used by cyber criminals to launch cyber-attacks. Malware variants are continuing to evolve by using advanced obfuscation and packing techniques. These concealing techniques make malware detection and classification significantly challenging. Novel methods which are quite different from traditional methods must be used to effectively combat new malware variants. Traditional artificial intelligence (AI) , specifically machine learning (ML) algorithms, are no longer effective in detecting all new and complex malware variants. Deep learning (DL) , an approach which is quite different from traditional ML algorithms, can be a promising solution to the problem of detecting all variants of malware.

R. Komatwar and M. Kokare el.al [17], Deep learning-based approach is starting to be used as a new paradigm to eliminate the shortcomings of existing malware detection and classification approaches. Deep learning has been used extensively in different areas including image processing, computer vision, human action recognition, driving safety, facial emotion recognition and natural language processing. However, it has not been used sufficiently in the cyber security field, especially in malware detection. Deep learning is a subset of artificial intelligence which works based on artificial neural networks (ANN). Deep learning uses several hidden layers and learns from examples. To increase the model performance, there areseveral deep learning architectures used recently such as deep neural networks (DNN), deep belief networks (DBN), recurrent neural networks (RNN), and convolutional neural networks (CNN).

A. F. Agarap el.al [18], Malware is a rapidly increasing menace to modern computing. Malware authors continually incorporate various sophisticated features like code obfuscations to create malware variants and elude detection by existing malware detection systems. The classification of unseen malware variants with similar characteristics into their respective families is a significant challenge, even if the classifier is trained with known variants belonging to the same family. The identification and extraction of distinct features for malware is another issue for generalizing the malware detection system. Features that contribute to the generalization capability of the classifier are difficult to be engineered with modifications in each malware. Conventional malware detection systems employ static signature-based methods and dynamic behavior-based methods, which are inefficient in analyzing and detecting advanced and zero-day malware.

B. Anderson el.al [19], The internet has become a key aspect of our daily lives. Although making our lives convenient, the internet has made innocent users vulnerable to attacks. The rise of the internet and the emergence of social networks have triggered exponential growth in malware. The evolution of malware starts as a hobby of technical enthusiasts and is now pinned with the main motive of making money. Malware detection and classification is one ofthe most significant problems in the area of cybersecurity. Signature-based methods areeffective against known malware, but they are ineffective against advanced and unknown malware. Malware authors introduce evasion techniques like obfuscation, encryption, packing, etc. on existing malware to elude detection, leading to more new malware. Malware files with the same malicious behavior belong to the same malware family.

StatCounter el.al [20], Android malware poses serious security and privacy threats to the mobile users. Traditional malware detection and family classification technologies are becoming less effective due to the rapid evolution of the malware landscape, with the emergence of so-called zero-day-family malware families. To address this issue, our paper presents a novel research problem on automatically identifying the security/privacy related capabilities of any detected malware, which we refer to as Malware Capability Annotation

(MCA). Motivated by the observation that known and zero-day-family malware families share the security/privacy related capabilities, MCA opens a new alternative way to effectively analyze zero-day-family malware (the malware that do not belong to any existing families) through exploring the related information and knowledge from known malware families.

Table 1.1 Analysis Table

Title	**Summary**	**Advantages**	**Techniques Used**
Memory efficient parallel bloom filters for string matching, in Networks Security, Wireless Communications and Trusted Computing [1]	Malwares are being produced at an unprecedented scale with hundreds of new entities targeting users across all of technology as malware developers explore new ways or exploit old ones to evade detection and defeat analysis	The results obtained shows there needs to be an improvement in mobile security as access to these obfuscation mechanics are available to the public and could be manipulated even with access to a trial version of these obfuscation tools	Obfuscation techniques
Gpu-based parallel signature scanning and hash generation [2]	Nowadays Cloud technology is booming and trending technology in the IT sector. In cloud computing the major agenda is to deliver services to the users in the	This model offers significant advantages over the previous host-based antivirus, including better detection of malware in the cloud.	Heavy Detection Technique and Light detection Technique

	most convenient way.		
Gpu-to-gpu and host-to-host multi pattern string matching on a gpu [3]	Banking Malware, has become a popular and ever more prevalent mechanism to monetise malware development.	The advantage of the new DGA based C2 hostname generation is to extend the lifetime of the malware samples by providing more C2 hostnames.	Obfuscation techniques
Efficient implementations of the approximate string matching on the memory machine models [4]	The simplest way for a computer user to secure his computer is to install antivirus software on his computer. As antivirus software becomes more sophisticated and powerful, evading the detection of antivirus software becomes an important part of malware.	Various approaches that antivirus termi-Nators can use antivirus software.	Self- defense techniques

The approximate string matching on the hierarchical memory machine, with performance evaluation [5]	Malware Analysis is the top trend in the security industry. The number of new malware samples and toolkits for automated malware generation are growing exponentially, whereas the analysis capacity and knowledge are going down.	The main advantage of using RDMA is the ability to read memory segments freely without passing through the CPU scrutiny, so we neither have permission limitation nor any another OS limitation.	Initial Decision Analysis (IDA)
Signature based malware detection for unstructured data in Hadoop [6]	Antivirus is most widely used to detect and stop malware and other unwanted files. Cloud antivirus is a malware detector architecture where virus definitions and other behaviors of suspicious files is analyzed on	Although cloud computing has many advantages but the biggest disadvantage is that a user cannot access files if they are not connected to the internet	N-Version protection technique

	cloud and controlled by a lightweight Agent on client system.		
Dynamically adaptive framework and method for classifying malware using intelligent static, emulation, and dynamic analyses [7]	Hackers use malware to gain access to target computers. Malicious payloads are usually generated using tools such as Metasploit. As a means of defense, the target computers deploy anti-virus solutions to detect these malicious payloads and protect the victim machines.	The results of the research work will be the best antivirus evasion tool that will be recommended to penetration testers for use during engagements and the most effective antivirus solution that will be recommended to end users to use on their devices.	Signature based, behavior based and heuristic based techniques.
Efficient detection of malicious worms with different analysis methods and techniques [8]	Countering the proliferation of malware has been for recent years one of the top priorities for governments, businesses, critical	A significant advantage of behavioural based detection over signature based is that the level of packing,	Malware, AV bypass, Antivirus Systems, DetectionTechniques, Payloads, Antivirus Evaluation.

	infrastructure, and end users.	encryption, polymorphism or metamorphism employed by the malware does not, in most cases, change its behaviour.	
Evaluation of automated static analysis tools for malware detection in portable executable files [9]	There are different types of embedded portable iDevices which can be used in criminal activities. The most commonly used gadget in the field of embedded portable iDevices is the iPad.	Benefit investigators in selecting the best data acquisition tool, depending upon the requirement of iDevice forensic case.	Techniques used to acquire data from iPad include jail breaking, using inbuilt operating system utilities and using forensic tools
A static heuristic approach to detecting malware targets [10]	Types of audit, probabilities of transitions between the functioning states of information of protection tools with discrete work time.	By these errors, average time of operation without breaking, intensity of appearing of error, average time of backup and	Audit, error, discrete work time, performance, graph, backup.

		intensity of backup of error is calculated between two errors.	
Case studies on intelligent approaches for static malware analysis [11]	Existing antivirus products employ diverse types of techniques to detect malware or any suspicious activities. The majority of such techniques rely on signature-based detection algorithms.	Hybridization of cybercrime investigation models with existing antivirus products to make an extension to their benefits to the entire community.	Antivirus products; Malware detection; Signature-based technique; Proprietary software; Anti-cyber crime
Static malware detection & subterfuge: Quantifying the robustness of machine learning and current anti-virus [12]	A recent report indicates that a newly developed mali-cious app for Android is introduced every 11 seconds. To combat this alarming rate of malware creation, we need a	SIGPID has been designed to extract only significant permissions through a systematic, 3-level pruning approach.	Data mining techniques

	scalable malware detection approach that is effective and efficient. In this paper, we introduce SIGPID, a malware detection system based on permission analysis to cope with the rapid increase in the number of Android Malware.		
Fast algorithms for mining association rules [13]	SMS (Short Message Service) is a messaging service in the mobile communication environment. In sending messages via SMS, message security is essential in order to maintain the integrity of the authenticity of the messages. Messages are secured not only in the aspect of	This algorithm has an advantage that ECDSA efficient in using memory for 160 bit ECDSA is as safe as 1024 bit RSA, so it is suitable to be implemented on a device with limited memory like mobile devices.	ECDSA, SMS, Digital Signature, Android, Boolean Permutation, SMS Authentication

	confidentiality, but also on how the delivered message is not altered by someone. To overcome These problems, digital signature is required.		
Study on the Financial Aspects of Network Security: Malware and Spam [14]	For most people, mobile devices have become a digital wallet that they can trust to securely hold everything from contacts and appoint-ments to banking and retail transactions.	AndroTotal, a scalable antivirus evaluation system for mobile devices, creates reproducible, self-contained testing environments for each antivirus application and malware pair and stores them in a repository, benefiting both the research community and Android device users.	Network communication or interprocess communication (IPC) anomalies.

Exploring permission-induced risk in android applications for malicious application detection [15]	Though popular for achieving full operation functionality, rooting Android phones opens these devices to significant security threats. RootGuard offers protection from malware with root privileges while providing user flexibility and control.	RootGuard improves the security of rooted Android phones, effectively mitigating attacks by malware with root privileges without affecting app performance and at the same time achieving low overhead. Future work will extend RootGuard's ability to collect additional context information, including the sequence and the pattern of system calls, so as to further facilitate user decision making and prevent malicious app behaviors.particularly by those using native code.	Security Server.

A New Malware Classification Framework Based on Deep Learning Algorithms [16]	Malware samples are first converted into grayscale images and given to the DL system. After the image acquisition section is fulfilled, the proposed method extracted high-level malware features from malware images by using the convolution layers of the proposed hybrid architecture.	1. DL models can automatically generate high-level features from existing features. 2. DL reduces the need for feature engineering. 3. DL can handle unstructured data efficiently	Malimg and Malevis datasets.
Malware images: Visualization and automatic classification [17]	The test results presented that the proposed method can effectively extract distinctive features for each malware type and family for classification.	Cloud computing brings many advantages to malware detection including easy access, more computational power and much bigger databases	Malware, malware classification, malware detection

Intelligent dynamic malware detection using machine learning in IP reputation for forensics data analytics [18]	Code obfuscation and packing techniques make the malware detection a very challenging task. This paper proposed a novel deep learning architecture to effectively detect malware variants.	Secondly, the proposed model was evaluated by state-of-the-art methods. The results obtained here also disclose and approve the advantage and supremacy of the proposed method overloading methods in the literature.	Deep neural networks, transfer learning, deep learning.
Intelligent Vision-Based Malware Detection and Classification Using Deep Random Forest Paradigm [19]	Malware is a rapidly increasing menace to modern computing. Malware authors continually incorporate various sophisticated features like code obfuscations to create malware variants and elude detection by existing malware detection systems.	Advantage is that vision-based analysis does not need static disassembly or dynamic execution of binaries, unlike other traditional malware analysis techniques.	Malware variants, malware visualization.

A3CM: Automatic Capability Annotation for Android Malware [20]	Android malware poses serious security and privacy threats to the mobile users. Traditional Malware detection and family classification technologies are becoming less effective due to the rapid evolution of the malware landscape, with the emergence of so-called zero-day-family malware families.	The first Android malware dataset with the capability ground truth. Our method achieves satisfactory performance in inferring the capability vectors of known Android malware,small-size-families' malware and zero-day-families' Android malware, respectively.	Android malware, security/privacy-related capability.

2.2 Research Gap

Antivirus software is designed to find known viruses and oftentimes other malware such as Ransomware, Trojan Horses, worms, spyware, adware, etc., that can have a detrimental impact to the user or device. Antivirus programs provide a way to protect against known threats. The effectiveness of an antivirus program is heavily dependent on how often it is updated. Therefore, it is important to have the antivirus program scheduled to update daily. Most antivirus programs rely on a library or database of known viruses that they use to compare with programs on a user's device. If a match is found, the malicious program will either be deleted or placed into a quarantine area from which a user can decide to restore or delete the program manually.

One limitation of traditional antivirus programs is that they only provide protection against known threats. Therefore, if someone cooks up new malicious code, an antivirus program may fail to detect it when a scan is done.

2.3 Problem Statement and Objectives

With Anti-Virus software, it is very easy for many programs to fall into a trap of reacting to new viruses rather than preventing them. This is because many programs are not effective against new strains of malware, spyware, etc. Hackers engineer many of their malware strains specifically to bypass popular programs. Antivirus is a kind of software used to prevent, scan, detect and delete viruses from a computer. Once installed, most antivirus software runs automatically in the background to provide real-time protection against virus attacks.

1. Protection from viruses and their transmission

   Antivirus detects any potential virus and then works to remove it. Keep in mind thatall this is mostly done before the virus gets to harm the system. So, this means that mostof the viruses are countered way before they get to do any harm to your systems. An antivirus may combat many viruses in a single day without your knowledge.

2. Block spam and ads

   Nowadays, most desktop browsers block pop-ups and unwanted advertisements as a matter of course, but what about blocking on Android? Even if you’re using a smartphone, there are ways to stop annoying and sometimes harmful pop-up advertisements.

3. Defense against hackers and data thieves

   To protect our devices from any harmful hackers we can add some password to it. Also that password gets encrypted so it cannot be visible to anyone.

For Eg : Protect your devices and accounts from intruders by choosing passwords that are hard to guess. Use strong passwords with at least eight characters, a combination of letters, numbers and special characters.

4. Ensures protection from removable devices

Whenever we plug in any external device like USB in our computer or laptop we should always ensure it's virus free and also we should check for antivirus in our device. If any such virus enters our device from any external device it can be detected by the antivirus.

5. Protects our data and files

The process of providing, administering, and monitoring security across all information repositories and objects inside an organisation is referred to as enterprise data protection. It’s a wide phrase that encompasses a variety of technologies, rules, procedures, and frameworks for ensuring data protection, regardless of where it’s consumed or kept within an organisation.

The aim of enterprise data protection is the adoption and maintenance of data protection privacy guidelines and regulations inside an organisation. Information standards and practices may change depending on how they are used and how crucial they are. Multi-factor authentication, data encryption and restricted access may be used to protect highly classified information.

In general, enterprise data privacy works to safeguard your organisation against information loss while also assuring the security of all devices using info. It is given with the help of popular information security solutions like firewalls and antivirus, as well as info security rules and standards for controlling and managing the entire process.

6. Supercharge our PC

   We can keep our softwares up-to-date and optimize the files of the device to avoid getting affected by the virus. And also we can clear the cache files which are not unused by us in a long interval of time. When we do multiple tasks at the same time we often forget to close the unwanted files or softwares which are running in the background which utilizes lots of space, battery and capacity of the device. Which ends up slowing down the device.

7. Firewall protection from spyware and phishing attacks

   Basically, a good firewall provides protection from prying eyes. It stops thieves and intruders from accessing your computer, laptop, workstation or server. A good firewall can protect your computer from malicious ‘worms’.

   A firewall also prevents confidential information from being sent out from your computer without your permission. This could be your passwords, bank details and other personal information.

Chapter 3

Proposed System

3.1 Algorithm

Step 1. At App side, sender encrypts the message using the receiver's public key.The public key of the receiver is publicly available and known to everyone. Encryption converts message into a cipher text.

This cipher text can be decrypted only using the receiver’s private key.

Step 2. The cipher text is sent to the receiver over the communication channel.

Step 3. On the mobile OS side, mobile os decrypts the cipher text using his private key. The for anyone to determine the receiver’s private key. After decryption, cipher converts back into a readable format.

Step 4. Device Root Check in which reverse engineering is used.

Step 5. Whitelisting and Blacklisting IP is used in which data is encrypted through the process of hashing.

3.2 Details of System:

3.2.1 Software Requirements:

1. Operating System: Windows
2. Jupyter Notebook
3. Python
4. Pandas
5. Numpy
6. Visual Studio Code
7. Node.js
8. Angular
9. Android Studio

3.2.2 Hardware Requirements:

1. Processor: intel i5 (7 th gen)
2. Ram: 8GB
3. Graphic Card: 4GB

3.3. Design Details:

3.3.1 System Flow Diagram:

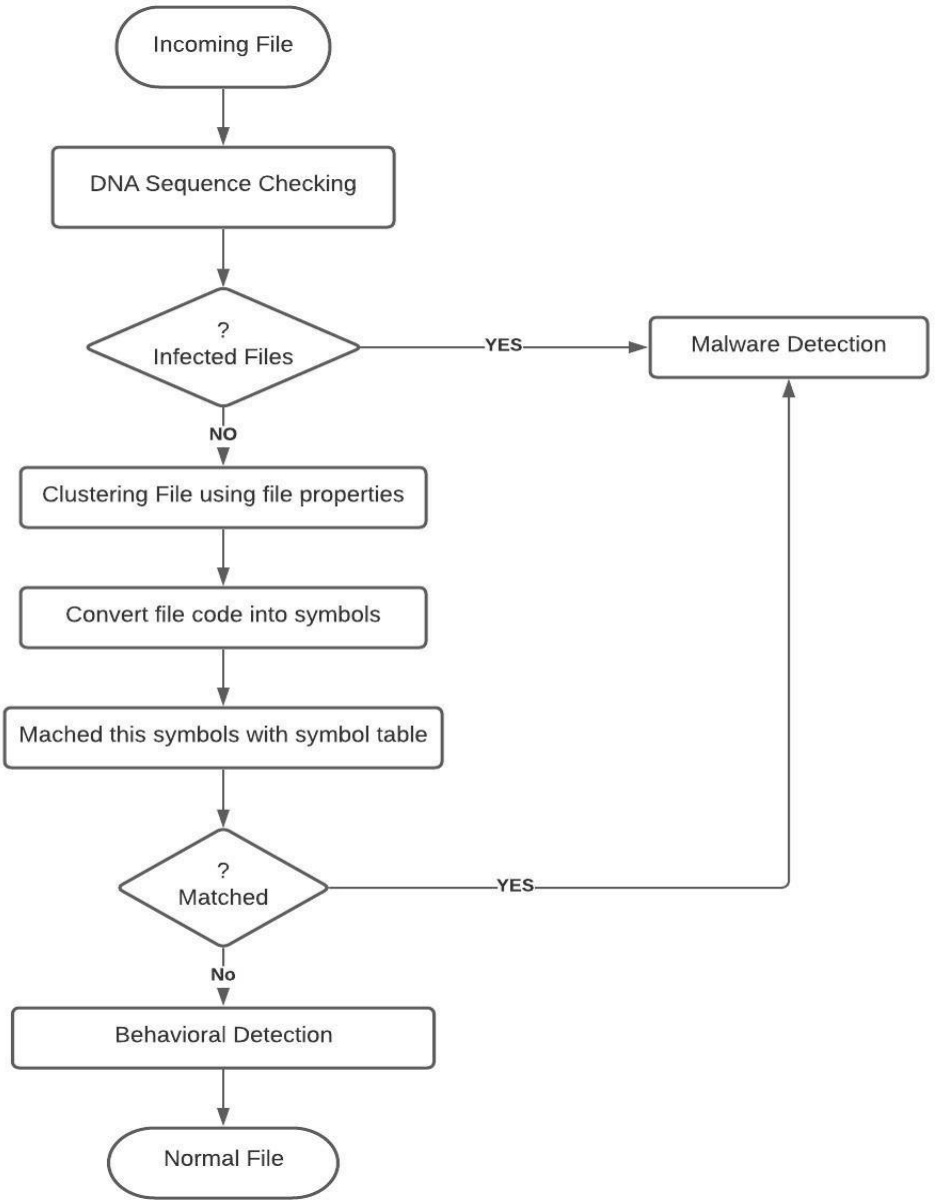

Fig 3.1 Malware detection system

System flow charts display how data flows in a system and how decisions are made to control events.

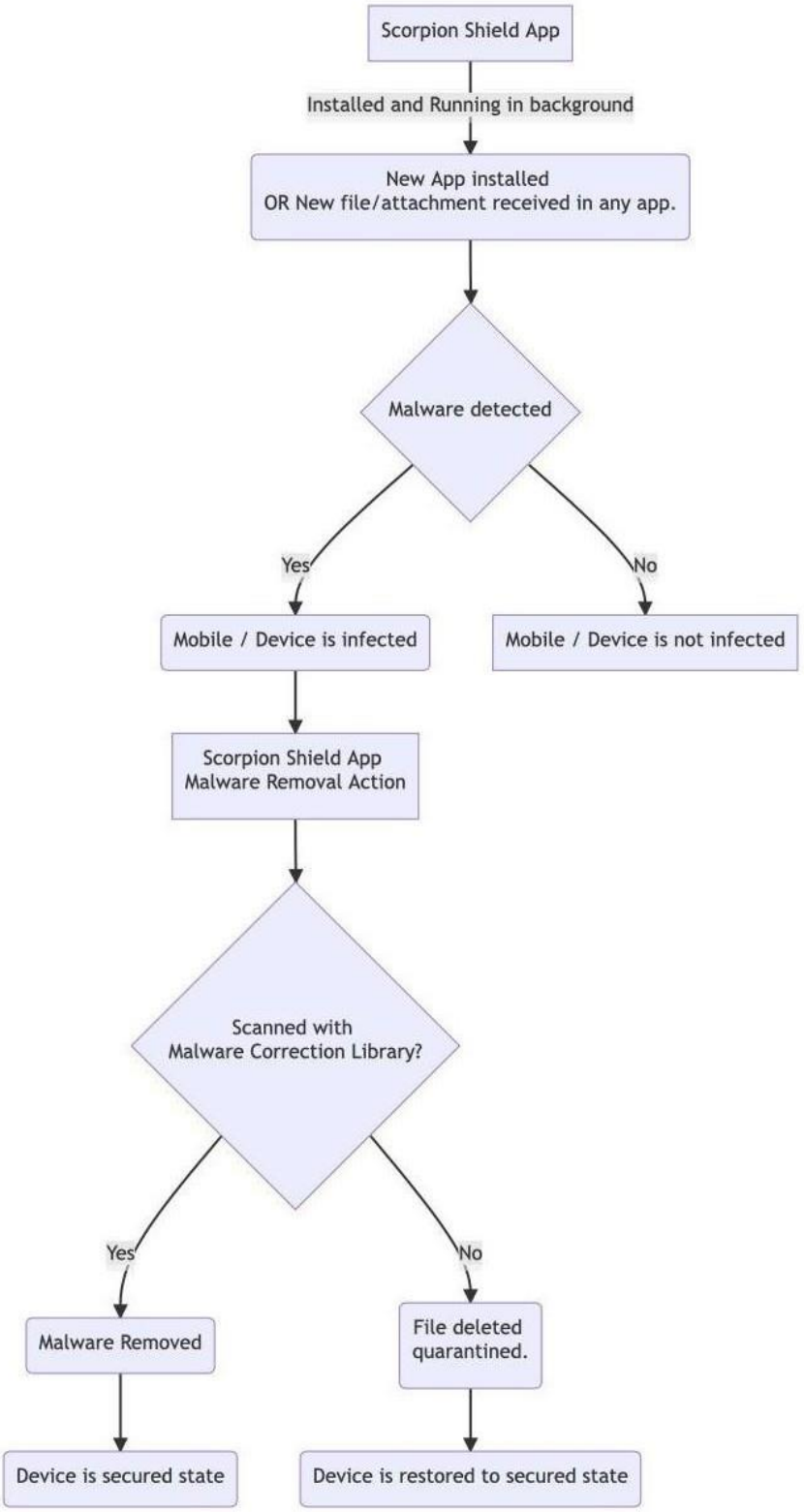

Fig 3.2 Malware correction system

3.3.2 Block Diagram:

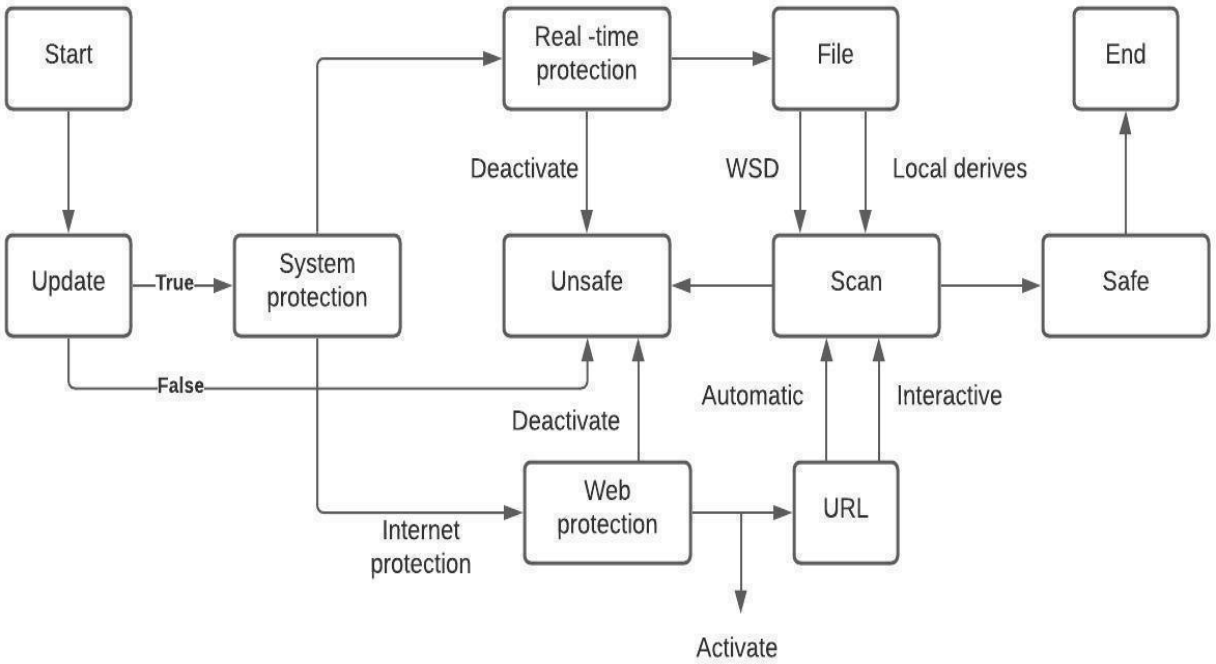

Fig 3.2 System Block Diagram

A block diagram is a diagram of a system in which the principal parts or functions are represented by blocks connected by lines that show the relationships of the blocks. They are heavily used in engineering in hardware design, electronic design, software design, andprocess flow diagrams.

A block diagram displays a system in which the principal parts or functions are represented by blocks connected by lines that show the relationships of the blocks.

A block diagram is a specialized, high-level flowchart used in engineering. It is used to design new systems or to describe and improve existing ones. Its structure provides a high-level

overview of major system components, key process participants, and important working relationships.

3.4 Methodology

App will begin by checking os programs and comparing them to known types of malware. It will also scan your mobile operating system as well as installed applications for behaviors that may signal the presence of a new, unknown malware. All these methodologies will be implemented in our app. Also, not only detection but also protection and advance notifications will be enabled with this app.

Heuristic-based detection

Heuristic Virus Checking is a methodology of virus detection. Anti-virus software makers develop a set of rules to distinguish viruses from non-viruses. Should a program or code segment follow these rules, then it is marked a virus and dealt with accordingly. This allows detection of any virus, and theoretically, should be sufficient to deal with any new virus attacks.F-secure virus software uses this method in addition to scanning, although not very many software packages available today utilize heuristic virus checking. Heuristic-baseddetection is considered the most common form of virus detection that uses an algorithm to differentiate thesignature of known viruses against a potential threat. It can unearth viruses that have not yet been discovered, as well as known viruses that have been modified or disguised, and released into the wild again. The only downside is that it can also generate false positive matches, meaning an antivirus scanner may report a file as being infected when it actually isn't. While these "false positives" are minimal, they're not uncommon.

Antivirus software utilizes several methodologies in scanning, detecting, and protecting computers and systems from viruses. As understanding increases about the vectors malicious code uses to attack and how antivirus software protects computer systems from the viruses,

people will be able to more effectively help in creating an environment that is secure and virus free.

Interception methodology

Interception software detects virus-like behavior and warns the user about it. How to detect virus-like behavior? Use heuristics again. Many viruses will perform some suspicious action, like relocating themselves in memory and installing themselves as resident programs.

Signature-based detection

Signature-based detection searches for the specific digital code of a virus (you can think of it as a virus' fingerprint) and if it finds it, quarantines or deletes it. Once a virus has been identified, it can be added to a signature database, which is kept locally or in the cloud to be accessed when scanning a system for threats moving forward. However, this process requires at least one user or system to be attacked by the malicious software and recognize it before everyone else can be protected against it. Put simply, it's not very useful for brand new threats.

Behavioral detection

Behavioral detection is a more modern technique for tracking down known and unknown viruses. It generally looks at what software does rather than examining what a piece of software is. For example, it checks for viruses that attempt to shut down or bypass your antivirus solutions on the system and once found, quarantine or delete them subsequently.

Cloud antivirus detection

Cloud antivirus needs an Internet connection to collect information, which is uploaded to, and processed by, a server in the cloud. It generally spares your computer additional processing by running all detection on the server.

3.5 UML Diagram:

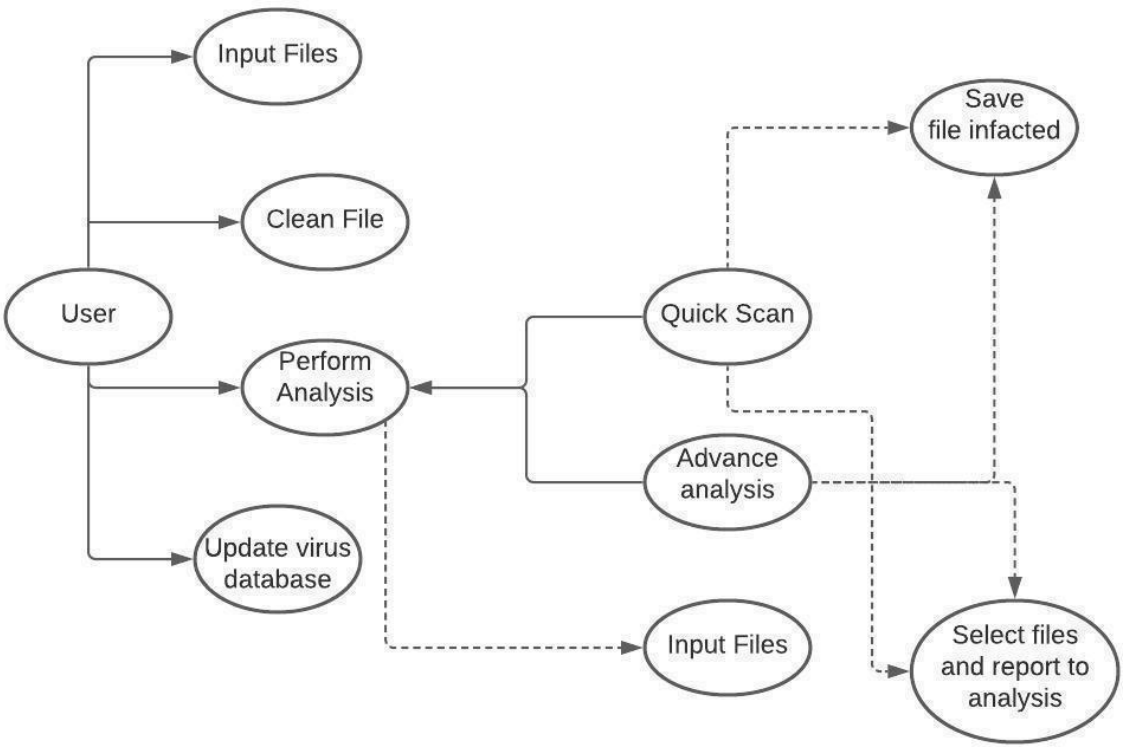

Fig 3.3 UML Diagram

A UML diagram is a diagram based on the UML (Unified Modeling Language) with the purpose of visually representing a system along with its main actors, roles, actions, artifacts orclasses, in order to better understand, alter, maintain, or document information about the system.

UML diagrams model the behavior of a system and help to capture the requirements of the system. UML diagrams describe the high-level functions and scope of a system.

UML diagrams can be used as a way to visualize a project before it takes place or as documentation for a project afterward. But the overall goal of UML diagrams is to allow teams to visualize how a project is or will be working, and they can be used in any field, not just software engineering.

3.5.1 Use Case Diagram

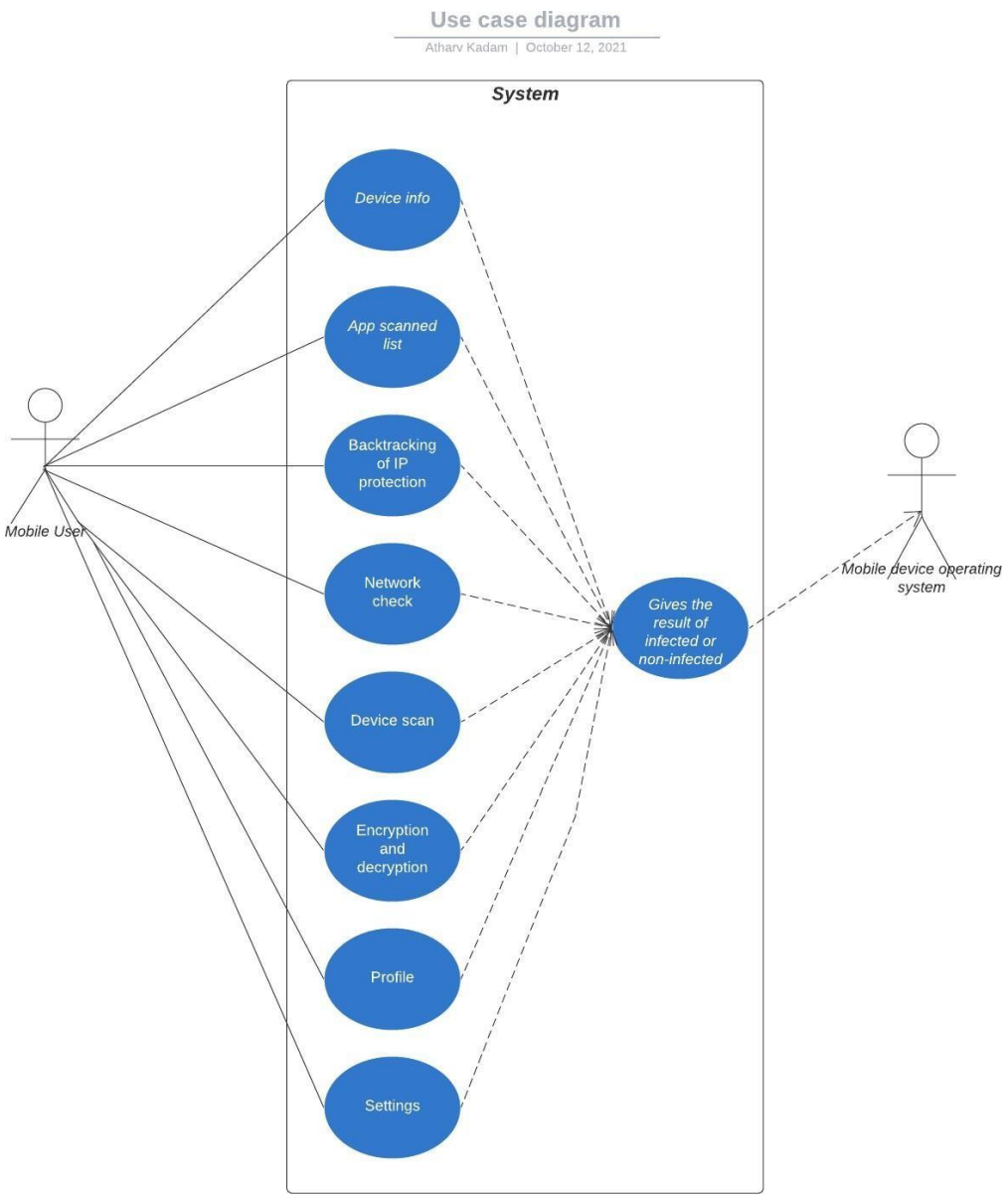

Fig 3.4 Use Case Diagram

Use-case diagrams describe the high-level functions and scope of a system. These diagrams also identify the interactions between the system and its actors. The use cases and actors in

use-case diagrams describe what the system does and how the actors use it, but not how the system operates internally.

3.5.2 Class Diagram

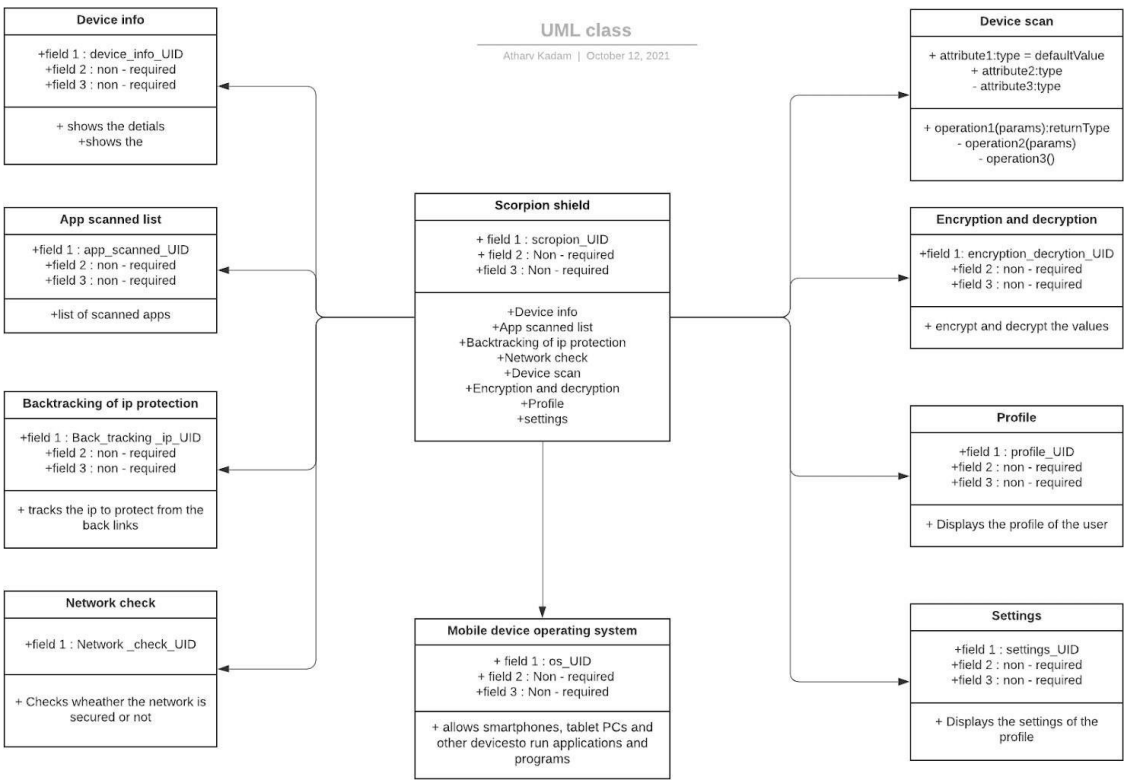

Fig 3.5 Class Diagram

Class diagram is a static diagram. The class diagram is widely used in the modeling of object oriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages. Class diagram shows a collection of classes, interfaces, associations, collaborations, and constraints.

The class diagram is the main building block of object-oriented modeling. It is used for general conceptual modeling of the structure of the application, and for detailed modeling, translating the models into programming code. Class diagrams can also be used for data modeling.The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed.

3.6 Gantt Chart

Fig 3.6. Gantt Chart

Gantt charts help teams to plan work around deadlines and properly allocate resources. Project planners also use Gantt charts to maintain a bird's eye view of projects. They depict, among

other things, the relationship between the start and end dates of tasks, milestones, and dependent tasks.

Chapter 5

Results

Initially, when the app is started we have to grant permission for Location, Read and write storage, read and write contacts and system alert window. Below are the figures describing these 4 permissions.

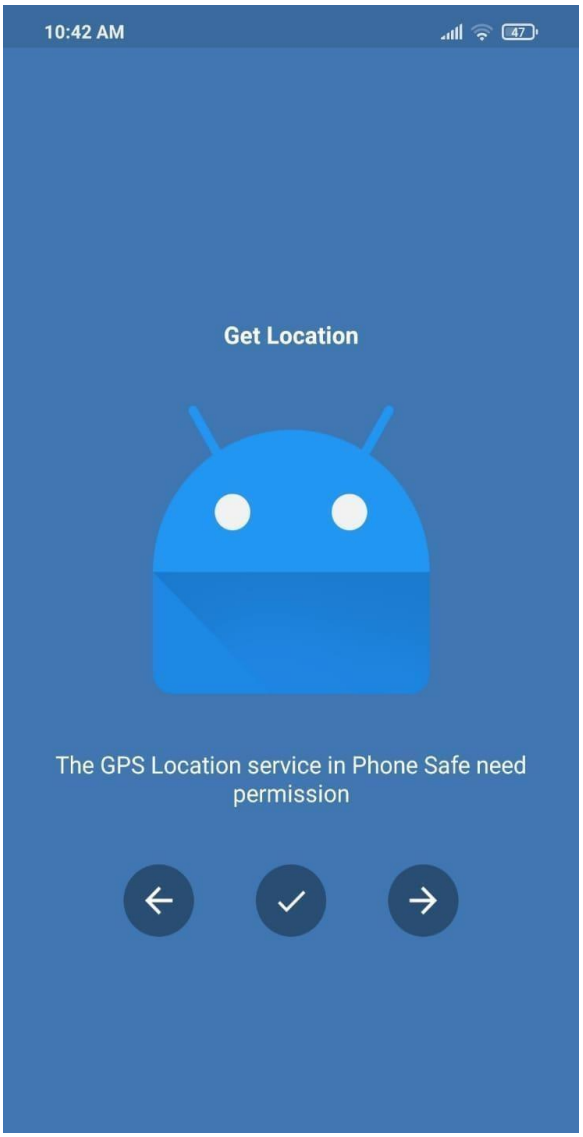

Fig 5.1. Application asking permission for location access

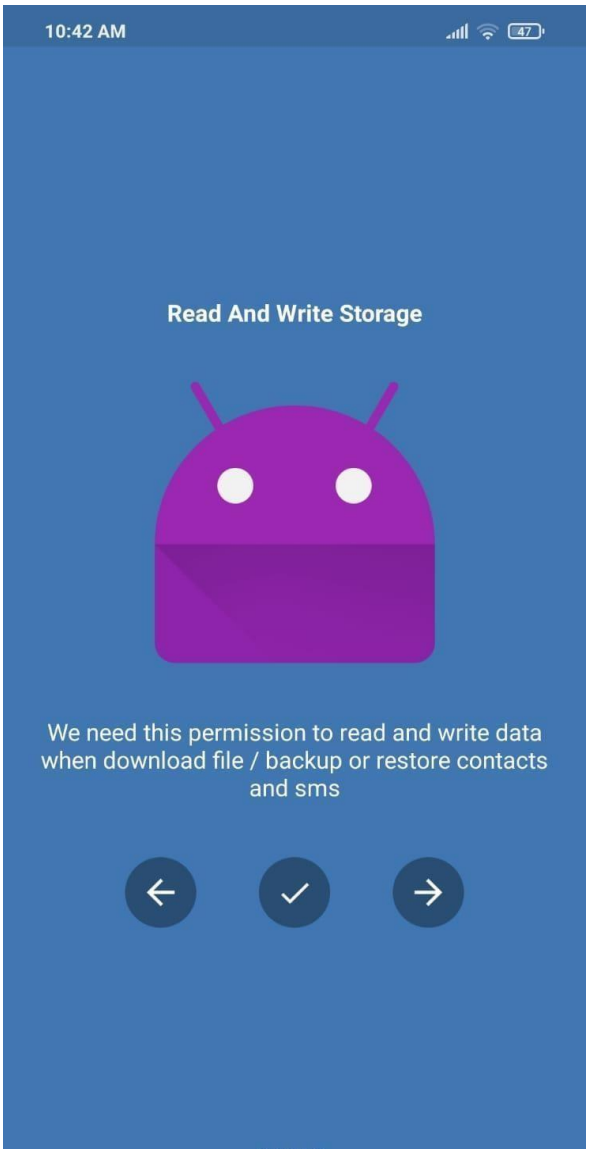

Fig 5.2. Application asking permission for read and write storage access

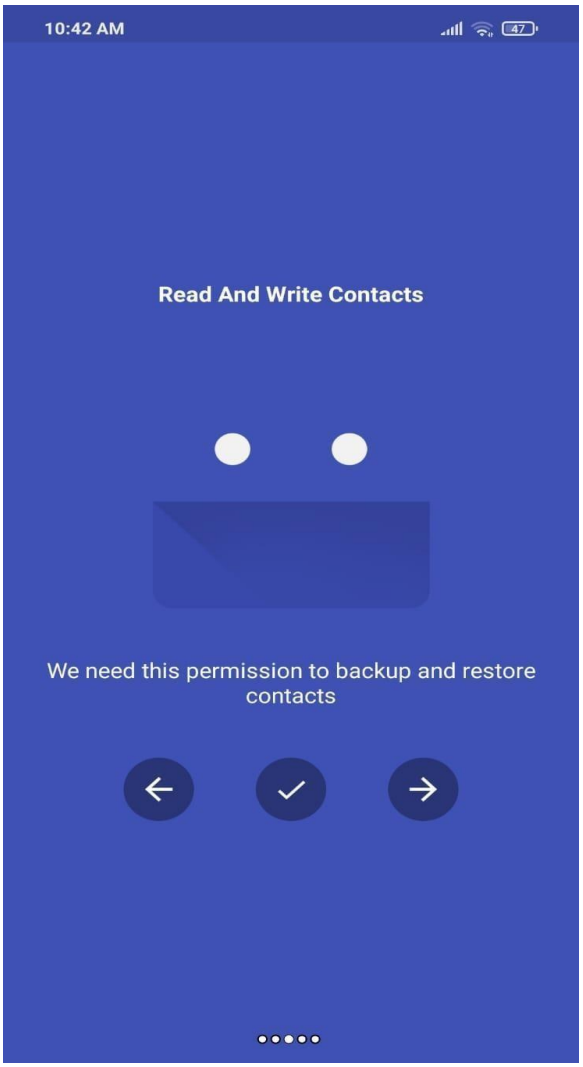

Fig 5.3. Application asking permission for read and write contacts access

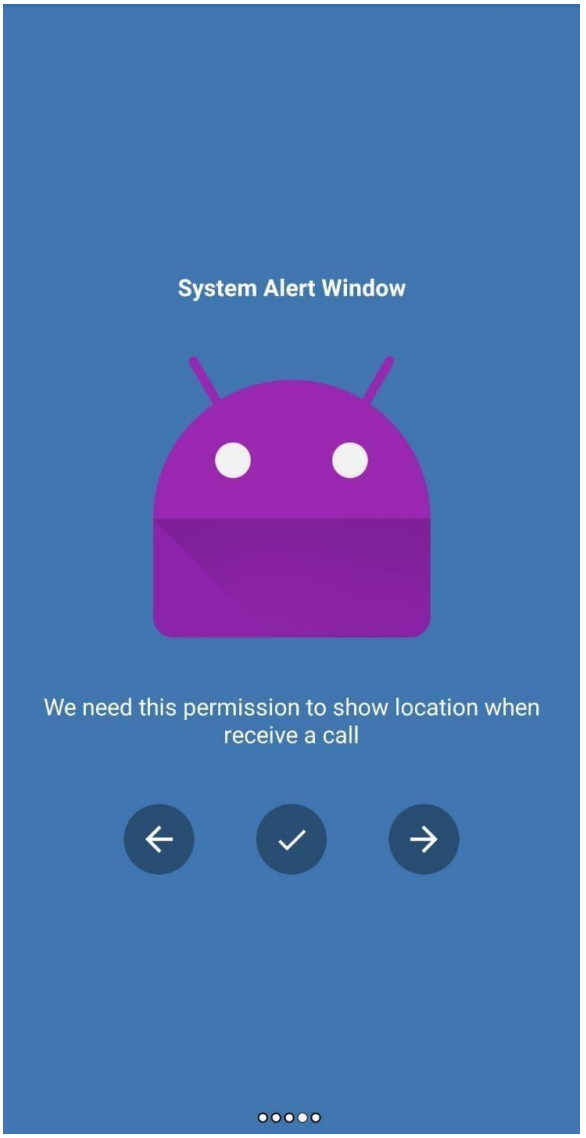

Fig 5.4. Application asking permission for system alert window

Fig 5.5 and Fig 5.6 displays the Index and 6 features of our application which are Virus scan, Anti theft, App locker, Wifi security, Call blocker and battery saver.

Virus scan : Virus scans search through your system to locate and remove any malicious threats on your device. You'll find most antivirus software guards against malware. This can include threats like viruses and worms, as well as, spyware, Trojans, ransomware, and adware.

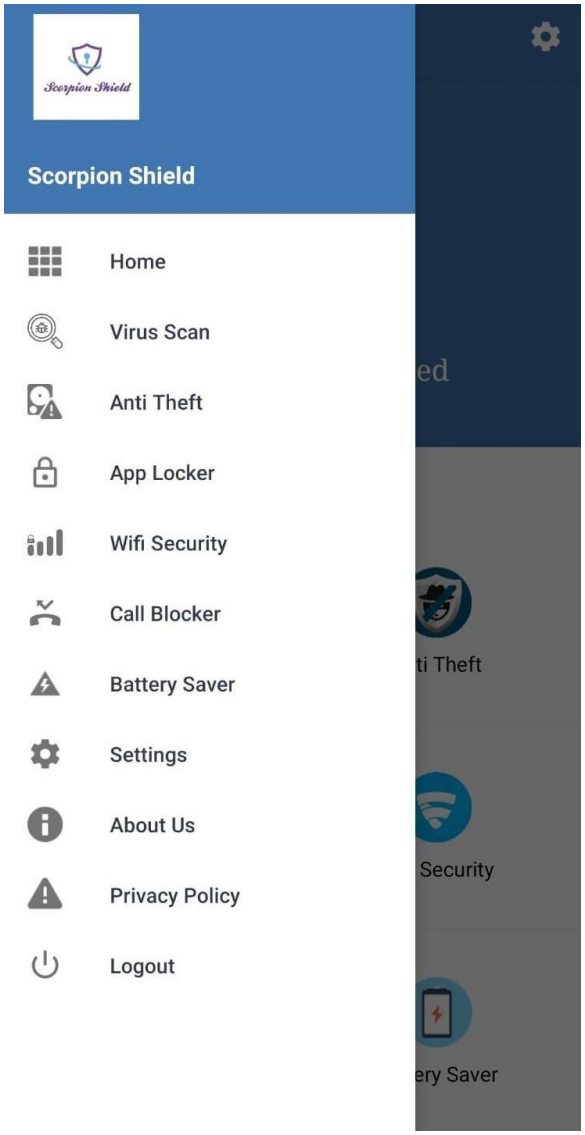

Fig 5.5. Index including all the features also about theprivacy policy and logout

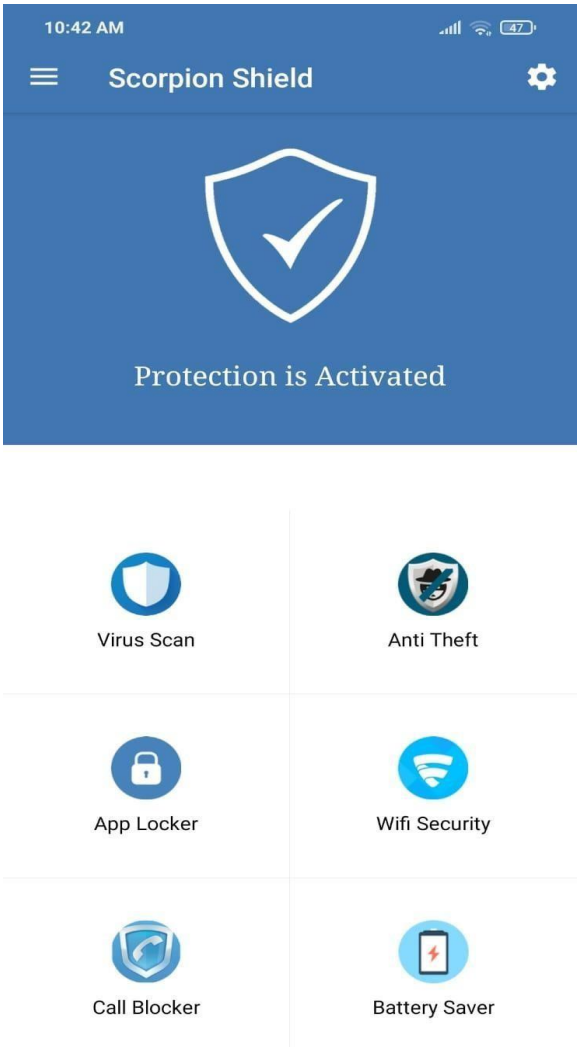

Fig 5.6. All the 6 features of the antivirus(scorpion shield) application

Starting with the first section of Virus scan, as the name suggests it will scan for viruses in the system but on three different levels-

1. Full scan
2. SD Card scan
3. Application scan

Full scan :

Full scan does the job of scanning the entire system for the harmful viruses in each and every internal and external files and folders. As you can see when we start the scan we can see two different sections on the user interface. On the left side we can see the progress of the ongoing scan and on the right side detected issues encountered in the scan are seen. A predefined, in-depth scan of your system that checks your storage drives and memory for malware. Quick scan may not detect some malware, but it can still inform you about a virus if your computeris infected. Full Scan requires much more time and OS resources but it detects all known viruses. We recommend performing a Full Scan every week.

SD Card Scan :

This scanning does the job of checking for issues in the attached SD card in the system. All of this scanning looks the same as the full scan.

Application scan :

This section scans each and every internal and external applications in the system and in the results categorizes every application as medium or high risk application. Antivirus is a kind of software used to prevent, scan, detect and delete viruses from an application. Once installed, most antivirus software runs automatically in the background to provide real-time protection against virus attacks.

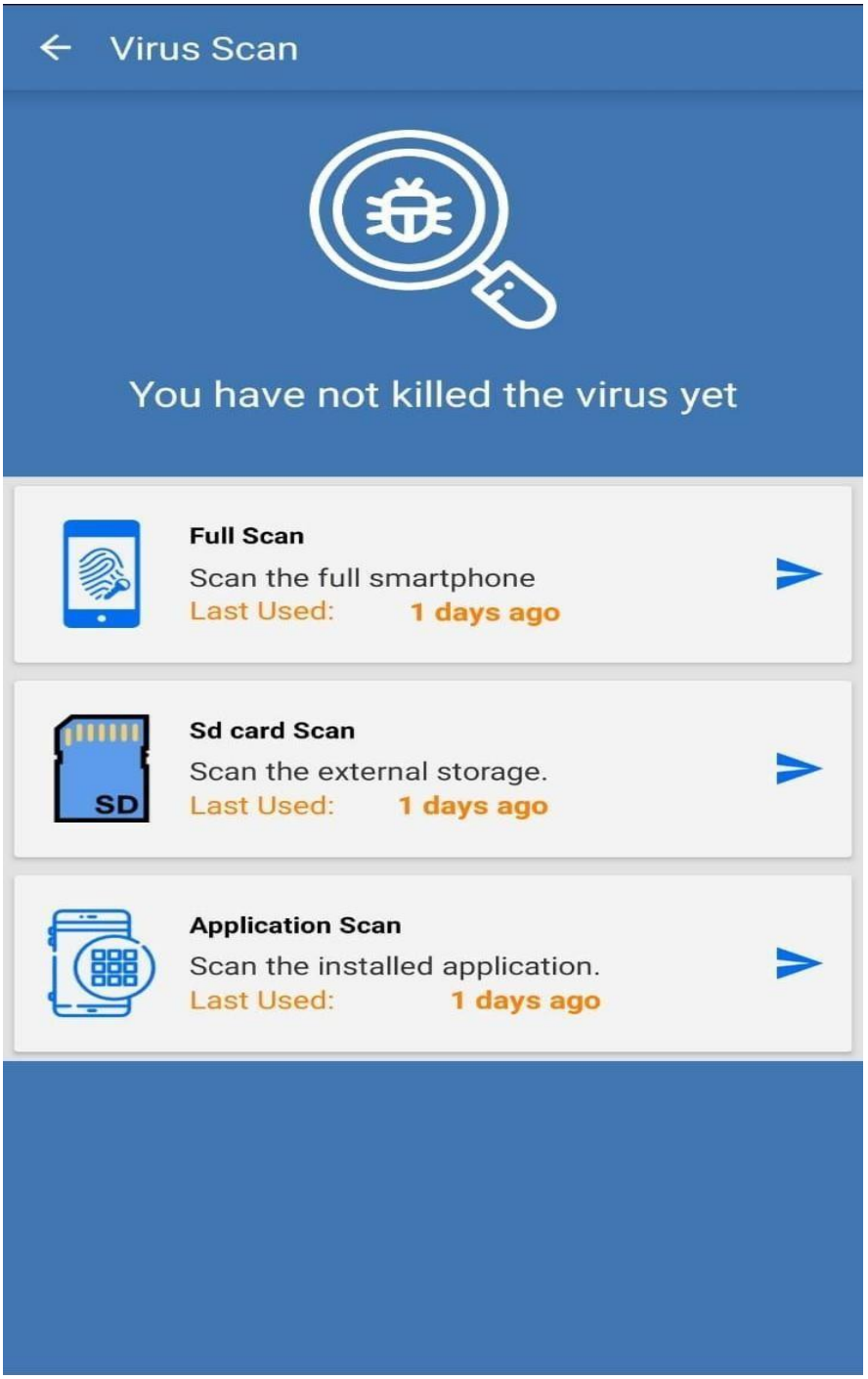

Fig 5.7. The 3 scans including Full Scan, SD card scan and Application scan

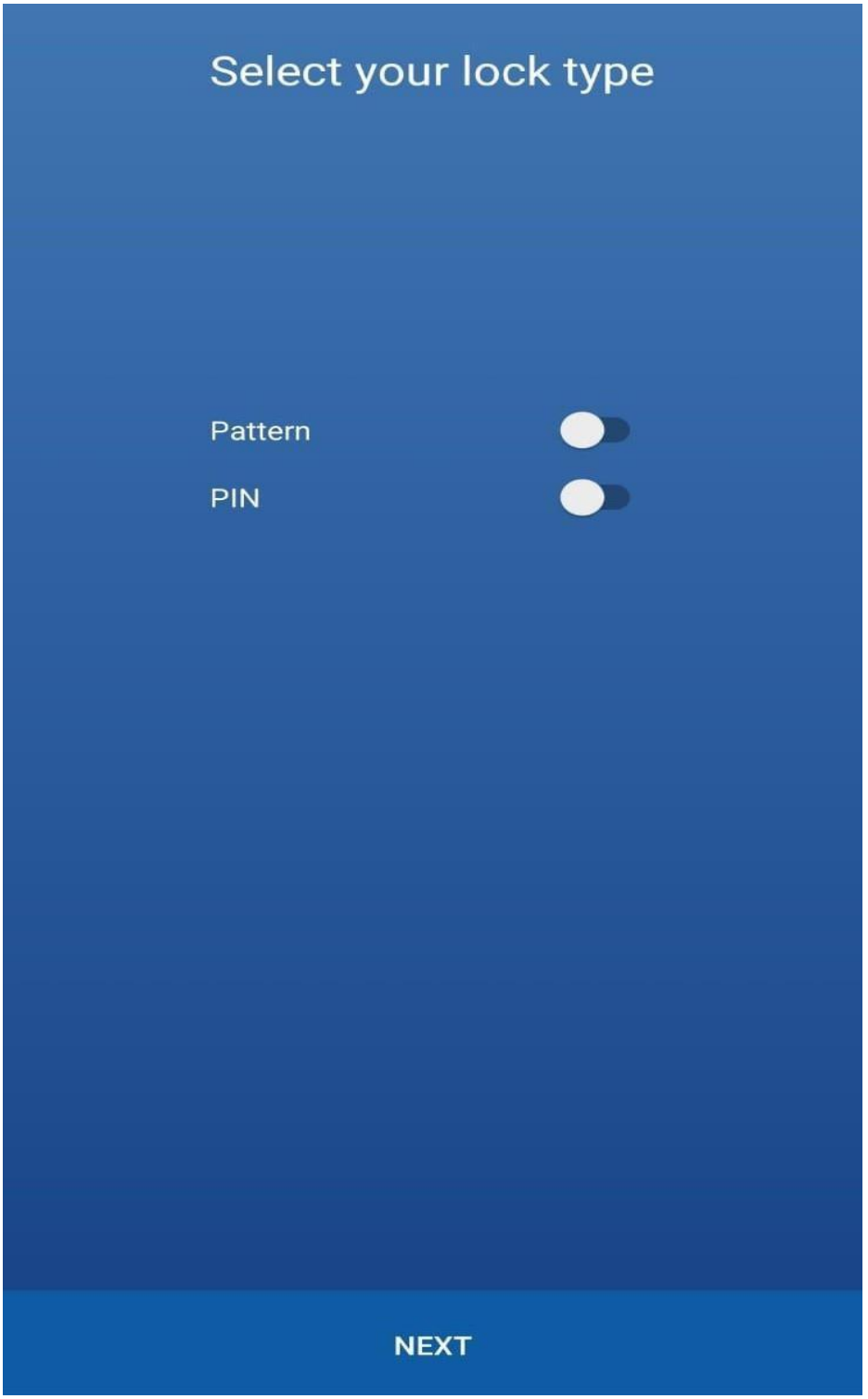

Fig 5.8. App locks of 2 types pattern and pin

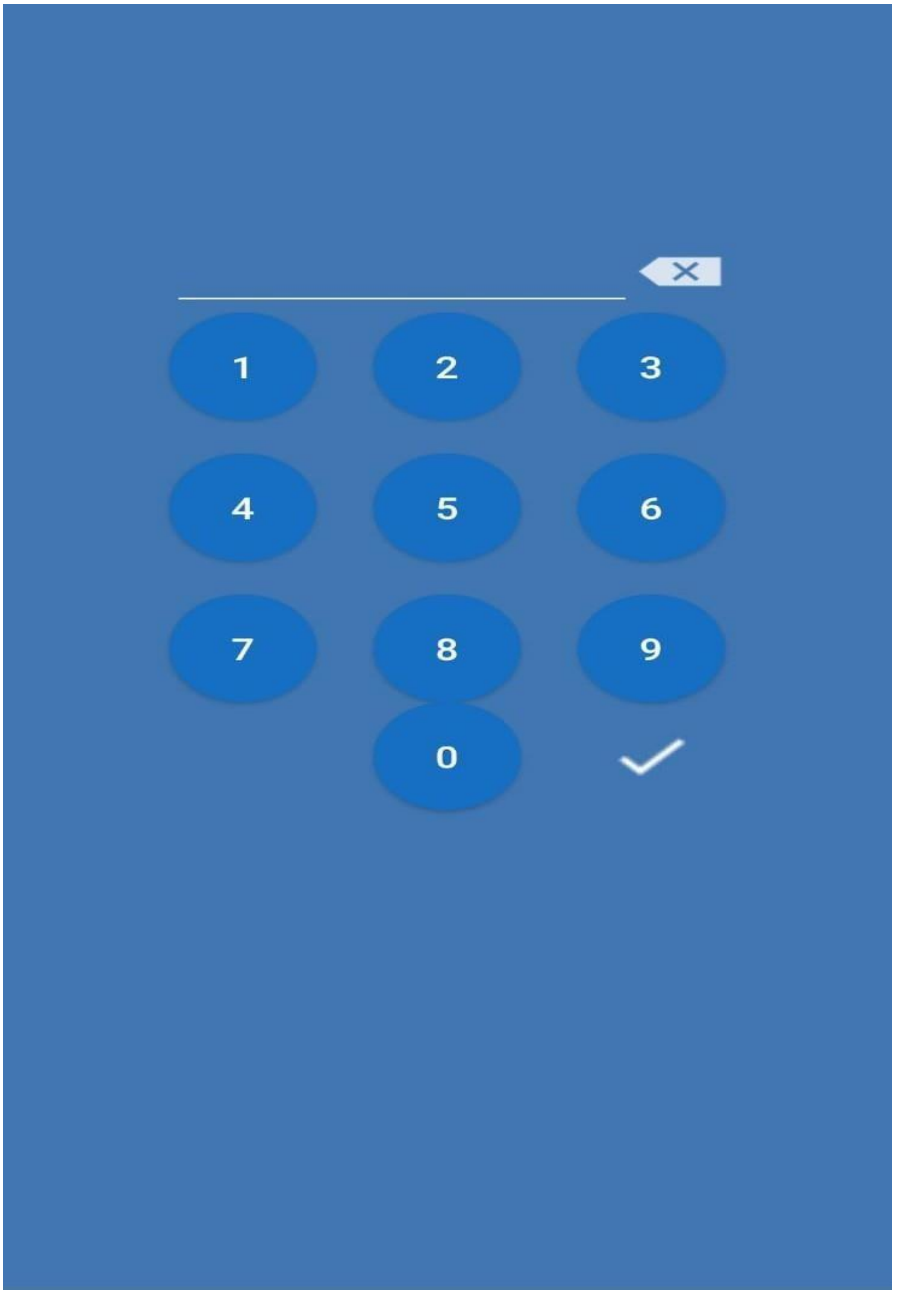

Fig 5.9. The option for adding pincode

Pincode for locking the applications.

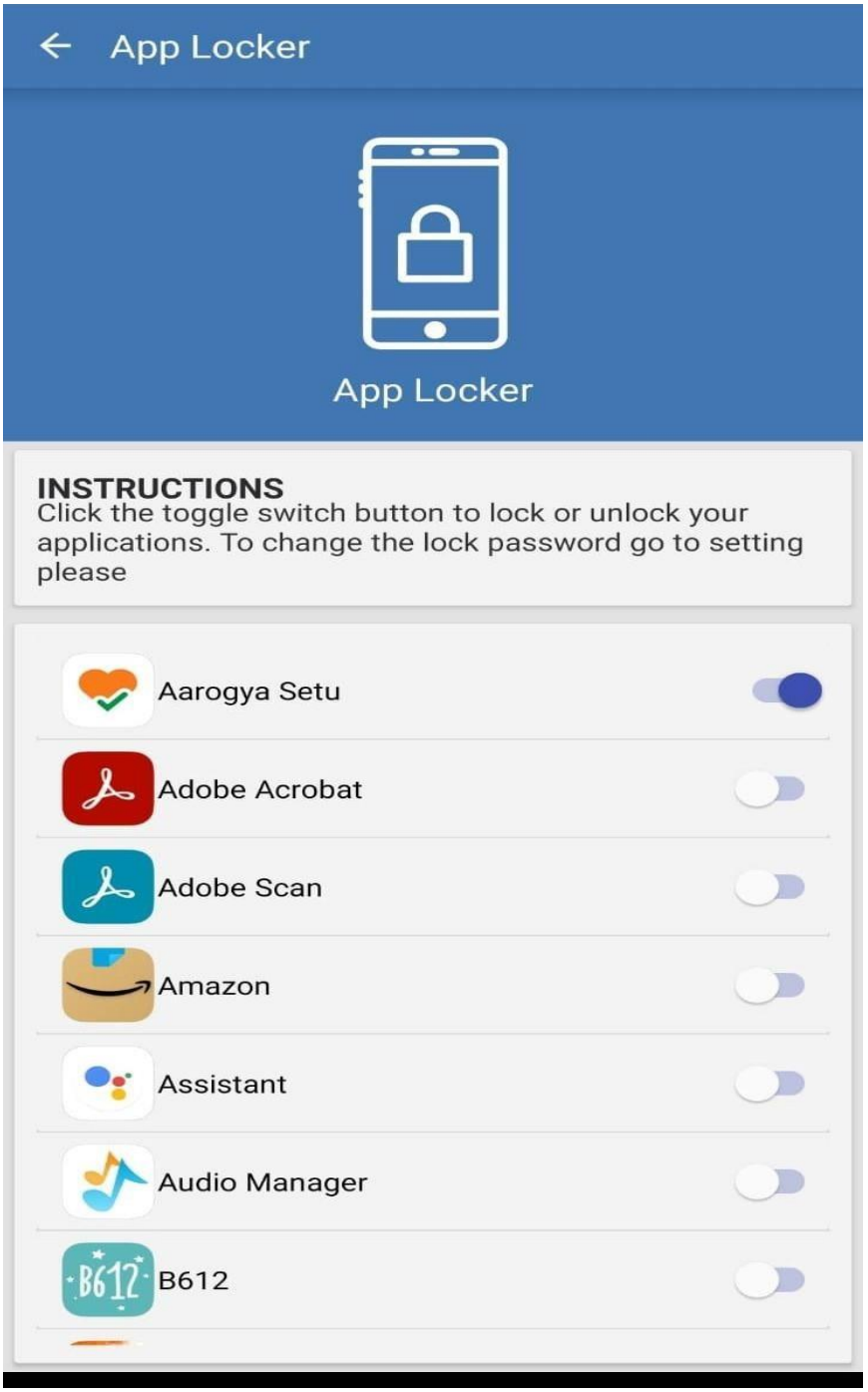

Fig 5.10. Application to be locked using pincode

Remote lock screen functionality will lock the system if it received a message of locking it due to security concerns. App locker Pattern and PIN supported- This can be used to lock the system's screen by either setting the password or PIN.

Fig 5.11. Wifi Security for providing privacy protection, data protection and hotspot

Wifi Security :- Privacy protection- Warns the system before connecting to an unknown hotspots in the range in order to not lose the system data data protection Allows it to connect to the hesteats only if it's safe.

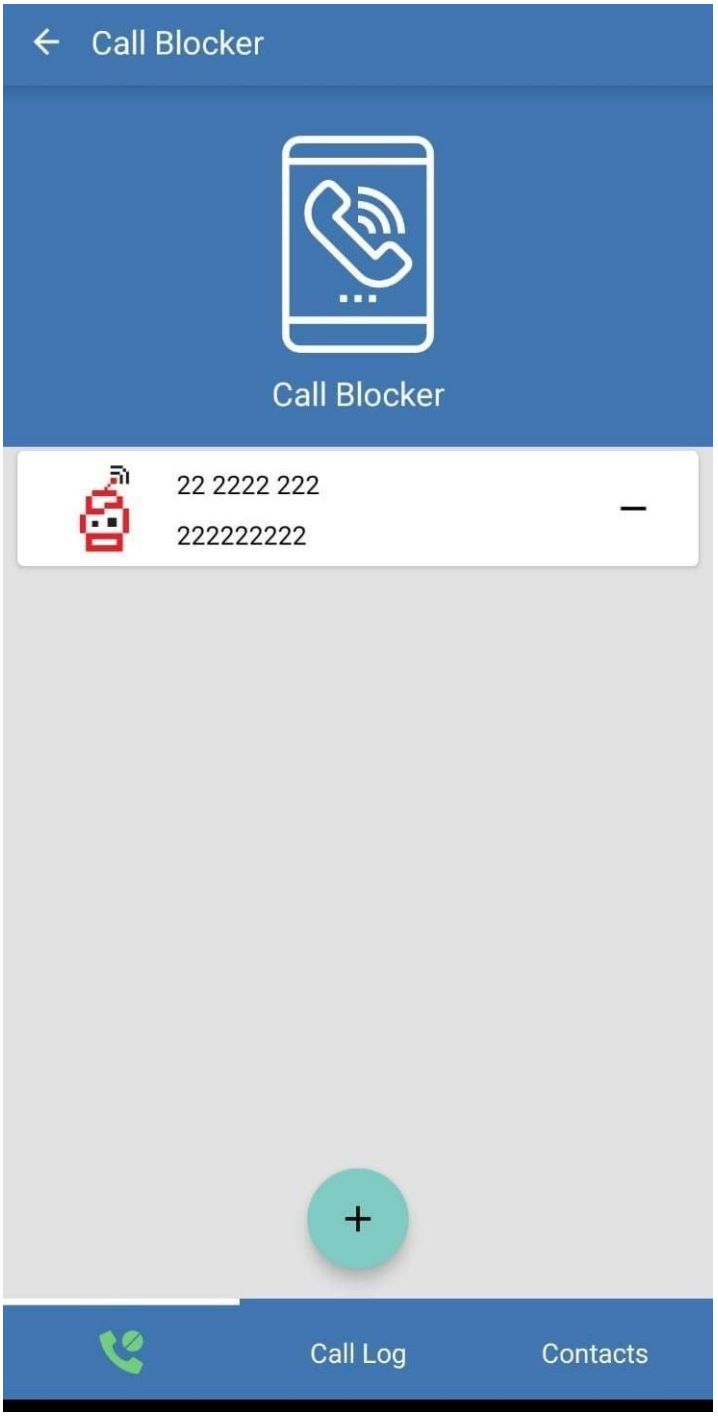

Fig 5.12. Calls that are blocked are displayed

Call blocker:- Blocks the manually added contact numbers in the system

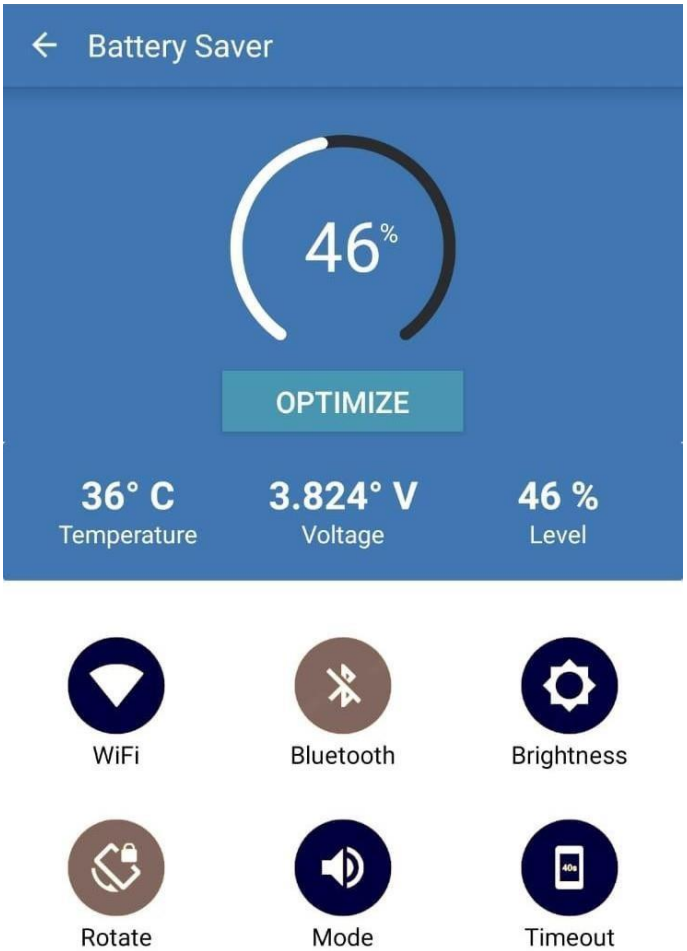

Fig 5.13. Battery optimization and service based application

Battery saver which optimizes battery, wifi, bluetooth, brightness, rotate, mode and timeout when the screen should go on sleep.

Battery saver: Turns on the battery saver by turning off the applications in the system such as Bluetooth hotspot etc so that consumption of power is reduced.

Chapter 6

Conclusion

Antivirus software acts as the final line of defence for PC and other devices, which means it can protect or at least mitigate threats to the devices when every other security software fails.

Due to swift internet technology development, malicious viruses spread through the developed network. The existing traditional detection antivirus cannot kill new viruses and unwanted malicious files. These antiviruses need improved features to overcome virus problems. Faced with all these situations, this project proposes a new antivirus architecture based on cloud computing. In this engine, different techniques will kill the remaining viruses that a traditional antivirus cannot. This model offers significant advantages over the previous host-based antivirus, including better detection of malware in the cloud.

References:

[1] Y.-Z. Li, “Memory efficient parallel bloom filters for string matching, in Networks Security, Wireless Communications and Trusted Computing”, 2009. International Conference on, vol. 1, pp. 485–488.

[2] B. Fechner, “Gpu-based parallel signature scanning and hash generation”, in Architecture of Computing Systems (ARCS), 2010 23rd International Conference on, pp. 1–6.

[3] X. Zha and S. Sahni, “Gpu-to-gpu and host-to-host multi pattern string matching on a gpu”, Computers, IEEE Transactions on, vol. 62, no. 6, pp. 1156–1169, 2013.

[4] K. Nakano, “Efficient implementations of the approximate string matching on the memory machine models” in Networking and Computing(ICNC), Third International Conference on, 2012, pp. 233–239.

[5] D. Man, K. Nakano, and Y. Ito, “The approximate string matching on the hierarchical memory machine, with performance evaluation” in Embedded Multicore Socs (MCSoC), IEEE 7th International Symposium on, 2013, pp. 79–84.

[6] A. K. Sahoo, K. S. Sahoo, and M. Tiwary, “Signature based malware detection for unstructured data in Hadoop” in Advances in Electronics, Computers and Communications (ICAECC), 2014 International Conference on. IEEE, pp. 1–6.

[7] M. Vincent, A. Mesdaq, E. Thioux, A. Singh, and S. Vashisht, “Dynamically adaptive framework and method for classifying malware using intelligent static, emulation, and dynamic analyses,” Oct. 27 2015, uS Patent 9,171,160.

[8] B. Rajesh, Y. J. Reddy, and C. Chakradhar, “Efficient detection of malicious worms with different analysis methods and techniques” virus, vol. 5, no. 4, 2016.

[9] A. P. Namanya, J. Pagna-Disso, and I. Awan, “Evaluation of automated static analysis tools for malware detection in portable executable files” in 31st UK Performance Engineering Workshop 17 September 2015, 2015, p. 81.

[10] M. Zakeri, F. Faraji Daneshgar, and M. Abbaspour, “A static heuristic approach to detecting malware targets” Security and Communication Networks, vol. 8, no. 17, pp. 3015–3027, 2015.

[11] T. Mithal, K. Shah, and D. K. Singh, “Case studies on intelligent approaches for static malware analysis,” in Emerging Research in Computing, Information, Communication and Applications. Springer, 2016, pp. 555–567.

[12] W. Fleshman, E. Raff, R. Zak, M. McLean, and C. Nicholas, et al. “Static malware detection & subterfuge: Quantifying the robustness of machine learning and current anti-virus,” arXiv preprint arXiv:1806.04773, 2018.

[13] R. Agrawal, R. Srikant, et al. Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB, volume 1215, pages 487–499, 1994.

[14] Bauer, J. ITU Study on the Financial Aspects of Network Security: Malware and Spam – Final Report July 2008.

[15] W. Wang. Exploring permission-induced risk in android applications for malicious application detection. Information Forensics and Security, IEEE Transactions on, 9(11):1869–1882, 2014.

[16] R. Lyer. A New Malware Classification Framework Based on Deep Learning Algorithms, IEEE Transactions on, volume 9, 2021

[17] R. Komatwar and M. Kokare. Malware images: Visualization and automatic classification, IEEE Transaction on, volume 10, 2020

[18] A. F. Agarap. Intelligent dynamic malware detection using machine learning in IP reputation for forensics data analytics, IEEE Transaction on, volume 2020

[19] B. Anderson. Intelligent Vision-Based Malware Detection and Classification Using Deep Random Forest Paradigm, IEEE Transaction on, volume 7,2019

[20] StatCounter. A3CM: Automatic Capability Annotation for Android Malware, IEEE Transaction on, IEEE Transaction on, volume 5,2021

Acknowledgement

We would like to express a deep sense of gratitude towards our guide Prof.Akshata Raut, Computer Engineering Department, her constant encouragement and valuable suggestions. The work that we are able to present is possible because of her timely guidance.

We would like to pay gratitude to the panel of examiners for the time, effort they put into evaluating our work and their valuable suggestions from time to time. We would like to thank Project Head of the Computer Engineering Department, Prof. Janhavi Sangoi for her support and coordination.

We would like to thank Head of the Computer Engineering Department, Prof. Ashwini Save for her support and coordination. We are also grateful to the teaching and non-teaching staff ofthe Computer Engineering Department who lend their helping hands in providing continuous support.

HIMANSHU DHANDE

DIVYA KARWANDE

ATHARV KADAM