HNSJ Humanitarian & Natural Sciences Journal
ISSN: (e) 2709-0833
www.hnjournal.net

Peer-Reviewed Journal
Indexed at: NSP - SJIF
Arabic Impact Factor 2020 = 0.44

RESEARCH ARTICLE

# A BRIEF SURVEY OF MOBILE ANDROID MALWARE AND CLASSIFICATION TECHNIQUES TOLLS

Maha Adam Gumaa[1]

[1] Computer Sciences, ALNeelain University

Email: angleeee@live.com

HNSJ, 2021, 2(11); https://doi.org/10.53796/hnsj21114

**Published at 01/11/2021** **Accepted at 20/10/2021**

## Abstract

The invention of smartphones is one of the most important achievements in the twenty-first century, as smartphones play an important role in our daily life in various fields such as social networks, education, banking services and reservations such as airline tickets, hotels, browse the Internet and many more. The Android operating system has become one of the most widely used platforms these days, the rapid increase in the use of Android and free applications has contributed to a significant increase in building applications loaded with malware that cause damage to devices or steal sensitive information for users. this study aims to discuss a brief survey of mobile android malware and classification techniques tools, and also discuss the types of malware that infect the Android system, as well as the life cycle of malware, and finally the techniques of malware detection that used for analysis of Android system.

**Key Words:** Android, Malware Type, Android Malware Detection, Static Analysis, Dynamic Analysis and Hybrid Analysis, Machine Leering algorithms.

عنوان البحث

# نبذة موجزة عن البرمجيات الضارة التي تصيب نظام الأندرويد وأدوات وتقنيات تصنيفها

مها آدم جمعة[1]

[1] علوم الحاسوب، جامعة النيلين، السودان

بريد الكتروني: angleeee@live.com

HNSJ, 2021, 2(11); https://doi.org/10.53796/hnsj21114

تاريخ النشر: 2021/11/01م تاريخ القبول: 2021/10/20م

## المستخلص

يعد اختراع الهواتف الذكية من أهم الإنجازات في القرن الحادي والعشرين ، حيث تلعب الهواتف الذكية دورًا مهمًا في حياتنا اليومية في مختلف المجالات مثل الشبكات التواصل الاجتماعية والتعليم والخدمات المصرفية والحجوزات مثل تذاكر الطيران والفنادق و تصفح الإنترنت وغيرها الكثير. أصبح نظام التشغيل اندرويد أحد أكثر المنصات انتشارا و استخدامًا هذه الأيام ، وقد ساهمت الزيادة السريعة في استخدام اندرويد ومجانية التطبيقات الي زيادة كبيرة في بناء تطبيقات مصابة بالبرامج الضارة التي تسبب تلفًا للأجهزة او تكون سببا في سر قة معلومات حساسة للمستخدمين . تهدف هذه الدراسة إلى مناقشة مسح موجز لأدوات تقنيات تصنيف البرمجيات الخبيثة لأجهزة الأ ندرويد المحمولة ، وكذلك مناقشة أنواع البرامج الضارة التي تصيب نظام الأندرويد ، بالإضافة إلى دورة حياة البرامج الضاره ، وأخيرًا تقنيات اكتشاف البرامج الضارة التي تستخدم في تحليل نظام الأندرويد.

## 1. Introduction

Smartphone’s are used everywhere, by everyone, for everything; they are the most recent technological trend of our lives today. Today's social life necessitates that we be constantly connected to the internet via smart phones; additionally, smart phones are rapidly being integrated into enterprises, government agencies, and even the military; in fact, smart phones are used by everyone on the planet from all ages and for various usages. All of these are reasons for the wide development of Smartphone’s hardware and software. Smart phones are based on several platforms; one of the most popular is Android. Its operating system free and open source is based on the Linux kernel is mainly designed for devices with touch - screen smart such as telephones and tablet computers, developed the Android system by the open alliance for mobile phones which is operated by Google. The Android system user interface is mainly based on direct processing by using tactile gestures that are largely compatible with realistic movements, such as clicking, wiping and pinching, in order to manipulate objects on the screen panel, in addition to the virtual keyboard for text entry. Google developed touch devices, as well as Android TV devices for televisions, Android Auto for cars, and Android Wear for watches. Each is developed with a special user interface [3]. Types of Android systems are also used on laptops, game consoles, digital cameras, and other electronic devices. Android has the largest installation base among all operating systems of any kind. It is the best-selling operating system on tablets since 2013, while on smartphones it is dominant by any standard. Globally, Android is estimated to have around 86% of the smartphone market worldwide, which is surprising when you consider that in 2009 it had a 3.9% share figure 1 shows Share of smartphone quarter.

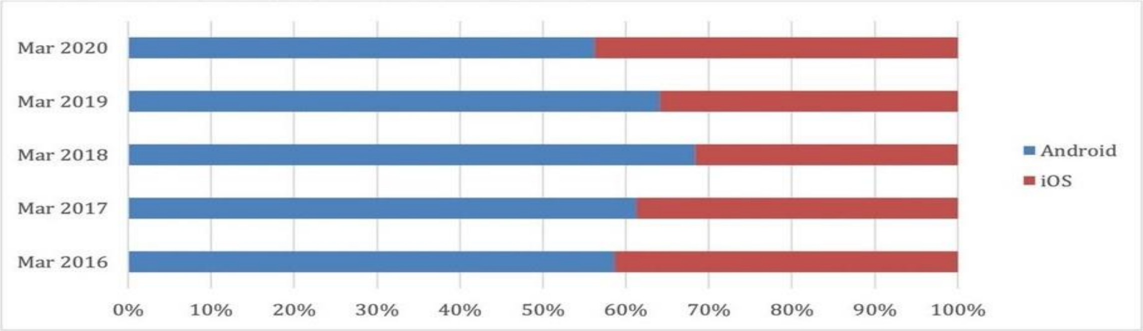

Figure 1. Share of smartphone activations quarter.

The popularity of Android has enabled the application marketplace to grow dramatically, and the black market presence has also grown rapidly, where paid applications are modified for free download and from untrusted websites or stores. When a smartphone user uses his phone, he may be exposed to various information security threats, which can disrupt the operation of the Smartphone and transmit or modify the user data. As a result, Android applications must ensure the privacy and integrity of the data they handle. There are several countermeasures and studies being conducted to ensure app privacy and integrity by detecting and preventing Malware threats in mobile devices. Some of these are signature-based antivirus scanners that detect known Malwares efficiently, while others rely on a detection and classification method in which they classify source code to detect Malware, even if it has no background in mobile applications. These countermeasures and studies differ in terms of accuracy and mobile resource consumption.

## Definition of Android Malware

Mobile malware is malicious software that is specifically designed to target mobile devices such as Smartphone's and tablets in order to access private data. Although mobile malware is not as common as malware that attacks traditional workstations [5], it is a growing threat because many companies now allow employees to connect to corporate networks using personal devices, potentially bringing unknown threats into the environment.

### 1.1 Malware Life Cycle

Malware for mobile platforms in general and Android in particular reproduce the behavior of viruses encountered on desktop [1] [2]. Their life cycle is structured around seven main phases.

- **Creation:** Step in which the programmer designs and implements all malicious code that will be included in the malware.
- **Gestation:** Stage during which the malicious application infiltrates and settles in the system that it wants to infect. It remains inactive throughout this stage. This is why its presence remains totally unknown for the user.
- **Reproduction or infection:** The malware reproduces a significant number of times before manifesting in this phase. The author of the malware seeks to remotely control devices and access private data. Malware spreads via file sharing or social engineering techniques on Android. It uses SMS, Bluetooth, WI-FI as communication means and often disguise themselves as a normal application.
- **Activation:** Some malware activates their destruction routine when certain conditions are satisfied (internal countdown reaches for example). The activation be done remotely. The purpose of this phase is to appropriate gradually all device resources.
- **Discovery:** The user notices strange behavior and suspects the presence of a malicious application. This strange behavior may include performance losses, current changes in the Web browser home page or the unavailability of certain system functions. Anti-viruses often assist the user in detecting malicious actions in sending alerts to the device owner. However, the furtive character of certain malware may extend, even complicate this phase.
- **Assimilation:** Antiviruses update their virus database after the discovery of new malware. If possible, a fix or antidote is also proposed to eliminate this threat
- **Elimination:** This is the phase when the antivirus discovering the malware prompts the user to remove it. It marks the death of the malware.

## 1.2 Types of mobile malware

Malicious software can be defined as mobile malware. This software is designed for mobile operating systems. Mobile malware is any code that is inserted, changed, or removed from a mobile application in order to harm or impair the intended system's operation. Malware comes in a variety of forms, including Adware, bots, spyware, viruses, Trojans, worms, rootkits, and Ransom ware are all examples of malware [3]. The majority of mobile malware is designed to disable or harm a mobile device, allowing a malicious user to remotely control the smartphone or steal personal information stored on the device. For many years, mobile malware has posed a threat to smartphones [3].

A. **Adware** which stands for advertising-supported software that automatically delivers advertisements. Common examples of adware are Pop-up ads and advertisements on websites. Often, software and applications offer free versions which bundle adware**.** Most adware is sponsored or written by advertisers and serves as a revenue-generating tool, but some adware is specifically designed to deliver advertisements, and it is not uncommon for adware to come bundled with spyware that can track user activity and steal personal information highly sensitive information. Adware/spyware bundles are significantly more dangerous and damaging than adware on its own due to the additional capabilities of spyware.

B. **Bot** a mobile bot is a type of malware that runs automatically after being installed on a device by a user. It gains complete access to the device and its contents and begins communicating with and receiving commands from one or more command and control servers. A botmaster is a cybercriminal who adds and manages infected devices to a network of mobile bots (botnet).

C. **Spyware** is a very common type of malware infection on mobile devices. It's malware that allows attackers to access all of your phone's information, including contacts, calls, texts, and other sensitive data, as well as hijack your microphone and camera. This data is collected and sent to a remote server. It is frequently attached to free software downloads or user-clicked links. Peer-to-peer (P2P) file sharing has increased the prevalence of spyware and its consequences**.**

D. **Trojans** are malware disguised as legitimate software and apps. They can be used to harvest your sensitive data, spy on your activity, delete files, gain access to your device, download other malware, and more. Requires users to activate it. In mobile devices, cybercriminals typically insert Trojans into non-malicious executable files or apps on the device. The user activates the Trojan virus when he or she clicks or opens a file. Once activated, Trojans can infect and deactivate other applications or the device itself and paralyze the device after a certain period of time or a certain number of operations. Banking Trojans target both international and regional banks by using fake versions of legitimate mobile apps or through phishing campaigns [6, 9].

E. **Worms** is a type of malware that infects other devices while remaining active on infected systems. Cybercriminals can transmit worms through short message service (SMS) or Multimedia Messaging Service (MMS) text messages and typically do not require user interaction to execute commands**.**

F. **Viruses** are a type of malware that has the ability to replicate and spread to other computers. Viruses frequently spread to other computers by attaching themselves to various programs and executing code when a user opens one of those infected programs. Viruses can also spread Web application vulnerabilities through script files, documents, and cross-site scripting [9].

G. **Rootkits** covert method of bypassing security restrictions to gain unauthorized access to the system. In simpler words [4], a backdoor is a piece of code that allows others to go in and out of a system without being detected.

**Ransomware** is a type of malware that locks the data on a victim's device or the device itself, Typically, encryption is used, and payment is demanded before the data or device is decrypted and access is restored to the victim. Unlike other types of attacks, the victim is usually notified of the occurrence of an exploit and given instructions on how to recover the data. Cybercriminals often demand payment in a crypto currency such as Bit coin, so that the cybercriminal's identity remains unknown.

## 2. Android Malware Classification Techniques

The goal of malware analysis is to understand how malware works and how to detect and eliminate it. Malware detection, as a profession, includes various techniques and principles, and a general classification with two major categories has been proposed**.**

### 4.1Android Malware Analysis Techniques

**4.1.1 Static Analysis** is the technique that deals with the features which are extracted from the suspect file or Appellation without executing. It is a preliminary analysis technique that entails extracting useful information from the suspect file [6]. The most common method of evasion is known as an Update Attack, in which malicious content is downloaded and installed as section of an update. Static analysis techniques are incapable of detecting this. As extracted from AndroidManifest.xml, the most common features of static analysis are permission and API calls [7]. The authors [15] presented an android malware detection approach based on static features of the Android applications such as Standard Permissions with Application Programming Interface (API) calls, Non-standard Permissions with API-calls, this classifier achieved Android malware detection accuracy of 99.6%

**4.1.2 Dynamic Analysis** detection detects malware based on the malware's execution behavior. Within that case, detection is accomplished by monitoring the execution of Android malware activity during runtime [8]. Contains a formalized paraphrase. [10] The authors proposed a dynamic malware detection framework for Android. Generated a system call capture system that collects and extracts system call traces from all applications during their run-time interactions with the phone platform. Following that, all of the collected system call data is aggregated and analyzed in order to detect and classify Android application behavior and achieved 96% accuracy.

**4.1.3 Hybrid Analysis** is an advanced security tool that applied to detect malware. Its combination of static and dynamic analysis methods Researchers sometimes prefer to use hybrid analysis, which combines static and dynamic analysis capabilities, to improve malware detection [6]. The authors of [9] presented a hybrid approach for detecting android malware based on both static and dynamic analysis. Collecting app behavior data in the runtime system calls of android applications in a dynamic way and processing the data in a static and offline measure. And then offline compare them to both the malware and benign collected pattern sets to classify the unknown application.

### 4.2 Android Malware Detection techniques

Today, intrusion detection systems (IDS) are the heartbeat of network management and a critical component of any institution's network security strategy. An IDS monitors the network for malicious activity and policy violations and reports that information to determine whether the unusual activity is a security risk or another type of anomaly. Detection techniques can be divided into three categories its signature-based (SB), anomaly-based (AB), and specification-based (SPB) detection.

**4.2.1Signature-based detection** is a malware detection method in which at least one byte of the software is compared to an existing signature of previously known malicious software, which is stored in a database known as Blacklist. [11] The idea is that most malware will be recognized through patterns or signatures. The inability of signature-based intrusion detection systems to detect unknown attacks is one of their most significant limitations. Malicious actors can simply change their attack sequences within malware and other types of attacks to avoid detection. Data transmission can also be encrypted to avoid detection by signature-based tools entirely. Furthermore, APTs are frequently carried out by threat actors who change their signature over time. This is the most widely used malware detection method [13].

**4.2.2Anomaly-based detection** or behavior intrusion detection system (IDS) goes beyond identifying specific attack signatures to detect and monitor malicious or unusual patterns of behavior. This particular system uses statistical, AI, and machine learning techniques to analysis massive amounts of data and network traffic and identify anomalies [11]**.** By monitoring behaviors that may be associated with attacks, the probability of detecting and mitigating a Malware program before the network is compromised increases.

With an anomaly-based, that everything that does not match the existing normalized baseline—for example, a user attempting to log in outside of normal business hours, new products being added to a network without authorization, or a flood of new IP addresses attempting to connect to a network—raises a potential red flag.

**4.2.3Specification-based detection** is a combination of the anomaly-based detection approach and signature-based malware approach. It's also monitors for any deviation, but instead of detecting the event that occurred of specific attack patterns, it looks for deviations in their behavior from the normal specification. A policy governs the events that occur from the program to the operating system.

For any given event, this policy specifies actions such as "allow," "deny," or "log." Some browsers, for example, have a policy of not instantly executing any file downloaded from a webpage that is not on the Whitelist. Such specification policies are extremely effective in preventing device infection via methods such as "drive-by downloads."

## 3. Machine Leering algorithms

It's part of artificial intelligence involving the design and development of algorithms and techniques that enable devices to possess the property of (learning). There are two types of learning: inductive and deductive. Big data inductively inferred general rules and judgments. The primary goal of automated learning is to extract useful information from data [14]. There are many type of algorithms we will explain some of them as follows:

**5.1 Support Vector machine (SVM)** is a supervised machine learning algorithm which can be used for classification or regression tasks. It is, mainly used in classification tasks. In the SVM algorithm, each data item is plotted as a point in n-dimensional space (where n is the number of features), with the value of each feature being the value of a specific coordinate [14-16]. And after that, we perform classification by locating the hyperplane that best distinguishes the two classes figure 2 shows a SVM example.

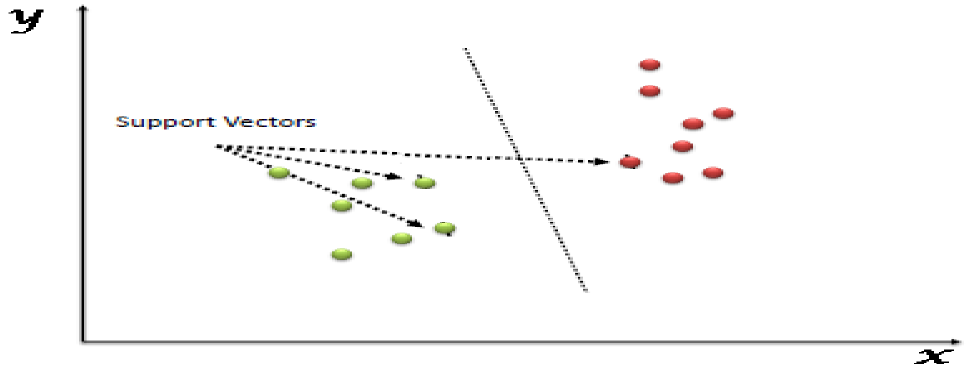

Figure 2: Support Vector machine example.

Existence and uniqueness of solutions co-ordinates are used to calculate support vectors. The SVM classifier is a frontier that best distinguishes between the two classes (hyper-plane/line).

**5.2 Random Forest (RF)**, also known as random decision forests, are an ensemble learning algorithm for classification, regression, and other tasks that involve training a large number of decision trees and then predicting the class that is the mode of the classes (classification) or the mean/average prediction (regression) of the individual trees. Random decision forests compensate for the proclivity of decision trees to outperform their training set. [16] figure 3 shows a RF example.

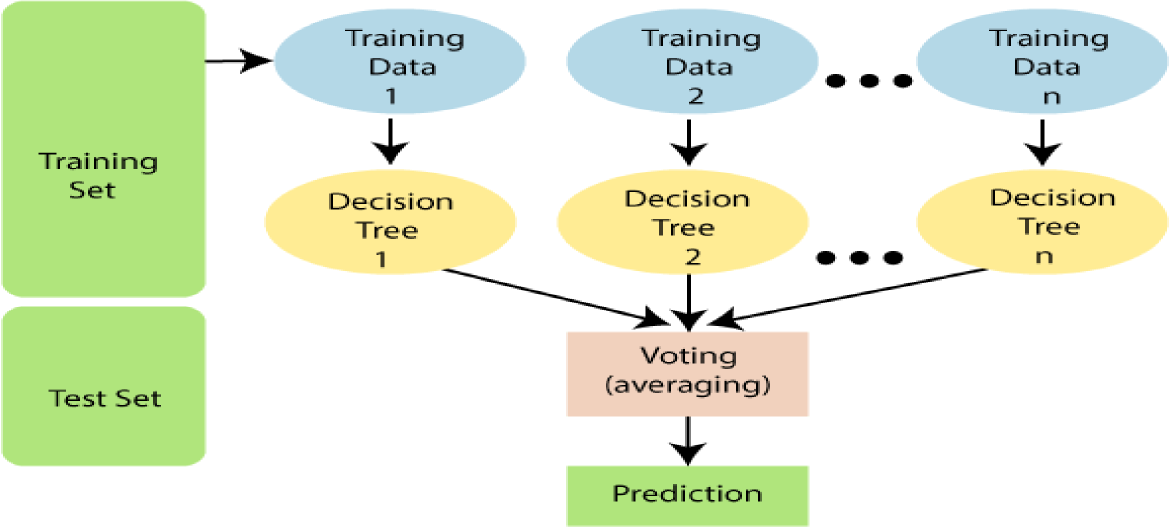

Figure 3: Random Forests example.

Random forests outperform decision trees in general, but their accuracy is lower than that of gradient boosted trees. Even so, data characteristics can have an impact on their performance.

**5.3 Naïve Bayes** is a probabilistic machine learning algorithm based on the Bayes Theorem, used in a wide variety of classification tasks**.** Assumption of the algorithm, that, all the input features are independent of each other and no correlation exists between them [14]. Being a probabilistic model, Naïve Bayes' outputs a posterior probability of belonging to a class given the input features.

$$P\ (A/B) = P\ (A/B_1, B_2, B_3, B_4 \ldots B_n)$$

$$P(A/B) = \frac{\mathrm{P(B/A\,)}P(A)}{\mathrm{P(B)}}$$

For each A possible outcomes or A number of classes. Of classes. Here,

P (A /B) is the posterior probability that given feature B belongs to class A and P (A) is the prior probability of the class A independent of the data, and P (B /A) is the likelihood which is the probability of the predictor given the class and P (B) is the prior probability of the predictor which is the normalizing factor.

**5.4 k-Nearest Neighbors (KNN)** algorithm is a simple, easy to implement supervised machine learning algorithm that can be used to solve both classification and regression problems**.** The output of k-NN classification is a class membership. A plurality vote of its neighbors classifies an element, with the element assigned to the class nearest k neighbors k is a positive integer, typically small. If k = 1, the element is simply assigned to the class of the element's single nearest neighbor [17]. The output of k-NN regression is the element's property value. This value is the mean of the values of the k closest neighbors [17]. A useful technique for both classification and regression is to assign weights to the contributions of the neighbors, so that the closer neighbors contribute more to the average than the farther ones. A common weighting scheme, for example, is to assign a weight of 1/d to each neighbor, where d is the distance between them. The neighbors are chosen from a set of objects for which the class for k-NN classification or object property value for k-NN regression is known [14]. This can also be considered of as the algorithm's training set, though no explicit training is required.

## 6. Conclusion

Due to the rapid and continuous development in the smart phone market and the significant increase in the number of applications and services provided to users, so that these devices integrate to the users every day activities. In this regard, malicious software (malware) has emerged as a major security concern in this domain. This study explains an overview of Android and discusses the types of malware that infect the Android system. And also explain the analysis approach. These three major approaches are static, dynamic and hybrid analysis approaches. Static analysis is a quick and low-cost analysis method that can be used to detect mobile malware. It examines a mobile program without requiring the program's code to be executed because it can detect mobile malware prior to the program's execution under inspection. Dynamic evaluation detects mobile malware after or during the execution of the program under inspection. Hybrid analysis combines both static and dynamic analysis approaches. And also discussed approaches to discovering malware and its type is signature-based (SB), anomaly-based (AB), and specification-based (SPB) detection. Finally, explain some of the machine learning algorithms.

## References:

[1]**]**Hachem, N., Mustapha, Y.B., Granadillo, G.G. and Debar, H., 2011, May. Botnets: lifecycle and taxonomy. In *2011 Conference on Network and Information Systems Security* (pp. 1-8). IEEE.h21

[2] Pandey, S.K. and Mehtre, B.M., 2014, April. A lifecycle based approach for malware analysis. In *2014 Fourth International Conference on Communication Systems and Network Technologies* (pp. 767-771). IEEE.

[3]Sahin, M. and Bahtiyar, S., 2020, November. A Survey on Malware Detection with Deep Learning. In *13th International Conference on Security of Information and Networks* (pp. 1-6).

[4]Suarez-Tangil, G., Tapiador, J.E., Peris-Lopez, P. and Ribagorda, A., 2013. Evolution, detection and analysis of malware for smart devices. *IEEE Communications Surveys & Tutorials*, *16*(2), pp.961-987.

[5] Zolkipli, M.F. and Jantan, A., 2011, March. An approach for malware behavior identification and classification. In *2011 3rd International Conference on Computer Research and Development* (Vol. 1, pp. 191-194). IEEE.

[6] RIASAT, R., SAKEENA, M., Chong, W.A.N.G., SADIQ, A.H. and WANG, Y.J., 2016. A Survey on Android Malware Detection Techniques. *DEStech Transactions on*

*Computer Science and Engineering*, (wcne).

[7] Alqahtani, E.J., Zagrouba, R. and Almuhaideb, A., 2019, June. A Survey on Android Malware Detection Techniques Using Machine Learning Algorithms. In *2019 Sixth International Conference on Software Defined Systems (SDS)* (pp. 110-117). IEEE.

[8] Damodaran, A., Di Troia, F., Visaggio, C.A., Austin, T.H. and Stamp, M., 2017. A comparison of static, dynamic, and hybrid analysis for malware detection. *Journal of Computer Virology and Hacking Techniques*, *13*(1), pp.1-12.

[10] Bhatia, T. and Kaushal, R., 2017, June. Malware detection in android based on dynamic analysis. In *2017 International Conference on Cyber Security And Protection Of Digital Services (Cyber Security)* (pp. 1-6). IEEE.

[11] Chakravarty, A.K., Raj, A., Paul, S. and Apoorva, S., 2019. A study of signature-based and behaviour-based malware detection approaches.

[12] Michael Rezek,(December 9, 2020),what is the difference between signature-based and behavior-based intrusion detection systems, Retrieved, 2021, Mar25 Fromhttps://accedian.com/blog/what-is-the-difference-between-signature-based-and-behavior-based-ids/

[13] Shanthamallu, U.S., Spanias, A., Tepedelenlioglu, C. and Stanley, M., 2017, August. A brief survey of machine learning methods and their sensor and IoT applications. In *2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1-8). IEEE.

[14] Wang, H., Ma, C. and Zhou, L., 2009, December. A brief review of machine learning and its application. In *2009 international conference on information engineering and computer science* (pp. 1-4). IEEE.

[15] Singh, A.K., Jaidhar, C.D. and Kumara, M.A., 2019. Experimental analysis of android malware detection based on combinations of permissions and API-calls. *Journal of Computer Virology and Hacking Techniques*, *15*(3), pp.209-218

[16] Osisanwo, F.Y., Akinsola, J.E.T., Awodele, O., Hinmikaiye, J.O., Olakanmi, O. and Akinjobi, J., 2017. Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, *48*(3), pp.128-138.

[17] Baqersad, M., Mohammadafzali, M., Choubane, B., Holzschuher, C., Hamedi, A. and Ali, H., 2018. Application of laser macrotexture measurement for detection of segregation in asphalt pavements. *Journal of Transportation Engineering, Part B: Pavements*, *144*(3), p.04018032.