Security Platform for Mobile OS

Project ID: - 19-001

BSc. (Hons) in Information Technology
Specializing in Cyber Security

Department of Information Systems Engineering
Sri Lanka Institute of Information Technology

Submitted on 13th May 2019

Security Platform for Mobile OS

Project ID:- 19-001

Authors:

Student ID	Name	Signature
IT16009400	**Brayan Benett A.S**	
IT16026544	**Vinushanth K**	
IT16034396	**Sam Abisherk R**	
IT16073388	**Ranjitha L**	

Supervisor

…………………………………….

Mr. Amila Nuwan Senarathne

Co Supervisor

…………………………………….

Mr. Kavinga Yapa Abeywardena

DECLARATION

We declare that this is our own work and this project does not incorporate with acknowledge 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.

……………………………… ………………………………

Brayan Benett A.S Vinushanth K

……………………………… ………………………………

Sam Abisherk R Ranjitha L

ABSTRACT

Evolution of human is the corner stone for everything that we see, feel and use today. History of Phone is the greatest living example we can see. All started from devices which transmitted signals from one place to another followed by wired devices which transmitted voice to a limited distance. With the invention of new technologies and improved mechanisms wireless signals helped in the field of telecommunication. Mobile phones aka hand held phones are the epitome of telecommunication is a true statement now than ever before.

In the 21st century mobile phones are the extended arms of almost every human being. Nowadays mobile phones are not only used in voice communication or text messages. It acts as a palmtop computer with almost all the capabilities. Modern devices have all the features that helped it to become the ultimate source of data for an individual. It was easy for an individual to keep all his data intact with him in his hands. But with great power comes the great responsibility. Almost all the details about the person is saved in his/her mobile phone which act as a single point of failure.

There are vulnerable points which can be exploited to acquire the personal and sensitive data from the device in order to gain unethical advantage over an individual. Bluetooth, Wi-Fi and human errors are some of those vulnerable points which can give out sensitive data to the perpetrator. In our research we intend to propose solutions for the main above-mentioned issues which will result in a secure trustable Android platform for secure operating system that can be very much helpful for the general public and professionals to safeguard their information.

TABLE OF CONTENT

LIST OF FIGURES

1. INTRODUCTION

1.1 BACKGROUND CONTEXT

The purpose behind this report is to give a brief review on the entire project of Security Platform for Mobile OS. The Report intend to depict the purpose and complete statement for the improvement of this application. Particularly, this report focuses on the technologies which will be utilized in the development of the solution, overall system architecture, requirements and constraints. Furthermore, it illustrates how each individual components of each member will be taking part into producing integrated final application to achieve the goal of the project. In addition to that, it is set up to focus on the background work done up to now and to state the path that intend to take to complete the Security Platform for Mobile OS to partners of the Research such as, the research supervisors, CDAP group, examination chiefs and the team members.

This document briefly describes the architecture and the functionalities provided by Security Platform for Mobile OS with prevention of data leakage from mobile phones through different ways and secure communication through Wi-Fi and Bluetooth. The product will be a customized Android OS which will have specialized features to prevent accidental data leakage in mobile phones. Following factors are divided as individual research components among the four members of the Security Platform for Mobile OS Team.

- Data leakage prevented Keyboard;
- Rogue Access point detection;
- Malware detection for Bluetooth;
- Malware detection for WIFI-Direct;

The implementation of the Secure Mobile OS happens in 3 stages.

1. Collecting samples for training ML model.

The valid samples of different diversity must be collected for each scenario. The more the sample data the more the accuracy is. The samples must be then grouped into data sets and classified according to their nature.

2. Train the Machine Learning Model.

The grouped data sets must be then fed to the Machine learning model in order to create an algorithm which will result in giving the desired accurate outcome for the system. Different Models and algorithms will be created for each instance of Data leakage as stated before. Thus each scenario will be singularly focused and the Models created will be more accurate and give much efficient and accurate result.

3. Intergrade the sub parts with Android OS to create a customized OS

Each system created individually focuses on one main task, these individual components must be Assembled as one unit which has all these functions.

Mobile phones are one of those inventions that helps human being in every possible way. In present days mobile phone act as a single go to point for all the important needs. Whether it could be communication or data storing. All the most important details such as personal health information, Credit card or Bank account details, addresses, email addresses and so on. And it needs to be understood that loss of assets these days are not only refers to the loss of money and properties it also refers to the sensitive data that we possess.

When we consider security breaches and data loss in mobile phones it needs to be noted that Bluetooth and Wi-Fi are a concern to our data. In the meantime, sensitive data can be disclosed to unauthorized individuals through insecure internet or wireless connection and even accidental transportation of data. In particular it has to be taken as a serious issue that an attacker can gain many important data through exploiting vulnerabilities in Bluetooth and wireless connectivity which will result in the loss of privacy and valuable assets. There have been several techniques been used in the past to decrease the possibility of these attack. Even though if we consider android platform which is used by majority have only a basic level of protection given to these channels.

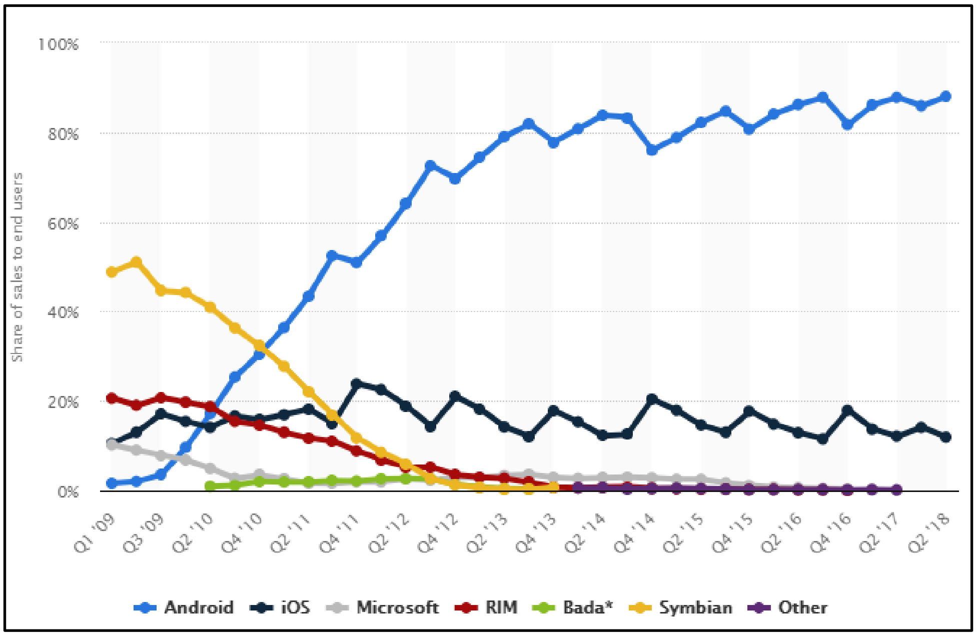


Figure 1– Global mobile OS market share

According to Kaspersky lab, almost 88% of the total mobile OS market share belongs to android. As android kernel is a modified version of Linux kernel and Linux being an open source the security measures that have been taken in android platform for Bluetooth security, Wi-Fi access points security, Wi-Fi direct security and accidental data leakage prevention is very minimum to none.

We propose to secure Android Mobile OS by implementing Machine learning technology in certain aspects of

- Wireless connectivity
- Bluetooth connectivity
- Keyboard system

which will result in a secure connection via Wireless access points, Secure file transection and connectivity via Bluetooth and prevent the accidental leakage of sensitive details to an unauthorized individual. On the whole our research product can be a comprehensive secure platform for mobile OS.

1.2 ADDRESSING THE LITERATURE

I. Data Leakage prevention

There have been several researches conducted on the topic of data leakage and its prevention. Even though there are no much researches done on accidental data leakage to unauthorized parties, general concept of data leakage prevention has been discussed widely. Most of them are industrial level standards.

A. Detecting Data semantic: A data leakage prevention approach [1]

The leakage of sensitive files and data to outsiders can be a disastrous thing for a multinational organization or to an individual. The leakage in data such as trade secret, company banking details, individual ID numbers and health records can affect the stake of a company, the privacy of an individual and the security of the assets.

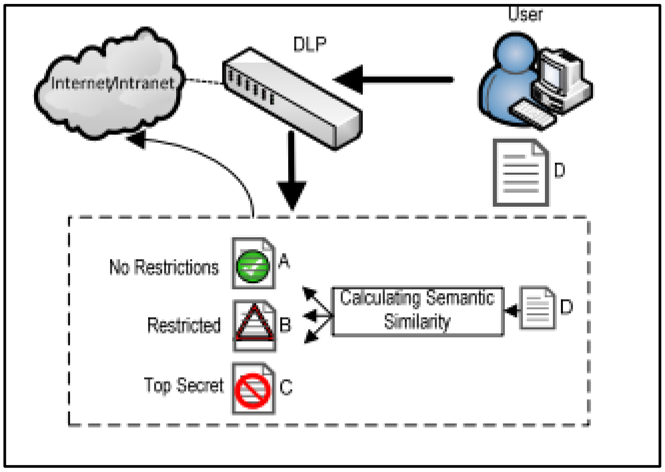
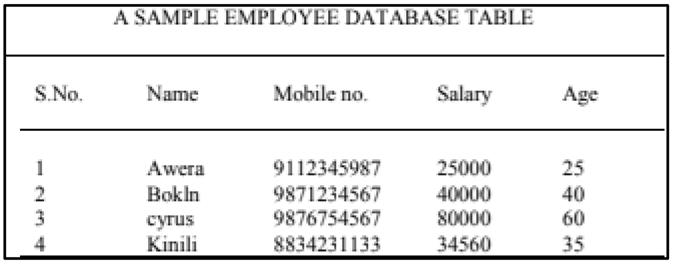


Figure 2 - DLP model

In this DLP model they use statistical data analysis to identify the sensitive data semantics. They use a very famous weighting function Term Frequency- Inverse Document Frequency (TF-IDF) which is used to retrieval of information and text mining to measure the amount of information in a document or a file. They are aiming to separate the details in the document according to pre-defined topics which has a predefined secrecy level. And once the detected, it can be blocked, quarantined or alerted according to the predefined set of instructions. And they propose to implement the DLP as a hardware appliance or a software agent which can be installed in the clients Device.

B. Automatic detection of sensitive attribute in PPDM [2]

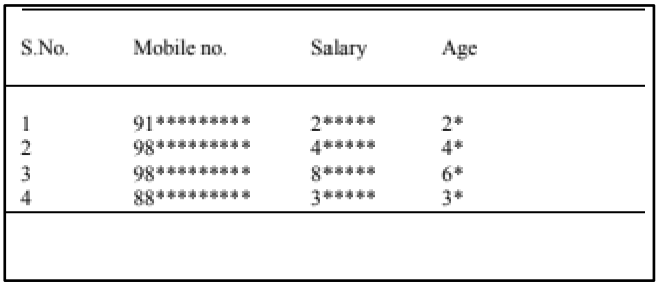
Here they propose to suppress the sensitive data by data mining in order to prevent it from being leaked. They propose to do the data mining to find out the sensitive attributes in a database and hide the values in order to protect the privacy.



A SAMPLE EMPLOYEE DATABASE TABLE

S.No.	Name	Mobile no.	Salary	Age
1	Awera	9112345987	25000	25
2	Bokln	9871234567	40000	40
3	cyrus	9876754567	80000	60
4	Kinili	8834231133	34560	35

Completely modified table to protect the employee privacy



S.No.	Mobile no.	Salary	Age
1	91*********	2*****	2*
2	98*********	4*****	4*
3	98*********	8*****	6*
4	88*********	3*****	3*

Figure 3 – Sample Database Table

First client query is thoroughly analyzed to find out the sensitive attributes. Attributes can be defined as sensitive with regards to the previously defined threshold value by the data owner. When considering a database sensitivity weight will be assigned to every attribute and the query will be analyzed for the total sensitivity value. If the total of the sensitivity of the required attributes exceed the threshold it will carry out the operations which were pre-defined. By this method they propose to eliminate the accidental leakage of sensitive data from a database to an outside unauthorized party.

C. Sensitive data leakage detection in pre-installed applications of custom Android firmware. [3]

In this research study they came up with a sensitive information leakage analysis system for android based devices by analyzing the over 290 custom ROMs which are already existing. The system has three main modules.

- APK extractor: - This extractor helps to extract the pre-installed applications from the custom ROMs. To extract, a batch script is being used to get all the applications which can be pre-installed in any formats such as zip.
- APK analyzer: - This particular module first analyzes every extracted application for sensitive paths and their entry and exit points. It checks the entry point and exit point in order to be used by the path matcher later. And also, this APK analyzer analyze the data flow of each applications and filter out the flows that relates to sensitive sources and critical sinks. These source-sink methods help to find out the actual sensitive data.
- Path Matcher: - In this module path matcher builds all the possible links for the entry points and exit points which were found by APK Analyzer to detect the availability of data leakage in the flow. These flows can be present anywhere.

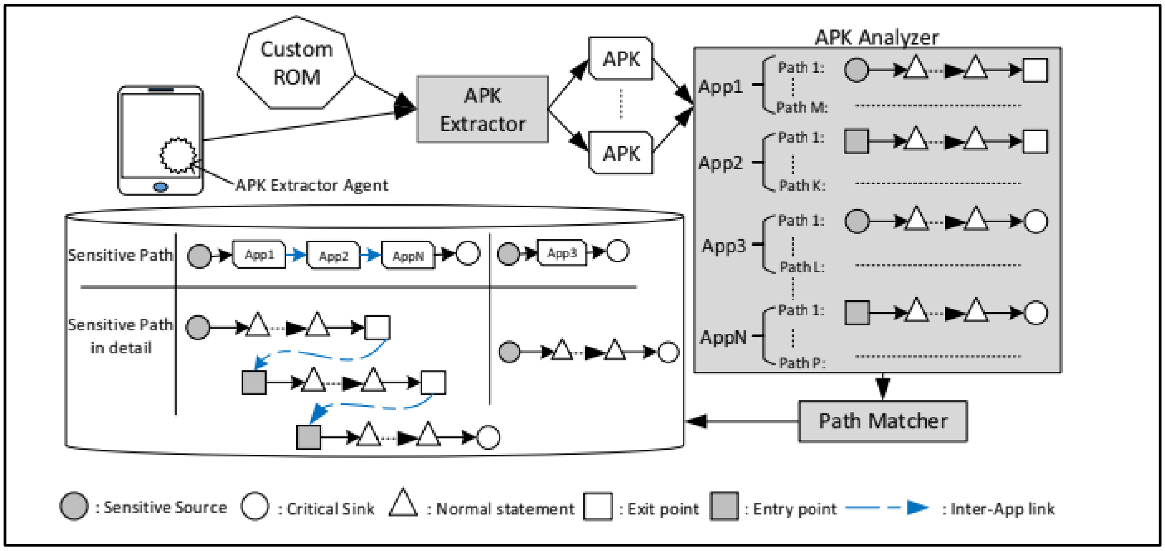


Figure 4 - Information leakage analysis system

As a result of these research they found 4 out of 290 custom ROMs had data leakage in the pre-installed system applications.

II. Rogue access point detection

A. Active User-side Evil Twin Access Point Detection Using Statistical Techniques (TMM, HDT). [4]

This research proposes a new lightweight end user based evil twin detection solution. This technique does not rely on fingerprint checking of suspect devices nor require a known authorized AP/host list. Thus, this solution is most beneficial for user who are in the move or has high mobility outside the organization.

This research focuses mainly on accomplishing most from Intrinsic communication and to identify the properties of the evil twin attacks. Furthermore, we propose two statistical anomaly detection algorithms for evil twin detection that is Trained Mean Matching (TMM)and Hop Differentiating Technique (HDT). In particular, our HDT improves TMM by removing the training requirement. HDT is resistant to the environment change such as network saturation and RSSI fluctuation.

B. CETAD: Detecting Evil Twin Access Point Attacks in Wireless Hotspots [5]

In this paper, we propose a mechanism CETAD leveraging public servers to detect such attacks. CETAD only requires installing an app at the client device and does not require to change the hotspot APs. CETAD explores the similarities between the legitimate APs and discrepancies between evil twin APs, and legitimate ones to detect an evil twin AP attack. Through our implementation and evaluation, we show that CETAD can detect evil twin AP attacks in various scenarios effectively. As most of the most solutions are designed for infrastructure network rather than for client devices. This research mainly focuses on developing a solution on client side. Thus, this research focus on designing a plug-and-play mechanism to detect evil twin AP attacks that only requires to install software at the client device.

There are many challenges that has to be faced when designing a client-side mechanism to detect evil twin AP attacks. First, the client has no information or access about the hotspot architecture as it has only limited resources. Second, many scenarios have to be considered while developing as all the hotspots use various Wi-Fi setup. Third, adding custom hardware, e.g., routers or servers, is not an option as it would limit the applicability and universal acceptance. We overcome these challenges and design a detection mechanism that we call CETAD (Client end Evil Twin Access point Detector).

When multiple AP's connect to the same ISP when they are legally configured to form a hotspot. Thus, they share same SSID and similar network parameters such as ISP names, Global IP address, Round trip timer, temporal network behavior. But a evil twin AP will use a different network setup. This research mainly focuses on these parameters to identify whether the APs belong to same group or not. Even though CETAD is designed for client devices, it can be extended to detect evil twin APs in an infrastructure network as well.

C. A Hidden Markov Model Based Approach to Detect Rogue Access Points[6]

This research focuses on using Hidden Markov Model to detect the Rogue Access Points in a WLAN. This approach identifies a RAP by observing the traffic characteristics of the end hosts in a network. Hidden Markov Model represents the probability of transitions between the different security states of an access point. HMM functions as follows, HMM is trained based on some training data set which consists of information obtained from related to packet traces. These packet traces are gleaned from traffic which includes normal internet activities.

Once the HMM is trained different traffics are monitored in the end user side. By observing the inter arrival time of the packets HMM decides whether an access point is authorized or not.

III. Secure Bluetooth

A. Bluetooth Security Protocol Stacks

Protocol stacks are defined as the combination and implementation of security protocols of hardware/software. Protocol stacks defines the connectivity between devices according to the standards. [7]

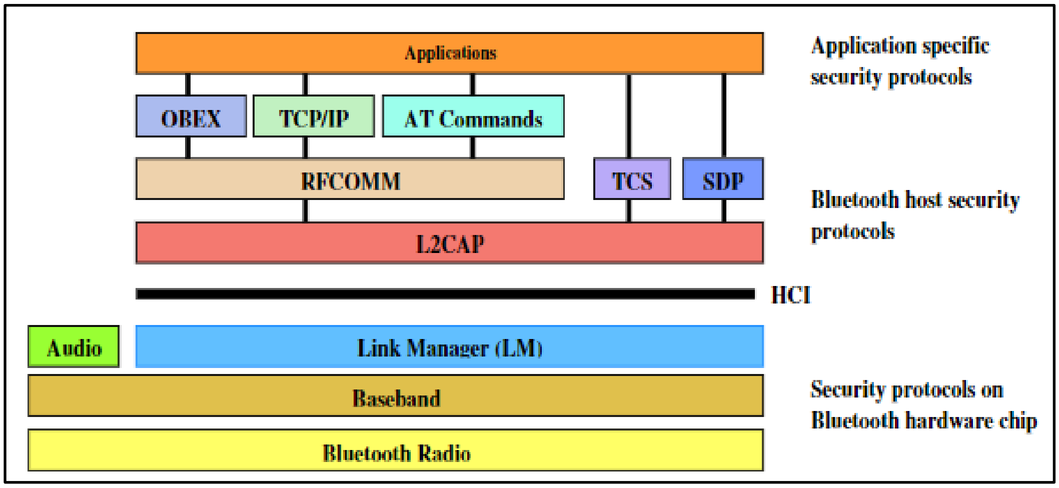

Applications
OBEX
TCP/IP
AT Commands
RFCOMM
TCS
SDP
L2CAP
HCI
Audio
Link Manager (LM)
Baseband
Bluetooth Radio
Application specific security protocols
Bluetooth host security protocols
Security protocols on Bluetooth hardware chip

Figure 5 - Bluetooth Security Protocol Stacks

OBEX- Object Exchange Protocol used to transfer objects such as files, contacts and calendars as client-server model.

SDP- Service Discovery Protocol discovers all the services that are around the RF range proximity and it validate the characteristics of the available discoverable device services.

The above-mentioned protocols are limited to some security extends according to their requirements. We are proposing a solution for the protocol stacks with an overall security of the inbound and outbound connectivity of the Bluetooth technologies by implanting a firewall which will contain the below stated features,

- Considering all the protocols in Bluetooth channels which will monitor and filter the inbound traffic for malicious packets with signature and behavioral based detections.
- Alerting the user for malicious behaviors of incoming files and automatically quarantine them.
- User confirmation will be granted for the particular device to permanently block the attacker device from being connected again
- Logs all Bluetooth events.
- We have the choice to decide on the trusted remote devices.
- Ability to validate the device types with the Bluetooth addresses.

B. Bluetooth Improved Link Key Generation

Several protection measures have been implemented at one of a kind protocol degree, however the safety of the protocols depends on the configuration of the user's device according to their decision of Turing one the discoverable and connectivity options. We can divide the discoverability and connectivity into three modes of operations for the security purposes,

Silent: In this more only the traffics will be monitored and no data connection will be accepted.

Private: According to this mode the device will be only discoverable for the devices which is already having the Bluetooth device address of the requested device. The Bluetooth device address is unique for the particular device.

Public: It is a discoverable device which is open for connectivity.

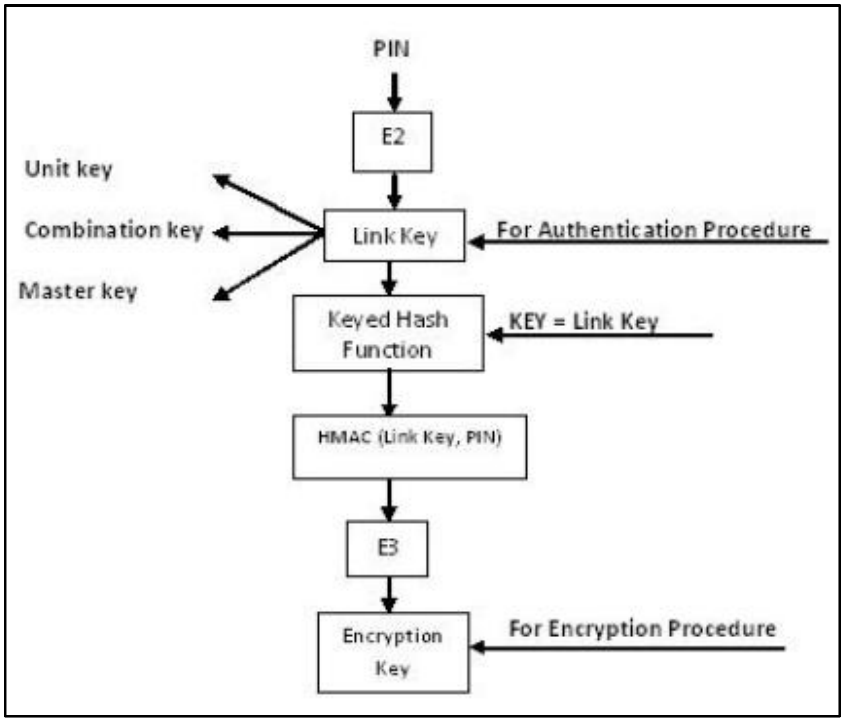


Figure 6- Link Key Generation

By discovering the unit key of the devices Man-in-the-Middle attack is possible. According to the researcher perspective the two devices connected with the unit key has to enter the PIN for the authentication purposes, considering that the file traffic will have a secure connection with the aid of Keyed Hash Function/Algorithm. The is known to the specific devices which is considered as Link Key by them. Then the Link key will go through the Keyed Hash Algorithm and PIN will be combined as E3 algorithm to obtain the Encryption key. The PIN also know to the only the connected devices so the untrusted and fake devices can be eliminated and cannot generate the Encryption Key for accessing the devices the each other. By this process the Man-in-the-Middle attack will be eliminated according the researcher's perspective. [8]

This paper proposes the encryption architecture of the Bluetooth key management technique, with above mentioned PIN encryption method and by validating the Bluetooth addresses with the aid of the embedded inbuilt Bluetooth firewall technology. Bluetooth addresses which is displayed in as 6 bytes and written in hexadecimal format and separated with colons. It will eliminate fake devices more accurately than previous method and the Man-in-the-Middle attack will be failed to exploit for the attacker.

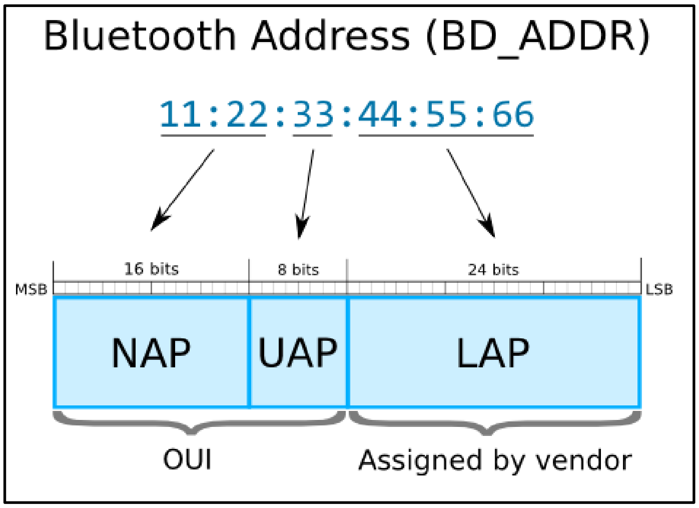


Figure 7 – Bluetooth Address

NAP- Non-significant Address Part value is used in Frequency-hopping spread spectrum (FHSS) frames.

UAP- Upper Address Part value is used for seeding the various Bluetooth specification algorithms.

LAP- Lower Address Part value uniquely finds a Bluetooth device as part of the Access Code in all transmitted traffics.

OUI- Organizationally Unique Identifier

There is an open source tool available for Bluetooth address lookup which provide the information of the vendor, the device type and manufacturing details. The tool can used by inputting the Bluetooth addresses. Our study will sum up by embedding this information gathering tool with the address of Bluetooth, which will be added alongside with our Bluetooth firewall technology for validating the device details and logging purposes.

IV. Secure Wi-Fi Direct

There have been no many researches done regarding the security issues of peer-to-peer technology connection and contingencies done. Some of the similar research topics and their abstracts are below.

A. Secure Device-to-Device Communications over Wi-Fi Direct- A Short-Authentication-String-Based Key Agreement Protocol [9]

This research mainly focused on designing protocol for pair wise key agreement to involve minimal mutual authentication or human interaction. This protocol basically provides an ideal security level in regarding to the required bits that must be mutually authenticated.

Short Authentication based key agreement protocol utilizes a cryptography commitment scheme. This commitment scheme allows to hide a chosen value through a single commit action. Any device which obtains the pair of value is able to reveal the committed value via an open operation. This commitment value doesn't leak any information about the hidden value. This cryptographic function helps to achieve an efficient commitment scheme.

The main goal of this research is to implement an Android application by involving Secure Key establishment functionality into a conventional Wi-Fi Direct application. This solution would allow the user to establish pairwise secret keys which can be used to encrypt the data transmitted through Wi-Fi direct connection.

B. Wi-Fi protected Service (WPS) [10]

Wi-Fi Direct devices are required to implement Wi-Fi Protected Setup (WPS) to support a secure connection with minimal user intervention. In particular, WPS allows to establish a secure connection by introducing a PIN in the P2P Client, or pushing a button in the two P2P Devices. Following WPS terminology, the P2P GO is required to implement an internal Registrar, and the P2P Client is required to implement an Enrollee. The operation of WPS is composed of two parts. The first part which is referred as “Phase 1”, the internal Registrar is in charge of generating and issuing the network credentials such as security keys, to the Enrollee. WPS is based on WPA-2 security and uses Advanced Encryption Standard (AES)-CCMP as cypher, and a randomly generated PreShared Key (PSK) for mutual authentication. The second part, denoted as “Phase 2”, the Enrollee (P2P Client) disassociates and reconnects using its new authentication credentials. In this way, if two devices already have the required network credentials (this is the case in the Persistent group formation), there is no need to trigger the first phase, and they can directly perform the authentication.

1.3 RESEARCH PROBLEM AND GAP

We have come a long way from the invention of mobile phones. If we say that mobile phones are the extended arms of human is truer than ever before. As its so portable and less costly than a fully fetched desktop or a laptop pc, Mobile phones are the go to choose for the majority. Even though the importance of mobile phones increase day by day the security aspects of those devices are merely a question mark.

Data leakage is something to be worried in every situation. Furthermore, accidental data loss is a growing concern for individual and organizations. This can even topple a legacy within minutes. So, the outgoing texts should be monitored to detect sensitive data. Like wise Bluetooth is a growing concern in the devices as the growth of Bluetooth devices are imminent. Malwares can be distributed through Bluetooth channel as it is not monitored and filtered. More than that the devices that are being connected through Bluetooth should be also monitored and logged.

The Usage of WIFI connected equipment inside an organization has increased to a great extend in the past few years. When considering WIFI connected devices mobile phones play a major role. With this comes the disadvantage of security breach through WIFI in mobile phones. A Rogue Access point or an Evil twin is an access point in a local WAN network inside of an organization which is most likely to be published by someone under any circumstances of that organization itself. Rogue Access Points can have same identity just like the legitimate access points, also they have strongest signal strength compared to the legitimate points which automatically leads the users to join the Ap’s without the knowledge of their nature. This opens the security to vulnerability. Information leakage can happen as soon as possible the connection is established. So, the prevention to be implemented should consider also the factors to provide a solution.

WIFI Direct is a greatly underappreciated technology that allows many WIFI enabled devices to seamlessly connect to each other and exchange data without the need for a central wireless router to organize the traffic and relay data packets. Especially when optimized using Net Spot, the easiest native wireless site survey software for Mac and Windows, WIFI Direct can make many tasks that would otherwise be complicated simple. **Wi-Fi Direct** is developed by the Wi-Fi Alliance, Wi-Fi Direct promises to deliver the speed of a traditional Wi-Fi network. Two devices are able to communicate directly, without the need of an internet connection.

1.4 RESEARCH OBJECTIVES

1.4.1 Component Objective

DLP Keyboard

- A keyboard that will be used as the default system keyboard for the OS.
- Keyboard with the function of detecting the sensitive data in the text field.
- Notification mechanism for the OS to notify the user about the sensitive information in the text.
- Trained Machine learning algorithm which can be used to detect sensitive information in text passages.

Secure Bluetooth

- A secured Bluetooth connectivity in mobile devices.
- A Bluetooth technology which blocks malware and unwanted malicious packets from entering in to the devices.
- An algorithm which does the filtration process automatically.
- Many complex operations such as behaviour analysis will be handles easily.
- No human effort is needed in malware identification
- It doesn’t require additional hardware
- Only open source tools will be used.

Secure Wi-Fi Direct

- Implanting a firewall in Wi-Fi direct which will monitor the incoming traffic for malicious files with the help of Machine learning algorithm.
- Keeping the logs for all the activities done through Bluetooth channels.

Implementing Rogue access point detector

- Implementing a method in mobile phones to detect whether the wireless network is a genuine one or Evil twin/Rogue wireless access point and filtering at the time of connectivity (Rogue wireless network Detection). Validating details from the access points with the help of Machine learning.

1.4.2 General Objective

- **Objective 1: Ease of use**
  The above-mentioned implementation has to be done in the system level itself so as the user experience will be easy.

- **Objective 2: Minimum use of available resources.**
  Implemented new methodologies are forced to use minimum amount of resources which won’t affect the usual system process.

- **Objective 3: Analyzation of data**
  Intended to analyze most of the currently available malware signatures and behaviors to create a data set which will result in a more optimized output.

- **Objective 4: Knowledge acquisition**
  As we step into an unknown territory, at the end all the members could gain a full knowledge about android architecture and modern technologies.

2. RESEARCH METHODOLOGY

2.1 Methodology

This particular part of the document explains how we are going to carry out our research. An idea about what are the technologies and tools that we are going to use is also mentioned in this part.

Pre-requirements

We need a set of devices and tools as the pre-requirements in order to carry out our process.

- Machine with Linux or Mac (64-bit environment): - In order to compile Android
- Sandbox: - It needs to be precautioned to safeguard the running system from malwares as we have to work with malwares to get data set.
- Machine Running Windows 7 or latest: - To run an emulator
- Android Emulator
- Mobile Phone with Android OS: - To check the customized ROM
- Network Simulator: - In order to create fake access points.

System Architecture

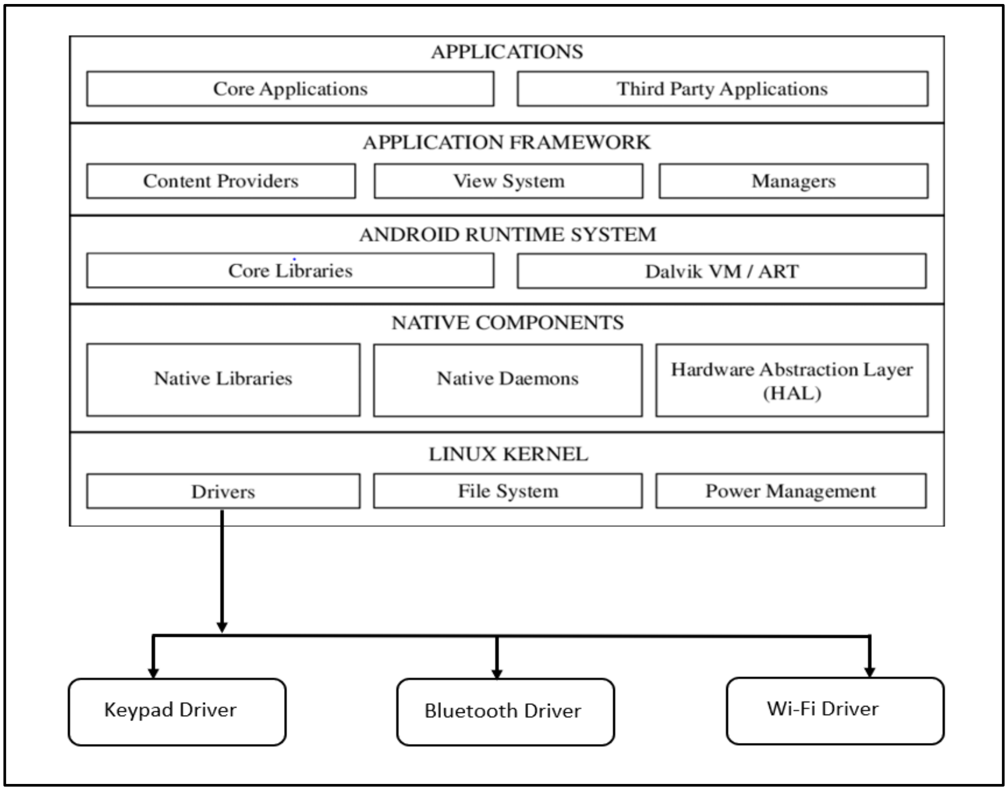


Figure 8 – System Architecture

The above figure shows the system architecture of a Android device. As shown the components we are considering reside under the Linux kernel and within the Drivers. In order to make changes to the Keypad Driver, Bluetooth driver and Wi-Fi Driver we need to make changes to the kernel. Thus, we use Android studio to create custom ROM.

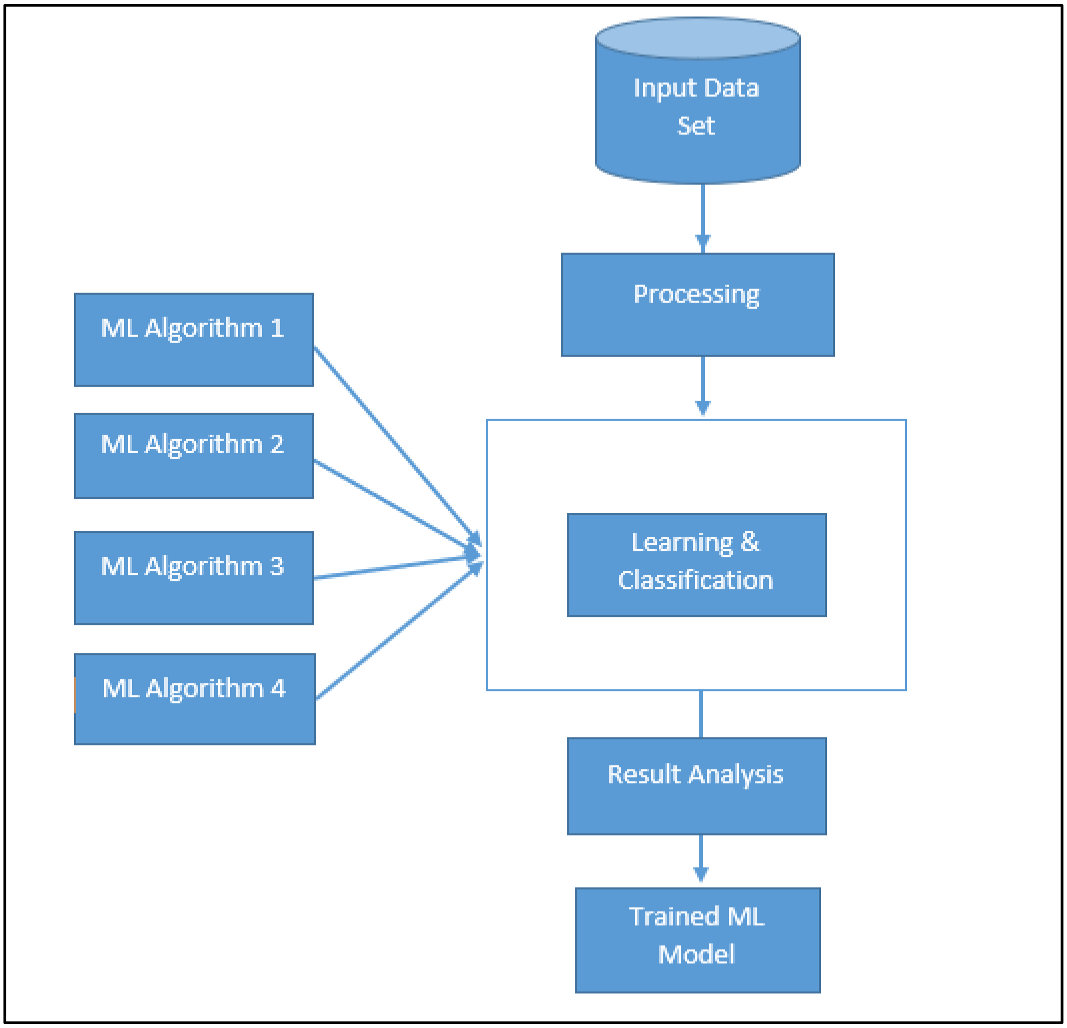


Figure 9 – Machine learning Architecture

In the Machine learning architecture as the figure depicts, all the finalized data set will be taken to processing state where critical attributes will be assigned for learning process. All processed data will undergo any of four machine learning algorithms (Eg: -Linear Discriminant Analysis, Classification and Regression Trees, K-Nearest Neighbors, Learning Vector Quantization, Support Vector Machines) and the learning outcome will be analyzed in Result analysis phase. In result analysis phase it will be decided whether the outcome is reliable or not and the finally the highest accurate algorithms will be selected to be used in every respective module. [11]

- Implementation of ML algorithm in Keypad Driver

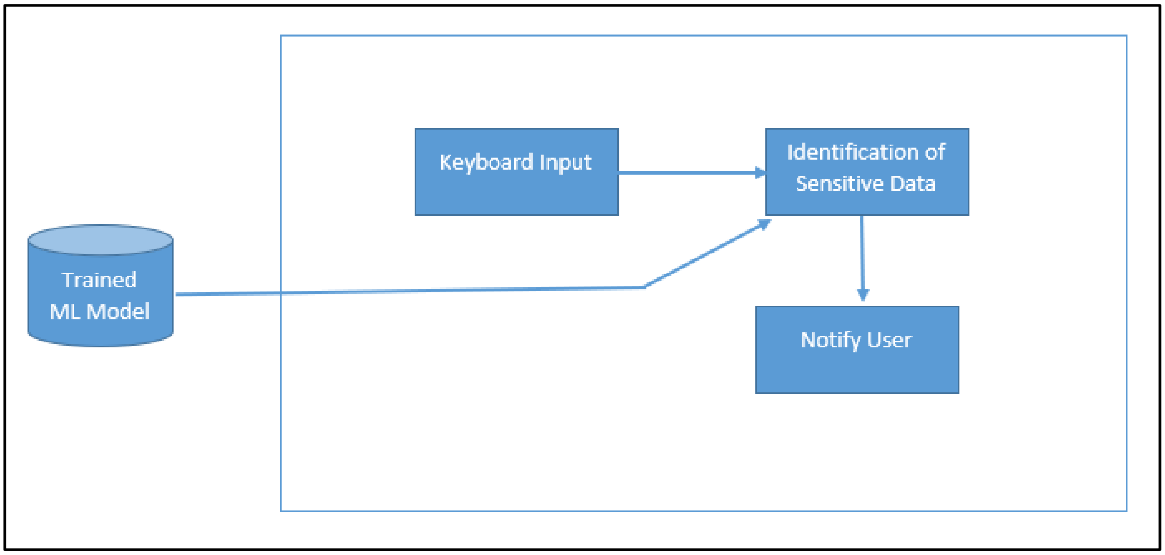


Figure 10 – ML algorithm in Keypad Driver

- Implementation of ML algorithm in Bluetooth Driver

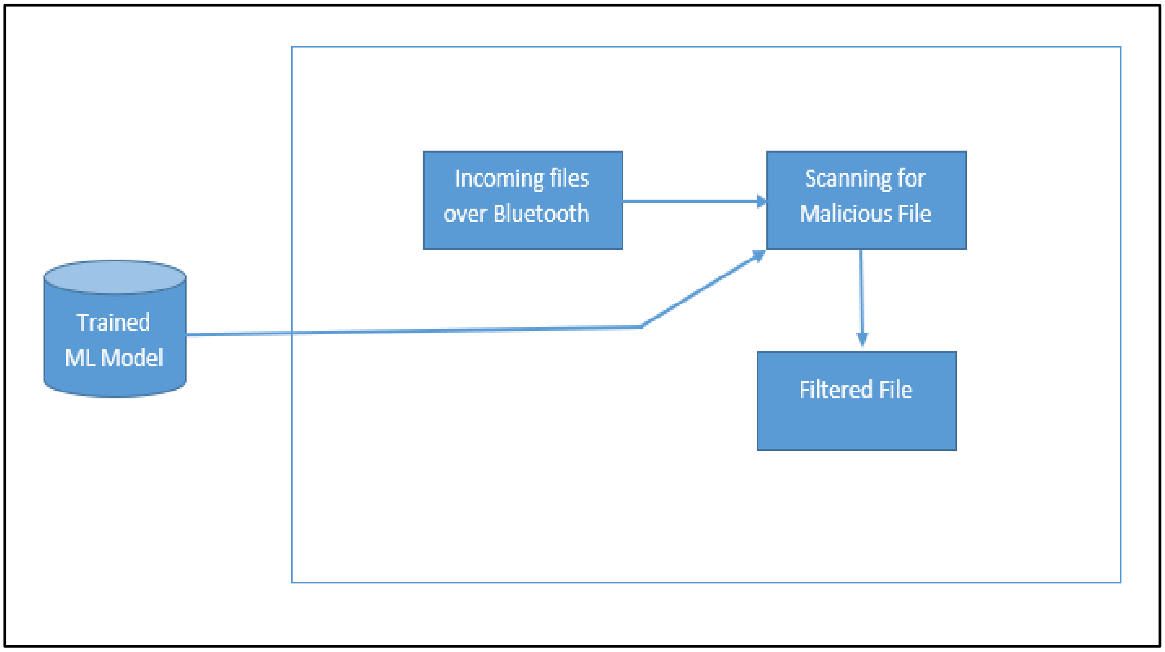


Figure 11 – ML algorithm in Bluetooth Driver

- Implementation of ML algorithm in Wi-Fi Driver

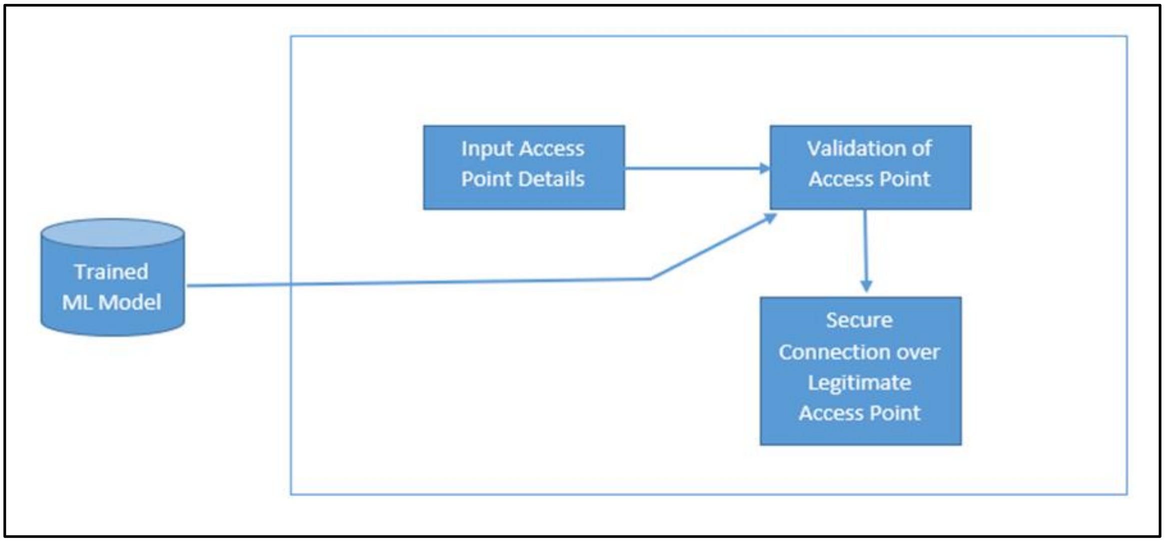

Input Access Point Details
Validation of Access Point
Trained ML Model
Secure Connection over Legitimate Access Point

Figure 12 – ML algorithm for Rogue access detection

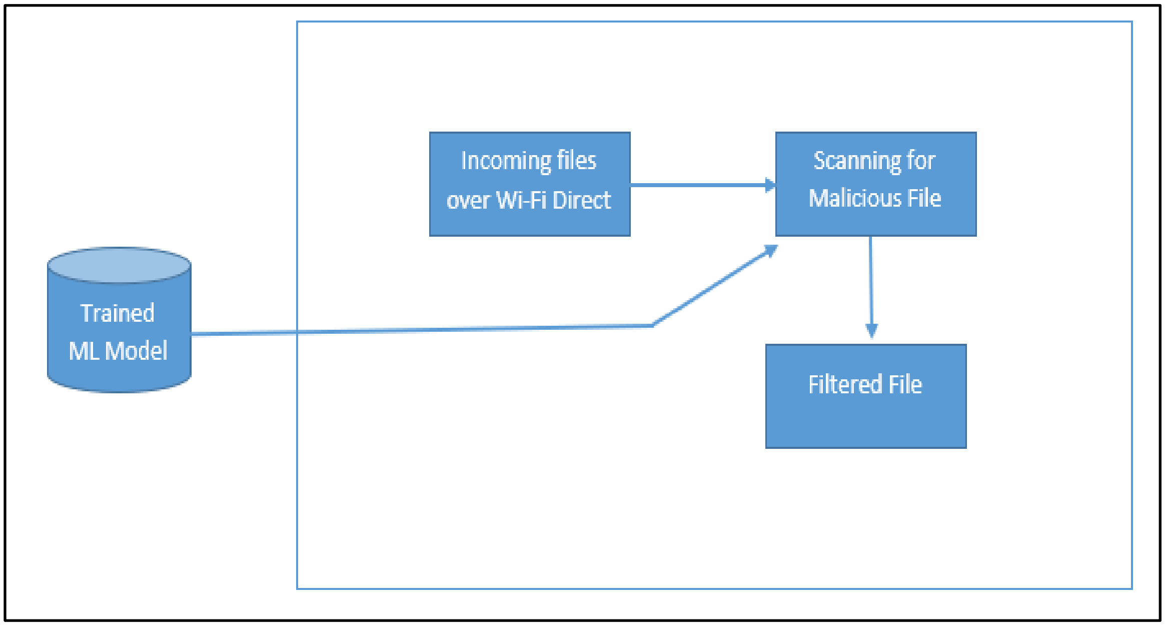


Figure 13 – ML algorithm for Secure Wi-Fi direct

As the above figures shows, most suitable ML algorithm will be implemented in every components in order to bring out the best results. With that a log management system will be implemented for both Bluetooth and Wi-Fi direct to keep the log of devices and files that are transferred and connected.

Malware Detection

Malware detection is the process of scanning the incoming for malicious detections, as we proposed our Secure Bluetooth and Secure Wi-Fi Direct technologies having the ability of detecting the malicious incoming files. For the malicious detection we are having two phases of detection. One is signature-based analysis which is as the normal antivirus detections then behaviors/anomalies-based detections which is as Next Gen Antiviruses detection. According the developed malware detection for the downloads of files through Bluetooth and Wi-Fi direct we could able to verify the files with signatures and their behaviors.

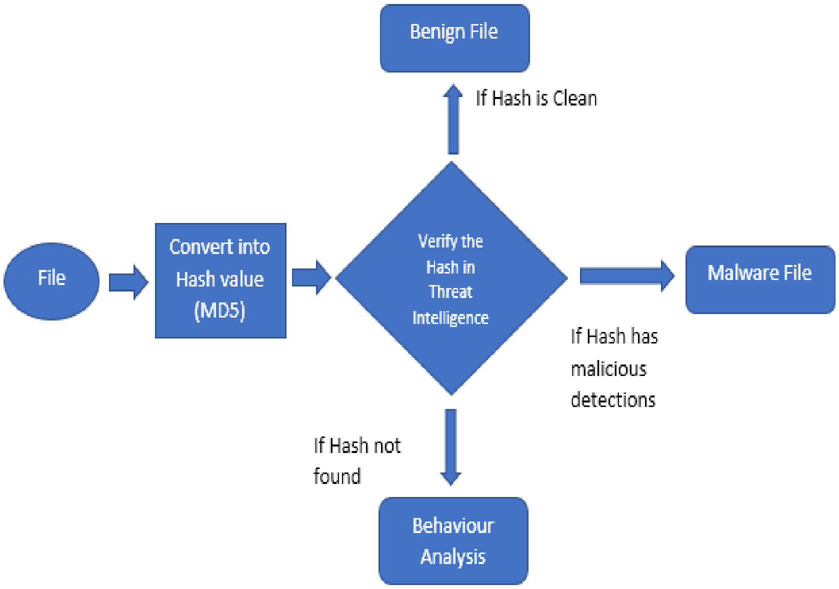

Benign File
If Hash is Clean
File
Convert into Hash value (MD5)
Verify the Hash in Threat Intelligence
Malware File
If Hash has malicious detections
If Hash not found
Behaviour Analysis

Figure 14: Machine Learning for Malware Detection

When file comes through Bluetooth/ Wi-Fi Direct the file will be first converted for it hash value, for this conversion we are using the MD5 hash function and we will be getting the MD5 hash value for the specific incoming file. Next the converted hash value (MD5) will go through the process of verification of hash value in the Global Threat Intelligence. In GTI we can get three outputs from the verification process they are, File hash can be clean, the hash can have malicious detections and the hash cannot be found in GTI. By checking the hash value in GTI we will be able to find out for that specific hash has any malicious detections then that file will be a malware or malicious file. If the hash is clean, then that file will be benign file. If that hash is not found in GTI then there comes an issue, for that we are having the behavior analysis detection.

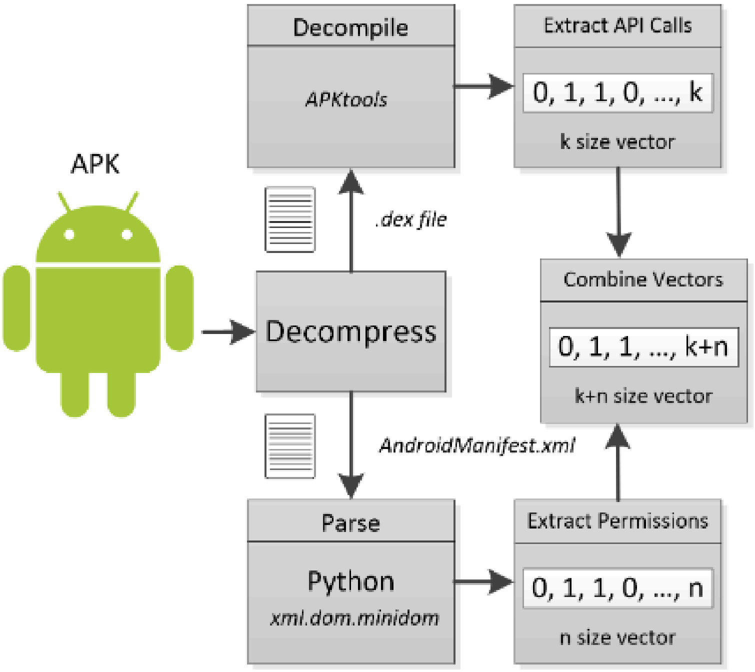


Figure 15: Malware Detection Cycle

The malware detection process of behavior analysis for the downloads of apk files through Bluetooth and Wi-Fi Direct will be examined by extracting the features from the apk files. The feature extraction process will consider the application's permissions and API calls. The combination that were taken for the feature extraction process are combination of permissions and API calls, {permissions + API} and analyzing the best combination is suitable for the malware detection accuracy. According to the information gathered from the research paper Detecting Android Malware by Using A Machine Learning Ensemble Method [16] Combination of permissions and API calls will be an improved method for accuracy in detection of malwares. With the conclusion of above-mentioned researchers our framework will do the extraction of permissions and API calls as a single feature. The permissions will be extracted from the AndroidManifest.xml. AndroidManifest.xml file is existed with every application in its root directory. It contains the essential/important information about the whole application and gives the details to the android system. So, in order to get these essential information APK files should be accessed. The APK contains the essential information such applications code, resources, certificates, assets, meta-data, libraries and manifest. After extraction the Manifest.xml files all the permissions will be converted into binaries.

2.2 Technology and Tools

Android Studio

Android Studio is the official integrated development environment for Google’s Android operating system, which will be used for the developing purposed for the Custom Rom and for the use of running as the emulator.

SwiftKey

SwiftKey is a virtual keyboard app developed by Touch-type for Android and iOS devices. SwiftKey uses a blend of artificial intelligence technologies that enable it to predict the next word the user intends to type. SwiftKey learns from previous SMS messages and outputs predictions based on currently inputted text and what it has learned.

Gboard is a virtual keyboard app

Gboard is a virtual keyboard app developed by Google for Android and iOS devices. Gboard features Google Search, including web results and predictive answers, easy searching and sharing of GIF and emoji content, a predictive typing engine suggesting the next word depending on context, and multilingual language support.

Azure Machine Learning Studio

Machine Learning Studio is a powerfully simple browser-based, visual drag-and-drop authoring environment where no coding is necessary.

Amazon Sage Maker

Amazon Sage Maker furnishes each engineer and information researcher with the capacity to build, train, and deploy AI models rapidly. Amazon Sage Maker is a completely overseen administration that covers the whole AI work process to mark and set up your information, pick a calculation, train the model, tune and streamline it for sending, make forecasts, and make a move. The models get to creation quicker with significantly less exertion and lower cost.

Genymotion

Genymotion is a great alternative to Android Studio's default emulator. It is intended to be for development purposes. One great benefit of using Genymotion is that it literally lets you select any Android phone to simulate and choose to install any Android version — from Android 4.4 KitKat to Android 7.0 Nougat.

VirusTotal

VirusTotal is one of the best global threat intelligence tool which aggregates most of the antivirus products and online scan engines to check for the malicious detections. This useful in detecting false positives as we can get clear view of the detection by comparing them within all the antivirus results.

Cuckoo Sandbox

Cuckoo Sandbox is the leading open source automated malware analysis system. It gives detailed report by highlighting the behavior of the file when the executed in the isolated environment for the testing purposes.

Dataset acquisition.

Datasets for the training of machine learning model were acquired according to each and every component.

- Sensitive data samples to build and train model for Data leakage prevented keyboard.
- Malware samples from Virus total to train ML model for Secure Bluetooth.
- Malware samples from Virus total to train ML model for Secure Wi-Fi.
- Access point samples and behavior samples from online forums for network details and GIT repositories (https://github.com/AbertayMachineLearningGroup/network-threats-taxonomy/tree/master/Datasets) to train ML model for Rogue access point detention.

From the samples, we will be using 75% for the training and rest 25% will be used for testing the model. The selected samples will be analyzed and extracted to reduce the redundancy and to increase the diversity of the samples. The extracted features will be pre-processed before applied to the machine learning algorithm to reduce redundancy and increase accuracy. The pre-processed data will be applied for multiple machine learning algorithm to choose the efficient model. Each models will be tested recurrently in order to get the best outcome.

3. RESULTS AND DISCUSSION

3.1 Results

3.2 Research Findings

3.3 Discussion

4. FUTURE WORKS

Since we have developed solutions for the problems we found only to android ecosystem, in future we are hoping to extend the protection to all the available mobile platforms starting from iOS ecosystem. And also, with the availability of more data sets the detection of sensitive data can be even more fine-tuned. Another future plan is to optimize the workload and to implement the algorithm into the mobile itself rather than relying on the cloud and API service to get the results as the specifications of mobile phones and the capability of the mobile phone processors increase day by day.

5. RESEARCH CONSTRAINTS

4.1 Platform Limitation

All the above implementations are tested and optimized for android platform. Can only be used for Android Versions after “Android KitKat”.

4.2 Limed resources on using Machine Learning

Need higher measure of assets for Machine Learning model preparing and need to follow similar undertaking for multiple users to prepare datasets. There is high requirement for resources such as developing platforms, require datasets, testing.

4.3 No researches available

This project which consist of four components which are less researched and there is no research exist as a whole. Related research materials have been combined according to its components.

4.4 Accuracy level should be higher than 70% for the predictions.

To call the detections successful accuracy rate should be above 70%.

4.5 Privacy concerns

Since we are using samples of sensitive information, it’s a must to maintain the privacy of the data which we are using for the training and testing.

6. REFERENCES

[1]S. Alneyadi, E. Sithirasenan and V. Muthukkumarasamy, *Detecting data semantic: A data leakage prevention approach*. 2015.

[2]P. Kamakshi and D. Babu, *Automatic Detection of Sensitive Attribute in PPDM*. 2012.

[3]N. Tan Cam, V. Pham and T. Nguyen, *Sensitive data leakage detection in pre-installed applications of custom Android firmware*. 2017.

[4] CANG, C., SONG, Y. AND GU, G. *ACTIVE USER-SIDE EVIL TWIN ACCESS POINT DETECTION USING STATISTICAL TECHNIQUES*

[5] MUFASA, H. AND XU, W.CETAD: *Detecting Evil Twin Access Point Attacks in Wireless Hotspots*

[6] SHIVARAJ, G., SONG, M. AND SHETTY, S. *A Hidden Markov Model Based Approach to Detect Rogue Access Points.*

[7]N. Minar and M. Tarique, *BLUETOOTH SECURITY THREATS AND SOLUTIONS: A SURVEY*. 2012.

[8]W. Iqbal, F. Kausar and M. Arif, *Attacks on Bluetooth Security Architecture and Its Countermeasures*. 2017.

[9] SHEN, W., YIN, B., COO, X. AND CHENG, Y.*Secure Device-to-Device Communications Over Wi-Fi Direct*

[10] CAMPS-MUR, D., GARCIA- SAAVEDRA, A. AND SERRANO. *Device to Device communication with WiFi Direct:Overview and experimentation*

[11]J. Le, "A Tour of The Top 10 Algorithms for Machine Learning Newbies", *Towards Data Science*, 2019. [Online]. Available: https://towardsdatascience.com/a-tour-of-the-top-10-algorithms-for-machine-learning-newbies-dde4edffae11. [Accessed: 09-Mar- 2019].