# Android Malware Detection

**Mrs. Indira Bhattachariya[1], Mr. Jinang Vora[2], Ms. Manasi Patil[3] , Mr. Priyesh Sharma[4]**

*[1]Assistant Professor, VESIT, Mumbai University, Mumbai, Maharashtra, India.*
*[2]MCA Final Year Student, VESIT, Mumbai University, Mumbai, Maharashtra, India,*
*[3]MCA Final Year Student, VESIT, Mumbai University, Mumbai, Maharashtra, India,*
*[4]MCA Final Year Student, VESIT, Mumbai University, Mumbai, Maharashtra, India.*

---

**Abstract -***The Android operating system is now the most common, which use the operating system of mobile devices, tablets, and smart TV 72.83% of the world market. Therefore, it is widely used by the Android operating system, to attract the attention of those who like the wrong intention, and those who want to violate the privacy of the user who is the creator of everything. Therefore, there is a need to improve existing malware detection, and every new version of Android phone should try, this is the current latest Android version-android 11(Red Velvet Cake) and the beta version of Android OS-12. Automated detection methods, such as antivirus software, are essential to protect Android-enabled devices on the market. Android 11 and limited file system access via the mobile app and more, has long been trying to limit malware and measures. Machine learning is considered algorithms (Random Forest, Support Vector Machine, Gaussian Naive Bayes, and K-Means), and arbitrary forests are used very, very efficiently.*

***Key Words***: **Machine learning, Malware, Random Forest.**

## 1.INTRODUCTION

An Android app is always the main competitor to the mobile apps available on the market. No later than June 2021, another 2,714,499 pieces of software will be added to the Google Play store. No later than April 2021, Android will occupy 72.83% of the global market. Because of the constant popularity of Android has become the most targeted operating system, it would seem. The number of Android-enabled devices has increased over the past few years, partly due to increased usage in the company and the financial services industry. The program often presents the process of processing confidential financial and personal information as part of mobile banking, social media, and communication.

Norton Anti-virus (av) is a malicious "program" that is designed specifically to gain access to a computer or in any other way cause it pain, usually without the owner's knowledge. After that, Norton describes various types of malware, such as spyware, Trojans, viruses, worms, Trojans, and adware. And in 2021, according to Kaspersky Lab, the Security Network, which reports that 1,451,660 mobile phone installation packages, 25,314 packages related to mobile banking Trojans, and 3,596 packages for mobile ransomware Trojans will be found.

Trojan checks the presence of it that came in.tencent. mcg package of the device, i.e. the mobile version of PUBG.

```
try {
    if (isInstalled("com.tencent.ig")) {
        moveToFirst = this.mPubgList.add("com.tencent.ig");
    }
    if (isInstalled("com.pubg.krmobile")) {
        moveToFirst = this.mPubgList.add("com.pubg.krmobile");
    }
    for (int i2 = 0; i2 < this.mPubgList.size(); i2++) {
        try {
            stringBuffer = r28;
```

This setup drives the need to improve safety and reliability, a large part of the market. This study explores whether it may be that malware can be found by analyzing permissions associated with an Android phone, banner ads for adding previous experience on small test datasets, and similar machine learning algorithms (ML).

## 2.BACKGROUND

The influx of mobile devices and applications suppresses the need for mobile security research. Android apps are still distributed in the Android Package Kit (APK) format, based on traditional ZIP compression. Reinstalling is a major threat because malicious engineering with APK files is easy when provided with open-source tools that are readily available. In the traditional attack model, re- while dynamic methods require a specific type of sandbox or simulation environment to perform a data collection application. Heuristic methods apply legal-based infiltration to malicious or malicious applications.

### 2.1Android Architecture

The Android software stack, provides a layered way to support Android apps. The Android app is compiled from source code, data files, and utility files using the Android Software Development Kit into an APK, an Android package, which is a file archive with an .apk . The APK file contains all the necessary content for the Android app and is a file used for the app installation. All components of the application must be listed in a single AndroidManifest.xml file that resides in the APK archive.

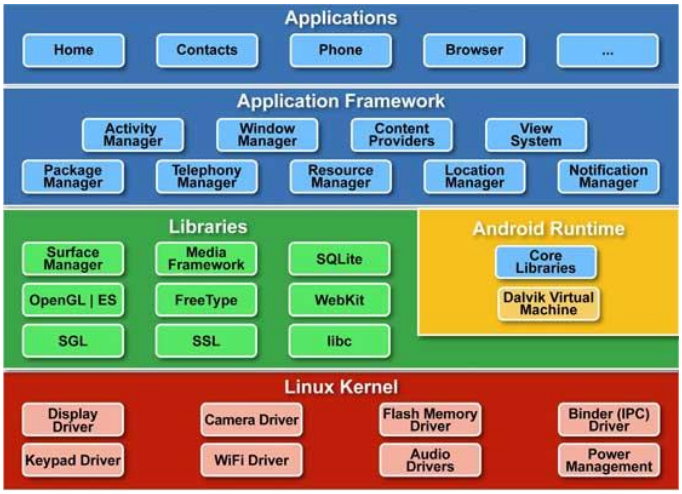



Figure 1: Android OS Architecture

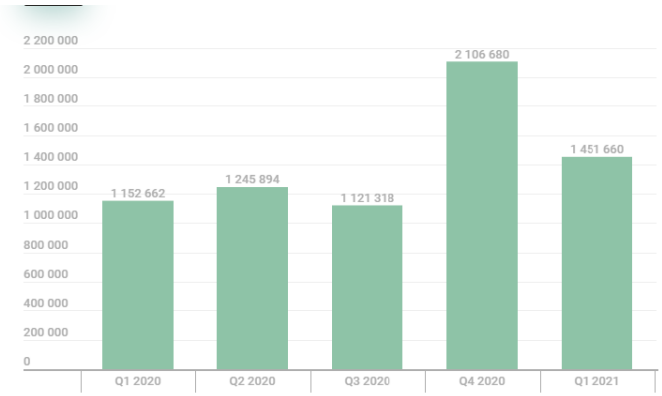
This study focuses on using the manifest file as an integration into selected machine learning algorithms. The manifest file also stores a large amount of information as permissions are required by the application. The exact content of the manifest file varies depending on application.

**Static Detection**

Kaspersky detected 1,451,660 malware installers in Q1 2021, a decrease of 655,020 from Q4 2020 but an increase of 298,998 year on year.

Strict analysis is the process by which a system is analyzed without being executed. The most common method of malware detection in AV programs is based on signature analysis.

Signature-based detection works by scanning a file and creating a unique identifier, a signature, for that particular file. The process of signature development varies between anti-virus programs. Normally, a file hash is created and compared to a table of hashes of known malicious files. If the hash file of the copied file is similar to a malicious file in the database, it may be considered malicious.




The most frequently updated items in the static analysis process are as follows: source code, assets, manifest files, thread patterns, and notorious files. The failure of based recognition signatures assures us that if no comparison signature exists, nothing can be said about the file in question.

Schmidt et al has developed a method for comparing app work calls with work calls for malicious samples using Prism, and Nearest Neighbor Algorithms (PART). Although they reported high accuracy in their static acquisition method, they stated that in the real world, the tasks they performed were not possible due to resource constraints. As a conclusion, they cited the need to do more work to create effective ways to find time.

There are also current books that support the steady discovery of malicious Android applications in machine learning in general, focusing on the research presented in this paper. Tam et al. addresses issues regarding the sandboxed nature of Android apps. To access the system, all applications must be licensed by the Android Permission System at the time of installation. Once the permissions have been installed and the kernel has been granted and enforced, applications can communicate independently through the application via API calls. Unfortunately, these rules also apply to anti-virus services and prevent applications from being tested by other applications. For this reason, signature-based acquisition methods are very effective.

**2.2Android Malware**

Mobile devices have become increasingly focused on everyday performance patterns. We all picked up IMs on our phones, took Zoom calls, and probably checked the customer record on a nearby device. In a typical organization today, 60% of devices that contain or access business mobile data are mobile. So important organizations are stepping up to speed up mobile security to protect themselves from threats such as mobile malware and better protect data.

Threats become more complex and visible across all platforms, and even standard applications can be a threat without proper nutrition. We only need to look at the dangers of WhatsApp. To manage these threats, organizations need to better understand how all applications interact with your environment and their risks .The biggest problem nearby is the misconceptions about mobile security such as 'UEM is enough to protect mobile devices', 'completely secure portable OSs', and 'legitimate app stores are safe'. But the cell phone is just one last resort. As mobile devices interact with business areas more and more, it is inevitable that they will be involved in security breaches.

Our Cloud Security 2020 report highlights some of the significant security issues we have seen in 2020:

52% of organizations have encountered malware on a remote device, up from 37% in 2019.

Android devices have a 5.3x chance of having a malicious app, and one that can be infected with malware, installed on iOS devices.

Companies with at least one 4.4x malware service that should be affected by password leaks than other companies.

**2.3 Malicious apps infected with mobile malware in 2021 so far**

Here are some instances of mobile malware identified in 2021:

**Barcode Scanner:** Adware infected 10 million Android users by March 2021. App users have received an influx of ads hijacked by their devices. The strange thing about cracking is that none of the users have just installed the app, so malware may be acquired as an ad SDK. Google Play quickly removed this app from the store, however, it can still be on millions of devices.

**AndroidOS/MalLocker.B:** Microsoft has acquired a new version of the software - AndroidOS / MalLocker.B, which has been distributed on social media. Responsible malware players carry malware like popular apps, cracked games or video players. Once installed, the freeware blocks access to devices by displaying other screens and ransom payment instructions.

**Uyghur Community Hack:** In an initial social engineering effort, the cybercriminal group 'Evil Eye' targeted more than 500 people from the Uyghur community around the world on Facebook. In Uyghur community groups, hackers would share prayer apps with the Uyghur community keyboards. Once downloaded, fake apps were hacked on devices using two types of trojan Android malware: ActionSpy and PluginPhantom. On iOS devices, hackers download malware known as Insomnia.

**Crypto wallet Trezor:** cybercriminals create a fake crypto wallet application like crypto brand Trezor. Trazors does not have the app, and any online documents claim they do not have it. Hackers have used this to create a malicious app to attract unsuspecting users. Users who became victims of the attack submitted their credentials and their cryptocurrency was stolen, with one victim reportedly losing $ 660,000 in the attack.

**Clubhouse app**: a malicious version of the only invite audio app that came out in March 2021. The original app can only be downloaded when members share a link to join the conversation. The hackers used the white space in March 2021, creating malicious apps that made the Clubhouse app real. When users launch a fake Trojan app, malware creates an app data hack and asks the user to log in and users provide credentials to the cyber killers. Trojan - nicknamed "BlackRock" can steal victim login data with no less than 458 online services. Malware was also able to receive 2FA attack SMS messages and take control of the user device.

**2.4Machine Learning Methodology**

Some of the most advanced technologies that are used to detect malware for mobile devices. Bred are divided into 2 categories: static and dynamic processes and strategies.

**Static Techniques**

Serious methods, focusing on the original application code, to find and evaluate the application safely and without the need to launch the application. DroidRange, Flowdroid, and Debrin are other programs that use static analysis. Static policies are divided into several categories:

**Signature-based approach:**This method removes the semantic model for creating a unique signature. A statement is recognized as a threat if its name matches the name of the current operation. This method is very fast, but can be published together with mixed code. It also failed with invisible malware variants, which requires regular updates of malware signatures.

**Permit-based analysis:** Analyse the state of each workout, because once installed it requires user authorization to access the property. By default, the app is not allowed in the software, user, but access to the system. The user must allow the program access to all requested permissions. These are methods that focus on classification, using which the software requires permission, which speeds up its work, but does not allow it to be stored in a separate file.

**Dynamic Techniques**

At this time, strong process analysis is carried out by the program, the program is analysed, at the same time as indicated in the natural environment. TaintDroid, VetDroid, and DroidScope are some test systems that also fall into this category. Dynamic analysis should classify behaviour according to the accounting method.

**Anomaly Based:** Anomaly tracks various system and device parameters, characteristics. To detect the presence of malicious code, you need to monitor the program's behaviour. These include battery level, CPU usage, and so on. This behaviour is defined and applied by the algorithm that performs the split.

**Clothing Analysis:**In the process of analysis, which tracks the number of sources of confidential information and detects any leaks to the service. This tool will calculate the flow of confidential data, click. It works well with the flow of tracking information, but it can't perform motion and trail monitoring.

**Machine Learning Method**

Classification of malware, widely used machine learning techniques. Permissions designed to protect user and system data directly reflect the critical performance of applications. By analysing the licensing usage of hundreds of malicious and malicious applications, unfamiliar behaviours can be learned that can be used to distinguish malicious and malicious applications. Permit related APIs can be used by the system. In addition, random forests, the most popular method of machine learning for us used to distinguish computer-based and virtuous applications. Random forest is basically used for segregation and decommissioning. Contains a set of binary decision trees. With the training of many deciduous trees, the algorithm integrates the results from these trees by voting.

**2.5Algorithms used in Machine learning:**

**Random Forest**

Random forests are mainly used for classification and regression. It consists of several, binary decision trees. Learning multiple decision tree is an algorithm that combines the results of this tree, by voting strategies .The name, when divided into words, is one of those that is "forests", which is a group of trees, and its a word, for example, "randomly", because randomly selected.

**Support Vector Machine**

SVM presents data classification-an approach that is able to create a non-linear solution and classifies information about what a non-uniform distribution is. This will help you avoid any of the attributes, number ranges, dominating non-numeric operational positions mechanism attributes, and to avoid a number of temporary-tasks for calculating subtasks due to the inner product of feature vectors. An SVM learning algorithm can be used to build models that can do the job.

**Gaussian Naïve Bayes**

Naive Bayes property management, algorithms, and machine learning classifications based on Bayes ' theorem. This is a simple classification technique, but with a high degree of functionality. They are used when the data size is larger. Classification. Naive Bayes is a guided machine learning classification algorithm based on Bayes ' theorem.

**K-Means**

The K-means clustering algorithm shows promising results with a high degree of accuracy during testing using the random forest algorithm.

**3. CONCLUSION AND FUTURE WORK**

The widespread use of Android-enabled mobile devices in the global market is an attractive Android platform for malware development. An active and reliable way to detect malware in the Android environment is a complex problem that, as always, does not find its solution.

Future work will discuss the development of a well-developed and new algorithm for detecting malicious code. Integration techniques, as well as the use of additional ML algorithms in accordance with the sequential receipt, use, explicit provisions, features. Many other static data-related functions are real workbooks that can be easily integrated into the file with the phrase method to create a large set. As a result of this research, the next natural step would be to share your Android malware families that are based on permissions. Finally, contradictions, Machine training can be seen as a response to enemy attacks, the point is to show state functions to trick ML algorithms.

**REFERENCES:**

1. Github,https://github.com/
2. Kaspersky Security Network.
3. Kaspersky Labs, “Mobile Malware Evolution 2017,” Kaspersky Labs, 07 March 2018: https://securelist.com/mobile-malware-review- 2017/84139/
4. Symantec, “What is anti-virussoftware?,” Norton Antivirus, 2019: https://us.norton.com/internetsecurity-malware-what- is-antivirus.html
5. Google Developers, "Platform Architecture,"Google, 3September 2018.https://developer.android.com/guide/platform/.
6. Google Developers, "Application Fundamentals," Google, 2018: https://developer.android.com/guide/components/fundamentals
7. Kaspersky Labs, “Mobile Malware Evolution 2018,” Kaspersky Labs, 05 March 2019: https://securelist.com/mobile-malware-evolution- 2018/89689/.
8. https://www.wandera.com/calling-all-threat-hunters-mobile-malware-to-look-out-for-in-2021