“ Detection And Prevention Of Various Cyber Attacks On Mobile Network ”

Bachelor of Engineering
in
Computer Engineering

Submitted by

Ms. Nikita M. Chorghe Roll No. (16CE7025)
Mr. Akshay R. Jain Roll No. (16CE7017)
Ms. Shraddha S. Mali Roll No. (16CE7013)

Guided by

(Mr. Prathmesh Gunjgur)

D Y PATIL
RAMRAO ADIK
INSTITUTE OF
TECHNOLOGY
NAVI MUMBAI

Department of Computer Engineering
Ramrao Adik Institute Of Technology

Dr. D. Y. Patil Vidyanagar, Sector-7, Nerul, Navi Mumbai-400706.
(Affiliated to University of Mumbai)

April 2020

Detection And Prevention Of Various Cyber Attacks Through Mobile Network

B.E. Project Report

Submitted in partial fulfillment of the requirements

For the degree of

Bachelor of Engineering

in

Computer Engineering

by

Ms. Nikita M. Chorghe (16CE7025)

Mr. Akshay R. Jain (16CE7017)

Ms. Shraddha S. Mali (16CE7013)

Supervisor

Mr. Prathmesh Gunjgur

Department of Computer Engineering

Dr. D. Y. Patil Group's

Ramrao Adik Institute Of Technology

Dr. D. Y. Patil Vidyanagar, Sector 7, Nerul, Navi Mumbai 400706.

(Affiliated to University of Mumbai)

April 2020

Ramrao Adik Institute of Technology

(Affiliated to the University of Mumbai)

Dr. D. Y. Patil Vidyanagar,Sector 7, Nerul, Navi Mumbai 400706.

CERTIFICATE

This is to certify that, project ‘A’ titled

“Detection And Prevention Of Various Cyber Attacks On Mobile Network”

is a bonafide work done by

Ms. Nikita M. Chorghe (16CE7025)

Mr. Akshay R. Jain (16CE7017)

Ms. Shraddha S. Mali (16CE7013)

and is submitted in the partial fulfillment of the requirement for the degree of

Bachelor of Engineering

in

Computer Engineering

to the

University of Mumbai

Supervisor

(Mr. Prathmesh Gunjgur)

Project Co-ordinator	Head of Department	Principal
(Mrs. Smita Bharne)	**(Dr. Leena Ragha)**	**(Dr. Mukesh D. Patil)**

Dissertation Approval for B.E

This is to certify that the project ‘A’ entitled ***“Detection And Prevention Of Various Cyber Attacks On Mobile Network”*** is a bonafide work done by ***Ms. Nikita M. Chorghe, Mr. Akshay R. Jain and Ms. Shraddha S. Mali*** under the supervision of ***Mr. Prathmesh Gunjgur***. This dissertation topic has been approved for the award of ***Bachelor’s Degree in Computer Engineering, University of Mumbai***.

Examiners :

1. ....

2. ....

Supervisors :

1. ....

2. ....

Principal :

....

Date : .../.../......

Place : ....

Declaration

We declare that this written submission represents our ideas in our own words and where other's 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 our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Ms. Nikita M. Chorghe	**16CE7025**	____________________
Mr. Akshay R. Jain	**16CE7017**	____________________
Ms. Shraddha S. Mali	**16CE7013**	____________________

Date : . . . /. . . /. . . . . .

Abstract

Mobile devices have become an indispensable part of our lives. With new inventions that enable our smartphones to perform new functions, our dependence on its usage cannot be overemphasized. These functions are limited not only to establish communication but also to store a wide range of personal and confidential data like passwords and PINs. Hence, securing mobile devices to prevent hackers from accessing this vital data has become highly essential.

A cyber-attack is a deliberate attempt to exploit hardware/software vulnerabilities and to capture, store, alter, misuse private data for personal gains. It is done with the help of malicious scripts, alter the software code or introduce backdoor traps, worms, viruses in the system and retrieve sensitive information which may lead to identity theft. Users also use mobile as a means of online transactions, storing financial information and passwords which makes it important to secure the network over which the communication is being carried out in order to prevent sniffing of data.

The proposed framework can be used to detect such attacks and thus mitigate such threats and secure user data.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

1.1 Overview

With a global increase in dependence on digital devices, there has also been a global increase in incidents of cyber-attacks on these devices, and with respect to such incidents, India has been no exception. Hence, it is the need of the hour to have a system in place that will not only detect but also ward off such attacks.

Smartphones connected to a network are vulnerable to a number of security threats such as malware that can easily breach the defense systems of IOS and Androids. Hence, an in-build strong defense mechanism is crucial to protect security breaches and prevent data and monetary thefts. Solely solutions capable of analyzing behavior at every layer for indications of attacks can protect the devices effectively. Attacks exploit the overall national and business organizations through malware, viruses and worms, faux websites, unauthorized net access, and different means of accessing private and organizational data.

In the present time, cyber-hacks are increasingly being launched for socio-political gains. Mobile networks, spanning over a large geographical area, the user is able to communicate from any location because of wide cellular coverage. Determining location details of users requires the use of software and also a prompt response as it changes rapidly. Mobile operators providing information services to users are vulnerable to nameless attackers who intend to abuse the provider and their network. Common threats namely hacking of web sites, copyright infringement, and illegal communications exist. Mobile operators don't have any predefined guidelines to deal with such threats. Relatively simple solutions have been provided by Internet service providers (ISP's) which can be incorporated in the mobile environment. New threats exploiting mobile network continue to appear. Detection of cyber-attacks help organizations in protecting devices, apps and data from malicious attacks.
The results of data thefts from smartphones can be devastating and far-reaching. It can affect people not only at personal level but also society at large.

As per the latest statistics, roughly 200 billion devices will be connected to the internet by 2020. This alarming rise in number of devices connected to the internet gives equal opportunities to attackers to steal data and infect websites. Since 2013 there has been a theft of 3,809,448 records. In 2016, 95 percent attacks were focused government, retail and technology based industries.[5]

According to Symantec's Internet Security Threat Report 2018, malware variants for mobiles increased by 54 percent in 2017. In 2017, 27 percent of all existing apps were malicious apps found in the Lifestyle category. Next in line: Music and Audio, with 20 percent, followed by Books and Reference, with 10 percent. Nearly 60 million Americans have been affected by identity theft, according to a 2018 online survey by The Harris Poll. Thus, the proposed framework will play an important role in mitigating attacks as 95 percent of cyber-security breaches are due to human error.[6]

1.2 Objective

Every business organization or individual user is a part of the network. These organizations and individuals are constantly endangered by cyber attacks. Cyber risks pose horrifying threats to them with increasing financial frauds and Cyber terrorist act. Governments, multi-national corporations, people face persistent threats, experience frauds, extortion and defacing of vital IT infrastructure. With increasing use of smartphones, threats have also increased. Business organizations and individuals face huge losses due to hacking or malicious attack. Thus it becomes necessary to detect and prevent these attacks to make the mobile connection secure. It is a necessity to provide a secure environment to mobile users.

1.3 Motivation

With the uncontrollable growth of cyberspace around the world, India also has witnessed a significant rise in web activities. This has resulted into an exponential growth in accessing information and connectivity. This led to the empowerment of individuals and also an organization which ultimately increased cyber attacks.

In 2016, 3 billion Yahoo accounts were hacked in one of the biggest breaches of all time. According to 2017 statistics, there are over 130 large-scale, targeted breaches in the U.S. per year, and that number is growing by 27 percent per year. The average cost of a malware attack on a company is 2.4 million dollar. Ransomware damage costs exceed 5 billion in 2017, 15 times the cost in 2015. In 2017, 412 million user accounts were stolen from Friendfinder's sites.[7] Due to such monetary loss and confidential data theft it is necessary to mitigate these attacks.

1.4 Organization Of Report

The report is organized as follows: The introduction of the project is covered in Chapter 1. The prerequisite study and literature survey is covered in Chapter 2. Proposed work and methodology is discussed in Chapter 3. Planning and formulation of the project is defined in Chapter 4 along with Gantt chart. All the diagrams which needs to explain the system are defined in Chapter 5. Expected Results and outcomes are defined in Chapter 6. Conclusion remarks are given in Chapter 7. Chapter 8 explains about the future work. References are included in Chapter 9.

Chapter 2

Literature survey

2.1 Survey of Existing System

Following section is a survey of some of the significant work done by the researchers in this particular domain. It includes current knowledge including substantive findings as well as theoretical and methodological contribution related to the proposal. It gives the references and guidelines to develop better system.

Hamad Alrashede et al [1]. developed a new theoretical method to verify the base station within a particular area before authenticating the connection between the mobile phones and base station. It uses parameters which cannot be modified or manipulated by the attacker unlike in previously proposed methods. An algorithm based on AKA algorithm was proposed for validating the connection between base station and mobile equipment.

Arpita Gupta et al [2]. have proposed a better and more efficient A3 algorithm used in GSM security algorithms. With the number of mobile users increasing, security of mobile networks is of utmost importance. Several security flaws have been detected in GSM security algorithms. Thus an enhanced version of A3 algorithm has been proposed to revamp the GSM security by adding an element of encryption.

2.2 Limitation of Existing System

- Very few systems have been developed on these attacks.
- The proposed systems have not been implemented in the real world
- The developed systems are based on parameters which can be easily manipulated by the attackers.
- The results provided by the existing systems can be improved by using better implementation technique.
- Some systems have only presented detection mechanisms. Prevention mechanism are yet to be developed.

2.3 Problem Statement

Any attack launched on a any computer or mobile network with the intention of gaining unauthorized access for ulterior motives like stealing data or money, stalking, breaching privacy etc. can be labelled as a cyber-attack. Such attacks can be launched using viruses, worms, backdoor traps, malwares, etc. Such attacks can threaten and cause serious damage to everyone at personal, corporate, social, political, or national levels. Hackers can steal vital data and sell it to rivals on dark web.

Mobile devices being portable, multi-functional etc, the number of cell phone users are ever increasing. Thus there is a great need of mobile security. Taking into consideration these points, a system will be developed to make it easier for the users to detect different threats and efficiently avoid those risks.

2.4 Scope

Cyber-attack can prove to be threatening in more than one way. Cyber crimes can be performed against individual, society or organization. Such type of attacks can result in data theft, monetary loss and affect the mental health of an individual. The developed system will play a vital role in reducing the effects of such attacks on individual and society by mitigating the threats.

Table 2.1: Summary of Literature Survey

NAME	DEFINITION	SOURCE	TARGET	LAYER
IMSI ATTACK	IMSI-catcher is a device used for intercepting packets and tracking location data of mobile phone users	Fake Base station	User mobile phone	Network layer
USER IDENTITY THEFT ATTACK	Identity theft is a crime in which masquerader steals the information from a legitimate user and use that information to impersonate someone.	Hacking Data,Stealing devices	Children, Mega social media users, High-income earners, The elderly.	Network and Physical layer
Smishing Attack	In Smishing, the attacker sends text messages containing malicious links, phone numbers, etc. to the victim and the attacker aims to steal sensitive user data such as bank account details,user credentials, etc through this message.	Incoming text messages	User mobile phone	Network layer

Chapter 3

Proposal

3.1 Proposed Work

The OSI model is used as a standard 7-layer communication model incorporated within devices. There are certain vulnerabilities within each layer. It is these vulnerabilities that are exploited by the hackers.

1. **Network Layer**: It is the third layer of OSI model is Network Layer. The main functionality of Network Layer is to provide a link between nodes for data transfer. Data instead of being transferred as whole, is divided into small packets and is sent over the network.

   Issues in Network Layer:
   Multiple common routing protocols are used by Network layer to perform routing within the network. These protocols can be exploited via packet sniffing and DoS attacks such as Ping floods and ICMP attacks which can be performed remotely.

2. **Physical Layer:** Physical layer as the name suggests manages the physical connection between the nodes. It is done with the help of cables etc.

   Issues in Physical Layer:
   Since Physical layer majorly deals with hardware equipment, cabling etc, attacks suspend this service resulting in Denial of Service (DoS) attacks. It is done by damaging cable to hamper the wireless signals.

Over the years many developments have been made in order to improve the OSI model. However the base structure remains the same and hence these problems exist till date. The attackers take advantage of such factors and exploit the vulnerabilities for their own benefit.[8] Thus after scanning password files of the user, it generates a report stating the possibility of threat to the user. Based on the report generated, necessary steps will be taken as per the defined protocol in order to mitigate the risk. Before switching networks, a set protocols will run in order to ensure that the network is an authentic network and not a trap set by the attacker in order gain access to confidential data. The system will be able to uniquely identify the threat based on the methodologies used by the attacker to perform these attacks. Once identified the system will implement the algorithm to prevent them. As per the NIST standards for a cyber-security framework, the five functions included in the framework core are shown in Figure 3.1:

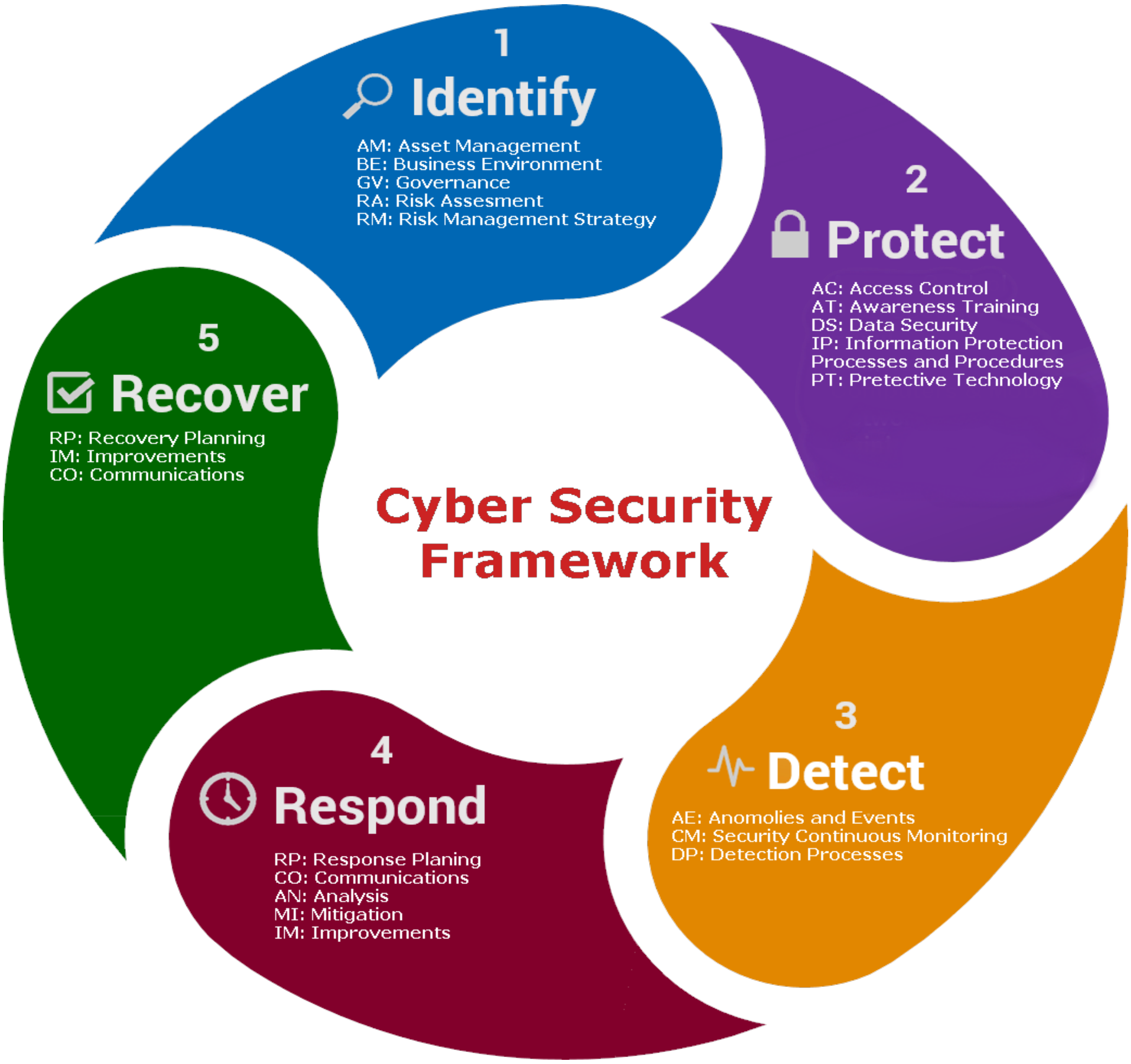


Figure 3.1: NIST Framework Architecture [18]

The five Functions are explained below:

1. **Identify**

   - Our system should identify the type of attack and what harm it will cause to our system. Identify is first function of the framework. It helps organizations to develop a better understanding on how to manage risks associated with the cyber threats.

2. **Protect**

   - The protect functions helps to reduce the impact of the cyber attack by incorporating the best security practices for data protection and overall system protection.

3. **Detect**

   - The Detect Function attempts to identify the presence of a cyber attack.
   - It performs detection at regular intervals.
   - It pre-defines a set of protocols to be followed in order to detect the presence of attack.

4. **Respond**

   - The respond function is a series of actions to performed after successful detection of a cyber attack in an attempt to successfully mitigate the attack.

5. **Recover**

   - It is set of steps to be followed in order restore any and all activities, services and capabilities damaged or affected by the attack.

3.1.1 IMSI attack:

Mobile communications cannot be trusted completely. The basic ideology behind the cellular network is that, the mobile device sends an authentication request to the nearby mobile towers providing the strongest signal for communication. IMSI stands for International Mobile Subscriber Identity. The IMSI catcher device is one of the most effective threats capable of compromising the security of communication by compromising user privacy.Each sim has a unique IMSI number comprising of Mobile Network Code (MNC), Mobile country code (MCC) and Mobile subscriber identity (MSI).

Within a geographical area, multiple mobile towers are present. Since the mobile device move within a geographical area, it keeps on reconnecting to a new mobile station which provides a better signal. An authentication protocol is followed before any mobile device is granted access to connect to a mobile tower. The device sends its IMSI number to the base station for authentication. After executing security algorithms and authenticating the device, it is granted connection to the base station.

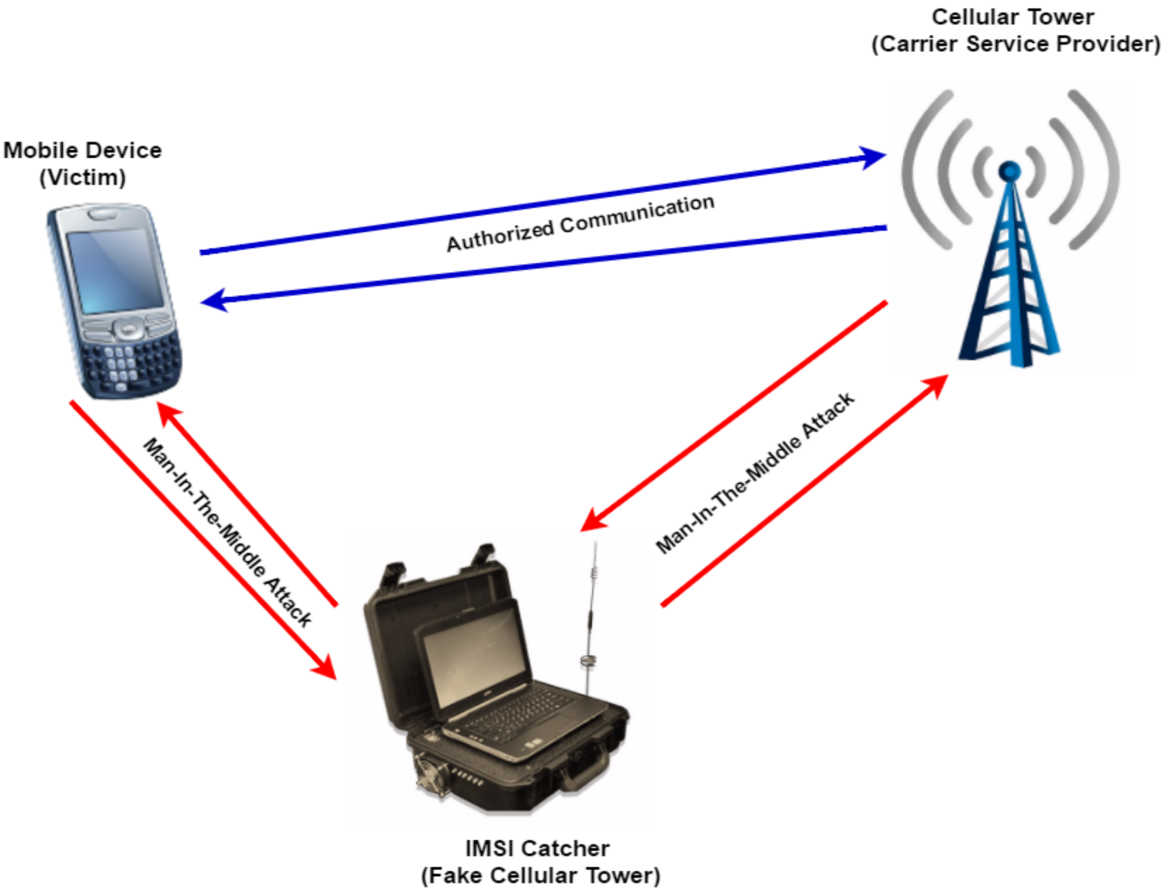


Figure 3.2: IMSI Attack [4]

In IMSI attack as shown in Figure 3.2, the attacker performs a man in the middle attack.

The attacker acts as fake base station for the mobile device. This results in the device sending its IMSI number to the attacker instead of base station. The attacker captures the IMSI number. Now the attacker acts as a fake mobile device and authenticates itself using the captured IMSI number. Once authenticated attacker performs man in the middle attack and captures each and every data packet thus proving to be a threat to user privacy. The attacker can eavesdrop on calls and record them, sniff SMS messages to redirect them, track location of the user, retrieve data from the target phone such as pictures, document etc. The proposed framework captures the current location using Loaction Id(LAC) provided by OpenCellId. It stores a database of all the present cell towers in that location. During it authentication, it checks whether the same cell towers are present or not. Based on the detected information, it further authenticates or terminates the connection request.
Algorithm 1 represents the steps for calculating the sum of cell-id's present in a specific location. The process of authentication, which consists of 2 parts:

- Verifying no intruder is using a fake cell-id to intercept the connection.
- Authentication of mobile device and base station in order to continue the process of communication.

is depicted in Algorithm 2.

Algorithm 1: CID-based cellprint generation algorithm [1]

```
Start
Read the current LAC
Find all Base Station belonging to current LAC
R = $\sum CIDs$
cellprint = R
End
```

Algorithm 2: Proposed Algorithm with AKA [1]

```
Start MS and note the location area
Read LAC and cellprint for the current location area
MS wil calculate the cell id for all Base stations and save in R
R = $\sum CIDs$
IF R = cellprint THEN # (MS compares between R and cellprint)
    MS sends its IMSI to BS
    BS sends the IMSI to authentication center AuC
    AuC generates the authentication vectors (AV) by using RAND and K$_i$
    SRES = A3(RAND, K$_i$)
    Kc = A8(RAND, K$_i$)
    AV (SRES, Kc, RAND) sends back to the BS
    BS sends the RAND to MS
    MS calculates the challenge response by using the RAND received from BS
    RES = A3(RAND, K$_i$)
    MS sends RES back to BS
    IF RES = SRES THEN # RES received from MS and SRES received from AuC
        connection successful
    ELSE
        connection failure
    ENDIF
ELSE
    Intruder id (IID) = R – cellprint
    Reject Connection Request
ENDIF
```

IMSI Flowchart

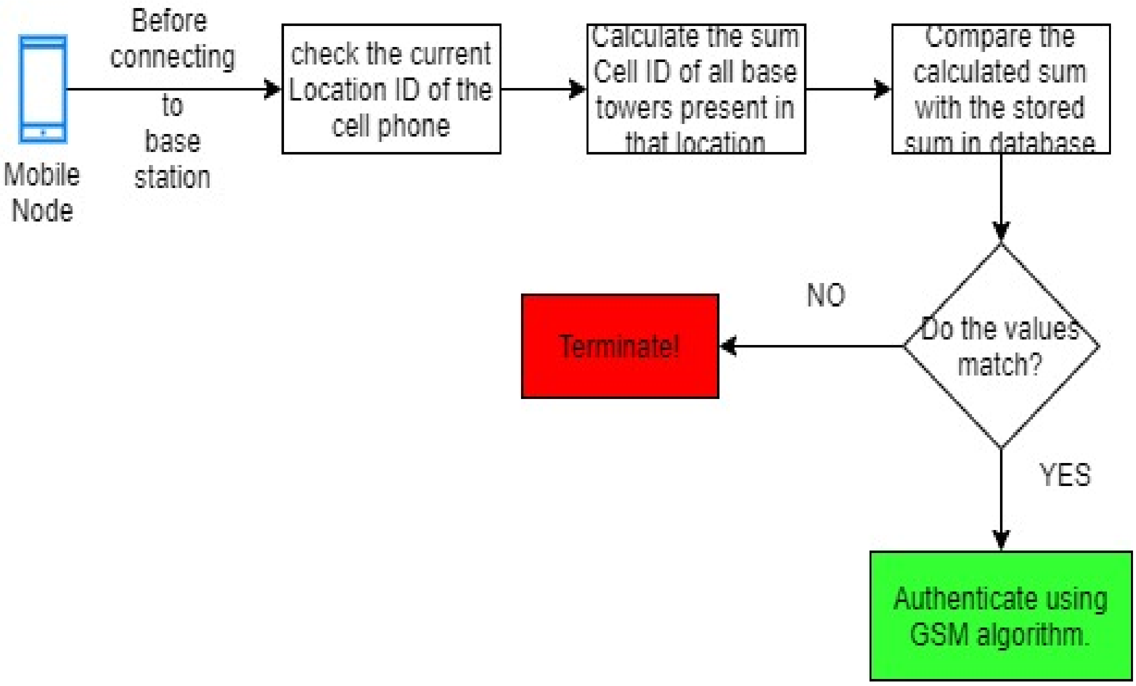

Before connecting to base station
Mobile Node
check the current Location ID of the cell phone
Calculate the sum Cell ID of all base towers present in that location
Compare the calculated sum with the stored sum in database
Do the values match?
NO
Terminate!
YES
Authenticate using GSM algorithm.

Figure 3.3: IMSI Flowchart

Each mobile device needs to connect to a base station in order provide cellular network to the user. Therefore the proposed algorithm as shown in Figure 3.3 performs a series of steps before the connection is granted in order to eliminate the threat of IMSI attack.

An open cellular data set from OpenCelliD was used which contained all the necessary Information such as Location Id(LAC), Mobile Country Code(MCC), Mobile Network Code(MNC), Cell-Tower Id(CID), type of network etc. Based on this data set a new data set was derived consisting of the LAC and sum of CID for that particular LAC. Each time a device request a connection to a base station:

- The current location of the device is captured.
- The sum of CID for that LAC is calculated.
- The value is matched with sum stored in the database.

Only if the values match, GSM algorithms(A3,A5,A8) algorithms are performed for further authentication and privacy. If the values do not match, the request to connect to a new base station is terminated based on the possibility of an attacker trying to perform IMSI attack.

A3 Algorithm

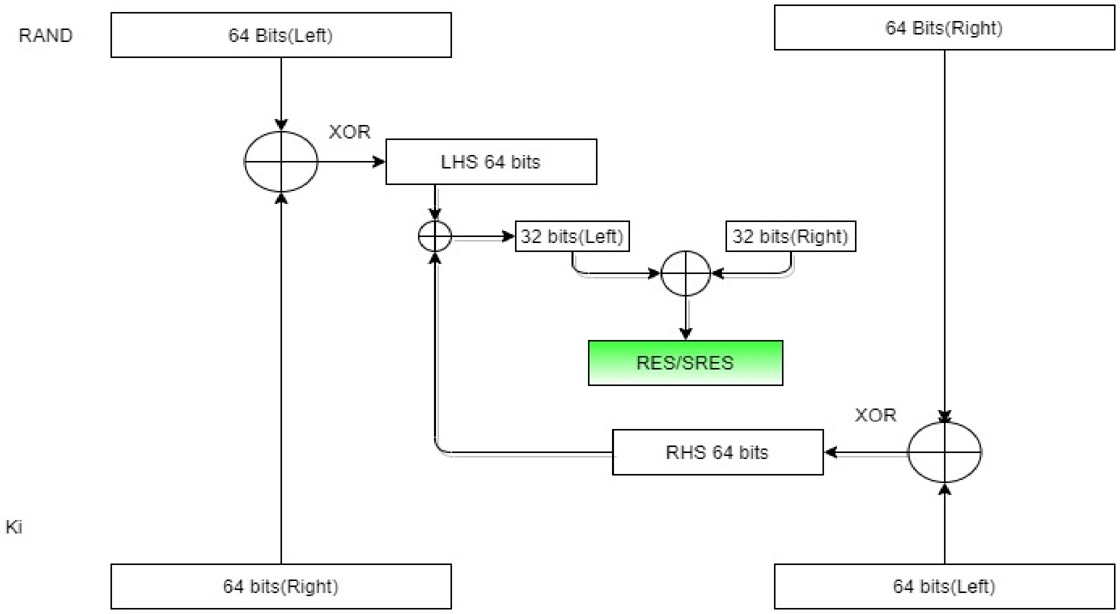


Figure 3.4: A3 Algorithm [2]

Authentication Key (K_i) and Random Number(RAND) are two 128 bit numbers used in GSM security algorithm used for authentication and encryption purposes. RAND is generated by the base station and provided to the mobile device. K_i is generated by the sim card. The proposed algorithm shown in Figure 3.4 suggests:

- Split both RAND and K_i into two halves of 64 bits each.
- XOR left 64 bits of RAND with right 64 bits of K_i.
- XOR right 64 bits of RAND with left 64 bits of K_i
- XOR both the results generated and split the 64 bit generated result into two halves of 32 bits.
- XOR the two 32 bits to generate RES/SRES
- RES is generated at the mobile device and SRES is generated at the base station.
- Only if both the values match , further authentication is carried out.

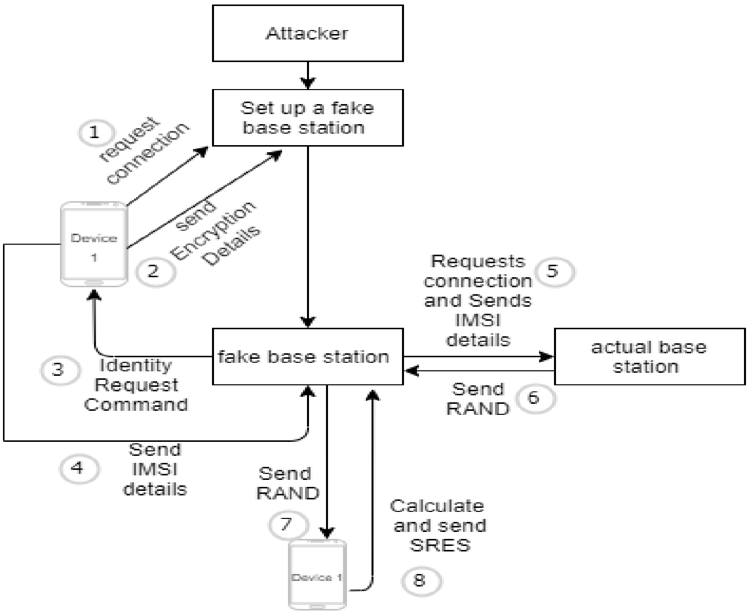


Figure 3.5: IMSI Attack graph

As shown in Fig 3.5, after the Fake Base Station receives the SRES from the mobile, it need not verify the SRES. It authenticates the connection to the mobile. From this point, the attacker can then intercept all the data coming to and from the mobile by assigning a TMSI to the mobile.

3.1.2 Identity Theft Attack:

Identity theft is a misdeed in which masquerader steals the information from a legitimate user and use that information to impersonate someone. Your passwords are stored in system using special algorithms known as hashing , Hackers try to access that file from system. Most of the time password is kept for important files. Leading to the immediate impact of certain significant losses such as cash and debt securities, individual victims of identity theft can have significant intangible costs, including damaging their reputation and credit report, which could prevent them from getting a loan or even getting a job. Depending on the circumstances, identity theft can take years to recover.

The businesses of their employees have become an enemy to steal their identity and face significant costs associated with credibility and trust. Once the business's reputation has been damaged, it is necessary for that business to be able to use it through additional security measures to assure customers that it will not do so again. It's necessary to know whether the password has been hacked or not.

Figure 3.6: Identity Theft Attack [20]

A cyber-attack is a deliberate attempt to exploit hardware/software vulnerabilities and to capture, store, alter, misuse private data for personal gains. It is done with the help malicious scripts, alter the software code or introduce backdoor traps, worms, viruses in system and retrieve sensitive information which may lead to identity theft. Users also use mobile as a means of online transaction, storing financial information and passwords which makes it important to secure the network over which communication is done.

System can be used to detect such attacks and thus mitigate such threats and secure user data.We have developed a hacked password checker(an example shown in Fig 3.6), which will work as an alert system for users. It will check whether the password has been hacked or not. It examines the password and if that password matches the list of passwords that have leaked then we will report to the user.

Identity Theft Flowchart:

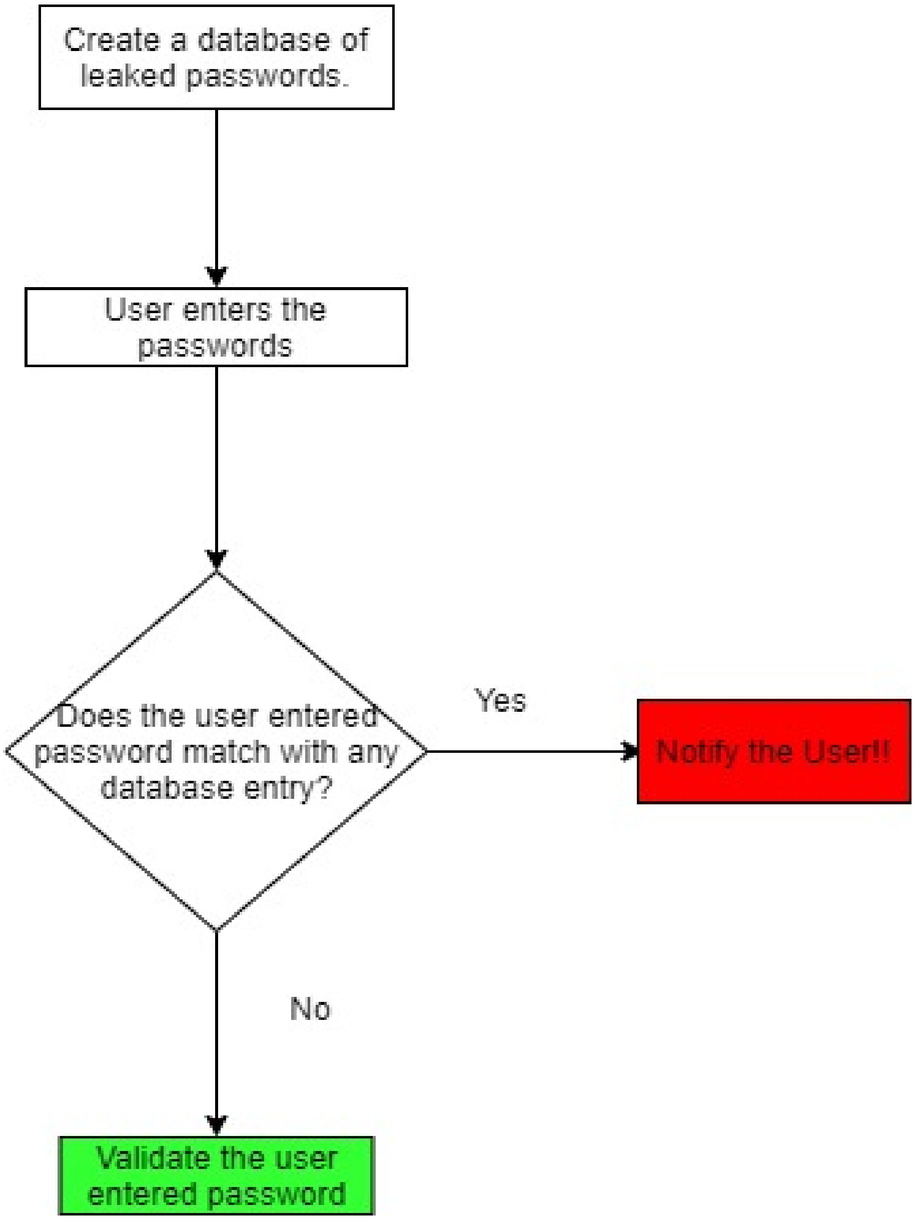

Create a database of leaked passwords.
User enters the passwords
Does the user entered password match with any database entry?
Yes
Notify the User!!
No
Validate the user entered password

Figure 3.7: Identity Theft Flowchart

Figure 3.7 explains the flow of prevention of id theft attacks. The database of the leaked passwords is created. The user enters his/her password. The system will check if that password matches the user entered password if that password matches then the user will be notified or else the user's password is not leaked.

Identity Theft Attack Graph:

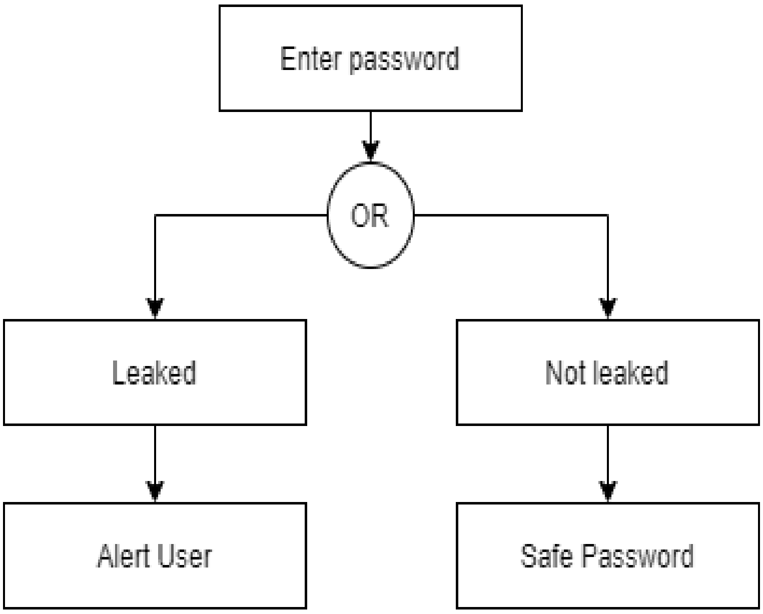


Figure 3.8: Identity Theft Attack Graph

Procedure : We have developed a hacked password checker, which will work as an alert system for users. It will check whether the passwords have been hacked or not. It examines the password and if that password matches the list of passwords that have been leaked, it will report to the user.

1. A leak data set is being created. A new associated set of data is created when each leaked password is converted to hash value.That hash value is stored.
2. The system takes an input from the user. The user enters his/her password. User entered password is converted into hash value.
3. If that hash value matches with any one of the values from the file, that password is leaked. The system will notify the user that his/her password has been previously hacked.
4. If the hash value does not match then the user-entered password is not leaked and can be safely used.

3.1.3 Smishing Attack:

Phishing was introduced from the word fishing which indicates the activity of catching fish using a bait. The word fishing is combined with the term phreak, which refers to hacking of computer systems to obtain free calls, and transformed into phishing to simply indicate hacking by phishing.

Phishing is a type of cyber-attack whose objective is to steal people's confidential information using a bait. The term smishing is, on the other hand, is obtained by combining the two words, SMS and Phishing and was used by David Rayhawk of McAfee for the first time on August 25, 2006. Attackers intend to steal the personal information of the victims using the content they send in SMS messages rather than other media.

Figure 3.9: Smishing Attack [11]

Smishing (a depiction shown in Fig 3.9) is an attack where the target is mobile devices, in which the attacker sends text messages containing malicious links, phone numbers or E-Mail IDs to the victim and the attacker aims to steal sensitive user data like bank account details, passwords, user credentials, credit card details, etc through this message. Through this message, the attacker prompts the user to click on the link or contact the phone number or E-mail ID provided in the SMS. Victims can also be tricked into downloading malicious apps that can be used to seize messages or quietly collect personal data. An overview of Smishing attack is shown in Fig 3.10.

Flowchart for Smishing Attack:

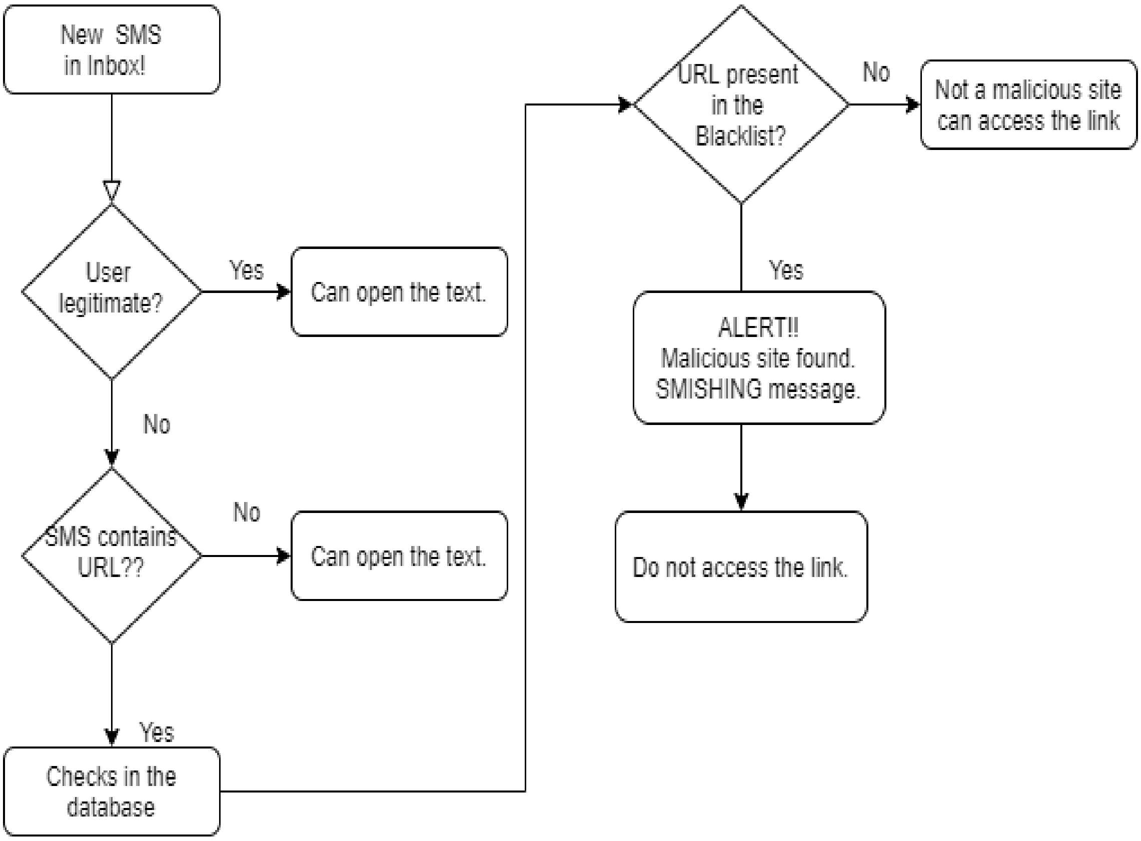


Figure 3.10: Flowchart for Smishing Attack

Some of the techniques used by researchers to prevent smishing are briefly discussed below.

1. **Content-Based Filtering:**

    - The content present in the message is assessed for suspected URLs, E-Mail IDs, Phone Numbers and Keywords using this technique. Smishing content filtering involves examining the text present in the SMS.
    - The smishing message is categorized based on the contents which are present in the text.
    - Sometimes content-based filtering is performed based on some set of rules which is also called as rule-based classification.

2. **Blacklisting:**

    - In this approach, a list of suspicious IP addresses and URLs are maintained by trusted sources which are then used to identify fraud websites.
    - This approach is used by various browsers that communicate with trusted servers to obtain a list of blacklisted URLs.
    - As blacklisting cannot detect zero-day phishing attacks,blacklists need to be frequently updated.
    - The approach works well as long as the list is regularly updated and this method yields low false-positive results.

3. **Whitelisting:**

    - In this approach, a list of authorized URLs are stored which is named as whitelist and this can be used for identification of genuine websites. As,checking genuine websites for malicious features is avoided by using this method,it reduces the complexity of the classification method.
    - Whitelisting alone cannot be used for smishing detection because it does not identify the malicious features of the URLs.

4. **Heuristic methods:**

    - Some researchers have used a heuristic-based approach in which features are discriminated from both the legitimate SMS and smishing.
    - SMS are extracted and a training dataset is build based on the extraction. The most malicious features identified based on the classification will be finally used for smishing message classification.
    - When the users receive a new SMS, then the machine learning algorithm predicts the message based on the learning from the training dataset.
    - This method gives high accuracy but if the maliciousness check is not conducted on the URLs then can it lead to false-positive results.

5. **URL Based methods:**

    - In this approach, the URL present in the text message is further analyzed to inspect the behavior of the URL.
    - The malicious behavior of the URL is inspected and the messages containing malicious URL is categorized as smishing message

Some of the ways you can use to prevent or avoid Smishing attacks:

- Avoid clicking on any UNKNOWN messages with links.Moreover,first think about who has sent you the message. Do you know the sender?
- Check for spelling errors and grammar mistakes.
- If you have received any messages in regards to your business assets or the partnerships that you have with them and/or the bank that is associate with them, call the respected organisation to see if it is a legal request before responding.
- Visiting the sender's website itself rather than providing information in the message is always a better option.
- Never provide financial or payment information on anything other than the trusted website itself.
- If the message starts with " Dear user, congratulations, you have won...." It is a clear sign for Smishing and it is known to us that in reality nothing in life is free".
- Don't click on links from unknown senders or those you do not trust.
- Be wary of "act fast," "sign up now," or other pushy and too-good-to-be-true offers.
- Install a mobile-compatible antivirus on your smart devices.
- Check for messages that contain the number "5000" or any number that is not a phone number. This is a strategy where scammers have masked their identity so their location and identity are not traceable.

In this proposed system, a Smishing detection model is developed which uses a combination of content-based approach and URL based detection to identify smishing messages. The system analyzes the presence of URL, Phone Number, E-mail ID and malicious keywords present in the message by assessing the contents of the message. The system also uses a machine-learning algorithm to investigate the smishing keywords present in the message.The below Figure 3.11 represents the system architecture of the proposed approach.

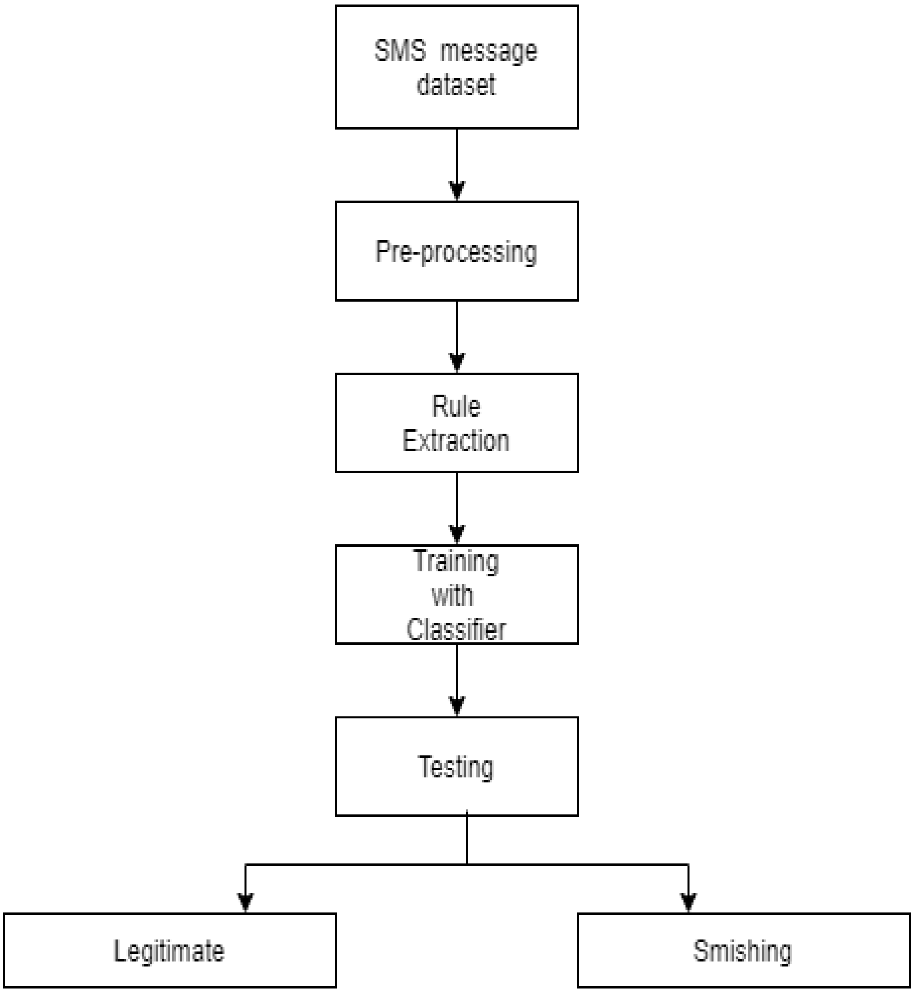


Figure 3.11: Process for Smishing Attack

- **Random forest algorithm :**
  Random forest is a supervised learning algorithm which is used for both classification as well as regression. But most of the times, it is used for classification problems. As we know that a forest is made up of trees and more trees means more dense forest. Similarly, random forest algorithm creates decision trees on data samples and then collects the prediction from each of them and finally selects the best solution by means of voting. It is a grouped method which is better than a single decision tree because it reduces the overfitting by averaging the result.The random forest is quick to train and highly accurate.

- **Feature Extraction:**
  Feature extraction is the process of transforming original data to a data set where the number of variables is reduced, which contains the most discriminatory information. This will reduce the data dimensionality, remove redundant or irrelevant information, and transform it to a form more appropriate for classification.

- **The Machine Learning Modeling process:**
  The process of modeling means training a machine learning algorithm to predict the labels from the features, tuning it for the business need, and validating it on holdout data.The output from modeling is a trained model that can be used for inference, making predictions on new data points

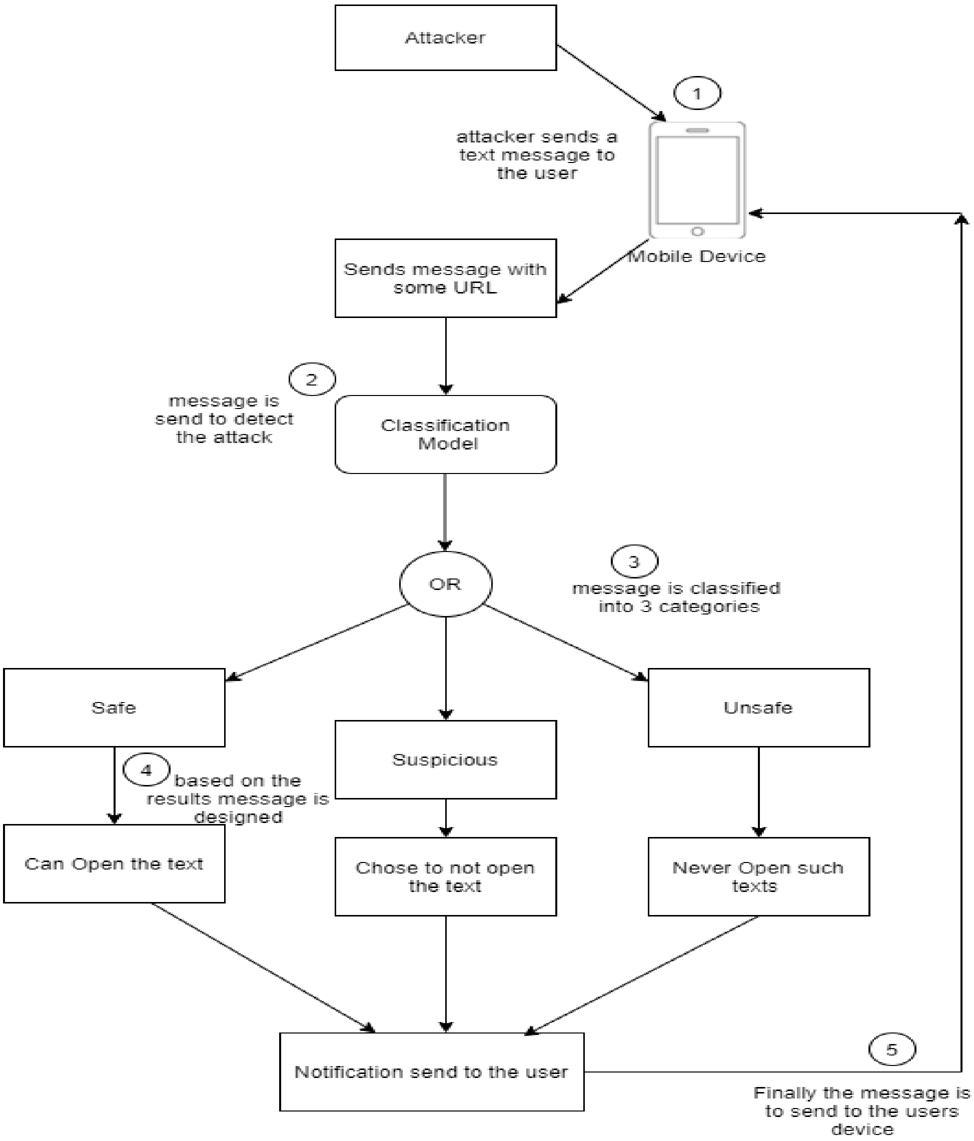

Attacker
1
attacker sends a text message to the user
Mobile Device
Sends message with some URL
2
message is send to detect the attack
Classification Model
OR
3
message is classified into 3 categories
Safe
Suspicious
Unsafe
4
based on the results message is designed
Can Open the text
Chose to not open the text
Never Open such texts
Notification send to the user
5
Finally the message is to send to the users device

Figure 3.12: Attack Graph for Smishing

3.2 Proposed Methodology

[9] NIST Cybersecurity Framework proposed by the US National Institute of Standards and Technology provides guidance for organizations in the United States for improving the ability of detection, prevention and responding to various cyberattacks.
Cyberattacks can have disatrous effects at a personal, organizational or national level. They can cause electric blackout, national security breach. Sensitive data such as medical records can be stolen. Entire cellular and computer system can be disrupted. As per the White House's Office of Management and Budget revealed that, of 96 federal agencies it assessed, 74 percent were either "At Risk" or "High Risk" for cyber attacks. [10]
Thus, we have developed a framework as per the NIST guidelines.

1. **IDENTIFY:**

   Our system should identify the type of attack and what harm it will cause to our system.

   - **Identity Theft Attack:** Passwords are stored in system using special algorithms known as hashing , Hackers try to access that file from system. Most of the time password is kept for important files. Leading to the immediate impact of certain significant losses such as cash and debt securities, individual victims of identity theft can have significant intangible costs, including damaging their reputation and credit report, which could prevent them from getting a loan or even getting a job. Depending on the circumstances, identity theft can take years to recover. Businesses of their employees have been victims of identity theft and face significant costs associated with credibility and trust. Once the business's reputation has been damaged, it is necessary for that business to be able to use it through additional security measures to assure customers that it will not do so again. It's necessary to know whether the password has been hacked or not.
   - **IMSI Attack:** An IMSI attack consists of recording the IMSI number of the targeted device. It acts as a fake base station with which the targeted device tries to establish an communication. The targeted device sends the encrypted information to the IMSI catcher. Once this is done, the IMSI catcher acts as mobile device and tries to authenticate itself and establish a connection with the base station using the encrypted information. After the authentication is complete, the IMSI catcher can sniff all the information and communication occurring between the targeted device and other devices. Each time the system tries to hop and connect to a new Base Station, a series of steps will be performed in order to detect if there is any possibility or threat of IMSI attack.

- **Smishing Attack:** Smishing is a cyber-security attack,in which Short Message Service (SMS) are used to steal personal credentials of mobile users. The trust level of users on their smart devices has attracted attackers for performing various mobile security attacks like Smishing.Users incline to click on the URLs in the SMS without realizing the possibility of smishing attacks.Also, users are in a habit of downloading and installing applications without thinking of the possible malicious behavior of the installed applications.The system will verify whether the link in the malicious or safe.

2. **PROTECT:** Necessary protocols and important safeguard measures to be followed in order to ensure the delivery of infrastructure services is outlined by the Protect function.

   - **Identity Theft attack:** Our developed system will work as an alert system for users. It will check whether the password has been hacked or not. It examines the password and if that password matches the list of passwords that have leaked then we will report to the user.
   - **IMSI Attack:** In order to protect against IMSI attack, a database of all existing base stations and their details is stored. Before authenticating the connection to the base station, the algorithm compares the required values of the base station with the values in the data-set. It authenticates the connection only if both the values match.
   - **Smishing Attack:** Extensive use of smartphones and increase in dependency of users on smartphone applications for performing various tasks are few reasons that led attackers to shift their focus to mobile devices.Also, the user believes that with two-factor authentication method, only trusted messages will be delivered to their devices.It was estimated that out of 10, seven people do not take any action against unwanted messages.The phishers will not only get money but also acquire information about contact numbers, mobile device versions, photos, etc.Hence protecting our system from such attacks is necessary.The proposed approach is a security model that protects the user from Smishing messages by blocking this messages and delivering only the normal messages to the user.

3. **DETECT:**

   How will our system sense that the attack has occurred. Guidelines to be followed for detection of occurrence of cybersecurity event is defined by Detect function.

   - **Identity Theft attack:** The user has to enter the password, that password will be converted to a hash value. This hash value will be checked whether it matches one of the values from the list of the leaked password. If both the values are similar then the password is leaked password. The system will generate an alert message.
   - **IMSI attack:** The proposed system captures the current location details of the mobile device. Based on the current location details, it extracts the sum of cell id's of base station in that location from the database. At the time of negotiating a connection with the base station, it first compares the sum of cell id's stored in the database and the sum of cell id's of base station currently present in that location. If the values do not match, it raises a suspicion of IMSI attack and prohibits any further authentication with the base station.
   - **Smishing Attack:** Multiple reports evidently indicate that Smishing attacks have exponentially increased over the last few years. Once the user opens the suspicious message which contains the link,the system will check whether the link is malicious by using certain algorithms.The proposed system uses Machine Learning and features are extracted and computed from the SMS to take the appropriate decision. The advantage of using the machine learning based technique is that it can detect the fake message coming from any source.The Smishing detection is a type of binary classification problem where a message can be divided into two categories i.e. Smishing and Legitimate.The smishing message is a harmful spam message that steals personal information.

4. **RESPOND:**

   All the necessary measures in order to contain the impact of the cybersecurity attack is defined by the Response function. **Identity Theft attack:** The system will alert the user about the password leakage. When the user gives its password as input, the system will check in its database if that password is present or not. If that input password is present then the password is hacked.

   IMSI attack: If IMSI attack is detected, it tries to find out the newly added base station by calculating the difference between the stored and newly calculated sum of cell id's. It then blocks that cell id with the possibility of it being an IMSI catcher.

   Smishing attack: The system will alert the user about the nature of the link. The alert messages will help the user to distinguish the sites and hence the user will not get trapped in the attacker's net.

5. **RECOVER:**

   All the necessary measures in order to restore all the services or capabilities affected by the attack is defined by Recover function.

   - **Identity Theft attack:** Personal information is precious so it is necessary to carry out certain steps if password has been leaked. First of all reset the password.Weak passwords or overusing the same password can have serious consequences. Your data can be compromised even if the password is strong. Password security may not completely prevent your data from being exposed, but these best practices can help minimize your risk. So set password which is strong and cannot be easily hacked.
   - **Smishing Attack:** First of all,the device should be immediately disconnected to get offline. The criminal could be in the process of installing malware on your computer. So if you have a wire connection, simply unplug the internet cable. If your device is wireless, disconnect it from the wifi network.Next, you should update your antivirus software and do a full scan of your computer to wipe out malicious viruses or malware.

3.3 System requirements

DETAILS OF HARDWARE AND SOFTWARE REQUIREMENTS:

For accessing this system,Internet connection is required and below mentioned specifications are required.

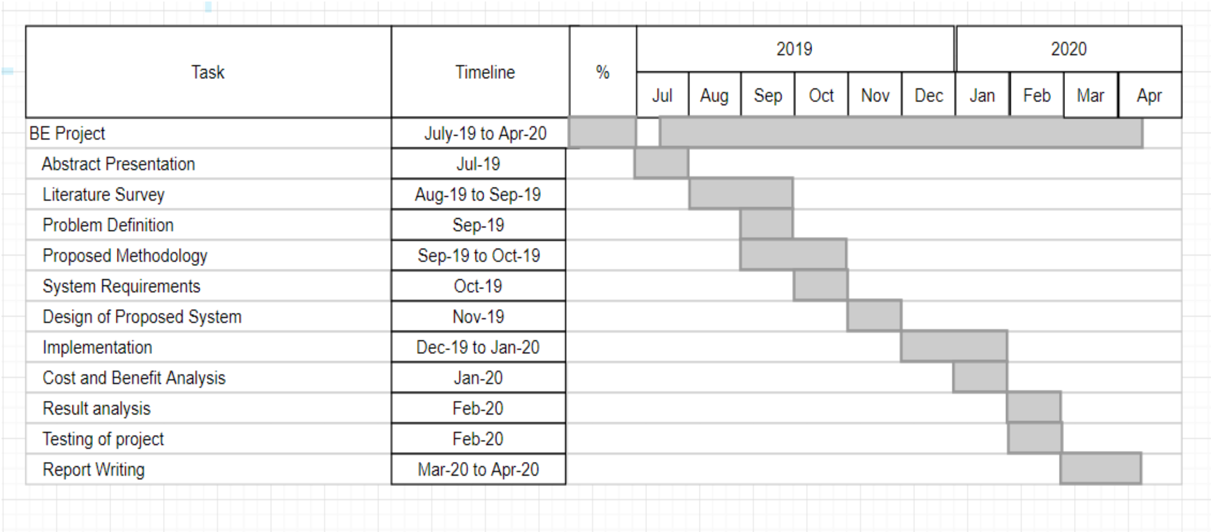
Software Required :

- Windows 10
- Python 3.6
- Hardware Required:
- i5 7^{th} Generation processor
- Graphic card: NVIDIA Getforce GTX

Chapter 4

Planning and Formulation

4.1 Schedule for Project



Task	Timeline	%	2019						2020			
			Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
BE Project	July-19 to Apr-20	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
Abstract Presentation	Jul-19		▨									
Literature Survey	Aug-19 to Sep-19			▨	▨							
Problem Definition	Sep-19				▨							
Proposed Methodology	Sep-19 to Oct-19				▨	▨						
System Requirements	Oct-19					▨						
Design of Proposed System	Nov-19						▨					
Implementation	Dec-19 to Jan-20							▨	▨			
Cost and Benefit Analysis	Jan-20								▨			
Result analysis	Feb-20									▨		
Testing of project	Feb-20									▨		
Report Writing	Mar-20 to Apr-20										▨	▨

Figure 4.1: Gantt Chart

4.2 Detail Plan of Execution

4.2.1 Requirement Analysis Phase

It is the initial phase of the project. This phase involves the collection and understanding of all the things which are necessary for the execution of the project. It involves analysis of all the software which we are using for the execution ex. Atom.

4.2.2 Analysis Phase

It involves analysis of the dataset. For analysis of information regarding Cell Towers, dataset provided by opencellid.org is used. Multiple datasets have been used which are updated regularly.
We are using Kaggle common password dataset for the detection of the leaked password.

4.2.3 Designing Phase

It involves the creation of the design of a system.A use case diagram at its simplest is a representation of a user's interaction with the system .Use case diagram explains about the use cases.

4.2.4 Coding phase

It involves actual implementation of the project.creating all the sub modules and integrating it is involved in this step. Separating the training and testing dataset is also involved in this phase.design of encoder and decoder is involved in this step.

4.2.5 Testing Phase

Testing of all the sub modules is involved in this phase.checking of whether each module is working properly or not is also involved in this phase. checking the efficiency of the project by trying various inputs for the improvement is also involved in this step.

4.2.6 Document writing or Report writing

It is the last phase of the project. It involves detailed writing of report.Writing of all the subtopics is involved in this phase.

Chapter 5

Design Of The System

5.1 Design Diagrams with Explanation

5.1.1 Use Case Diagram

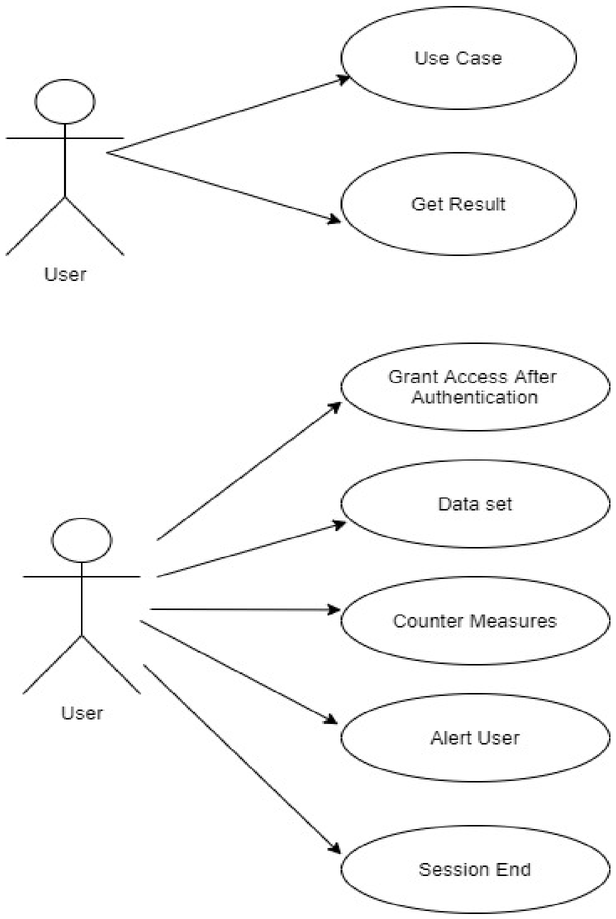


Figure 5.1: Use Case Diagram

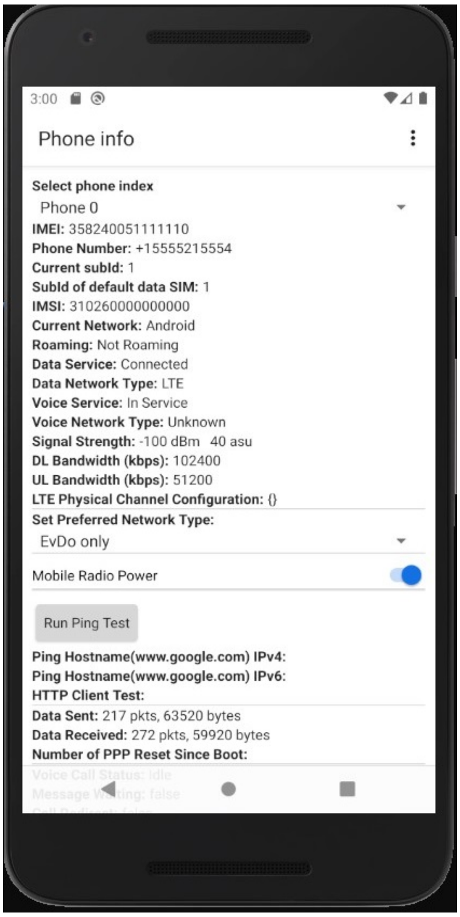
A use case diagram at its simplest is a representation of a user's interaction with the system. Use case diagram as shown in Figure 5.1 are usually referred to as behavior diagrams used to describe a set of actions (use cases) that some system or systems (subject) should or can perform in collaboration with one or more external users of the system (actors).

Chapter 6

Results And Discussion

6.1 Implementation Details

The figures below explain the flow of Imsi attack detection and GSM security authentication before the device establishes a connection with the cell tower.

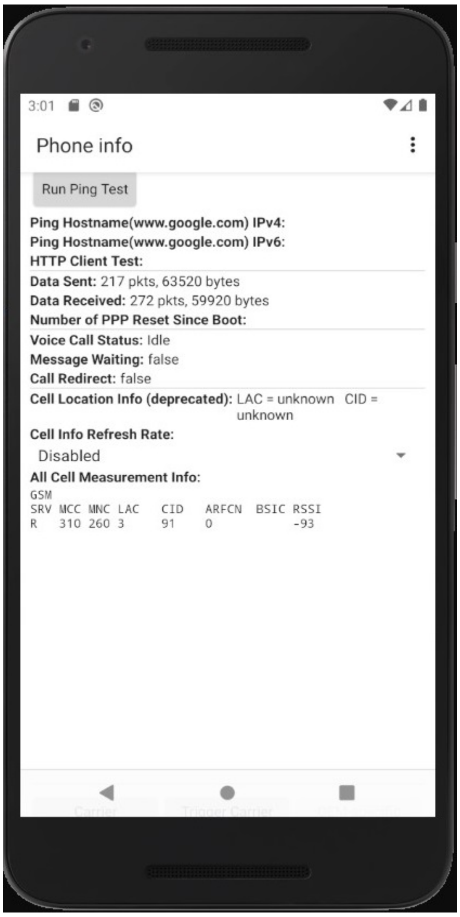


Select phone index
Phone 0
IMEI: 358240051111110
Phone Number: +15555215554
Current subId: 1
SubId of default data SIM: 1
IMSI: 310260000000000
Current Network: Android
Roaming: Not Roaming
Data Service: Connected
Data Network Type: LTE
Voice Service: In Service
Voice Network Type: Unknown
Signal Strength: -100 dBm 40 asu
DL Bandwidth (kbps): 102400
UL Bandwidth (kbps): 51200
LTE Physical Channel Configuration: {}
Set Preferred Network Type:
EvDo only
Mobile Radio Power

Ping Hostname(www.google.com) IPv4:
Ping Hostname(www.google.com) IPv6:
HTTP Client Test:
Data Sent: 217 pkts, 63520 bytes
Data Received: 272 pkts, 59920 bytes
Number of PPP Reset Since Boot:
Voice Call Status: idle
Message Waiting: false

Figure 6.1: Extracting current cell tower connection details

Fig 6.1 displays the various details such as IMEI, IMSI number etc. of a cell phone connected to a network.



Ping Hostname(www.google.com) IPv4:
Ping Hostname(www.google.com) IPv6:
HTTP Client Test:

Data Sent: 217 pkts, 63520 bytes
Data Received: 272 pkts, 59920 bytes
Number of PPP Reset Since Boot:

Voice Call Status: Idle
Message Waiting: false
Call Redirect: false

Cell Location Info (deprecated): LAC = unknown CID = unknown

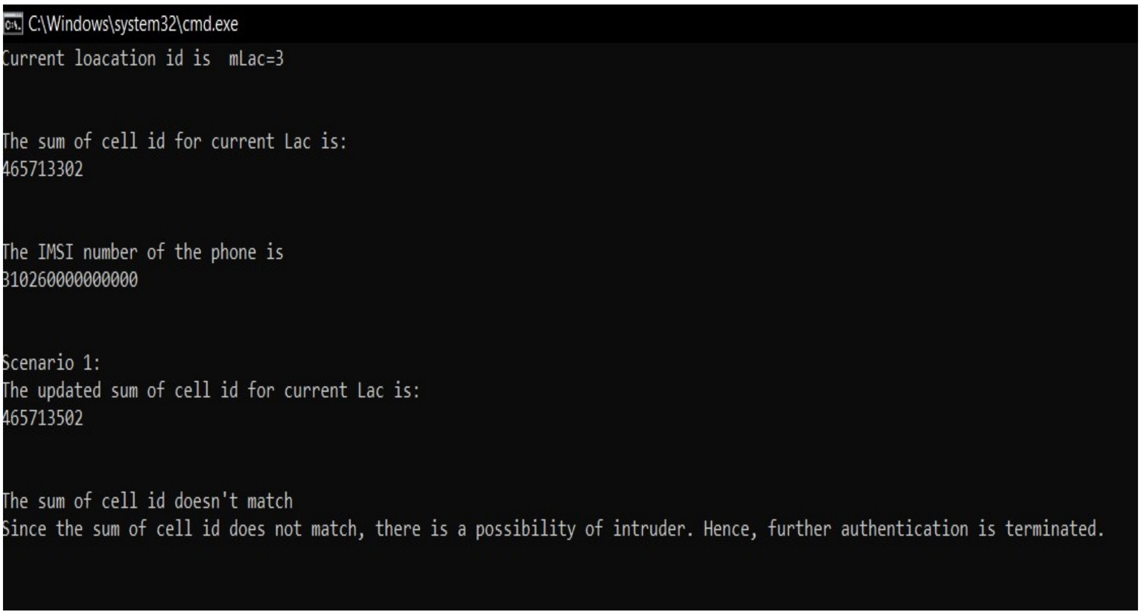
Cell Info Refresh Rate:
Disabled

All Cell Measurement Info:

```
GSM
SRV MCC MNC LAC    CID    ARFCN  BSIC RSSI
R   310 260 3      91     0           -93
```

Figure 6.2: Extracting current cell tower connection details

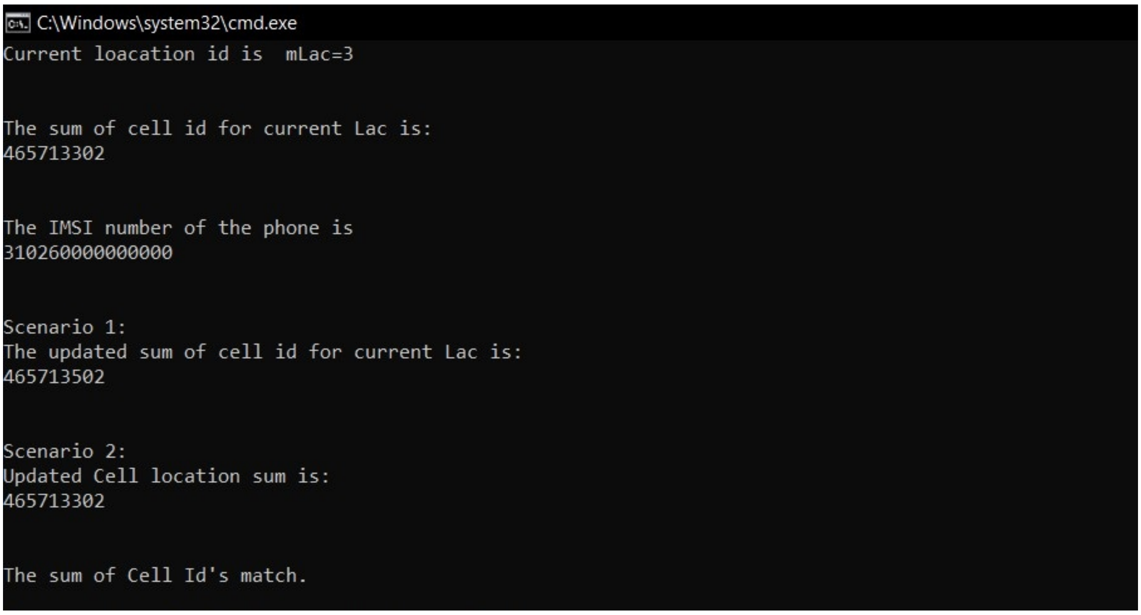
Fig 6.2 displays the details of the current cell tower to which the phone is currently connected.



```
Current loacation id is  mLac=3


The sum of cell id for current Lac is:
465713302


The IMSI number of the phone is
310260000000000


Scenario 1:
The updated sum of cell id for current Lac is:
465713502


The sum of cell id doesn't match
Since the sum of cell id does not match, there is a possibility of intruder. Hence, further authentication is terminated.
```

Figure 6.3: Identity Theft Detection Scenario 1

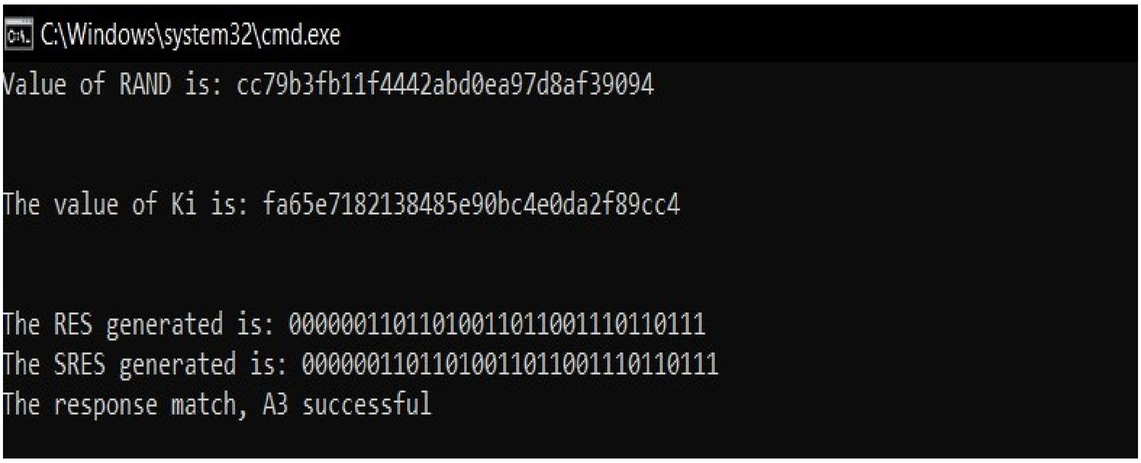
Fig 6.3 depicts the scenario where in an intruder is detected as the sum of cell-id’s do not match.



```
Current loacation id is  mLac=3


The sum of cell id for current Lac is:
465713302


The IMSI number of the phone is
310260000000000


Scenario 1:
The updated sum of cell id for current Lac is:
465713502


Scenario 2:
Updated Cell location sum is:
465713302


The sum of Cell Id's match.
```

Figure 6.4: Identity Theft Detection Scenario 2

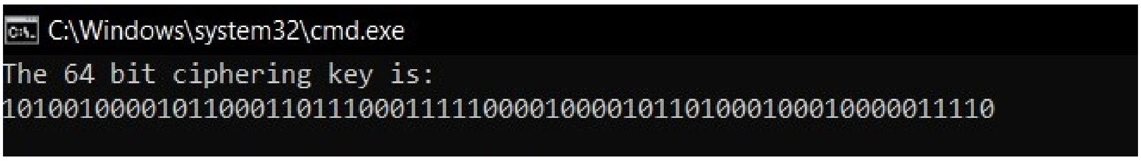
Fig 6.4 depicts the scenario wherein the sum of cell-id’s match.



```
Value of RAND is: cc79b3fb11f4442abd0ea97d8af39094


The value of Ki is: fa65e7182138485e90bc4e0da2f89cc4


The RES generated is: 00000011011010011011001110110111
The SRES generated is: 00000011011010011011001110110111
The response match, A3 successful
```

Figure 6.5: A3 algorithm

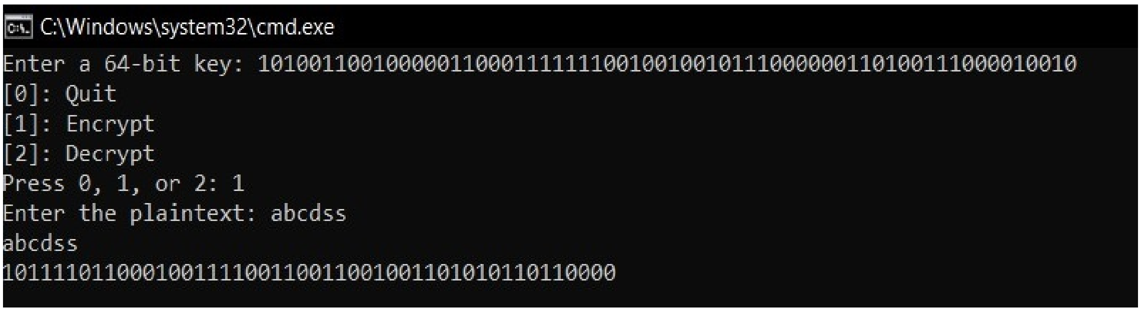
After the cell-id's match, GSM authentication is carried out. Fig 6.5 shows the A3 authentication process where SRES and RES are matched for further authentication.



```
The 64 bit ciphering key is:
1010010000101100011011100011111000010000101101000100010000011110
```

Figure 6.6: Key genreation using A8 algorithm

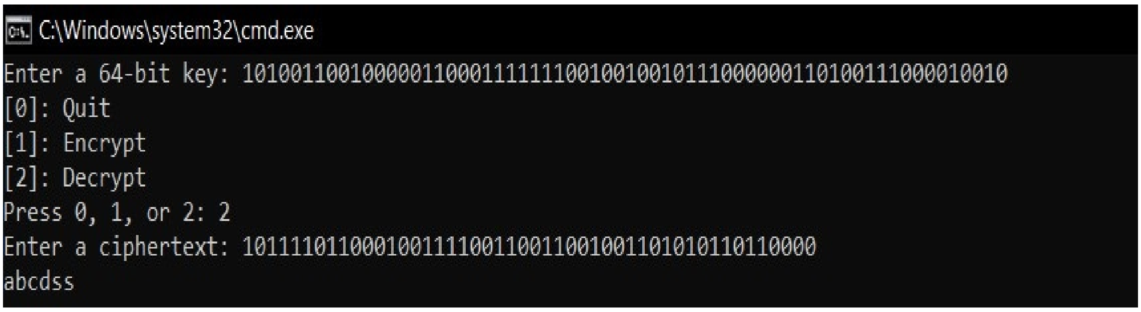
Once A3 authentication is successful, 64 bit cipher key is generated for encryption/decryption purpose using A8 algorithm as shown in Fig 6.6.



```
Enter a 64-bit key: 1010011001000001100011111110010010010111000000110100111000010010
[0]: Quit
[1]: Encrypt
[2]: Decrypt
Press 0, 1, or 2: 1
Enter the plaintext: abcdss
abcdss
101111011000100111100110011001001101010110110000
```

Figure 6.7: Encryption using A5 algorithm

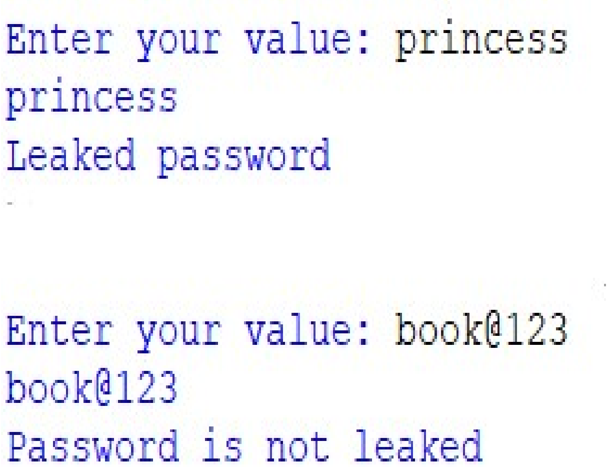
Fig 6.7 displays the encrypted result of user text based on A5 algorithm.



```
C:\Windows\system32\cmd.exe
Enter a 64-bit key: 1010011001000001100011111110010010010111000000110100111000010010
[0]: Quit
[1]: Encrypt
[2]: Decrypt
Press 0, 1, or 2: 2
Enter a ciphertext: 101111011000100111100110011001001101010110110000
abcdss
```

Figure 6.8: Decryption using A5 algorithm

Fig 6.8 displays the decrypted result of cipher text based on A5 algorithm.



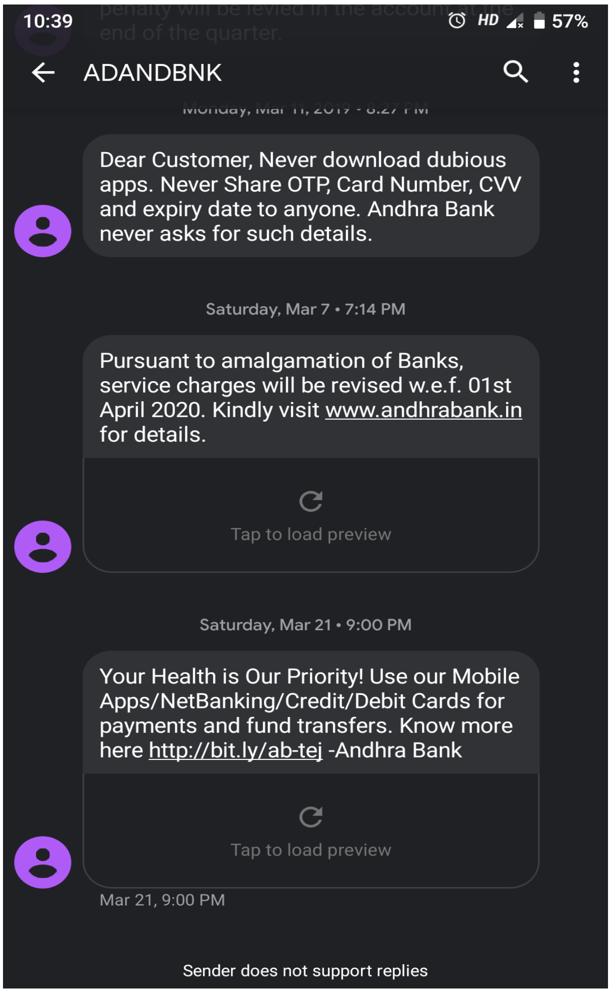
```
Enter your value: princess
princess
Leaked password


Enter your value: book@123
book@123
Password is not leaked
```

Figure 6.9: Identity Theft Detection

There are two possible scenarios:

- User password is leaked
- User password is not leaked

as shown in Fig 6.9.



penalty will be levied in the account at the
end of the quarter.

Monday, Mar 11, 2019 • 8:27 PM

Dear Customer, Never download dubious apps. Never Share OTP, Card Number, CVV and expiry date to anyone. Andhra Bank never asks for such details.

Saturday, Mar 7 • 7:14 PM

Pursuant to amalgamation of Banks, service charges will be revised w.e.f. 01st April 2020. Kindly visit www.andhrabank.in for details.

Saturday, Mar 21 • 9:00 PM

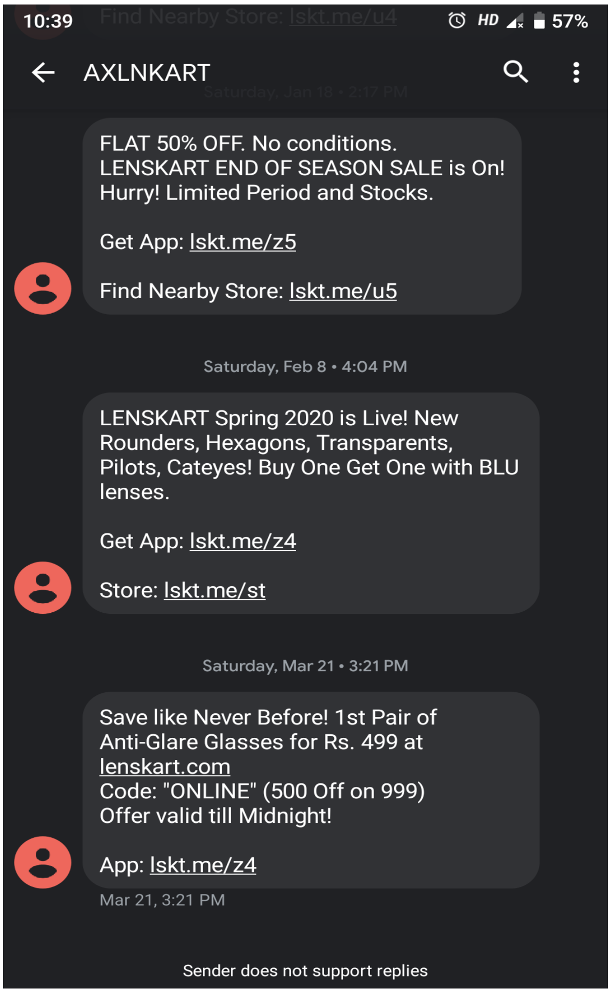
Your Health is Our Priority! Use our Mobile Apps/NetBanking/Credit/Debit Cards for payments and fund transfers. Know more here http://bit.ly/ab-tej -Andhra Bank

Mar 21, 9:00 PM

Sender does not support replies

Figure 6.10: Smishing

Figure 6.10 is an example of a typical SMiShing attack. Nowadays,most of the users are conducting banking transactions through smart phones, many SMiShing messages claim to be coming from a financial institution. Whenever the messaged is received from their bank,many of the users tend to not think twice before acting. Attackers always use legalized sounding language and even some branding to assist their pretext.



Find Nearby Store: lskt.me/u4

Saturday, Jan 18 • 2:17 PM

FLAT 50% OFF. No conditions.
LENSKART END OF SEASON SALE is On!
Hurry! Limited Period and Stocks.

Get App: lskt.me/z5

Find Nearby Store: lskt.me/u5

Saturday, Feb 8 • 4:04 PM

LENSKART Spring 2020 is Live! New Rounders, Hexagons, Transparents, Pilots, Cateyes! Buy One Get One with BLU lenses.

Get App: lskt.me/z4

Store: lskt.me/st

Saturday, Mar 21 • 3:21 PM

Save like Never Before! 1st Pair of Anti-Glare Glasses for Rs. 499 at lenskart.com
Code: "ONLINE" (500 Off on 999)
Offer valid till Midnight!

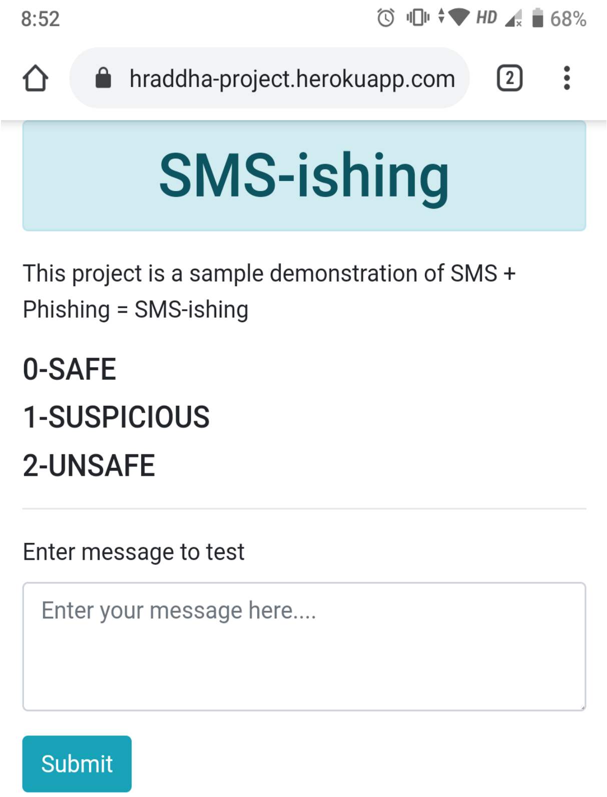
App: lskt.me/z4

Mar 21, 3:21 PM

Sender does not support replies

Figure 6.11: Smishing

Fig 6.11 shows another example of a Smishing attack.



8:52 68%

hraddha-project.herokuapp.com

SMS-ishing

This project is a sample demonstration of SMS + Phishing = SMS-ishing

0-SAFE

1-SUSPICIOUS

2-UNSAFE

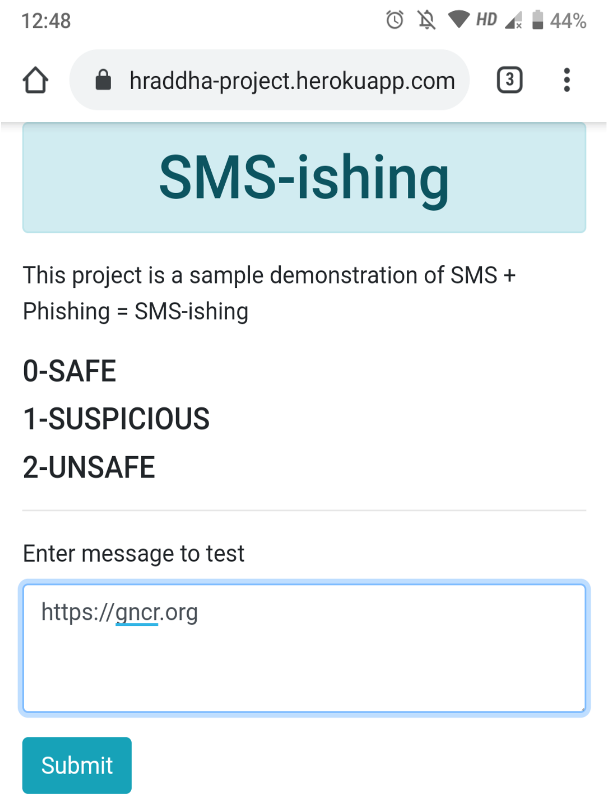
Enter message to test

Enter your message here....

Figure 6.12: Smishing

URLs of websites are separated into 3 classes as shown in Fig 6.12:

- Benign(SAFE): Safe websites with normal services
- Spam(SUSPICIOUS): Some of the webistes attempt to flood the user with advertising or sites such as fake surveys and online dating etc.
- Malware(UNSAFE): Website created by attackers to install malwares,viruses on your system,gather sensitive information, or gain access to private computer systems.



12:48

44%

hraddha-project.herokuapp.com

SMS-ishing

This project is a sample demonstration of SMS + Phishing = SMS-ishing

0-SAFE

1-SUSPICIOUS

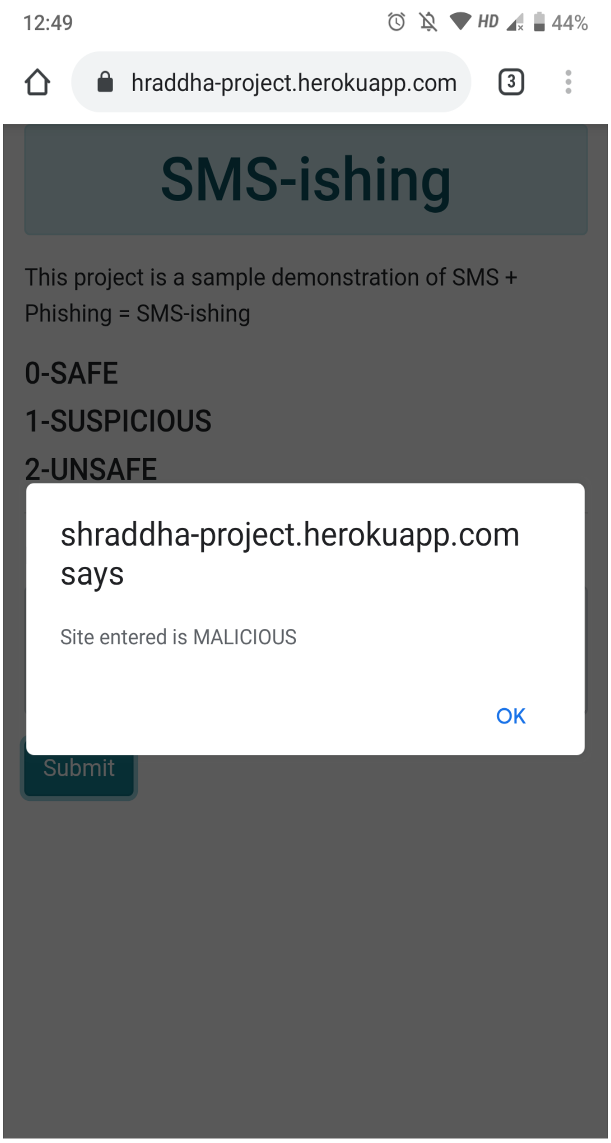
2-UNSAFE

Enter message to test

https://gncr.org

Figure 6.13: Input message

Upon getting raw text message (Fig 6.13), the link is extracted from it and it is given as input to Feature_extraction.py to extract features from that link and those features are converted into numpy array. Conversion to numpy array is must because our Machine Learning model expects the input data in numpy format. Upon conversion, we give it as input to our ML model and the predicted output is sent back to 'index.html'. We then process that JSON in JavaScript to display corresponding message



SMS-ishing

This project is a sample demonstration of SMS + Phishing = SMS-ishing

0-SAFE

1-SUSPICIOUS

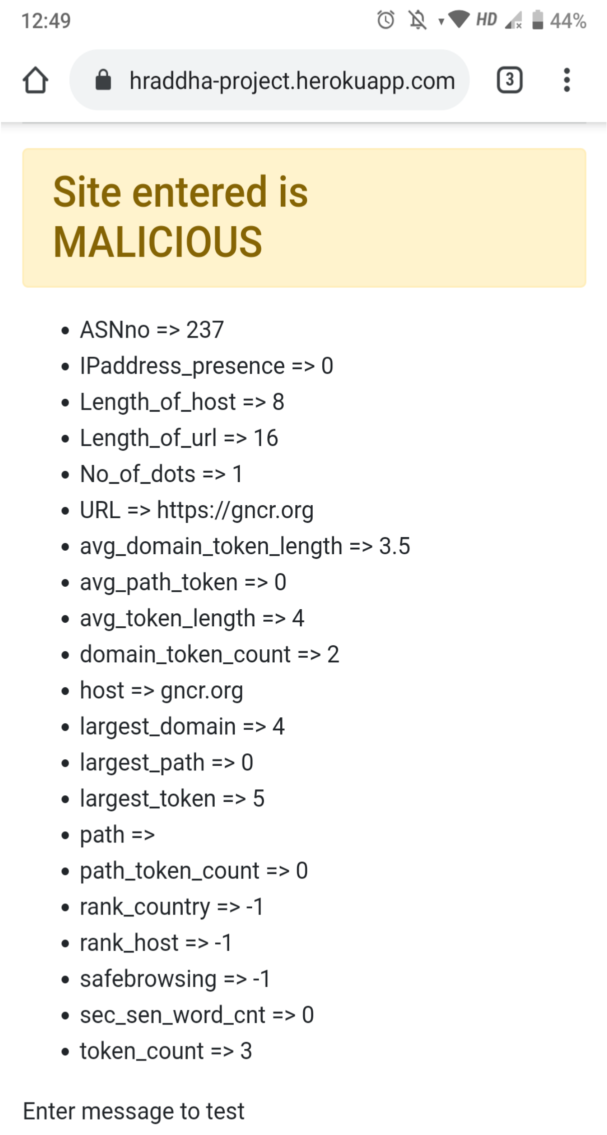
2-UNSAFE

shraddha-project.herokuapp.com says

Site entered is MALICIOUS

Figure 6.14: Notification by the system

As seen in Fig 6.14, the system then produces a pop-up message which tells us whether the site is malicious or legitimate. The system is takin user SMS as input then the system pre process the SMS input and extract relevant features from it it gives this pre processed input to pre-trained model for prediction for model we have used random forest classified with 10 trees then the prediction is sent to user on screen flask is used for this purpose.



12:49 44%

hraddha-project.herokuapp.com

Site entered is MALICIOUS

- ASNno => 237
- IPaddress_presence => 0
- Length_of_host => 8
- Length_of_url => 16
- No_of_dots => 1
- URL => https://gncr.org
- avg_domain_token_length => 3.5
- avg_path_token => 0
- avg_token_length => 4
- domain_token_count => 2
- host => gncr.org
- largest_domain => 4
- largest_path => 0
- largest_token => 5
- path =>
- path_token_count => 0
- rank_country => -1
- rank_host => -1
- safebrowsing => -1
- sec_sen_word_cnt => 0
- token_count => 3

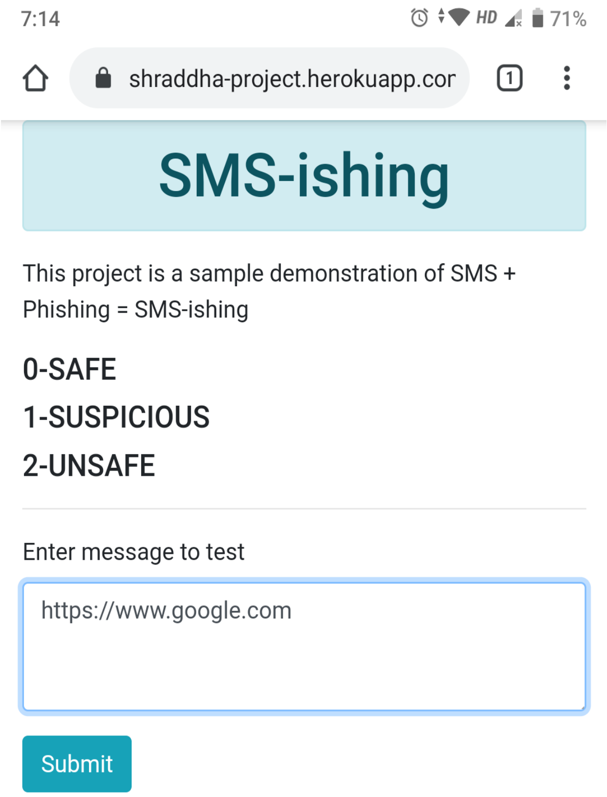
Enter message to test

Figure 6.15: Features of the site

Some of the features extracted by the system for every website we give as input are shown in Fig 6.15.

- URL Length - Long URLs are generally used by attackers to hide the doubtful parts. Hence, the length of the URL can be calculated. If the length of the URLs turn out to be more than 52 characters then the URL is considered as phishing.

- Google Index - This feature tests whether a website is stored in Google's index or not. When a site is indexed by Google, it is displayed on search results. Usually, phishing web pages are simply accessible for a short period and as a result, many phishing web pages may not be found on the Google index.



7:14 71%

shraddha-project.herokuapp.cor

SMS-ishing

This project is a sample demonstration of SMS + Phishing = SMS-ishing

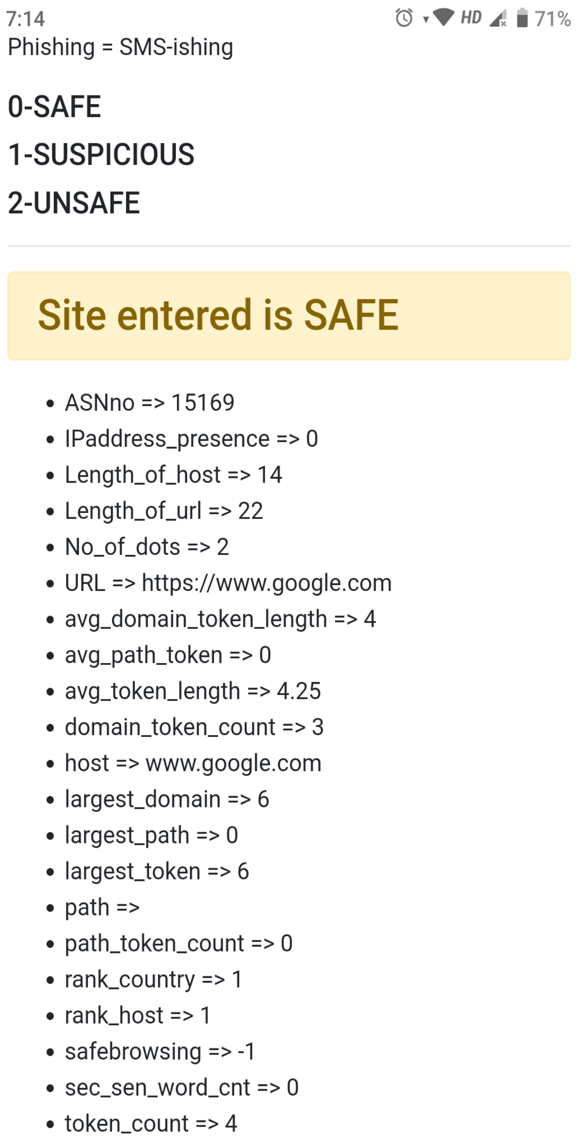
0-SAFE
1-SUSPICIOUS
2-UNSAFE

Enter message to test

https://www.google.com

Figure 6.16: Features of the site

Fig 6.16 shows another example of user input.



Phishing = SMS-ishing

0-SAFE
1-SUSPICIOUS
2-UNSAFE

Site entered is SAFE

- ASNno => 15169
- IPaddress_presence => 0
- Length_of_host => 14
- Length_of_url => 22
- No_of_dots => 2
- URL => https://www.google.com
- avg_domain_token_length => 4
- avg_path_token => 0
- avg_token_length => 4.25
- domain_token_count => 3
- host => www.google.com
- largest_domain => 6
- largest_path => 0
- largest_token => 6
- path =>
- path_token_count => 0
- rank_country => 1
- rank_host => 1
- safebrowsing => -1
- sec_sen_word_cnt => 0
- token_count => 4

Figure 6.17: Features of the site

Fig 6.17 shows that the entered site is Safe.

6.2 Result Analysis

6.2.1 IMSI Attack

In order to verify the working of IMSI detection, databases provided by OpenCellId have been used. These database contain information such as Location Id, Cell tower Id, Mobile Country Code, Latitude, Longitude, type of connection provided by the cell tower, etc. Since there was no access available to a live dataset of cell towers, two static databases have been used which contain different set of data. The system will check for the current Location Id of the phone. It compares the sum of Cell tower ID within that Location and with the help of database. It checks the Sum stored in the updated database and if the values match, GSM authentication is carried out else it suspects an intruder and suspends cell tower connection.

6.2.2 Identity theft Attack

Database contains 14,341,564 unique passwords, used in 32,603,388 accounts which has been leaked.The database of the leaked passwords is created. The user enters his/her password. The proposed system checks if that password matches password that user has entered, if that password matches then the user will be notified or else the user's password is not leaked.

6.2.3 Smishing Attack

In this work, the dataset provided by Kaggle which is a Open-source platform and one of the most popular websites amongst Data Scientists and Machine Learning Engineers. The datset is a collection of malicious and safe websites and is used to evaluate messages, determine the distinguishing features and detect smishing messages.There are total 1 Lakh entries in the dataset which are used for training the model.

6.2.4 Performance Analysis

Attack	Attack detected	Attack Not Detected
IMSI Attack	1.001 s (Fig 6.3)	10.666 s (Fig 6.4)
Identity Theft	3.8921 s (Fig 6.9)	4.2201 s (Fig 6.9)
Smishing Attack	4.1286 s (Fig 6.14)	6.2241 s (Fig 6.17)

6.2.5 Performance Metrics

IMSI Attack

- **Coverage :** The system can be used to detect the attack over various geographical regions as it contains database of cell towers of more than 150 countries [25].

- **Detection :**

  $$A = \frac{1}{n} \sum_{i=1}^{n} a_i = \frac{a_1 + a_2 + \cdots + a_n}{n} \tag{6.1}$$

  Here, a_1,a_2,a_3 etc. represent the time values to detect the attack. The mean time to detect the attack is around 4s.

- **Identification :** The system pinpoints the cell-id which is being used by the attacker.

- **Authentication :** The system will perform 2 way authentication between base station and mobile before exchanging confidential information and entire authentication process completion takes about 11s.

Identity Theft Attack

- **Mean-Time-to-Detect :** Detect and respond phase will be accurate due to the updation of the data time to time.Attack is detected in 3.8921s.

- **Time required to inform user about the password :** In our system if user want to check if password is hacked or not the system reply within few second whether password is hacked or not.

- **Measurable :** Data that has been collected is accurate and complete is will get updated from time to time.

Smishing Attack

- **Confusion Matrix :** Confusion Matrix is used in Machine Learning algorithms specifically in classification problems for performance measurement where output can be two or more classes.It is a table with 4 different combinations of predicted and actual values.

Table 6.1: Confusion Matrix

	Actually Positive (1)	**Actually Negative (0)**
Predicted Positive (1)	True Positive (TP)	False Positive(FP)
Predicted Negative(0)	False Negative (FN)	True Negatives(TN)

- **Model Accuracy:** Model accuracy in terms of classification models can be defined as the ratio of correctly classified samples to the total number of samples:

$$\begin{aligned} Accuracy &= \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \\ &= \frac{83,670}{1,00,000} \end{aligned} \tag{6.2}$$

Hence for the system,the Model Accuracy is 83.67%

6.3 Cost and Benefit Analysis

The Constructive Cost Model (COCOMO) is a software cost estimation model developed by Barry W. Boehm. The model uses a basic regression formula containing parameters derived from project history data and current project characteristics. There are three types of COCOMO model, we will be using the basic cost estimation model. Basic COCOMO includes an effort to develop software (and costs) as a function of system size. Program size is expressed in thousands of lines of code (SLOC). COCOMO operates in three stages of software projects. The phase of our project is Organic projects where

“ small ’groups with “ good’ experience deal with ”less than rigorous” requirements. The COCOMO equations that take the form as shown in below equations 6.1,6.2,6.3.

$$Effort\ Applied(E) = a_b(KLOC)_b^b[man - months] \tag{6.3}$$

$$Development\ Time(D) = c_b(Effort\ Applied)_b^d[months] \tag{6.4}$$

$$People\ required(P) = \frac{Effort\ Applied}{Development\ Time\ [count]} \tag{6.5}$$

where, KLOC is the estimated number of delivered lines (expressed in thousands) of code for project. The coefficients a_b, b_b, c_b and d_b are given in the following Table 6.3:

Software project	a_b	b_b	c_b	d_b
Organic	2.4	1.05	2.5	0.38
Semi-detached	3.0	1.12	2.5	0.35
Embedded	3.6	1.20	2.5	0.32

Constant values for COCOMO Model : Basic COCOMO is ideal for a quick estimate of software costs. However it doesn’t cost a difference on Hardware

6.4 Process model

The Iterative model : The waterfall model gets its name because of the cleaning effect from one section to another as shown in the following Figure. In this model each stage is well-defined to start with the end point, by posting the target in the next section. This model is also referred as a linear sequence model or software life cycle.

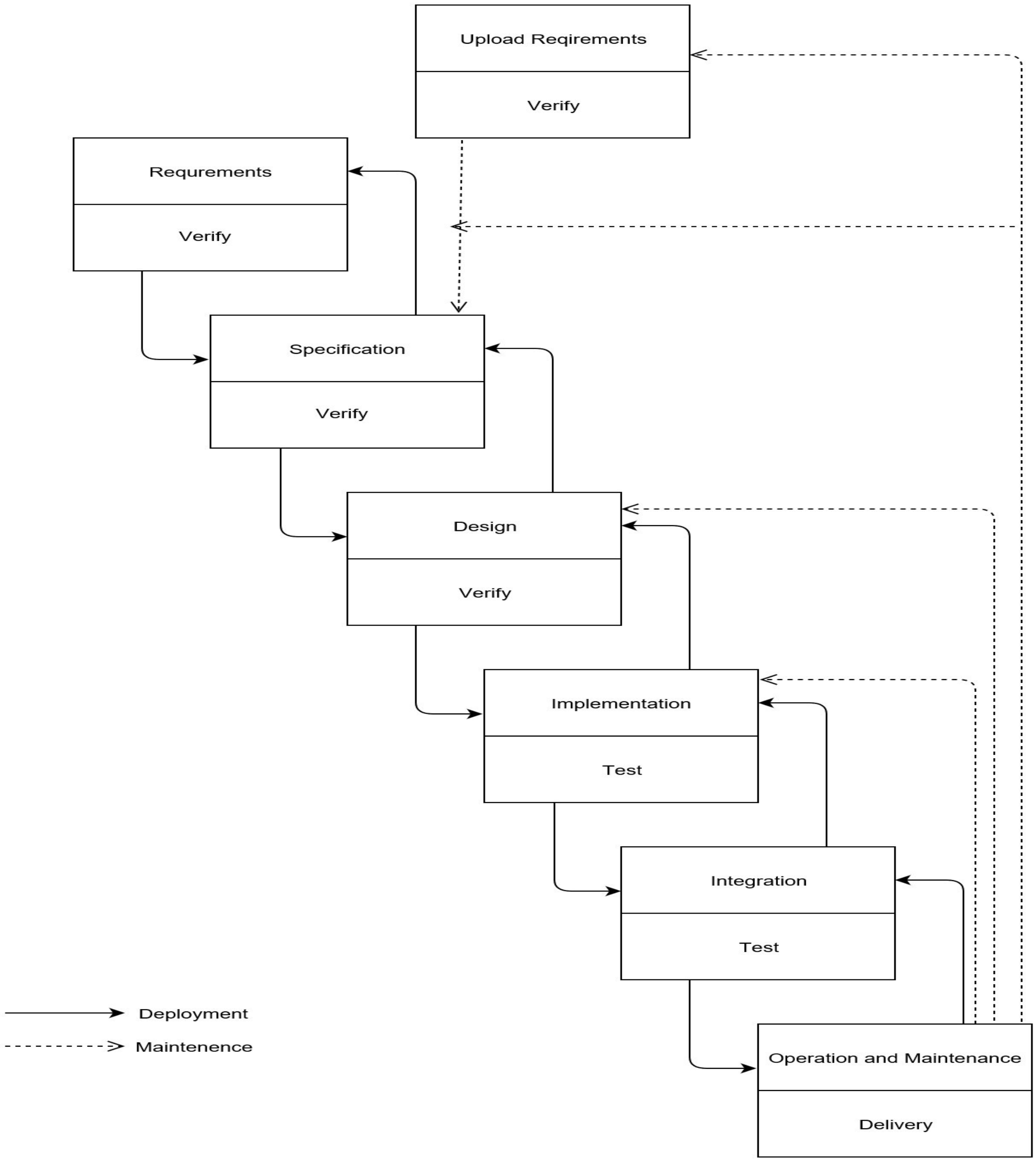


Figure 6.18: Scheme illustrating waterfall model

The model has six different stages, namely:

1. Requirements:First phase involves what is need to be designed and its function The specifications of the input and output are studied and noted.
2. System Design: System design is prepared by studying the requirement specification from the first phase. It defines overall system architecture and also hardware and software requirement.
3. Implementation: After System design is created,small programs called units are developed and afterwards they are integrated in the next phase.Testing is carried out for each developed unit.This testing is called unit testing.
4. Integration and Testing:After testing of each unit in impementation phase all the units are integrated. Integrated testing is done so that the user does not face any problem.
5. Deployment of System: Once the testing is done, the product is deployed in the customer environment.
6. Maintenance: After the product is deployed at client ends it should be maintained. Client is provided with regular maintenance and any changes required in future.

Advantages of the Iterative Model :-

- Phases of iterative model do not overlap with each other.
- Since development is linear it is easy to manage development process.

Disadvantages of the Iterative Model :-

The waterfall model is the oldest and most widely used. However, many projects do not usually follow their respective sequences. This is due to the environmental problems associated with its strict format.

- This model is used only when the requirements are very well known,clear and fixed.Not suitable for projects in which requirements are not fixed.

- System or products is available only at the end of development process.
- It is very difficult to modify system requirements in the middle of the development process.

Chapter 7

Conclusion And Future Work

Even though all people are not victims of cyber crimes,they are also at high risk.Crimes not always occur by physically visiting that place,nowadays it is done with the help of computer.As the technology is increasing criminals don't have to go outside to commit crime,nor they have to rob banks going out.They have everything they need in a single device like computer,laptops,mobile etc. They dont use guns nowadays but digital devices.

Cyber threats are a big deal. Cyberpace has grown worldwide. India has also seen a sharp rise in internet activity. A cyber attack uses malicious code to convert computer code, logic or data, resulting in unsafe data effects that can disrupt data and lead to cyber crimes, such as information and identity theft,IMSI attcks etc.

The proposed system could be potentially used on mobile networks.The interface will be capable of identifying different attacks and preventing them by following different pre-defined protocols for different attacks.

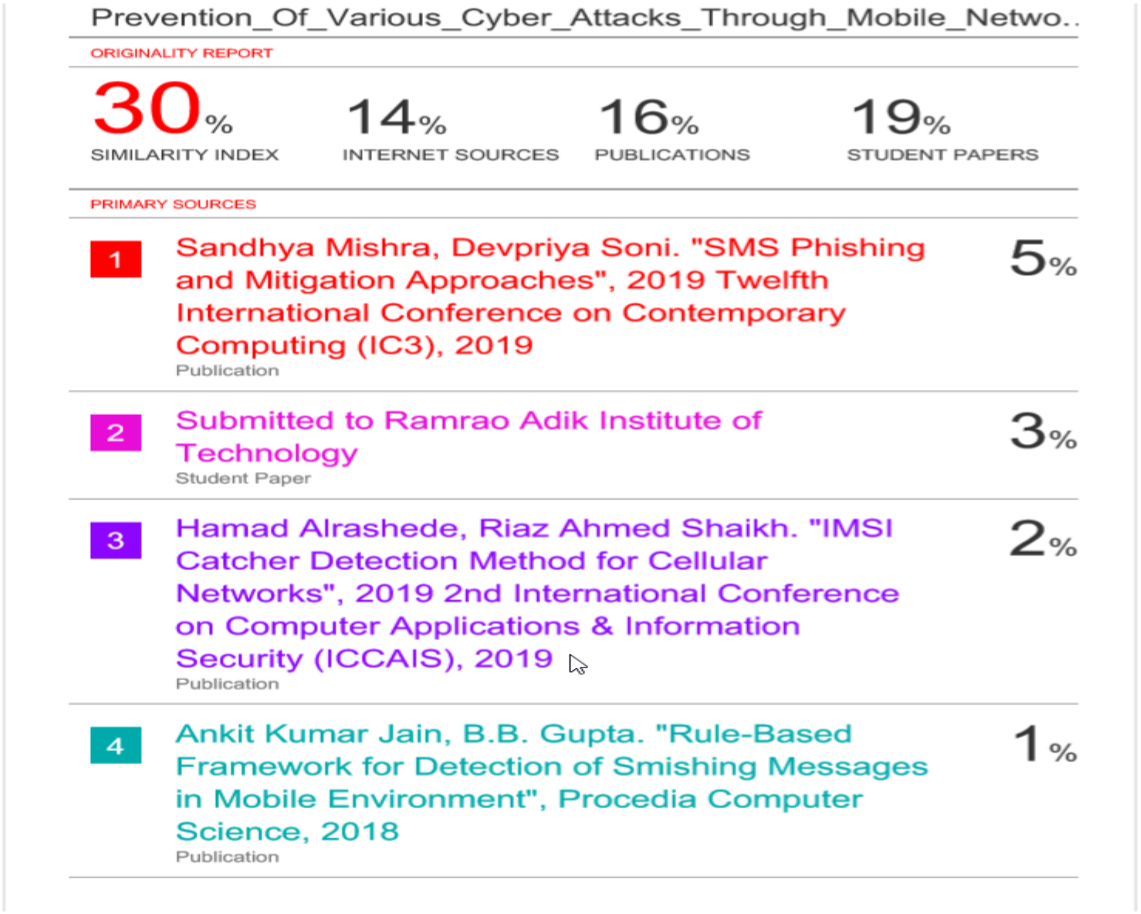
The system can be further enhanced and several other functionalities can be added so to make a more secure and functional interface. Algorithms for other similar type of attacks can be incorporated in the framework so that the system can be further expanded.//

5G is a new type of network: a platform for innovation that will not only improve Broadband mobile services today, but will also expand mobile networks to support greater diversity of devices and services and connect new industries with improved functionality, efficiency and cost. This evolution will be giving rise to new and different types of attacks. Various other vulnerabilities will be exploited by attackers using new or different tools and methods. Thus with the rise of new exploits, new prevention mechanisms can be incorporated within this framework.

The system recently works on 4G network, In future it can be expected to function on 5G networks also as there are many advantages of 5G networks compared to the previous versions.

Appendix A

Plagiarism Report



Prevention_Of_Various_Cyber_Attacks_Through_Mobile_Netwo..

ORIGINALITY REPORT

30%	14%	16%	19%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Sandhya Mishra, Devpriya Soni. "SMS Phishing and Mitigation Approaches", 2019 Twelfth International Conference on Contemporary Computing (IC3), 2019 Publication	5%
2	Submitted to Ramrao Adik Institute of Technology Student Paper	3%
3	Hamad Alrashede, Riaz Ahmed Shaikh. "IMSI Catcher Detection Method for Cellular Networks", 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), 2019 Publication	2%
4	Ankit Kumar Jain, B.B. Gupta. "Rule-Based Framework for Detection of Smishing Messages in Mobile Environment", Procedia Computer Science, 2018 Publication	1%

Appendix B

Paper Publication

Nikita Chorghe,Akshay Jain,Shraddha Mali,Prathmesh Gunjgur “Identity Theft Prediction Using Game Theory,” ICACC 2020 (International Conference on Automation, Computing and Communication),April 2020. [Paper Submitted]

Identity Theft Prediction Using Game Theory

Nikita Chorghe[1,4,*], *Akshay* Jain[2,**], *Shraddha* Mali[3,***], and *Prathmesh* Gunjgur[4,****]

[1]Department of Computer Engineering, Ramrao Adik Institute of Technology Nerul, Navi Mumbai, India.
[2]Department of Computer Engineering, Ramrao Adik Institute of Technology Nerul, Navi Mumbai, India.
[3]Department of Computer Engineering, Ramrao Adik Institute of Technology Nerul, Navi Mumbai, India.
[4]Department of Computer Engineering, Ramrao Adik Institute of Technology Nerul, Navi Mumbai, India.

Abstract. Digital devices have become an integral part of each and every person's life. The range of use of these devices is increasing daily. Over decades, the number of users have increased from thousands to millions, and are still increasing. Due to its multi-functional features, its importance is now being recognized more than ever. Initially, they were used only for calling and texting, however; nowadays, they are also being used to store important data such as account numbers, card numbers, credentials, private pictures, passport copies, etc. The most common form of Identity Theft attack is through stealing passwords. Once the password is stolen, user privacy is lost and the data is compromised. Thus, a system consisting of a database that comprises of leaked passwords collected from various social sites and common passwords as a part of a dictionary attack used by hackers has been created by us. When a user enters his/her password, it runs it through the database and checks for a match. This document emphasizes on how game theory can be utilized in predicting the possibility of a successful attack and discusses important concepts such as the various components of game theory and Nash Equilibrium.

1 Introduction

With widespread digital growth occurring globally, the threats associated with these digital devices are also evolving. All digital devices are prone to cyber-attacks. With India emerging as a digital nation, the safety of digital devices can be compromised due to multiple vulnerabilities, and they can be exploited.

Mobile devices connected to a network can easily become a target of cyber-attack. Hence, it is important to secure these networks. The main motive of malware is to penetrate IOS, Windows and Android defense system. The security mechanism must be an integral part of the system and must work in coordination to identify threats and safeguard user data. Cyber-attacks are carried over the cyber-space, with social or political gains being the main objective. Digital devices, other than being used as a means of communication are also used for storing personal and corporate data. Virus, malicious codes, etc. can result in disruptive consequences which can lead to data loss and endanger privacy of the users.

Identity theft is one of the most common attacks used by an attacker to steal user data. It is also called as Identity Fraud wherein a hacker obtains important pieces of private confidential data such as Email accounts, pan card numbers, pins and passwords, in order to assume someone else's identity.

This data can serve various purposes to the hacker such as to obtain certain credit, goods or other services registered in the name of the impersonated person. The hacker can also provide false identification to the police in order to avoid warrants or arrests and to prevent a criminal record.

As per the latest statistics, approximately 200 billion devices will be connected to the internet by 2020. This alarming rise in number of devices connected to the internet gives increased opportunities to attackers to steal data and infect websites. Since 2013 there has been a theft of 3,809,448 records. In 2016, 95% attacks were focused on government, retail and technology based industries [6].

2 Related Work

K.Veena *et.al.* have proposed a new methodology that identifies the assorted identity of any user and determines if synthetic identity theft attack has occurred. They took three styles of information: Input, Normal, and Target data-set. They used varied identities that maybe text and string information. Various identities are classified in three-class as 100 percent that represent high identity with accurate data, 75% represents medium identity with partial data and 0% low identity with the wrong information. The expected values

*e-mail: nikitachorghe98@gmail.com
**e-mail: aj2812@gmail.com
***e-mail: shraddha.mali1812@gmail.com
****e-mail: prathmesh.gunjgur@rait.ac.in

are 0% or 100% for the various identities and normal value ranges from 0% to 100%. The neural networks are trained with the on top of values. The progress is obtained for the epoch values, time, performance, gradient, and validation checks [1].

Sharmistha Dutta *et.al.* have proposed the method that deals with the credit card application crimes. The techniques are used to remove identity theft. They have proposed new data mining techniques. Mainly two algorithms are used they are Communal Detection and Spike Detection for fraud detection. Communal algorithm identifies the communal data and Spike detection algorithm is used to detect spikes in the duplicates.The system uses resilience concept which is the multi-layered data mining based approach [2].

Philip A.K.Lorimer *et.al.* proposed a framework which focuses on Validation of social profile. Using the proof-of-concept study their proposed framework detects abnormal behavior in social profiles. Numerical values result shows that if the matching threshold in a decision tree is set properly than system identifies the compromised accounts by avoiding the heavy central processing [3].

Very few systems have been developed for identity theft attack detection. The process of identification of identity theft attack is based on different parameters for which various databases have been compiled. These parameters can be easily manipulated

1. They have proposed a method that deals with credit card information theft by using two algorithms: communal data and spike detection algorithm. Major advantage is that they used multi layered data mining approach [1].

2. They have also developed framework that focuses on validation of social platform [2].

3. A major drawback of all the systems is that if by chance a user's password is hacked, the using that password data can be easily be accessed. Thus, there was a need to mitigate this risk by preventing the users from already using passwords that have been leaked [3].

3 Proposed Work

The objective to develop the OSI model was to provide a set of design standards to equipment manufacturers so that they can communicate with each other. It consists of 7 layers. Attackers target the vulnerabilities present in the network and physical layers.

Over the years, many developments have been made in order to improve the OSI model. However, the base structure remains the same. Hence, these problems exist till date. A cyber-attack is an intentional exploitation of resources, hardware/software and networks for personal or monetary gains. Using malware, infecting systems by viruses and worms, using backdoor traps or brute force attacks are some of the common means and methods to conduct cyber-attacks. Attackers perform these attacks to gain access to confidential information or resources of an organization, or to steal data that can be later sold on dark web.

The main motive of attackers behind Identity Theft is to steal the the personal or financial information of a person with the sole purpose of obtaining that person's identity and make transactions or purchases.

Table 1. Issues in Physical and Network Layers

Network layer	**Physical layer**
Various data routing paths are provided by Network layer for its communication.	Actual physical connectivity is defined by Physical layer
Data is transferred in the form of packets using different logical network paths in a sequential order controlled by the network layer.	physical layer defines the hardware equipment, cabling, wiring etc.
Network layer utilises multiple routing protocol on the network.	It majorly consists of hardware equipment such as cables, routers etc.
Various attacks such as ICMP, packet sniffing and DOS attacks are performed majorly over the Internet.	DOS attacks are performed by cutting wireless network cables.

Table 1 refers to various issues present in the Physical and Network layer of the OSI model. These vulnerabilities are exploited by the attackers to gain access to confidential data.There are different types of theft are shown in Table 2.

Table 2. Different Types of Thefts [4]

Theft Type	**Description**
Criminal	A criminal uses the identity of other person in order to evade an arrest and conviction records.
Medical	This type of theft is usually done in order to obtain free medical services.
Financial	Among all the types of thefts, this is one of the most common and frequently committed theft. Another person's identity is used for goods, credits and services.
Child Identity Theft	The perpetrator uses child's name and other confidential information with the intention of avoiding arrests, obtaining loan or a job etc.

Identity theft is a crime in which an attacker steals the information from a legitimate user and uses that information to impersonate someone. This leads to the immediate impact of losing something important like money. Identity Theft victims can face abstract costs such as defamation of character, prevention in securing credit cards or position in a company, etc. There are very less chances of recovering from Identity theft after such circumstances.

Participation in various events is only possible because of the identities assigned to the individuals and the privileges granted to them. Some of these are driving, attending school, donating blood, using credits, taking loans etc. In cases of account hacks, the hacker exploits the privileges granted to the victim by misusing resources such as writing cheques against the person's account or misusing stolen credit cards to buy utilities in the person's name. In the other cases, the hackers misuse individual's identity for credit cards or loan applications. Fraudsters are coming up with new and varied methods to assume the identity of other individuals. Thus it is necessary to develop a preventive measure in order to reduce the disastrous effects of identity theft [8].

Identity theft can occur in various forms such as insurance, credit and debit card or mortgage fraud. Most of these thefts are unreported because the victim is not aware about occurrence of such an event. The risk to financial institutions and many such institutions continues to grow exponentially. Hence, security improvements that provide technical solutions and adjust as per human behavior are required.

Employees of businesses and organizations are targets of Identity Theft as it leads to defamation and breach of trust. Sometimes users keep single password for multiple accounts. These can be social networking apps, email accounts or even bank accounts. If the hacker gets access to any one of the passwords, the most common practice followed is to use the same password to hack other accounts linked to the victim. Thus, the hacker through a single password can gain access to huge amount of personal data. Once the data is at hand, the hacker can demand for a ransom, or release the data on the internet. Apart from monetary loss, it can also lead to loss of life.

One of the possibilities is that the leaked passwords are used by an organization for centralized servers. Getting access to a server's password can be really troublesome for an organization. The hacker can demand for a ransom, can sell all the data to organization's competitors, or sell their client's private data on the dark net which can put the user at risk.

The proposed framework can be used to detect attacks and thus, mitigate such threats and secure user data. We have developed a hacked password checker, which will work as an alert system for users. The system will check whether the password has been hacked and if so, it will report to the user.

3.1 Procedure

The proposed work has a hacked password checker, which will work as an alert system for users. It will check whether the passwords have been hacked or not. It examines the password and if that password matches the list of passwords that have been leaked, it will report to the user. Algorithm 1 provides with a series of steps to be followed in order to prevent the user from using an already leaked password. This in turn will prove to be beneficial for the user by eliminating the user as a possible target of identity theft.

Algorithm 1: Identity Theft Detection

```
1.Start
2.Create a data set of leaked passwords.
3.Store all the entries after converting them
  into a hash.
4.Take an user input.
5.Convert the input into hash value.
6.Check the user input with each value stored
  in the database.
7.if The user input matches with any of the
  values stored in database then
    Alert the user the password has been
      leaked;
else
    Notify the user the password has not been
      leaked;
end
```

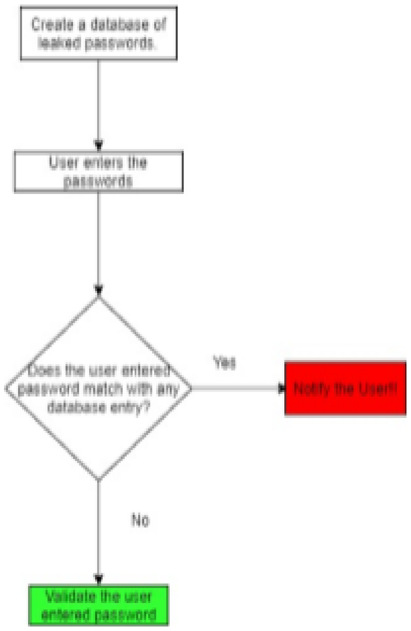

Create a database of leaked passwords.
User enters the passwords
Does the user entered password match with any database entry?
Yes
Notify the User!!
No
Validate the user entered password

Figure 1. Identity Theft Flowchart

The Figure 1 explains the flow of prevention of id theft attacks. The database of the leaked passwords is created. The user enters his/her password. The system will check if that password matches the user entered password if that password matches then the user will be notified or else the user's password is not leaked.

3.2 Game Theory

Game theory is a theoretical framework which is used to analyze social situations among the different competitors [5].

Game theory is used to study situations involving more than 1 participant in a scenario and in which not everything depends on one person. Game theory analyses the behavior of two or more participants involving rewards or punishments. For eg: In the game of chess, the opponent's moves continuously influences your strategy. It is also used to analyze situations in which multiple companies are competing against one another. Let us suppose, if one of the competitor reduces the price of its product, other companies might be forced to do the same. The modeling of strategic interaction between multiple competitors in a scenario containing rules and outcomes is called game theory [5].

Game theory can be a useful approach in order to solve cyber security problems that require rational decision making. An assumption in made in game theory that all the participants have complete information about each other and know all the rules and regulations. Game theory enhances the ability to anticipate the actions of the hackers. Recently, there were few attacks wherein the hackers stole sensitive data and demanded Bitcoins as a ransom to maintain the integrity and data security. Once the data is leaked, the hackers have complete control over it. They can compromise the security of the data by threatening the availability and confidentiality of it. Game theory is a preferable as it share similar concerns with cyber security in various aspects of their application. The move the hacker with respect to defender is not only contingent of the hacker's decision but also depends upon the defender's behaviour. Thus game theory can be used as a mathematical tool to deal with cyber security problems based on multi agent behavior. Thus the combination of game theory and cyber security can be used to solve real time problems [7].

3.2.1 *Nash Equilibrium*

The concept is named after the American mathematician John Nash. It is a scenario or an optimal outcome wherein the player has no motivation to change the chosen strategy. It is an concept which states that opponent's move, provides no incentive to any of the players to deviate from the initially chosen strategy. The player even after changing the course of action will receive no incremental benefit provided all other players do not change their methods. A game can contain multiple Nash Equilibria or none at all. It helps in determining a set of actions that all players must take in order secure the best possible outcome for themselves.

In order to better understand Nash Equilibrium, we can consider an example. Let us suppose there are two players: A and B. In this game, players can either choose strategy 1 which can win them Rs 100 or they can choose strategy 2 wherein they lose Rs 100.

It is common sense that both players A and B will choose strategy 1. Even after revealing the strategies of player A and B to each other, the likelihood of the players changing their currently selected strategy is zero. Thus there is no deviation form the originally chosen strategy even after knowing the opponent's move. Thus strategy 1 is Nash Equilibrium [5].

3.2.2 *Components of Game theory*

The Table 3 represents various components and its description of game theory.

Table 3. Components of Game Theory [5].

Components	Description
Game	Any scenario where in the result is influenced by the actions of multiple players.
Player	Any entity making deliberate decisions based upon the rules of a game.
Strategy	Series of actions that players will perform under a given circumstance which may or may not occur in the game.
Payoff	Payoff is anything that a player incurs by arriving at a consequence. It can be in any form such as money, utility etc.
Information Set	The data available at any given point in a game.
Equilibrium	The stage where all the players have reached an a outcome after making their decisions.

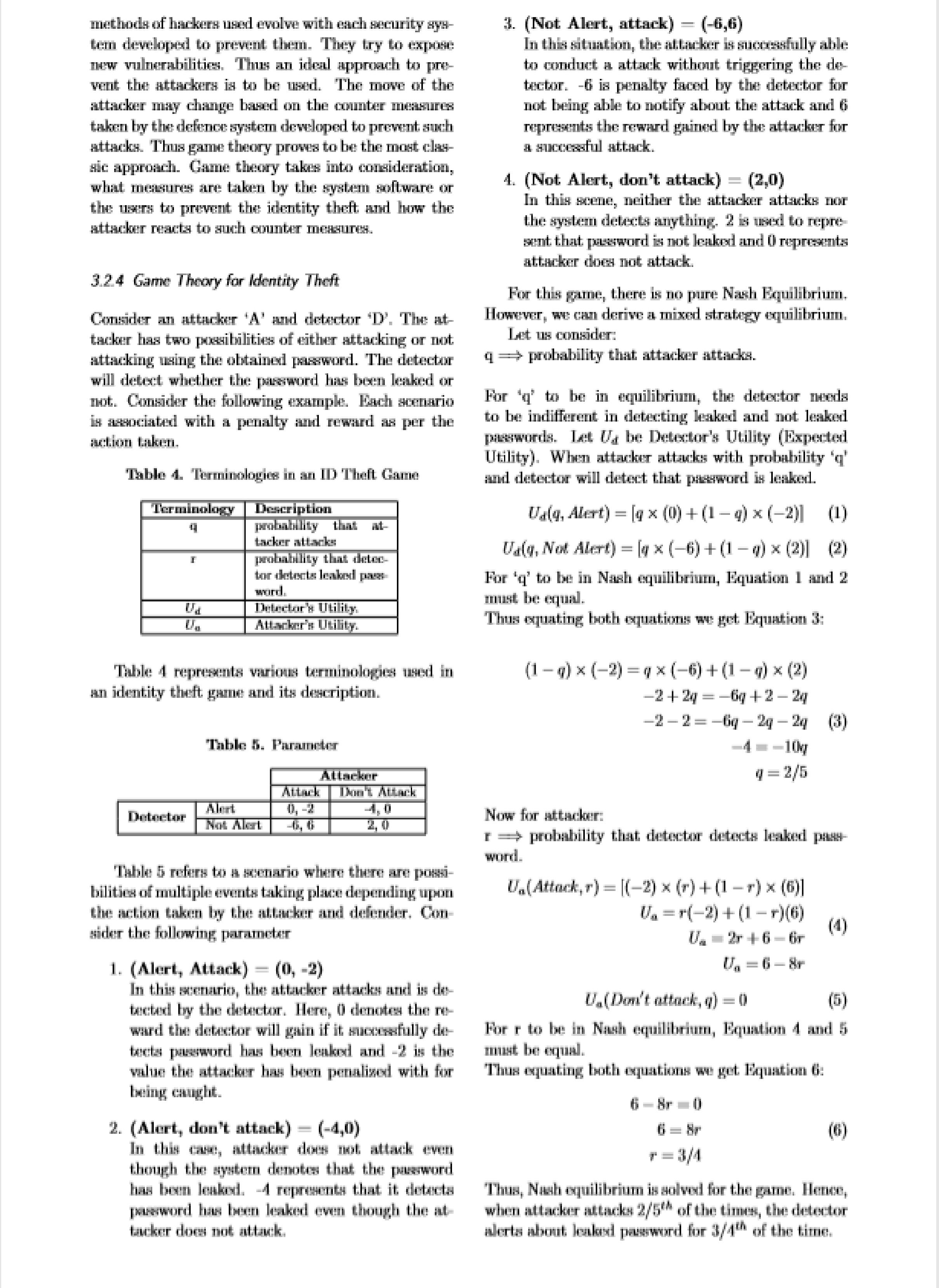
3.2.3 *Motivation to use Game Theory*

Recently, 17 million user records were stolen from Zomato's database. This information included email id's and passwords of users. This confidential data was sold on the popular Dark web for a mere $1000 [11].

In the year 2016, the Internet service company Yahoo reported two major breaches. Both the breaches are considered as one of the largest breaches in the history of Internet. The first breach affected roughly 500 million Yahoo users. Yahoo confirmed that second breach affected all 3 billion users of Yahoo. Verizon which was to buy Yahoo for $4.8 billion, bought it for a reduced price of $350 million proving to be a loss for Yahoo [9].

Any user can become a victim of Identity Theft. As per US Department of Justice, 17.6 million people in the United States experience some form of identity theft every year. In 2014, victims experienced a combined average loss of $1,343. In total, victims lost a massive $15.4 billion [10].

In order to reduce such catastrophic events, it is obligatory to develop a solution to prevent thefts. The



methods of hackers used evolve with each security system developed to prevent them. They try to expose new vulnerabilities. Thus an ideal approach to prevent the attackers is to be used. The move of the attacker may change based on the counter measures taken by the defence system developed to prevent such attacks. Thus game theory proves to be the most classic approach. Game theory takes into consideration, what measures are taken by the system software or the users to prevent the identity theft and how the attacker reacts to such counter measures.

3.2.4 Game Theory for Identity Theft

Consider an attacker 'A' and detector 'D'. The attacker has two possibilities of either attacking or not attacking using the obtained password. The detector will detect whether the password has been leaked or not. Consider the following example. Each scenario is associated with a penalty and reward as per the action taken.

Table 4. Terminologies in an ID Theft Game

Terminology	Description
q	probability that attacker attacks
r	probability that detector detects leaked password.
U_d	Detector's Utility.
U_a	Attacker's Utility.

Table 4 represents various terminologies used in an identity theft game and its description.

Table 5. Parameter

		Attacker	
		Attack	Don't Attack
Detector	Alert	0, -2	-4, 0
	Not Alert	-6, 6	2, 0

Table 5 refers to a scenario where there are possibilities of multiple events taking place depending upon the action taken by the attacker and defender. Consider the following parameter

1. **(Alert, Attack) = (0, -2)**
   In this scenario, the attacker attacks and is detected by the detector. Here, 0 denotes the reward the detector will gain if it successfully detects password has been leaked and -2 is the value the attacker has been penalized with for being caught.

2. **(Alert, don't attack) = (-4,0)**
   In this case, attacker does not attack even though the system denotes that the password has been leaked. -4 represents that it detects password has been leaked even though the attacker does not attack.

3. **(Not Alert, attack) = (-6,6)**
   In this situation, the attacker is successfully able to conduct a attack without triggering the detector. -6 is penalty faced by the detector for not being able to notify about the attack and 6 represents the reward gained by the attacker for a successful attack.

4. **(Not Alert, don't attack) = (2,0)**
   In this scene, neither the attacker attacks nor the system detects anything. 2 is used to represent that password is not leaked and 0 represents attacker does not attack.

For this game, there is no pure Nash Equilibrium. However, we can derive a mixed strategy equilibrium.

Let us consider:
q $\Longrightarrow$ probability that attacker attacks.

For 'q' to be in equilibrium, the detector needs to be indifferent in detecting leaked and not leaked passwords. Let U_d be Detector's Utility (Expected Utility). When attacker attacks with probability 'q' and detector will detect that password is leaked.

$$U_d(q, Alert) = [q \times (0) + (1-q) \times (-2)] \quad (1)$$

$$U_d(q, Not\ Alert) = [q \times (-6) + (1-q) \times (2)] \quad (2)$$

For 'q' to be in Nash equilibrium, Equation 1 and 2 must be equal.
Thus equating both equations we get Equation 3:

$$\begin{aligned}(1-q) \times (-2) &= q \times (-6) + (1-q) \times (2)\\ -2 + 2q &= -6q + 2 - 2q\\ -2 - 2 &= -6q - 2q - 2q\\ -4 &= -10q\\ q &= 2/5\end{aligned} \quad (3)$$

Now for attacker:
r $\Longrightarrow$ probability that detector detects leaked password.

$$\begin{aligned}U_a(Attack, r) &= [(-2) \times (r) + (1-r) \times (6)]\\ U_a &= r(-2) + (1-r)(6)\\ U_a &= 2r + 6 - 6r\\ U_a &= 6 - 8r\end{aligned} \quad (4)$$

$$U_a(Don't\ attack, q) = 0 \quad (5)$$

For r to be in Nash equilibrium, Equation 4 and 5 must be equal.
Thus equating both equations we get Equation 6:

$$\begin{aligned}6 - 8r &= 0\\ 6 &= 8r\\ r &= 3/4\end{aligned} \quad (6)$$

Thus, Nash equilibrium is solved for the game. Hence, when attacker attacks $2/5^{th}$ of the times, the detector alerts about leaked password for $3/4^{th}$ of the time.

4 Conclusion

Setting a password is not the only protective measure one must undertake to safeguard data. There is always a possibility of the password getting hacked or the password already being leaked. Hence, one can lower the risk of data theft by using passwords that have not been previously leaked or those which are not a part of common dictionaries used by hackers to perform brute force attacks. Thus, a system has been developed wherein a database of passwords leaked over the years from multiple social sites and those commonly used in brute force attack are compiled. This system notifies its user if their password is a part of the database. Analysis of the possibility of a successful data theft attack has been conducted with the help of Game Theory. The key concepts such as Nash equilibrium and mixed strategy equilibrium have been applied in order to predict the possibility of attack in multiple scenarios.

References

[1] Veena, K., & Meena, K. "Determination of performance to verify the synthetic identity theft by training the neural networks," IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2017.

[2] Sharmistha Dutta, Ankit Kumar Gupta and Neetu Narayan."Identity Crime Detection Using Data Mining," IEEE International Conference on Computational Intelligence and Networks (CINE), 2017.

[3] Philip A. K. Lorimer, Victor Ming-Fai Diec,and Burak Kantarci."Participatory detection of identity theft on mobile social platforms," IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2017.

[4] Tajpour, Atefeh, "Identity Theft and Fraud Type," International Journal of Information Processing and Management (IJIPM), 2013.

[5] Raoof, Omar & Al-raweshidy, Hamed."Theory of Games: an Introduction" 10.5772/46930.(2010).

[6] Identity Theft - 1 [Online]. Available https://www.cybintsolutions.com/cyber-security-facts-stats/

[7] Annapurna P Patil, Bharath S and Nagashree M Annigeri. "Applications of Game Theory for Cyber Security System: A Survey." International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 17 (2018) pp. 12987-12990, 2018.

[8] Helser, Susan, "Identity theft education: Comparison of text-based and game-based learning" (2016). Graduate Thesis and Dissertations. 15930. https://lib.dr.iastate.edu/etd/15930

[9] Identity Theft - 2 [Online]. Available: https://en.wikipedia.org/wiki/Yahoo!_data_breaches

[10] Identity Theft - 3 [Online]. Available https://www.csid.com/2016/09/real-cost-identity-theft/

[11] Identity Theft - 4 [Online]. Available bit.ly/zomatoleaks

Appendix C

Project Competition

References

[1] Hamad Alrashede,Ria.z Ahmed Shaikh, “IMSI Catcher Detection Method for Cellular Networks,” 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS),July 2019.

[2] Arpita Gupta,Prateek Singh Chandel, “Security Enhancement in GSM using A3 algorithm,”.International Journal of Computer Applications, December 2014.

[3] Muhammet Baykara,Zahit Ziya Gürel, “Detection of phishing attacks,”6th International Symposium on Digital Forensic and Security (ISDFS),25 March 2018.

[4] IMSI-1 [Online] Available: `http://iisecurity.in/blog/imsi-catcher/`

[5] CyberAttack Statistics-1 [online] Available: `https://www.cybintsolutions.com/cyber-security-facts-stats/`

[6] CyberAttack Statistics-2 [online] Available: `https://us.norton.com/internetsecurity-emerging-threats-html`

[7] CyberAttack Statistics-3 [online] Available: `https://www.varonis.com/blog/cybersecurity-statistics/`

[8] Vulnerabilities in Layers-1 [Online] Available: https://cybersecuritynews.co.uk/network-vulnerabilities-and-the-osi-model/

[9] NIST Framework - 1 [Online] Available: https://en.wikipedia.org/wiki/NIST_Cybersecurity_Framework/

[10] CyberAttack Statistics -4[online] Available: https://preyproject.com/blog/en/what-are-cyber-threats-how-they-affect-you-what-to-do-a

[11] SMS phishing -1 [Online] Available: https://www.csoonline.com/article/3411439/smishing-and-vishing-how-these-cyber-attacks-work-and-ho html

[12] SMS phishing -1[Online] Available: https://www.kaspersky.co.in/resource-center/threats/what-is-smishing-and-how-to-defend-against-it

[13] SMS phishing -3 [Online] Available: https://www.consumerreports.org/scams-fraud/smishing-silly-word-serious-fraud/

[14] SMS phishing -4 [Online] Available: https://www.csoonline.com/article/3411439/smishing-and-vishing-how-these-cyber-attacks-work-and-ho html

[15] SMS phishing -5[Online] Available: https://info.focustsi.com/it-services-boston/10_tips_to_avoid_smishing

[16] SMS phishing -6 [Online] Available: https://www.secureworldexpo.com/industry-news/5-smishing-attack-examples-everyone-should-see

[17] SMS phishing -7 [Online] Available: https://www.phishprotection.com/blog/recovering-from-a-phishing-attack/

[18] NIST -2 [Online] Available: http://www.myspaceintl.com/it--cyber-security-solutions.html

[19] SMS phishing -8 [Online] Available: https://www.webroot.com/blog/2019/09/16/smishing-explained-what-it-is-and-how-you-can-prevent-it

[20] ID Theft - 1 [Online] Available: https://bit.ly/idtheftimage

[21] SMS phishing - 9 [Online] Available: https://www.secureworldexpo.com/industry-news/5-smishing-attack-examples-everyone-should-see

[22] SMS phishing -10 [Online] Available: https://www.social-engineer.org/framework/attack-vectors/smishing/

[23] SMS phishing -11 [Online] Available: https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm

[24] SMS phishing -12[Online] Available: https://towardsdatascience.com/modeling-teaching-a-machine-learning-algorithm-to-delive

[25] IMSI - 2[Online] Available: https://drive.google.com/open?id=1Zl6jKckItoKf4j39KPxZPhIJQMKkihGU

[26] SMS phishing -13[Online] Available: https://medium.com/thalus-ai/performance-metrics-for-classification-problems-in-machi

Acknowledgement

We take this opportunity to express our profound gratitude and deep regards to our guide **Mr. Prathmesh Gunjgur** for his exemplary guidance, monitoring and constant encouragement throughout the completion of this report. We are truly grateful to his efforts to improve our understanding towards various concepts and technical skills required in our project. The blessing, help and guidance given by him time to time shall carry us a long way in the journey of life which we are about to embark.

We take this privilege to express our sincere thanks to **Dr. Mukesh D. Patil, Principal, RAIT** for providing the much necessary facilities. We are also thankful to **Dr. Leena Ragha** Head of Department of Computer Engineering, Project Co-ordinator **Mrs. Smita Bharne** and Project Co- coordinator **Mrs. Bhavana Alte**, Department of Computer Engineering, RAIT, Nerul Navi Mumbai for their generous support.

Last but not the least we would also like to thank all those who have directly or indirectly helped us in completion of this thesis.

Mr. Akshay Jain
Ms. Nikita Chorghe
Ms. Shraddha Mali