IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Malware Detection using Data Analysis Approach in Android Device

Kanchana Binjhade[1], Dr. Varsha Sharma[2]

[1]*Mtech scholar, SOIT UTD-RGPV,, Bhopal, India*

[2]*Assistant Professor, SOIT UTD-RGPV, Bhopal, India*

***Abstract*:** Android device and various applications are useful today in every sector. Different attack scenarios are generating towards security and threats. Different study on creating an efficient framework for designing and analyzing the android structure and multiple threats detection is performed. While dealing with multiple applications in mobile, there are background operations, which may use user's information. These applications save users information secretly and damage the information in smart phone. The previous algorithms and work performed by authors for anomaly app detection and prevention is based on limited criteria. This research has presented a Behavior based IDS (Intrusion detection system) which is enhanced with a powerful NLP (Natural language processing) neural network for detecting malicious behaviors of the Android mobile devices. The experimental results indicate that proposed IDS (Intrusion detection system) can detect an anomaly in the Android operating system effectively and enhance MADAM (Multi-Level Anomaly Detector for Android Malware) algorithm which deals with the Pattern analysis and behavior structure analysis of mobile application in android devices. An anomaly detection and prevention approach over mobile devices is performed. An experiment is performed with a number of regular applications used for different purpose. The observed results shows the efficiency of proposed algorithm over existing rule based and behavior based anomaly detection approach over mobile android OS.

Keywords: *Mobile Computing, Malware, malware detection, Mobile architecture, Android antivirus.*

Introduction

Android is an advanced operating system and a complete stack of software for mobile devices. Android APIs are an abundant system group services covered in an intuitive class file which provides easy access to several characteristics like location, network, telephone, media, camera, and so on. Android is established on the Linux, but does not operate a standard Linux kernel. Android core enhancement includes alarm drivers, Android shared memory drives, power management, core debugger and logger [9]. Malware is the combination of two words known as malicious & software, thus Malware is the software which puts the malicious and harmful effect on the software, operating system or other components. Several Malware and Malware detection techniques [13] are presented that provide a description of the dissimilar types of aggressions and types of Malware, similarly network-based Malware attack, standard Malware attack, etc. In network Malware, Malware is used as spyware, malicious effects on users' machines. Ordinary Malware such as autorun.inf system.inf etc. They are used to put the detrimental effect on the user's machine. The machine learning techniques that help detection and prevention alerts are the component that can be used in this segment. Where Android SDK (Software development kit) follows the architecture to study the detection of anomalies. There is a lot of research done that helps determine anomalies of Android Mobile or other software. The detection of Android mobile Malware has become as important as a large number of cloning and uses applications [16].

These classifications are made using the classifier techniques. In classification, data is divided into small blocks which are known as classes. In classification technique end users needs the information before of time that how these classes are explained. Credit risk applications and fraud detection are particularly well suited to this type of analysis. This approach basically employs neural network or decision tree-based classification algorithms. The data classification process involves first learning and then classification.

Rapid intrusion detection will facilitate intruder detection and damage limitation. Mobile phones usually contain confidential information, such as GPS, contacts, SMS, call logs that are vulnerable to Android phones. The IDS can stop malicious activity and allows proactive attack mitigation. Even if it was not possible to stop the intrusion, it is useful to understand that an intrusion occurred, no matter how it happened and what damage was caused [19].

1.1. Machine Learning Techniques

Machine learning is a data mining tool; which is connected with developments in techniques and methods that allow the system to learn. It is believed that a machine will learn when it changes its structure, program and data in such a way that it improves future performance. You can also ask why the machine should go through the learning process, despite the reason why it cannot be designed specifically to perform a desired task. It has been observed that some activities cannot be defined without examples. Sometimes it is possible to define only the pair of the input, but not the relationship that exists between them. Then, to understand the relationship between the input-input

pairs, the machine must go through the learning process. When learning, the machine adjusts the internal structure to produce the exact output. Now the question arises of what a machine must learn to produce the correct output. There are varieties of computer structures that a machine needs to learn, such as functions, programming logic, problem solving techniques, etc. [6]

The machine learning techniques that help detection and prevention alerts are the component that can be used in this segment. Where Android SDK (Software development kit) follows the architecture to study the detection of anomalies. There is a lot of research done that helps determine anomalies of Android Mobile or other software. The detection of Android mobile Malware has become as important as a large number of cloning and uses applications [16].

These classifications are made using the classifier techniques. In classification, data is divided into small blocks which are known as classes. In classification technique end users needs the information before of time that how these classes are explained. Credit risk applications and fraud detection are particularly well suited to this type of analysis. This approach basically employs neural network or decision tree-based classification algorithms. The data classification process involves first learning and then classification.

2. Malware Detection Techniques used in Android Device

Due to its open environment, Android is the most focused mobile platform by Malware that aims to steal personal information or control user devices; In addition Android has multiple third party application stores which makes it convenient for cybercriminals to repackage Android applications with piece of malicious code. The two most used approaches to the detection and analysis of malicious are static and dynamic analysis. Static analysis is based on inspection of the source or binary code to find suspicious patterns (Malware) within the code. This approach has been used by many antivirus companies. Dynamic analysis monitors and compares the execution behavior of an application (eg. system calls, file accesses, API calls) against malicious and / or normal behaviors through the use of machine learning techniques [8].

Malware Types to have a better understanding of the malicious ones, it is useful to classify them. Malware can be divided into different classes according to their purpose. The classes are the following:
Botnets are commonly used to send spam via email, participate in click fraud campaigns and generate malicious traffic for distributed denial of service attacks [10]. Trojan Malware class is used to define the types of Malware that aim to appear as a legitimate software. For this reason, the general dissemination vector used in this class is social engineering that is, making people believe that they are downloading the legitimate software [10]. Ransomware Malware aims to encrypt all the data in the machine and ask the victim to transfer money to obtain the decryption key. In general, a machine infected with ransomware is "frozen" because the user can not open any file and the image of the mobile is used to provide information about the requests of the attackers [10]. Rootkit functionality allows the attacker to access data with higher permissions than allowed. For example, it can be used to provide unauthorized administrative access to the user. Rootkits always hide their existence and very often are invisible in the system, which makes their detection and, therefore, their elimination incredibly difficult [10]. SMS Trojans use the SMS (text) messaging services of a mobile device to send and intercept messages. The user usually has no knowledge of the behavior. The infection occurs when an application with malicious code is installed. These applications range from legitimate applications that are recompiled with malcode, malicious direct applications with a false name and applications with fake download links.
These applications can be uninstalled using the mobile device's uninstall feature [10].

Intrusion Detection Approach:- Intrusion detection is two types' approaches: detection of signature and detection of anomalies. Signature detection is the technique used by most commercial systems. The detection of anomalies, in which the analysis looks for patterns of unusual activities, has been and continues to be investigated. The detection of the anomalies is used by a small number of IDS [8]. Signatures based detection activities of the analysis system that look for events corresponding to a predefined model or signature that describes a known attack. They collect network traffic and, therefore, precede the analysis. The analysis is based on a comparison of patterns (pattern matching). The system contains a database of attack patterns and will look for similarities with them and when a match is detected, the alert is activated. These systems are really effective in detecting attacks but they produce a large number of false positives.
Signature detection involves searching for network traffic for a series of bytes or packet sequences known as malicious. An anomaly-based intrusion detection system is an IDS to detect intrusions in the network and on the computer and incorrect applications by verifying the agility of the system and classifying it as normal or characteristic. The categorized is based on probing or rules, rather than patterns or signature and efforts to detect any type of misapply that falls out of normal system operation. An anomaly-based intrusion detection system is an intrusion detection system for detecting both network and computer intrusions and misuse by monitoring system activity and classifying it as either normal or anomalous. The anomaly detection focuses on identifying unusual behavior in a host or a network.

Machine learning can extract knowledge from large amounts of data, and then the extracted knowledge is used for prediction. Machine Learning is a technique to train machines from data training, here a machine is composed to use some algorithms through which you can make your own selections and give the result to the user. It is usually treated as a subfield of AI. Today ML is used for complicated classification of data and decision making. In general, it is the development of algorithms that allow the system to learn and make important decisions. It has excellent links with the mathematical optimization that provides the industry with the processes, the theory and the application area, and is designed in a series of calculation tasks where the planning and programming of specific algorithms is not practical [6].
Generally, ML is categorized into three subparts:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning.

3. Proposed Methodology

In this approach a hybrid algorithm is discussed which uses the multiple features from android host. A kernel, application, user and package are used for the host based approach and further server usage library is included for the data dictionary approach for instrument. Host kernel, application, user and package details, server URI, anomaly patterns, and NLP library

An Algorithm and MADAM architecture for detection of anomaly and Malware in android mobile are presented. MADAM Architecture, which is Host-based architecture and Malware detection platform. A multilevel and behavior based algorithm approach is followed by the author. Behavior-based pattern detection architecture was used by this technique. This architecture work on detection Rootkit, SMS Trojan, Spyware, Botnet, Ransomware, installer, Trojan as intrusion entity.

A different level of detection such as application installer and kernel level, further the activity running level and other given user activity level finding of Malware is presented in this paper by authors. They have done all type of global and activity monitoring in their execution. The presentation of the experiment was performed with android application build and execution..

In this research work for detection and Malware in Android mobile are presented. Enhanced MADAM Architecture, which is (Host and server based) Hybrid based and detect to the Malware. Behavior approach and pattern anomaly detection techniques are proposed method and including some new features such as server URI, Http monitoring, pattern extraction, NLP extraction, keyword impack, server packages, and dynamic package library, optimal log monitor [7].

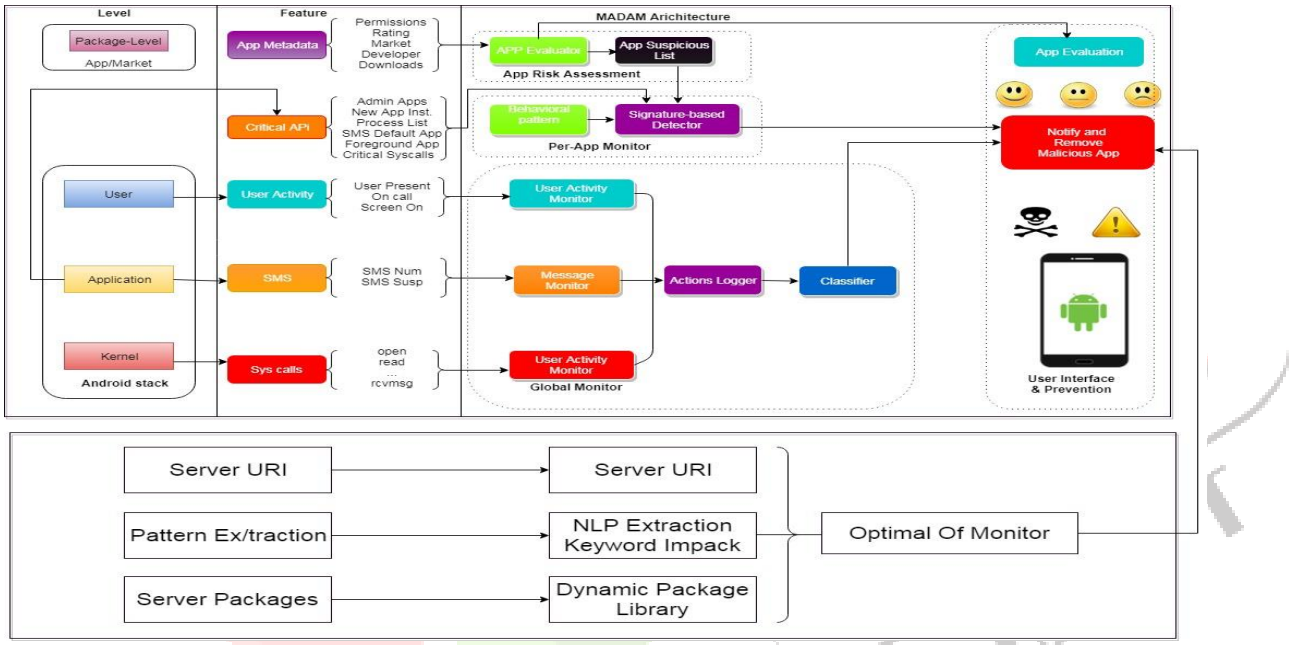

Fig.3.1 Enhanced MADAM Architecture

3.1 Algorithm

In this approach a hybrid algorithm which uses the multiple features from android host. A kernel, application, user and package are used for the host based approach and further server usage library is included for the data dictionary approach for instrument.
Input: Host kernel, application, user and package details, server URI, anomaly patterns, NLP library
Output: Anomaly usage application, result parameter

Algorithm steps:

Begin [
Fetching OS Information ()
{
Kernel Use ();
Package Details ();
}
Int n=numofApplication ();
ForeachApplication (1-n)
{
Fetching usage Statistics ();
Fetching Consumption ();
}
Retrieve server packages ();
Th1=host usage ();

```
String s [] ={Server patterns};
MatchOpr ()
{
ForeachApp ()
{
If (Th1==appdetails || pattern match=appdetails)
Return appinfo;
}
Foreach(appinfo 1-n)
{
Return rankingwiseappinfo();
}
}
}
] End;
```

4. Result:

The experimental results indicate that our IDS can detect anomalies of the Android system with relative accuracy and detection rate.

- Malicious Netflows:-

We presented lightweight IDS which are enhanced with a powerful NLP neural network for detecting malicious behaviors of the Android mobile devices.

- Accuracy:-

When an intrusion is indicated correctly, we have a "True Positive" (TP) fact. When a no intrusion is indicated and this assertion is correct, we have a "True Negative" (TN). When the IDS indicate an intrusion and this assertion is wrong a "False Positive" (FP) alarm is triggered. Lastly, when a non-intrusion is indicated and an intrusion is indeed in progress, we have a "False Negative" (FN) incident. FN is the worst case situation of every detection mechanism since it causes a false alarm. Given these terms, we evaluated our IDS using the accuracy value and the detection rate. Accuracy (ACC) is defined in the following equation as the number of intrusions over the total number of events [19].

$$ACC = \frac{TP + TN}{TP + FP + FN + TN}$$

- Detection Rate:

On the other hand, the detection rate (DR) is the possibility of an alarm given by all real intrusions [19].

$$DR = \frac{TP}{TP + FN}$$

Statistical Analysis

According to the aforementioned definitions, the results of our experiments are summarized in Table 4.1. The experimental results indicate that our IDS can detect anomalies of the Android system with detection rate to 96.9% and 97.0% respectively.

Table 4.1 Result Analysis of Existing Algorithm with Proposed Algorithm

Parameters	Existing Algorithm	Proposed Algorithm
Dynamic/Static	Both	Both
Rooting	Yes	Yes
detection Rate	96.9%	97.0%
Overhead	1.4%	1.4%

In the table given above shows the difference analysis between the existing technique and proposed based approach. It helps in understanding the efficiency of proposed technique.

Graph for the Detection Analysis

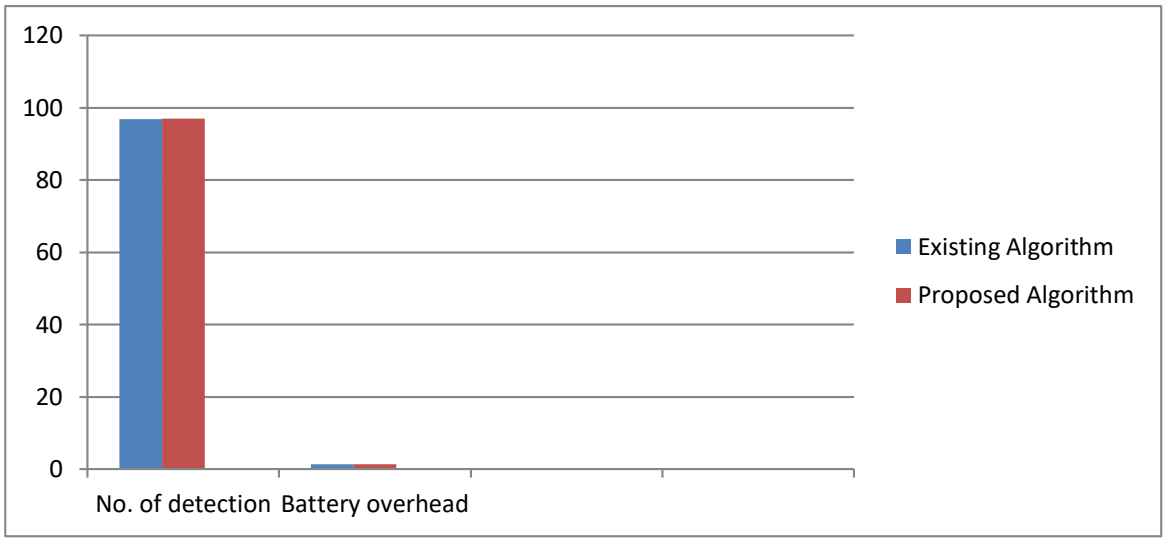

Fig. 4.1 Comparison between the existing intrusion application analysis and proposed solution.

As presented in the figure 4.1 Comparison between the existing intrusion application analysis and proposed solution used in the dissertation.

5. Conclusion and Future Work:

In the current scenario, there are several techniques to detect intrusions in mobile computing. Several services are provided upon request to the user. Such approach requires advanced functions to solve these problems. This research presented a concise explanation of the method used to detect the discomfort in the cell phone. The increase in computing and storage capabilities of Smartphone's has attracted more and more cyber attacks in terms of writing mobile Malware for dissimilar purposes. The popularity and advanced features of modern mobile devices attract the attention of hackers and cybercriminals. In this research, we presented behaviors based on IDS that were improved with a powerful MLP neural network to detect harmful behaviors of Android mobile devices. The experimental results indicate that our IDS are able to detect an anomaly in the Android operating system effectively. Specifically, the accuracy and detection rate of the proposed system reach 96.9% and 97% respectively.

6. Future Work:

As per the previous works and their limitations. A conclusion drawn that the further extension can be done in the following area.

1. Auto-Learning process of the data and Malware features need to be performed which can be further be done by efficient ANN technique such as KNN technique.
2. A security option can be opted while performing the network based analysis and real-time analysis detection over the algorithm.
3. An auto-learning, enhanced security based model is further going to be derived by us with network usage analysis in the proposed system.
4. An Enhancement of previous Anomaly detection approach is performed in proposed work.

References:

[1]. Fang, Z., Han, W., & Li, Y. (2014). Permission based android security: Issues and countermeasures. computers& security, 43, 205-218.

[2]. Mylonas A, et. al, "Assessing privacy risks in Android: a user-centric approach", 2013.

[3]. S. Poeplau, Y. Fratantonio, A. Bianchi, C. Kruegel, and G. Vigna, "Execute This! Analyzing Unsafe and Malicious Dynamic Code Loading in Android Applications," in NDSS'14, 2014.

[4]. A. S. Sayyad, T. Menzies, and H. Ammar, "On the Value of User Preferences in Search-based Software Engineering: A Case Study in Software Product Lines," in ICSE, 2013, pp. 492–501.

[5]. [WWW,document].Google.URL,https://developer.android.com/reference/android/content/pm/PackageManager.html; 2016 [accessed 01.06.16].Android Developer Package manager.

[6]. N. Omar, M. Albared, T. Al-Moslmi and A. Al-Shabi, "A Comparative Study of Feature Selection and Machine Learning Algorithms for Arabic Sentiment Classification", in: 10th Asia Information Retrieval Societies Conference, AIRS 2014, Kuching, Malaysia, December 3-5, 2014, pp. 429-443.

[7]. Andrea Saracino, Daniele Sgandurra, GianlucaDini and Fabio Martinelli," MADAM: Effective and Efficient Behavior-based Android Malware Detection and Prevention", IEEE 2015.

[8]. GianlucaDini , Fabio Martinelli , Andrea Saracino , and Daniele Sgandurra, "MADAM: a Multi-Level Anomaly Detector for Android Malware", Springer 2012.

[9]. Anshul Arora, Shree Garg, Sateesh K Peddoju, "Malware Detection Using Network Traffic Analysis in Android Based Mobile Devices", 4 Eighth International Conference on Next Generation Mobile Applications, Services and Technologies 2014 IEEE.

[10]. MahmoodDeypir," A New Approach for Effective Malware Detection in Android-based Devices", 13th International ISC Conference on Information Security and Cryptology (ISCISC2016) September 7-8, 2016; ShahidBeheshti University – Tehran, Iran.

[11]. Gates, C. S., Chen, J., Li, N., & Proctor, R. W. (2014). Effective risk communication for android apps. Dependable and Secure Computing, IEEE Transactions on, 11(3), 252- 265.

[12]. Kelley, P. G., Cranor, L. F., &Sadeh, N. (2013, April). Privacy as part of the app decision-making process. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 3393-3402). ACM.

[13]. Ruofan Jin, Bing Wang," Malware Detection for Mobile Devices Using Software-Defined Networking", 2013 Second GENI Research and Educational Experiment Workshop.

[14]. Geneiatakis, D., Fovino, I. N., Kounelis, I., &Stirparo, P. (2015). A Permission verification approach for android mobile applications. Computers & Security, 49, 192-205

[15]. PallaviKaushik, Amit Jain," Malware Detection Techniques in Android", International Journal of Computer Applications (0975 – 8887) Volume 122 – No.17, July 2015.

[16]. J. Chen, M. H. Alalfi, T. R. Dean, and Y. Zou, "Detecting android malware using clone detection," J. Comput. Sci. Technol., vol. 30, no. 5, pp. 942–956, 2015.

[17]. Kelley, P. G., Consolvo, S., Cranor, L. F., Jung, J., Sadeh, N., &Wetherall, D. (2012). A conundrum of permissions: installing applications on an android smartphone. In Financial Cryptography and Data Security (pp. 68-79). Springer Berlin Heidelberg.

[18]. Himgouri P. Barge, et. al, "Mobile malware detection through analysis of web application network behavior", International Journal of Computer Science and Mobile Computing Vol.3, December- 2014.

[19]. Sohkyoung (Michelle) Cho, "Intrusion Detection Systems vs. Intrusion Prevention Systems" , NIST Special Publication 800-94. 2008.

[20]. Panagiotis I. Radoglou-Grammatikis, Panagiotis G. Sarigiannidis, "Flow Anomaly Based Intrusion Detection System for Android Mobile Devices", 2017 6th International Conference on Modern Circuits and Systems Technologies (MOCAST)

[21]. Y. Zhou, X. Zhang, X. Jiang, and V. W. Freeh, "Taming information-stealing smartphone applications (on android)," in Proceedings of the 4th International Conference on Trust and Trustworthy Computing, ser. TRUST'11. Berlin, Heidelberg Springer-Verlag, 2011, pp. 93–107.