Mobile Spyware Identification and Categorization: A Systematic Review

Muawya Naser[1], Hussein Al Bazar[*2], Hussein Abdel-Jaber[2]
Email: m.aldalaien@psut.edu.jo, halbazar@arabou.edu.sa, habdeljaber@arabou.edu.sa
[1] Department of Cybersecurity, Princess Sumaya University for Technology, Amman, Jordan
*Corresponding author
[2]Faculty of Computer Studies, Arab Open University (AOU), Riyadh, Saudi Arabia

Keywords: *malware analysis, spyware detection, mobile phones threats, soft computing*

Received: March 23, 2023

Smartphones have revolutionized the way we live, work, and interact with the world. They have become indispensable companions, seamlessly integrating into our daily routines. However, with this pervasive usage comes a growing security concern. Mobile phones are increasingly becoming targets of cyber-attacks, with more than 26,000 attacks happening daily. Among these threats, spyware is one of the most prevalent and insidious threat. Researchers have explored various techniques for identifying and categorizing mobile spyware to address this issue. These efforts are crucial for enhancing the security of our mobile devices and protecting our sensitive data from prying eyes. In this paper, we have conducted a comprehensive survey of the existing techniques and summarized their strengths and limitations. Our analysis encompasses a range of approaches, from signature-based detection to machine learning-based classification. We also explore the latest advancements in behavioral analysis and intrusion detection systems. By consolidating this knowledge, we provide a valuable reference point for future research on mobile spyware detection and prevention. In conclusion, this paper highlights mobile security's critical role in our digital lives. It underscores the importance of ongoing research and innovation in mobile security to safeguard our personal information and prevent cyber-attacks.

Povzetek: Ta članek ponuja celovit pregled tehnik za odkrivanje mobilnih vohunskih programov, združuje znanje o pristopih, kot so podpisna detekcija in strojno učenje, ter poudarja ključno vlogo mobilne varnosti v digitalni dobi.

1 Introduction

Mobile phones have been observing increasing popularity over time. The number of mobile phone users was 7.26 billion (bn) in 2022, while they are envisaged to reach about 7.49 bn by 2025 [1]. Among these, almost six bn are Smartphone users, out of which 5.07 bn have access to the Internet [2]. Most of these users use Smartphones powered by Android Operating Systems [3]. In contrast, only some use other Operating Systems, such as iOS [4], [5]. Smartphones with internet access are always prone to cyber security threats. Among these threats, the most common are Trojans, Worms, Ransomware, and Spyware, as shown in (Figure 1) [6].

Trojans
Worms
Prevalent Threats to Mobile Phones
Virus
Spyware

Figure 1: Common threats to a smartphone.

As its name suggests, spyware is a malicious program aimed at spying on a user's equipment (in most cases, a Smartphone). It is like a ghost in your machine [7]. While residing unauthorizedly in the Smartphone, the ghost can adversely impact the phone. These harms can stretch from using the victim's camera and speaker/ microphone, pattern recording, keystroke logging, stealing banking, and other credentials to crypto mining using your phone. Almost 26,000 cyber-attacks are carried out daily, with spyware having a dominant share [8]. Most of these intrusions have economic motivation at the backdrop. In 2012, the US senate committee on Commerce, Science, and Transportation reported that every one of eleven Smartphones was affected by spyware [9]. This has surely increased significantly, as the total damage is now anticipated to reach $10 trillion (tr) by 2025 [10]. A survey shows that 85% of phones are being affected by spyware [11].

Past works

There are considerable advancements being made by the researchers in detecting, classifying, and combating these threats posed to Smartphones. Some of the researchers have tried to survey and overview the research findings of other researchers. Some of these works are discussed here.

B. Amro et al. have overviewed the existing malware detection techniques for mobile phones. The two major mobile phones operating systems (OS), Android and iOS, are considered for the research. Advantages and disadvantages of each technique have also been summarized [12].

In another research, Y.S.I. Hamed et al. have discussed cloud-based intrusion detection systems for mobile malware. The authors have concluded in the research that mobile isn't processing intensive device whereas IDS needs an intensive processing, therefore, cloud-based solutions are more viable ones [13].

Developments in deep learning for malware detection are surveyed by Z. Wang et al. The authors have brought all types of malwares for Android under investigation. Different aspects of deep learning for malware detection have also been delineated [14]. The classical Machine Learning (ML) algorithms for malware detection have been surveyed by R. Vinayakumar et al. in [15]. In this research, the classical ML algorithms of malware detection, classification, and categorization are evaluated using two datasets i.e. public and private. Different timescales are used to remove all dataset bias in the experimental analysis and come out with proposed model using image processing technique with optimal ML parameters to fill the gap of time-consuming problem of current malware detection algorithms and provides an effective zero-day malware detection solution.

M. Ashawa et al. have highlighted the malware detection techniques being used for Android phones. It has been inferred from the research that most of these techniques are inefficient in detecting malwares with obfuscation. At the end, the authors have presented a critical review of each of the malware detection techniques [16]. E. M. Karanja et al. have surveyed the literature regarding malware attack and their detection in Internet of Things (IoTs). In this research, characterization, propagation, and analysis tools of IoTs malware has been discussed [17].

The literature available, so far, has been focusing on general malware detection. Moreover, the current overview papers aren't specifically aimed at the malware of mobile phones. Along with that, it considers every type of attack like Denial of Service (DoS) attack, phishing and spoofing etc., but not that attack alone that entails spyware. Despite the severity of threat that the spyware alone presents, there is very little focus on overviewing its literature.

Rationale of this research

Neutralizing spyware, thus, becomes a critically important task ahead. Scientists have been regularly investigating new and more effective methods of combating spyware. On the contrary, spyware also changes its signatures [18]. However, there are a few broad categories of spyware detection. It is also worth noting that the broad categorization of spyware detection includes almost the same methods as any other type of malware. However, the details differ. These methods are (i) Static Methods, (ii) Dynamic Methods, (iii) Hybrid Methods, and (iv) Machine Learning.

Researchers have been regularly investigating spyware attack vulnerabilities in mobile phones. But these techniques have yet to be combined and analyzed to serve as a reference for further research. This paper has comprehensively surveyed the techniques used to detect and identify spyware in Smartphones. There are survey and overview papers in the literature (some of which have been discussed above) that focuses on every type of malware. Rationale of this paper is that it focuses on overviewing the techniques used for detecting and combating spyware specifically.

Research questions

- What is the state-of-art of spyware detection in mobile phones?
- What are the advantages and disadvantages of the state-of-art for spyware detection in mobile phones?

Aims of this work

- Exploring the recent methods of detecting spyware in mobile phones
- Presenting the advantages and disadvantages of the recent methods for spyware detection in mobile phones

This paper introduces a literature survey of the modern methods for spyware detection in mobile phones and also reveals the features and drawbacks of these modern methods.

The rest of the paper is organized as follows: section 2 discusses the background of spyware variants, dataset obtaining, the security architecture of Android phones, and methods of detecting spyware; Section 3 explores methodology used for this research; section 4 has overviewed the related literature in detail; discussions and conclusions are provided in sections 5 and 6 respectively.

2 Background

Before plunging into different spyware detection methods in the literature, a few things need to be discussed in the context of spyware types, Android security architecture, dataset acquisition, and the detection and identification methods of spyware.

A. Variants of spyware

Spyware is a type of malware used for spying and espionage purposes. Since it changes its signatures through obfuscation, the exact count of spyware variants is difficult to determine. However, a few are: 'SW.SecurePhone,' 'SW.Qieting', 'mSpy,' 'Flexispy,' 'GnatSpy,' and 'Android APT Spy' [19], [20]. Yet, all these spywares can be categorized broadly. Some of the infamous categories of Android spyware are as follows.

- ***Spybots:*** This type of spyware monitors user patterns, gathers information about different user's activities, and later these are transmitted to third parties without the user's consent. This can intrude on the user equipment, a useful application, or any browser extension, etc. [21]

- ***Cookies***: When cookies act as spyware, they transmit the user's web surfing behavior to unauthorized people. It is passive spyware that works based on existing web browser functions. [22]
- ***Systems monitors:*** Generally, system monitors are used for recording user actions with good intentions. It uses this record for any future system diagnostic. On the contrary, system monitors can publish these user activities to the public while acting as spyware. Keyloggers are examples of system monitors that steal user information [23].
- ***Browser hijackers****:* This spyware tries to change users' browser settings and preferences. Later, it changes the content on the website per the spyware author's will [24].
- ***Miners:*** This is emerging spyware that uses a hosted phone to mine Cryptocurrency. This runs in the background constantly and can adversely affect cell phone resources.
- ***Code for malware:*** Spyware also comes with covert code for installing malware like Trojans and viruses.
- ***Legitimate spyware:*** These are used for spying on intimate partners or children but can also serve dual purposes, e.g., 'find-my-phone' and 'Hello spy' etc. [25].

B. Android security architecture/features

Since Android systems dominate the smartphone market and most of the literature investigates spyware issues in Android, it is necessary to discuss Android security architecture. Some of the key features/components of Android Security architecture are as follows:

- The most important part of its security is the Linux kernel.
- Securing communication among different processes
- Leaving a signature on every application
- Permissions are of two types: granted by the user and defined by the application [26].
- Sandboxing all applications [27]
- Deep defense is also one of its important features.
- Security embedded in design [28]

C. Dataset acquisition

Dataset acquisition comes as a prerequisite for experimenting with spyware detection. Researchers have used various datasets acquired from different real platforms or using virtual environments. Some researchers have preferred using datasets acquired from real mobile phones, while others have used the datasets of virtual environments [29]. Some of the popular datasets are highlighted below.

- ***Derbin4000****:* This is a publicly available dataset of 4000 benign and the same number of malicious samples. It contains samples of many types of malware and can be filtered for spyware samples [30].
- ***AMD project:*** Belonging to AMD, this dataset consists of 24,553 malicious instances. This, too, can be filtered for obtaining the dataset of spyware samples as it has 71 families of malware [31].
- ***AAGM Dataset:*** This dataset has been collected by installing 1900 applications on smartphones. This is among the most viable datasets for spyware detection in Smartphones [32].
- ***M0Droid Dataset*****:** This is a dataset that has been obtained using the M0Droid tool. This dataset consists of data obtained on the kernel level of Smartphones. It has recorded signatures of different families of malware. One of these families is spyware [33].
- ***Self-Recorded Datasets:*** Some researchers prefer to capture data using sniffing tools. Some common tools are Wireshark [34], TCPdump [35], NetworkMiner [36] and Kismet [37]. Using these tools, researchers can generate traffic of their choice. For spyware specifically, these tools can help chase down new spyware variants.

D. Detection and identification methods of spyware

A variety of methods are used for spyware detection, the four prevalent methods used for spyware detection and identification are presented in Figure 2 and discussed briefly in this part as follows:

- ***Static method:*** This spyware identification method analyzes the spyware program to detect malicious parts. This malicious part of the program is later used to identify any future spyware intrusion. Reverse engineering is applied for future identification, and programs like those detected would be classified as spyware. Different tools are used to identify the malicious code: IDA Pro, Ollydbg, etc. [38]. Various popular techniques are used for static spyware analysis: Fingerprinting, File Format Inspection, assembly, etc., as shown in Figure 2.
- ***Dynamic method:*** The dynamic method has dynamism in detecting spyware. It makes decisions based on the function and behavior of the spyware. In this method, a model is trained to record the behavior of the spyware based on past data. It also can identify spyware on runtime. Techniques employed in this method mainly trace functions, their parameters, and control flow [39]. The major tools for running dynamic analysis include Sandbox [40], RegShot [41] and Process Explorer [42], etc.
- ***Hybrid method*****:** As the name suggests, the hybrid method takes advantage of both the static method and dynamic method by combining them. It first runs static analysis and then assesses if any sign of spyware behavior exists. It uses static and dynamic identification [38].
- ***Machine learning method:*** This spyware detection method utilizes ML algorithms to classify the encountered intrusion as spyware or not. A dataset of real spyware instances or virtually generated spyware

traffic is recorded. An ML classifier is then trained using certain features of this data. After training, the model is used for future spyware prediction. Some popular techniques used are Deep Learning (DL), Support Vector Machine (SVM), Random Forest (RF) and Naive Bayes (NB), etc. [43].

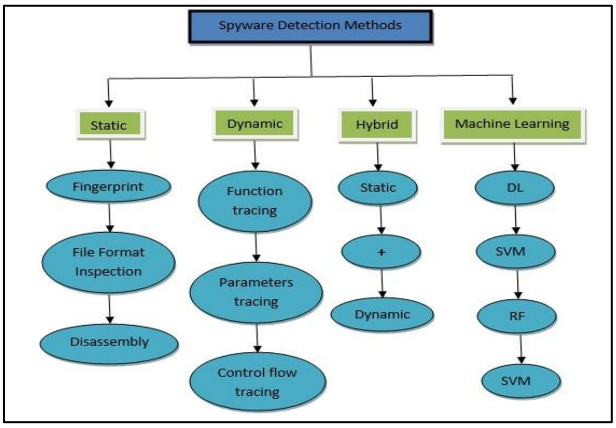

Figure 2: Overview of spyware detection methods.

3 Methodology

While conducting this research, IEEE Xplore [44], MDPI [45], ScienceDirect [46] and ACM [47] were consulted as the main sources. The keywords searched were ‘mobile spyware,’ ‘Smartphone spyware,’ ‘Android Spyware,’ and ‘Spyware detection.’ Most results obtained after searching these keywords overlapped, as shown in Table 1. For this research, papers published after 2015 were considered for two reasons: mobile phones have greatly updated their security and research work before 2015 may be of very less use now (for instance, N. Xu et al. have investigated spyware issues in 3G [48], while now is the era of 5G); the other is that there can be very least research found on mobile spyware of era before 2015 for there was no boom of Smartphones before 2015. This paper has arranged the cited works chronologically, as shown in Table2. A summary of techniques used has been given in Table 1.

Table1: Search results of different keywords.

Keyword	Top five relevant results
‘Mobile spyware’	D. Harkin et al. H. Abualola et al. M. Naser et al. F. Pierazzi et al. F. Fasano et al.
‘Smartphone spyware’	M. K. Qabalin et al. R. Zhang et al. D. Harkin et al. M. Naser et al. M. H. Saad
‘Android spyware’	F. Pierazzi et al. M. H. Saad et al. H. Abualola et al. P. Kaur et al. M. K. Qabalin et al.
‘Detection of spyware in mobile phones	D. Harkin et al. M. H. Saad et al. H. Abualola et al. M. K. Qabalin et al. R. Zhang et al.

4 Related work

The boom of mobile phones has been attracting the attention of researchers. Rich literature is available on it, such as [49], [50], [51], [52], [53], [54], [55], [56], [57], [58]. As so far generic malware detection is concerned, it too has enough material available, as [59], [60], [61], [62], [6]. However, mobile spyware detection, in specific, has very limited research on it. The primary reason is that there is research on general mobile malware detection rather than spyware detection. Another issue is that as Android systems dominate mobile customer count, so is the research arena. This survey, too, would follow the same pattern dominated by the research on detecting spyware in Android phones. Nonetheless, this paper would be inclusive of the research carried for any other mobile OS like iOS. The most recent works are discussed as below.

M. Naser et al. have adopted a novel approach to identify spyware in Android phones. They have applied three ML models on a novel dataset of spyware. There were 168,501 spyware instances in the dataset. The models applied were SVM, NB, and Fine Decision Trees (FDT). FDT was the most accurate classifier, with a value of 98.2% [63].

A comprehensive research in this direction has been published by E. Liu et al. This study is about tracing three main mechanisms of spyware actions: how it abuses Application Program Interfaces (APIs), how users’ personal information is being stolen through APIs, and evading detection systems by hiding the presence of the application. The authors have considered 14 popular consumer spyware applications. A total of eight malicious capabilities have been described for how spyware steals information, evades detection system, and persist in a phone. Each capability has also been proposed with a mitigation method. Some of the capabilities are using the camera obscurely, invisible access to the microphone, recording screen instances, and hiding the malicious app icon and instead using some popular and useful app icons. JADX was used for decompiling the source code [64]. The result section of this study shows different threats revealed in the experiments, their respective threat model, and the specific result of the vulnerability of any app to this threat [65].

M. K. Qablain et al. have adopted a two-pronged approach to investigating mobile spyware. The first ramification of their research is about acquiring a novel dataset of spyware, while the other is about detecting Android spyware. The dataset was acquired from five commercially available spyware applications: mSPY [66], UMobix [67], MobileSPY [68], FlexiSPY [69], and

TheWiSPY [70]. All the spyware applications functioned in full swing. The traffic generated by these applications was recorded by a packet sniffer called PCAPDroid. The dataset is in PCAP as well as CSV format. The data have class A for normal traffic, class B for spyware installation traffic, and class C for typical spyware traffic. Afterwards, ML models were trained to detect spyware traffic. The authors have used different ML models independently for each variant of spyware. Among these, RF has performed the best among all. Overall, the binary classifier has achieved an accuracy of 79% while the multi-class classifier 77% [71].

Another study has been added by F. Pierazzi et al. to identify spyware in Android phones. This research has threefold objectives: it first distinguishes between goodwares and spyware; then spyware is related to its specific family; at last, it also automatically selects key features to underpin if the malicious program is spyware or any other malware. Static analysis has been combined with ML/DL for carrying out experiments. A novel model has been proposed to incorporate many ML classifiers in one called Ensemble Late Fusion (ELF). VirusTotal [72] has been used to create the dataset. The dataset consisted of a total of 15000 malware. The model works so that predictions are made using six commonly used traditional and four DL classifiers. ELF then makes predictions based on the predictions made by traditional classifiers. Regarding, ELF has achieved an F1 score of 0.982 and an Area Under Curve (AUC) too of 0.982 for spyware vs. goodware prediction, an F1 score of 0.960, and an AUC of 0.963 in the case of spyware vs. other malware prediction [73].

D. Harkin et al. have carried out interesting research on spyware. The focus of their research is to compare which between Android and iOS is more susceptible to getting compromised. The authors have maintained that Android users are more prone to spyware victimization vis-à-vis iOS users. They have backed their view that Android offers more ‘openness’ while iOS is more ‘closed.’ Nine general spyware applications were considered, like MSpy and Trackview, etc. Based on five reasons, the authors have concluded that Android phones are more vulnerable to spyware attacks than iOS-supported phones. In the end, it has been concluded in the research that the main reason for the vulnerability of both Operating Systems (OS) is their design philosophy, where Android is permissive. At the same time, iOS is more reserved [74].

Literature has further been enriched by H. M. Salih et al. In this research, a fake game application was developed and installed on an Android phone for spying purposes. The spyware has three-fold mechanisms: an Android application for spying, a desktop application for controlling the victim's phone, database to store the victim’s information. On one side, the application steals information from the user. On the other hand, the desktop application phone can take control of many phone features. The authors have deduced this lesson from the research that names of spyware applications should be stored in a database, and every new application installed should be matched with it; Google and other giants should take serious actions against attackers; encryption of memory should also be ensured; anti-virus applications should be used [75].

An approach has been proposed by M. Conti et al. to identify spyware based on the network traffic it generates. The proposed technique has been called ASAINT (A Spy App Identification System based on Network Traffic). It has been tested on both Android and iOS. For carrying experiment, the researchers first set up a network with a gateway, an AP (Access Point), and many nodes. The traffic was captured using Wireshark for 73,33 hours. A total of 3365 instances were included in the final dataset. The identification was made using three ML algorithms: RF, Logistic Regression (LR), and K Nearest Neighbors (KNN). Regarding accuracy, RF gave a commendable f1 score of up to 0.92. LR had the best time efficiency of classification. The overall accuracy was about 85% [76].

F. Fasano et al. proposed a novel method of detecting Android spyware. In their work, they have proposed a temporal logic-based framework. This framework functions based on the formal method of model checking/ validation. The model accepts two inputs: a Labelled Transition System (LTS) and temporal formula. The result is true if the formula is verified and false if not. Mu-calculus [77] has been used for model checking. On the implementation side, a dataset of 80 applications from 26 categories was collected. Malicious copies of these apps were generated using Android Framework for Exploitation (AFE) [78] along with DroidChameleon [79]. This dataset was then experimented with for spyware detection using temporal logic. It gave an astounding result of 0.98 (98%) accuracy. [80]

S. Hutchinson et al. have experimented with forensic analysis of spyware. In the experiment, the researchers first considered a spyware application belonging to Android.spy.277.origin [81] family, obtained from GitHub. This application was installed on an emulator for permission, code, and traffic analysis. The traffic that the application generated was captured using Wireshark. First, its code was analyzed, especially the AndroidManifest.xml file; its required permissions were analyzed. Then, its manipulation of information like email and messages was monitored; at last, the application was installed on a real phone to see if Play Protect works. It was revealed that for the first time there was no problem in installing the app; for the second time, it was flagged as malicious by Play Protect, and for the third time, the application did not get installed. At the end of the research, the authors proposed a framework for forensic researchers for any future analysis. This framework includes Static, Dynamic, and network analysis of an application under investigation. This will give a clear picture of the application [82].

R. Zhang et al. have applied reverse engineering to exploit a vulnerability in Android phones. In the research, the authors have deployed AI to carry a stealthy attack, called Vaspy, on the phone using voice. The spyware imitates the activation voice for voice assistants. The spyware uses ML to select a suitable time for the attack. The spyware was tested against VirusTotal. It was further tested against three prevalent Android spyware miners:

Derbin, DroidAPIMiner, and MaMaDroid. The disguised spyware proved resilient against the detectors [83].

Another interesting aspect of mobile spyware has been unearthed by R. Chatterjee et al. In this research, the authors have delineated how some applications are overtly or covertly used for intimate partner violence. There are two faceted findings in the research. How many undetected applications are present for spouses' surveillance, and how do some surveillance applications happen to be dual purposed, i.e., legal and covert? For this purpose, such applications were searched for with many keywords, such as “track my wife,” and more than 27000 URLs (Uniform Resource Locators) were returned. Among these, more than 10000 applications were found, and an ML algorithm was trained to filter out the irrelevant applications. The model succeeded in achieving 93% accuracy. At last, 61 on-store and nine off-store applications were selected for in-depth analysis. As far as the existing anti-spyware applications are concerned, big names like AVG [84] and McAfee [85] even could not correctly classify what was manually labelled as spyware. Their detecting accuracy was a mere 3%. As a result of this research, as the authors claim, google has started improving its security [86].

M. H. Saad et al. have conducted another promising experiment. Authors of the research have developed a traffic intercepting malicious application, which they have called a ‘chameleon.’ When installed on an Android system, the developed application would act as a man in the middle. This spyware disease/application is designed in such a way as to intercept incoming SMS, incoming call, and outgoing call. Then the recorded information is transmitted to a cloud database. The authors have proposed a dynamic fuzz-based detection model to detect this spyware disease. The authors name the proposed spyware detection model ‘DroidSmartFuzzer.’ Further, the authors have constructed a real environment for detecting the behavior of spyware. At last, the obtained results have been empirically compared with real results. The DroidSmartFuzzer was tested against 20 spyware applications, some free and others proprietary [87].

An attention-grabber aspect of spyware has been targeted by H. Abulola et al. in their research. They have unveiled how a ‘notification listener’ can exploit an Android's phone security. The main applications targeted in the research are WhatsApp, Facebook Messenger, BBM, and SMS. An ‘SMS Backup’ application is installed and granted permission for fiction listening. These notifications were routed to be sent to the attacker's email. The authors were successful in exploiting the action listening capability of Android. The experiment shows that in Android 4.3, the capability can be exploited for all four services; in contrast, in Android 5.0, the capability can be exploited only for SMS and BBM notification. In the end, the authors have suggested that BBM should change its notification structure. At the same time, Android should look into its permission mechanism [19].

P. Kaur et al. have also added their part. The authors have proposed a novel hybrid approach for detecting spyware in Android phones. In their proposed methodology, a broadcast listener has been deployed to look for any new application installation or update to any existing application. Upon receiving any new or updated application, the broadcast receiver locates its .apk file and reengineers it. The researchers have considered various applications for their experiment. These applications have been scanned using existing antivirus software and validated using the proposed solution. The proposed solution analyses three aspects of an application and classifies an application as spyware or not. The three aspects considered for analysis are Description, Interface, and Source code analysis. Each of these three aspects has a certain weightage in decision-making. Source code analysis has got the highest weightage of 70%. The result shows that the proposed solution has, in some cases, performed better than the existing anti-viruses [88].

An attempt has been made by Z. Zhang et al. to enhance the security of cameras on Android phones. To do so, the authors have developed an application to spy-on-user using the phone's camera. This application will evade the three traditional ways of spying camera detection: API auditing, anti-spyware, and Mobile Device Management (MDM). Such an attack is called a transplantation attack. For this purpose, the authors repackaged the existing application with camera permission. The app was tested for 69 different phones from 8 different vendors with different Android versions; it gave a success rate of almost 46%, meaning that half of the phones worldwide are susceptible to transplantation attacks. To defend against such attacks, the authors have proposed two steps solutions: separating permission and group ID and implementing SEAndroid policy [89].

Table 2: A summary of spyware detection methods.

Research	**Technique**	**Result**
M. Naser et al.	ML models, specifically SVM, NB, and FDT, were used	Accuracy=98.2% (for FDT)
E. Liu et al.	Hybrid approach with source code analysis and behavior analysis by examining protocol etc.	Different for each of the fourteen apps concerning each threat
M. K. Qablain et al.	Obtained a novel dataset and applied many ML models to detect spyware	RF binary model performed best with an accuracy of almost 79%

F. Pierazzi et al.	Combination traditional ML classifiers and DL classifiers with ELF	For goodware vs. spyware, f1=0.982 and AUC=0.982
D. Harkin et al.	Behavior analysis based on protocol analysis	Concluded that Android systems are more to spyware than iOS.
H. M. Salih et al.	Unearthing the spyware behavior of a game	Deduced that security should be strengthened further to avert such incidences
M. Conti et al.	Three ML algorithms, RF, LR, and KNN, were used	The overall accuracy of 85%, whereas LR was the most time efficient
F. Fasano et al.	Temporal Logic	Accuracy=98%
S. Hutchinson et al.	Forensic analysis	Proposal of a framework for spyware researchers
R. Zhang et al.	Attacking mobile phones through voice assistant using AI	It successfully evaded anti-virus applications, and another vulnerability exposed
R. Chatterjee et al.	ML for classifying applications as malicious or not	Unearthed much-undetected spyware and much dual-purposed spyware
M. H. Saad et al.	Static analysis using Fuzz testing	Successfully detected 19 out of 20 spyware applications
H. Abulola et al.	Exploiting the notification listening capability of Android for different applications	A loophole in 'Notification Listener' was identified.
P. Kaur et al.	Hybridization of Description analysis, Interface analysis, and Source code analysis	Detected even those spywares which evaded the prevalent antiviruses
Z. Zhang et al.	Discovering security loophole in Android through transplantation attack	Achieved an overall 46% success

Table 1: Advantages and disadvantages of different techniques used for spyware detection.

Technique used	**Advantage**	**Disadvantage**
ML algorithms	Accurate and precise	There can be the problem of false positives and false negatives
Behavior-based techniques	Applications with spyware behavior can be easily recognized	Some spyware applications may not behave surreptitiously
Traffic Analysis	Traffic generated by spyware has certain characteristics that can lead to its detection.	Legitimate spyware applications also generate the same traffic.
Permission analysis	Spyware has specific permissions that can be traced	Some benign apps also need the same permissions

5 Discussions

This paper introduced the latest techniques used to detect spyware in mobile phones. The pros and cons of these techniques are also exhibited in this paper. These techniques include ML algorithms, behavior-based techniques, traffic analysis, and permission analysis. The ML algorithms are precise and accurate. However, they have issues regarding the false positives and false negatives. It can easily recognize the applications that used spyware behavior. As a con of behavior-based techniques, there exists spyware applications that could behave surreptitiously. The traffic analysis has the ability to detect the spyware from characteristics of the traffic generated by the spyware. Nevertheless, legitimate spyware applications generate the same traffic. The permission analysis is able to trace the specific permissions of the spyware. Nonetheless, there exists benign apps requiring the same permissions.

It can be inferred from the analyzed paper that most of the experiments were performed on datasets of pre-meditated spyware and virtual environments. In very few instances, data was collected from real environments. This is problematic because spyware in a real environment may vary much more than in a virtual environment.

Moreover, most of the literature focuses on the general malware of mobile phones. Spyware, in specific, is the very least targeted in research. Spyware should be focused on the most because of its stealthy nature and its covert way of action. Besides that, spyware issues in Android-based phones are investigated the most. That is because Android is the most targeted platform by spyware perpetrators. The reason Android is targeted the most by threats is its popularity and open nature. Researchers must realize that spyware issue should also be investigated in other platforms like iOS and IoTs as well.

In the so far literature, most identification methods follow behavioral detection. It has proved to be very effective. Other methods, too, need to be employed.

There is very least focus on cloud-based IDS among the researchers. So, they should focus on cloud-based IDS. Such systems give the mobile phones freedom from intensive processing which is the scarcest resource for a mobile phone [13].

Researchers must also consider fast detection solutions because of that a slower detection may provide enough time to an attacker to have the mission accomplished till the system detects it as a threat.

6 Conclusions

The increasing popularity of mobile phones is resulting in so many security challenges. Among these security challenges, spyware is the most prevalent one. It can harm the victim’s device directly by stealing information or opening the way to other malicious software. Researchers have been trying to curb this menace. Some research has so far been carried out on the issue. This paper surveyed many of the techniques for detecting spyware in mobile phones, analyzing the most recently proposed methods and techniques, the achieved results of each proposed method, and the most relevant were discussed here. This paper would serve as a reference point for the researchers of mobile spyware domain.

Acknowledgement

The authors extend their appreciation to the Arab Open University for funding this work through AOU research fund No. (AOURG-2023-011).

References

[1] P. Taylor, "Forecast number of mobile users worldwide 2020-2025," 18 Jan 2023. [Online]. Available: https://www.statista.com/statistics/218984/number-of-global-mobile-users-since-2010/.

[2] "The Mobile Economy," GSMA, 2022.

[3] "Android," 2023. [Online]. Available: https://www.android.com/. [Accessed 19 Feb 2023].

[4] "ios 16," 2023. [Online]. Available: https://www.apple.com/ios/ios-16/. [Accessed 19 Feb 2023].

[5] F. Laricchia, "Mobile operating systems' market share worldwide from 1st quarter 2009 to 4th quarter 2022," 16 Nov 2022. [Online]. Available: https://www.statista.com/markets/418/topic/481/telecommunications/. [Accessed 14 Jan 2023].

[6] V. KOULIARIDIS, K. BARMPATSALOU, G. KAMBOURAKIS and S. CHEN, "A Survey on Mobile Malware Detection Techniques," *IEICE Transactions on Information and Systems,* vol. E103D, no. 2, pp. 204-211, 2020.

[7] T. F. Stafford and A. Urbaczewski, "Spyware: The Ghost in the Machine," *Communications of the Association for Information Systems,* vol. 14, pp. 291-306, 2004.

[8] C. Brooks, "MORE Alarming Cybersecurity Stats For 2021 !," 24 Oct 2021. [Online]. Available: https://www.forbes.com/sites/chuckbrooks/2021/10/24/more-alarming-cybersecurity-stats-for-2021-/?sh=2521f4714a36.

[9] Hearing before the Committee on Conference, Science and Transportation, "impact and policy implications of spyware on consumers and businesses," u.s. government printing office, washington, 2012.

[10] S. Morgan, "Cybercrime To Cost The World $10.5 Trillion Annually By 2025," 13 Nov 2020. [Online]. Available: https://cybersecurityventures.com/hackerpocalypse-cybercrime-report-2016/.

[11] A. Ilumba, "Most Americans Think Big Tech Is Spying On Them," 24 Sep 2020. [Online]. Available: https://www.whistleout.com/CellPhones/Guides/americans-think-companies-are-spying.

[12] B. Amro, "MALWARE DETECTION TECHNIQUES FOR MOBILE DEVICES," *International Journal of Mobile Network Communications & Telematics,* vol. 7, no. 4/5/6, pp. 1-10, 2017.

[13] Y. S. I. Hamed, S. N. A. AbdulKader and a. M.-S. M. Mostafa, "Mobile Malware Detection: A Survey," *International Journal of Computer Science and Information Security (IJCSIS),* vol. 17, no. 1, pp. 56-65, 2019.

[14] Z. WANG, Q. LIU and Y. CHI, "Review of Android Malware Detection Based on Deep Learning," *IEEE Access,* vol. 8, no. 2020, pp. 181102-181126, 2020.

[15] R. VINAYAKUMAR, M. ALAZAB, K. P. SOMAN, P. POORNACHANDRAN and A. S. VENKATRAMAN, "Robust Intelligent Malware Detection Using Deep Learning," *IEEE Access,* vol. 7, no. 2019, pp. 46717-46738, 2019.

[16] M. Ashawa and S. Morris, "ANALYSIS OF ANDROID MALWARE DETECTION TECHNIQUES: A SYSTEMATIC REVIEW," *International Journal of Cyber-Security and Digital Forensics (IJCSDF),* vol. 8, no. 3, pp. 177-187, 2019.

[17] E. M. karanja, S. Masupe and J. Mandu, "INTERNET OF THINGS MALWARE: A SURVEY," *International Journal of Computer Science & Engineering Survey (IJCSES),* vol. 8, no. 3, pp. 1-20, 2017.

[18] S. Wang, J. Wang, Y. Song and S. Li, "Malicious Code Variant Identification Based on Multiscale Feature Fusion CNNs," *Computational Intelligence and Neuroscience,* vol. 2021, 2021.

[19] H. Abualola, H. Alhawai, M. Kadadha, H. Otrok and A. Mourad, "An Android-based Trojan Spyware to Study the NotificationListener Service Vulnerability," in *The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016),* 2016.

[20] P. Kohli, "Android APT spyware, targeting Middle East victims, enhances evasiveness," 23 Nov 2021. [Online]. Available: https://news.sophos.com/en-us/2021/11/23/android-apt-spyware-targeting-middle-east-victims-improves-its-capabilities/.

[21] M. Boldt, B. Carlsson and A. Jacobsson, "Exploring Spyware Effects," Semantic Scholar, 2004.

[22] E. Skoudi and L. Zeltser, Malware: Fighting Malicious Code, Pearson, 2004.

[23] S. Lysenko, K. Bobrovnikova, P. Popov, V. Kharchenko and D. Medzaty, "Spyware Detection Technique Based on Reinforcement Learning," in *International Workshop on Intelligent Information Technologies & Systems of Information Security*, 2020.

[24] A. Moshchuk, T. Bragin, S. D. Gribble and H. M. Levy, "A Crawler-based Study of Spyware in the Web," in *Network and Distributed System Security Symposium*, San Diego, California, USA, 2006.

[25] M. Almansoori, A. Gallardo, J. Poveda, A. Ahmed and R. Chatterjee, "A Global Survey of Android Dual-Use Applications Used in Intimate Partner Surveillance," in *Proceedings on Privacy Enhancing Technologies*, 2022.

[26] S. Srivatsa and sksrivatsa, "Android Security Issues," WUSTL, 2014.

[27] S. Acharya, U. Rawat and R. Bhatnagar, "A Comprehensive Review of Android Security: Threats, Vulnerabilities, Malware Detection, and Analysis," *Security and Communication Networks*, vol. 2022, pp. 1-34, 2022.

[28] R. Mayrhofer, J. V. Stoep, C. Brubaker and N. Kralevich, "The Android Platform Security Model," *ACM Transactions on Privacy and Security*, vol. 24, no. 3, pp. 1-35, Aug 2021.

[29] J. P. d. Wit, D. Bucur and J. v. d. Ham, "Dynamic Detection of Mobile Malware Using Smartphone Data and Machine Learning," *Digital Threats: Research and Practice*, vol. 3, no. 2, pp. 1-24, June 2022.

[30] D. Arp, M. Spreitzenbarth, M. Hubner, H. Gascon and K. Rieck, "DREBIN: Effective and Explainable Detection of Android Malware in Your Pocket," in *Proceeding of 17th Network and Distributed System Security Symposium, NDSS. 14*, 2014.

[31] Y. Li, J. Jang, X. Hu and X. Ou, "Android malware clustering through malicious payload mining," in *International Symposium on Research in Attacks, Intrusions, and Defenses*, 2017.

[32] A. H. Lashkari, A. F. A.Kadir, H. Gonzalez, K. F. Mbah and A. A. Ghorbani, "Towards a Network-Based Framework for Android Malware Detection and Characterization," in *15th International Conference on Privacy, Security and Trust, PST*, Calgary, Canada, 2017.

[33] M. Damshenas, A. Dehghantanha, K.-K. R. Choo and R. Mahmud, "M0Droid: An Android Behavioral-Based Malware Detection Model," *Journal of Information Privacy and Security*, vol. 11, no. 3, pp. 141-157, 2015.

[34] "Packet Capture," 18 Feb 2023. [Online]. Available: https://kismetwireless.net/docs/api/packet_capture/. [Accessed 18 Feb 2023].

[35] "NETRESEC," 18 Feb 2023. [Online]. Available: https://www.netresec.com/?page=NetworkMiner. [Accessed 18 Feb 2023].

[36] "TCPDUMP MAIN PAGE," 18 Feb 2023. [Online]. Available: https://www.tcpdump.org/manpages/tcpdump.1.html. [Accessed 18 Feb 2023].

[37] 18 Feb 2023. [Online]. Available: https://www.wireshark.org/. [Accessed 18 Feb 2023].

[38] J. Landage and P. M. P. Wankhade, "Malware and Malware Detection Techniques: A Survey," *International Journal of Engineering Research & Technology*, vol. 2, no. 12, pp. 61-68, Dec 2013.

[39] R. Tahir, "A Study on Malware and Malware Detection Techniques," *I.J. Education and Management Engineering*, vol. 2018, no. 2, pp. 20-30, 2018.

[40] "What Is a Sandbox?," 19 Feb 2023. [Online]. Available: https://www.proofpoint.com/us/threat-reference/sandbox#:~:text=In%20the%20world%20of%20cybersecurity,URLs%20and%20observe%20its%20behavior.. [Accessed 19 Feb 2023].

[41] anuxraw, "Regshot," 16 Jun 2019. [Online]. Available: https://github.com/Seabreg/Regshot. [Accessed 19 Feb 2023].

[42] Mark Russinovich, "Process Explorer v17.02," 19 Feb 2023. [Online]. Available: https://learn.microsoft.com/en-us/sysinternals/downloads/process-explorer. [Accessed 19 Feb 2023].

[43] M. Odusami, O. Abayomi-Alli, S. Misra, O. Shobayo, R. Damasevicius and R. Maskeliunas, "Android Malware Detection: A Survey," in *International Conference on Applied Informatics, ICAI*, 2018.

[44] 17 Feb 2023. [Online]. Available: https://ieeexplore.ieee.org/Xplore/home.jsp. [Accessed 17 Feb 2023].

[45] 17 Feb 2023. [Online]. Available: https://www.mdpi.com/. [Accessed 17 Feb 2023].

[46] 17 Feb 2023. [Online]. Available: https://www.sciencedirect.com/. [Accessed 17 Feb 2023].

[47] 17 Feb 2023. [Online]. Available: https://www.acm.org/. [Accessed 17 Feb 2023].

[48] N. Xu, F. Zhang, Y. Luo and W. Jia, "Stealthy Video Capturer: A New Video-Based Spyware in

3G Smartphones," in *Second ACM conference on Wireless network security*, 2009.

[49] D. Wu, G. D. Moody, J. Zhang and P. B. Lowry, "Effects of the design of mobile security notifications and mobile app usability on users' security perceptions and continued use intention," *Information & Management,* vol. 57, no. 5, 2020.

[50] A. Saranya and R. Naresh, "Efficient Mobile Security for E Health Care Application in Cloud for Secure Payment Using Key Distribution," *Neural Processing Letters,* no. 2021, 2021.

[51] A. C. Cinar and T. B. Kara, "The current state and future of mobile security in the light of the recent mobile security threat reports," *Multimedia Tools and Applications,* no. 2023, 2023.

[52] Z. Wan, L. Bao, D. Gao, E. Toch, X. Xia, T. Mendel and D. Lo, "AppMoD: Helping Older Adults Manage Mobile Security with Online Social Help," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies,* vol. 3, no. 4, pp. 1-22, 2019.

[53] M. Bahrini, G. Volkmar, J. Schmutte, N. Wenig, K. Sohr and R. Malaka, "Make my Phone Secure!: Using Gamification for Mobile Security Settings," in *Proceedings of Mensch und Computer*, 2019.

[54] H. Shahriar, M. A. Talukder and M. S. Islam, "An Exploratory Analysis of Mobile Security Tools," in *2019 KSU CONFERENCE ON CYBERSECURITY EDUCATION, RESEARCH AND PRACTICE*, 2019.

[55] P. Ratazzi, A. Bommisetti, N. Ji and W. Du, "PINPOINT: Efficient and Effective Resource Isolation for Mobile Security and Privacy," in *Mobile Security Technologies (MoST) 2015*, 2015.

[56] D. Quang, B. Martini and C. K.-K. Raymond, "The role of the adversary model in applied security research," *Computers & Security,* vol. 81, pp. 156-181, 2018.

[57] T. Moletsane and P. Tsibolane, "Mobile Information Security Awareness Among Students in Higher Education : An Exploratory Study," in *2020 Conference on Information Communications Technology and Society (ICTAS)*, Durban, South Africa, 2020.

[58] A. Balapour, H. R. Nikkhah and R. Sabherwal, "Mobile application security: Role of perceived privacy as the predictor of security perceptions," *International Journal of Information Management,* vol. 52, no. Jun 2020, Jun 2020.

[59] A. K. V. C. Attia Qamar, "Mobile malware attacks: Review, taxonomy & future directions," *Future Generation Computer Systems,* vol. 97, pp. 887-909, 2019.

[60] M. A. A. S. A. M. A. A. Moutaz Alazab, "Intelligent mobile malware detection using permission requests and API calls," *Future Generation Computer Systems,* vol. 107, pp. 509-521, 2020.

[61] A. S. Francesco Mercaldo, "Deep learning for image-based mobile malware detection," *Journal of Computer Virology and Hacking Techniques,* vol. 16, no. 2020, p. 157–171, 2020.

[62] Z. C. Q. Y. B. Y. L. P. Z. J. Shanshan Wang, "A mobile malware detection method using behavior features in network traffic," *Journal of Network and Computer Applications,* vol. 133, pp. 15-25, 2019.

[63] M. Naser and Q. A. Al-Haija, "Spyware Identification for Android Systems Using Fine Trees," *information,* vol. 14, no. 2, pp. 1-10, 2023.

[64] skylot, "jadx," 19 Feb 2023. [Online]. Available: https://github.com/skylot/jadx. [Accessed 2023 Feb 19].

[65] E. Liu, S. Rao, S. Havron, G. Ho, S. Savage, G. M. Voelker and D. McCoy, "No Privacy Among Spies: Assessing the Functionality and Insecurity of Consumer Android Spyware Apps," *Proceedings on Privacy Enhancing Technologies,* vol. YYYY, no. X, pp. 1-18, 2023.

[66] "The Best Phone Tracker," 19 Feb 2023. [Online]. Available: https://www.mspy.com/. [Accessed 19 Feb 2023].

[67] "Advanced Cell Phone Tracker," 2023. [Online]. Available: https://umobix.com/. [Accessed 19 Feb 2023].

[68] "Mobile Spy," 2023. [Online]. Available: https://www.mobile-spy.com/. [Accessed 19 Feb 2023].

[69] "FLEXISPY," 2023. [Online]. Available: https://www.flexispy.com/?a_aid=cb523eda. [Accessed 19 Feb 2023].

[70] "Monitor Kids," 2023. [Online]. Available: https://www.thewispy.com/. [Accessed 19 Feb 2023].

[71] M. K. Qabalin, M. Naser and M. Alkasassbeh, "Android Spyware Detection Using Machine Learning: A Novel Dataset," *sensors,* vol. 22, no. 15, pp. 1-25, 2022.

[72] Hispasec Sistemas, "VIRUSTOTAL," 2023. [Online]. Available: https://www.virustotal.com/gui/home/upload. [Accessed 19 Feb 2023].

[73] F. PIERAZZI, G. MEZZOUR, Q. HAN, M. COLAJANNI and V. S. SUBRAHMANIAN, "A Data-driven Characterization of Modern Android Spyware," *ACM Transactions on Management Information Systems,* vol. 11, pp. 1-28, April 2020.

[74] D. Harkin and A. Molnar, "Operating-System Design and Its Implications for Victims of Family Violence: The Comparative Threat of Smart Phone Spyware for Android Versus iPhone Users," *Violence Against Women,* vol. 27, no. 6-7, pp. 1-25, 2020.

[75] H. M. Salih and M. S. Mohammed, "Spyware Injection in Android using Fake Application," in *International Conference on Computer Science and*

Software Engineering (CSASE), Duhok, Kurdistan Region – Iraq, 2020.

[76] M. Conti, G. Rigoni and F. Toffalini, "ASAINT: A Spy App Identification System based on Network Traffic," in *The 15th International Conference on Availability, Reliability and Security (ARES 2020)*, Ireland, 2020.

[77] J. Bradfield and C. Stirling, "12 Modal mu-calculi," *Studies in Logic and Practical Reasoning*, vol. 3, no. 2007, pp. 721-756, 2007.

[78] subho007, "AFE," 27 Sep 2015. [Online]. Available: https://github.com/appknox/AFE. [Accessed 20 Feb 2023].

[79] V. Rastogi, Y. Chen and X. Jiang, "DroidChameleon: evaluating Android anti-malware against transformation attacks," in *8th ACM SIGSAC symposium on Information, computer and communications security*, 2013.

[80] F. Fasano, F. Martinelli, F. Mercaldo, V. Nardone and A. Santone, "Spyware Detection using Temporal Logic," in *5th International Conference on Information Systems Security and Privacy (ICISSP 2019)*, 2019.

[81] A. Bhatia, "ashishb/android-malware," 28 Apr 2016. [Online]. Available: https://github.com/ashishb/android-malware/blob/master/Android.Spy.277.origin/4f2c13cd7d1eb0ff87ed7805faf0b48f40b9f1aa1782ccaf0916bc7ec37360b6. [Accessed 20 Feb 2023].

[82] S. Hutchinson and U. Karabiyik, "FORENSIC ANALYSIS OF SPY APPLICATIONS IN ANDROID DEVICES," in *Annual ADFSL Conference on Digital Forensics, Security and Law*, 2019.

[83] R. ZHANG, X. CHEN, S. WEN, X. ZHENG and Y. DING, "Using AI to Attack VA: A Stealthy Spyware Against Voice Assistances in Smart Phones," *IEEE Access*, vol. 7, pp. 153542-153554, 2019.

[84] "The FREE antivirus you’re looking for," 2023. [Online]. [Accessed 19 Feb 2023].

[85] "McAfee," 2023. [Online]. Available: https://www.mcafee.com/. [Accessed 19 Feb 2023].

[86] P. D. H. O. S. H. J. P. D. F. Rahul Chatterjee, "The Spyware Used in Intimate Partner Violence," in *IEEE Symposium on Secuirty and Privacy (SP)*, 2018.

[87] A. S. a. G. I. S. Mustafa Hassan Saad, "Android Spyware Disease and Medication," in *2015 Second International Conference on Information Security and Cyber Forensics (InfoSec)*, Cape Town, South Africa, 2016.

[88] S. S. Parmjit Kaur, "Spyware Detection in Android Using Hybridization of Description Analysis, Permission Mapping and Interface Analysis," in *Proceedings of the International Conference on Information and Communication Technologies, ICICT 2014, 3-5 December 2014*, Bolgatty Palace & Island Resort, Kochi, India, 2015.

[89] P. L. J. X. J. J. L. L. Zhongwen Zhang, "How Your Phone Camera Can Be Used to Stealthily Spy on You: Transplantation Attacks against Android Camera Service (CODASPY '15)," in *5th ACM Conference on Data and Application Security and Privacy*, 2015.

[90] Z. Ling, J. Luo, K. Wu, W. Yu and a. X. Fu, "TorWard: Discovery of Malicious Traffic over Tor," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, 2014.