Network Behavior Analysis using Android Malware Detection

Nikita Kataria[1], Aayushi Singh[2], and Rishabh Kamal[3],
[1,2,3](Department of Information Technology, ABES Institute of Technology, Ghaziabad)

Abstract— *The rapid growth of smartphones has led to a new era of science and technology. Android and IOS are the most popular smartphone platforms that offer public marketplace. With the increasing use of smartphones, users are concerned about security break through and malicious behavior. The increasing rate of security threats and leakage of privacy are becoming more vulnerable and dangerous without the user's attention. In this paper, we approach malware by abstraction of program behaviors. We approach to protect mobile devices against attacks based upon detection principles, architecture and collected datasets.*

Keywords— *Android, Smartphones, Malware, PyCharm, Wireshark.*

I. INTRODUCTION

Characterization of data is done by collecting a large amount of data issued by applications.it can also be used for anomaly-detection system or a misuse detecting system. Some applications are constrained due to security reasons leading to cause greater threats to our android smartphones. Our work focuses on monitoring suspicious behavior at runtime and recognizing their malicious functions in android applications. The functions and methods frequently seen in malicious code are detected. Thus, malicious behavior could be the highlight using this technique. It is beneficial to monitor an application at runtime so as to understand how it interacts with the device by providing application programming interfaces (APIs). APIs are used to request services from the Operating System which includes a set of functions, procedures, and methods used by computer programs. How some software components like protocols, routines, and tools should act when subject to invocations by other components are specified by the API.

Wireshark- Wireshark is an open source packet analyzer used for data capturing [1]. Wireshark helps you see the network activities on a microscopic level. It supports decryption for many protocols. Wireshark has the most powerful display filters. It is a program that understands the structure of different networking protocols. Using Graphic User Interface (GUI), we can browse the network data that has already been captured. Wireshark has left behind many applications as it is free of cost unlike other applications without worrying about license keys. Plug-ins can be created for new protocols. In Wireshark the data can be captured from a file of already captured packet or from a live network connection. After capturing the data, we save the file in Wireshark and then copy that file to PyCharm. It is also used by Quality Assurance engineers to verify network applications.

PyCharm- JetBrains PyCharm is an IDE i.e., Integrated Development Environment used in computer programming. It works only for Python language and no other language. It is one of the best Python Oriented IDE. PyCharm supports web frameworks like Django, Web2py and Flask [2] It also has an integrated python debugger. Its features include unit testing, syntax and error handling, coding assistance and code analysis. PyCharm features can be extended as the developers can write their own codes in the API provided by PyCharm. There are a number of plugins which are compatible with PyCharm. Plugins other than JetBrains also work with PyCharm. It also includes version control integration and project & code navigation.

Kivy Framework- Kivy is used to create applications that use python libraries on all types of operating systems. As we know that python cannot be used to run Android applications so we use kivy launcher to run python programs on android devices. Kivy framework contains an intermediate language used to design custom widgets. This language is used to describe user interfaces and interactions. We first use python to create the base widget and then use kivy to construct User Interface. It can run on Linux, Android, Windows OS, iOS and Raspberry Pi [3]. It has a multi-touch mouse simulator.

MALWARE

Malware is a type of malicious software that causes damage to computer, client, server or any computer network intentionally. The various types of malwares are computer viruses, Worms, Trojan Horses, etc.

TYPES OF MALWARE

Spyware- It is a software that aims to gather information about a person or organization, without their knowledge, and send such information to another entity without the consumer's consent.

Adware- It is a software that automatically generates online advertisement in the user interface of the software or on a screen presented to the user during the installation process.

Trojan Horses- It is any malware which misleads users of its true intent. Trojans generally do not attempt to inject themselves into other files or otherwise propagate themselves.

Worms- A computer worm is a type of malware that spreads copies of itself from computer to computer. A worm can replicate itself without any human interaction, and it does not need to attach itself to a software program in order to cause damage.

CYBER SECURITY

Cyber security is a way through which we protect our program, system, network from unauthorize person and the aim of the hacker is either steal the information of the user or to make changes in the information.

IMPORTANCE OF CYBER SECURITY

It is very important to implement these measures and ensure that the network is protected from the attacker if we want to convert the system into digital platform. It is very important for domain such as medical, financial, security and government which have huge amount of data to process [5].

Sometime when user visit the malicious URL or the user system have some type of malware there is high chance the attacker may steal the information from the system.

By the increase of the virtual storage platform there is high chance of cyber attack

The perfect approach of cyber security is when there is various layer of security is there such as firewall vpns etc.

CYBER ATTACK

Cyber-attack is a kind of offensive move that targets the computer, its data or any computer device. A cyberattack may steal, alter, or destroy a specified targeted file or other data of the computer.

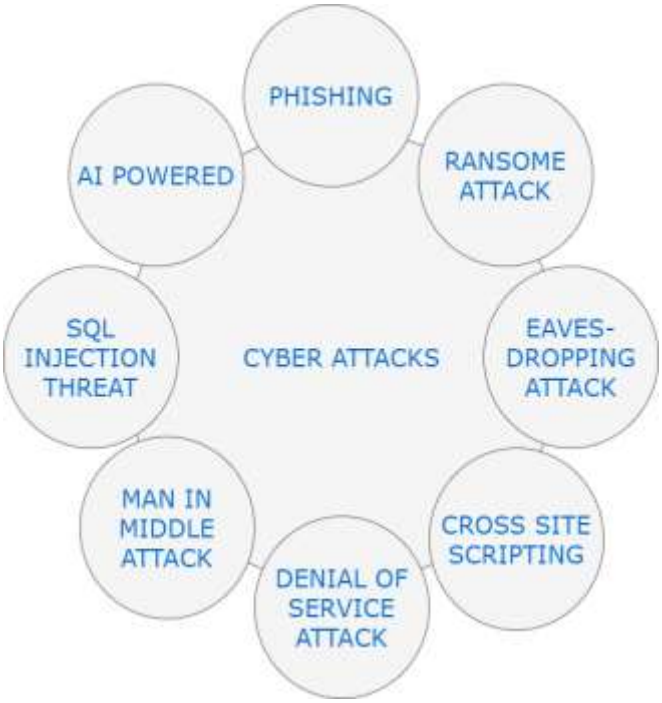

Fig 1: Types of Cyber Attacks

TYPES OF CYBER ATTACK

PHISHING- In phishing[4] the critical information such as user name, password, credit and debit card information is taken without the permission of the user. The hacker sends an email, message or any link to the user to steal the information

AI POWERED- AI is a technology [4] which is frequently use in various platform such automation of the cars, socio humanoid robots, hospitals, national security etc. Attack on

AI platform results a catastrophic effect such as shut down of power supply in hospitals during operations etc.

SQL INJECTION THREAT- Sql is a programming language through which user access the information by making queries [4]. Attacker access the database and make a baleful query to access the information

EVESDROPPING ATTACK- In this attack, hacker steals the information when user transfer the data over a network[5]. It usually happens when their connection between user and server is weak.

RANSOME ATTACK – It is one of the most dangerous attack in cyber security where attacker not only access the information but doesn't allow the user to access the data from database [4].

MAN IN MIDDLE ATTACK- In this attack the attacker involve himself between the communication of the user and server [4]. When server send the critical data to the user it accesses the information

CROSS SITE SCRIPTING – In this attack the hacker attaches the malware code with the commonly used website to steal the information [6].

DISTRIBUTION OF DENIAL AND SERVICE ATTACK- In this attack the attacker send the huge amount to request to the server which result the server crash and the authenticated user is unable to make the request to the server. There is no loss of data but restarting the server is expensive.

IMPACT OF CYBER ATTACK

- There is huge impact of cyber attack which cause a huge damage to the organization such as economical cost in which
- There is a loss of data and it also include of damage of the system
- There is a high chance that customer will not trust on the organization and no future customer will there and its also impact the reputation.

SOME MALWARE DETECTION APPROACHES

There are two types of malware approaches, namely-

- Static Approach
- Dynamic Approach

Static Approach- It is a process in which malware analysis is done without actually running without the code. It simply checks the functionalities of an application and execution and it include signature and permission-based approach

Dynamic Approach- It is a process in which malware analysis is done after running the code. It is more effective as it can detect the malicious behavior of an application which cannot be detected using a static approach.

It mainly includes Anomaly Based approach, Taint analysis and Emulation Based approach.

II. RELATED WORK

Previous works have addressed the problem of understanding the Android application behavior in several ways. There are several examples of inspection mechanisms for identification of malware applications for Android OS. Some of them are given below-

In this paper, Karami et al. [7] developed a system for automating the user interactions, which was a transparent instrumentation system to study various functionalities of an app. In addition to this, runtime behavior analysis of an application using input/output (I/O) system calls gathered by the monitored application within the Linux kernel was introduced.

In this paper, Bugiel et al. [8] proposed a framework named *XManDroid* used for security purposes that extends the monitoring mechanism of Android, so that application-level privilege escalation attacks at runtime based on a given policy can be detected and prevented. But this approach modifies the Android framework that has to be ported for each of the devices and Android versions in which it is intended to be implemented came up as a drawback. Unlike [7,8], it is not required to change the framework of Android smartphones if we feel like monitoring the network traffic.

Other authors have proposed different security techniques regarding permissions in Android applications. For instance-In this paper, Au et al. [9] introduced a tool which is used for extracting permission specification from Android Operating System source code.

Unlike other methods, Jeon et al. [10] implemented an application that does not intend to monitor smart phones. It aims at sorting the Android permissions by embedding a module inside each Android application i.e., he built a module which was used to control the permissions of Android applications.

[11], Suarez-Tangil et al. in [12] Faruki et al. in [13] and Sufatrio et al. [14] provides a general overview of the security threats in mobile devices and approaches to deal with malicious malware.

With the upgrading versions of Android OS, the number of firmware's are also increasing. This is the scenario in which the proposed infrastructure in this paper best fits.

III. PROPOSED SOLUTION

The method that has been used in this project is Permission-Based Analysis. As we know, all the mobile applications require permission to run in android. Permission requested by applications governs the access rights in Android systems. These permissions are mandatory so as to protect the system from getting encountered by viruses. Fortunately, users' data cannot be hampered or misused so easily so at the time of installation, users must allow all the access required by the application to get installed. We have mentioned all the permissions required by applications in the AndroidManifest.xml file.

ALGORITHM

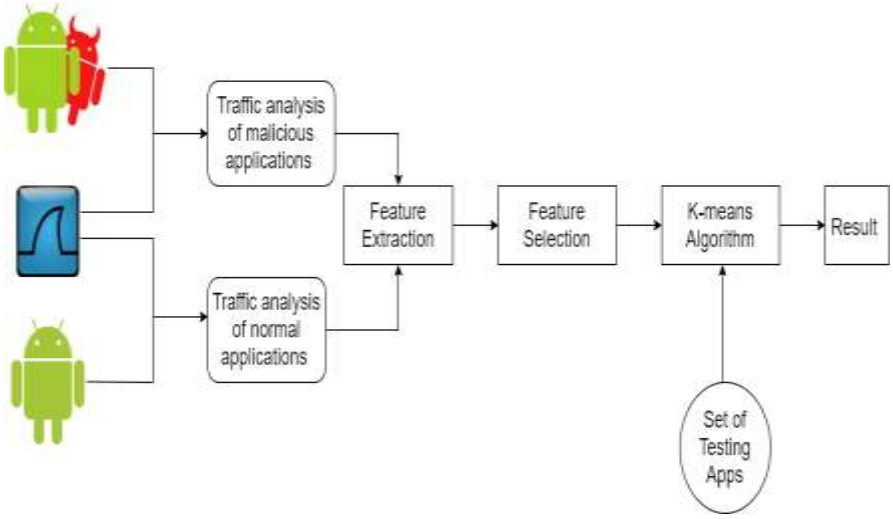

Fig 2: Process of Malware Detection

Differentiate the feature which is useful for malware detection and discard the other feature by using train.csv file

Follow the steps to capture the network

- Start the Wireshark for capturing the network and then save the file.
- Copy the above saved Wireshark file in PyCharm and then restart the file
- Run the FileCapture command to analyse the network traffic
- To check the whether the URL is malicious or not copy the URL and click on submit button. It will run dataset3.csv file to determine benevolent and malicious URL

IMPLEMENTATION

For removing the problems related to android malware detection we follow a step by step procedure, which is given as follows: -

Module 1:

Feature Extraction- In feature extraction we follow a procedure to obtain data from android application files. Firstly, we decompose the malware and good ware applications to extract the data. Then we retrieve information from this data. And lastly, by extracting permissions from each application, we build the datasets.

Feature Selection- Applying feature selection is very important as there are many adverse effects of the extracted features when we apply machine learning on android mobile devices because of some processing restrictions, battery and storage. This is why applying feature selection at primary stage is very important.

K-means Algorithm-

1. Select c centroids arbitrarily, for clusters k_i i.e., [1, c].
2. Assign data points to data points which are closest to the data points.
3. Calculate k_i of cluster k_i i.e., [1, c]
4. Repeat step 2 & 3 until no points change between clusters

1. Select c random instances as centroids of the cluster K_1; K_2; ……K_C.
2. For each training instance x:
   a) Compute Euclidean distance D (K_i, x), i=1...c. Find cluster K_q, closest to x.
   b) Assign x to K_q,.
3. Repeat step 2 until centroid of clusters K_1 ; ……K_CFor each test instance y:
4. Compute Euclidean distance D (K_i, y), i=1...c. Find cluster which is closest to y.
5. Classify y as normal instance using Threshold rule Threshold rule used for test instance y is as follows:

   Assign y→ if P ($^{z}/_{y}$)> Threshold;

URL Detection- It is the most common method to detect malicious websites or links. In this, we use machine learning to detect such malicious URLs by taking a set of training data and based on some properties and functionalities we classify them as malicious or benevolent URLs. The working of this module is shown in fig 3.

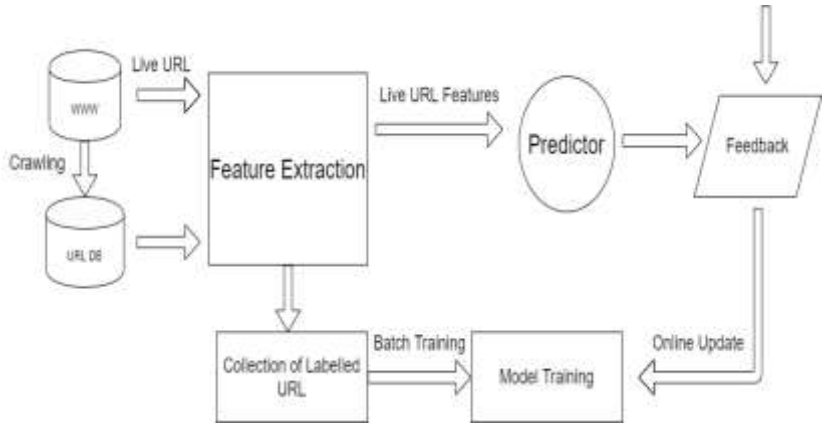

Fig 3: Process of URL Detection

IV. RESULT AND ANALYSIS

Firstly, we extract the required features of an application. These collected features are called datasets. By using machine learning algorithms and approaches, we sort these datasets to distinguish between malicious and benevolent applications. The table in the introduction shows the process of malware detection diagrammatically

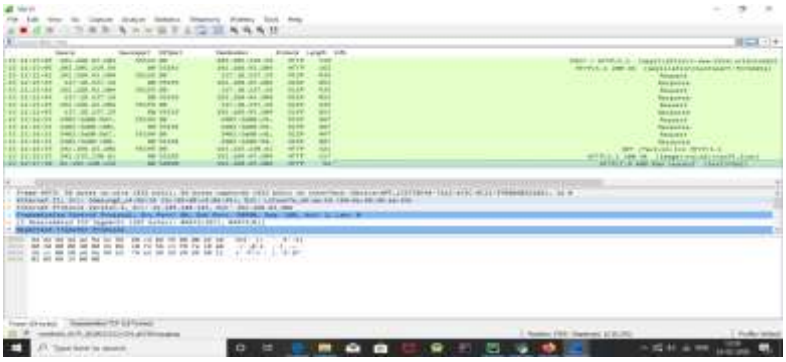

Fig 4: Benevolent URL

When system visit the benevolent URL there is no suspect retransmission which is shown in fig4

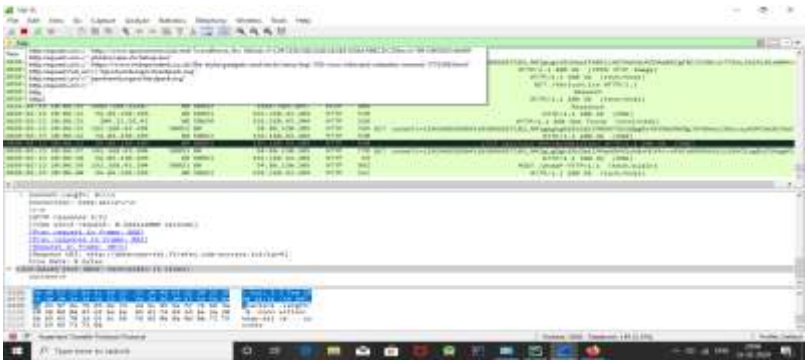

Fig 5: Malicious URL

When system visit malicious URL then suspect re-transmission takes place which is shown in fig5.

V. CONCLUSION AND FUTURE SCOPE

This paper is all about android malware and its detection techniques. We have discussed malware and the types of malware that are available and the approaches required to detect the malware. Android malicious software's is very wide in number because of its open nature. These malicious applications loiter the user data privacy, device integrity and are difficult to detect since they behave as genuine applications. This detection technique helps to detect the malicious software's and websites at runtime i.e., at the time of downloading a software or opening a website. It helps in maintaining the integrity and confidentiality of the users

VI. REFERENCES

1. *https://www.wireshark.org/docs/wsug_html_chunked/ChapterIntroduction.html*
2. *https://www.tutorialspoint.com/pycharm/pycharm_introduction.htm*
3. ***https://kivy.org/#home***
4. *https://www.mygreatlearning.com/blog/types-of-cyber-attacks-and-why-cybersecurity-is- important/*
5. *https://www.investopedia.com/terms/e/eavesdropping-attack.asp*
6. *https://owasp.org/www-community/attacks/xss/*
7. *Karami, Mohammad, Mohamed Elsabagh, Parnian Najafiborazjani, and Angelos Stavrou. "Behavioral analysis of android applications using automated instrumentation." In 2013 IEEE Seventh International Conference on Software Security and Reliability Companion, pp. 182-187. IEEE, 2013.Gaithersburg, Md, USA, June 2013.*
8. *Bugiel, Sven, Lucas Davi, Alexandra Dmitrienko, Thomas Fischer, and Ahmad-Reza Sadeghi. "Xmandroid: A new android evolution to mitigate privilege escalation attacks." Technische Universität Darmstadt, Technical Report TR-2011-04 (2011).*
9. *Au, Kathy Wain Yee, Yi Fan Zhou, Zhen Huang, and David Lie. "Pscout: analyzing the android permission specification." In Proceedings of the 2012 ACM conference on Computer and communications security, pp. 217-228. 2012.*
10. *Jeon, Jinseong, Kristopher K. Micinski, Jeffrey A. Vaughan, Ari Fogel, Nikhilesh Reddy, Jeffrey S. Foster, and Todd Millstein. "Dr. Android and Mr. Hide: fine-grained permissions in android applications." In Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices, pp. 3-14. 2012.*
11. *La Polla, Mariantonietta, Fabio Martinelli, and Daniele Sgandurra. "A survey on security for mobile devices." IEEE communications surveys & tutorials 15, no. 1 (2012): 446-471.*
12. *Suarez-Tangil, Guillermo, Juan E. Tapiador, Pedro Peris-Lopez, and Arturo Ribagorda. "Evolution, detection and analysis of malware for smart devices." IEEE Communications Surveys & Tutorials 16, no. 2 (2013): 961-987.*
13. *Faruki, Parvez, Ammar Bharmal, Vijay Laxmi, Vijay Ganmoor, Manoj Singh Gaur, Mauro Conti, and Muttukrishnan Rajarajan."Android security: a survey of issues, malware penetration, and defenses." IEEE communications surveys & tutorials 17, no. 2 (2014): 998-1022.*
14. *Faruki, Parvez, Ammar Bharmal, Vijay Laxmi, Vijay Ganmoor, Manoj Singh Gaur, Mauro Conti, and Muttukrishnan Rajarajan. "Android security: a survey of issues, malware penetration, and defenses." IEEE communications surveys & tutorials 17, no. 2 (2014): 998-1022.*