Spyware Detection using Temporal Logic

Fausto Fasano¹, Fabio Martinelli², Francesco Mercaldo²,¹, Vittoria Nardone³ and Antonella Santone¹

¹Department of Bioscience and Territory, University of Molise, Pesche (IS), Italy
²Istituto di Informatica e Telematica, Consiglio Nazionale delle Ricerche, Pisa, Italy
³Department of Engineering, University of Sannio, Benevento, Italy

{fabio.martinelli, francesco.mercaldo}@iit.cnr.it, vnardone@unisannio.it

Keywords: Malware, Android, Security, Formal Methods, Temporal Logic.

Abstract: In recent years smartphones have become essential in daily life. A user can perform several operations through a smartphone since they are increasingly similar to a personal computer. Furthermore, smartphones collect a large number of sensitive information. The most widespread mobile operating system is Android, this is the reason why malware writers target this platform. Malicious behaviours able to steal private information are called spyware. This paper aims to detect this kind of threat in mobile environment: we present a preliminary framework able to recognize Android spyware. It is based on model checking technique and it uses temporal logic formulae to identify malicious behaviours. We evaluate the proposed framework using a synthetic dataset obtaining a precision equal to 0.98 and a recall equal to 1.

1 INTRODUCTION

Mobile device currently permeate our everyday activity. From back transaction, to update the status on social networks, mobile devices allow us to perform a variety of activities. As a matter of fact, smartphone sales exceeded the current X86 PC platform in 2016, and this trend is expected to grow up in 2018¹.

Mobile devices quickly attracted the interest of the attackers, and it is easy to understand the reason why: if compared with PC platforms, in our smartphones are stored more and more sensitive and private information. Furthermore, smartphones manage the SIM card in which there is our credit, also for this reason this is an appealing attack surface for malicious software writers (Cimitile et al., 2018), (Mercaldo et al., 2016a).

Mobile operating systems producers tried to remedy to this rampant spread of malicious software targeting mobile platform.

For instance, Google with the aim to consent the publication of a new app on Play Store (the official market for Android users) requires a deep scan of the app aimed to find possible malicious activities. Indeed the new app must be submitted to Bouncer (Oberheide and Miller, 2012), an automatic application scanning system introduces in 2012 with following distinctive features, including:

- static analysis in search of known threats;
- it runs the software in a virtual emulator (QEMU) and identifies its behavior;
- it starts and tracks the behavior of the app for 5 minutes;
- it explores the app in every button.

Bouncer performs a static analysis using the antivirus software provided by VirusTotal (a service able to evaluate the application simultaneously with 60 different antivirus) but, considering the signature-based detection approach offered by current antivirus technologies, it is possible to mark a malicious sample as malware only whether their signature is stored into the antivirus repository (and consequently it is not possible to detect zero-day threat).

With regard to the dynamic analysis, the app is ran for a limited time window (5 minutes): in case the app does not exhibit the malicious behaviour in this period it passes this test. Furthermore, usually malware is able to understand whether it is executed on a virtual environment (in this case it will not perform the malicious action, to avoid the sandbox detection).

For these reasons, it is easy from malicious writers to elude the current detection (Canfora et al., 2018;
Spyware Detection using Temporal Logic

Cimitile et al., 2017; Mercaldo et al., 2016b; Canfora et al., 2015b).

The preferred target of mobile malicious software is represented by ourselves: this is the reason why usually mobile malware is able to secretly record phone calls, collect images, videos, text messages and even the GPS coordinates of the victims and send them to the attackers and, generally speaking, to spy the infected users (this is the reason why this kind of malicious software is called spyware).

This is the reason why in this paper we present a framework able to detect Android spyware. In particular, we develop a model checking based framework identifying this kind of threat. Our solution is behavioural based since it is able to detect the malicious spyware using temporal logic formulae. The considered logic rules are the formal specification of the malicious behaviour performed by a spyware sample. The framework models an android application as a labeled transition system starting from its bytecode. Then, using a model checker tool, it verifies the specified malicious behaviour against the model of the application. The output of the model checker, and thus of our framework, is binary: it is equal to true when the formula is verified on the model and false otherwise. Our method considers an application under analysis as spyware if the output of model checker is equal to true.

The paper proceeds as follows: next section introduces background concepts related to Model Checking and Mu-Calculus Logic exploited by the proposed framework, Section 3 describes our method aimed to detect Android spyware, Section 4 presents the performance evaluation of the proposed framework and, finally, conclusion and future work are discussed in Section 6.

2 MODEL CHECKING AND Mu-Calculus LOGIC

Verification of a software or hardware system involves checking whether the system in question behaves as it was designed to behave. Formal methods have been successfully applied to safety-critical systems (Santone et al., 2013) and in other domains such as biology (Ruvo et al., 2015; Ceccarelli et al., 2014).

One reason is the overwhelming evidence that formal methods do result in safer systems. In this paper we show that formal methods are extremely well-suited to spyware detection. First of all, in this section we recall some basic concepts.

Model checking is an formal method for determining if a model of a system satisfies a correctness specification (Clarke et al., 2001). A model of a system consists of a labelled transition system (LTS). A specification or property is a logical formula. A model checker then accepts two inputs, a LTS and a temporal formula, and returns true if the system satisfies the formula and false otherwise.

A labelled transition system comprises some number of states, with arcs between them labelled by activities of the system. A LTS is specified by:

- a set $S$ of states;
- a set $L$ of labels or actions;
- a set of transitions $T \subseteq S \times L \times S$.

Transitions are given as triples $\langle start, label, end \rangle$.

In this paper, to express proprieties of the system we use the modal mu-calculus (Stirling, 1989) which is one of the most important logics in model checking.

The syntax of the mu-calculus is the following, where $K$ ranges over sets of actions (i.e., $K \subseteq L$) and $Z$ ranges over variables:

$$\varphi ::= tt | ff | Z \mid \varphi \land \varphi \mid \varphi \lor \varphi \mid [K] \varphi \mid \langle K \rangle \varphi \mid vZ.\varphi \mid \mu Z.\varphi$$

A fixpoint formula may be either $\mu Z.\varphi$ or $vZ.\varphi$ where $\mu Z$ and $vZ$ binds free occurrences of $Z$ in $\varphi$. An occurrence of $Z$ is free if it is not within the scope of a binder $\mu Z$ (resp. $vZ$). A formula is closed if it contains no free variables, $\mu Z.\varphi$ is the least fixpoint of the recursive equation $Z = \varphi$, while $vZ.\varphi$ is the greatest one. From now on we consider only closed formulæ.

Domains of fixpoint variables, free and bound variables, can be defined in the mu-calculus in analogy with variables of first order logic.

The satisfaction of a formula $\varphi$ by a state $s$ of a transition system is defined as follows:

- each state satisfies $tt$ and no state satisfies $ff$;
- a state satisfies $\varphi_1 \lor \varphi_2$ ($\varphi_1 \land \varphi_2$) if it satisfies $\varphi_1$ or (and) $\varphi_2$. $[K] \varphi$ is satisfied by a state which, for every performance of an action in $K$, evolves to a state obeying $\varphi$. $\langle K \rangle \varphi$ is satisfied by a state which can evolve to a state obeying $\varphi$ by performing an action in $K$.

For example, $\langle a \rangle \varphi$ denotes that there is an $a$-successor in which $\varphi$ holds, while $[a] \varphi$ denotes that for all $a$-successors $\varphi$ holds.

The precise definition of the satisfaction of a closed formula $\varphi$ by a state $s$ ($\text{written } s \models \varphi$) is given in Table 1.

A fixed point formula has the form $\mu Z.\varphi$ ($vZ.\varphi$) where $\mu Z$ ($vZ$) binds free occurrences of $Z$ in $\varphi$. An occurrence of $Z$ is free if it is not within the scope of a binder $\mu Z$ ($vZ$). A formula is closed if it contains
Table 1: Satisfaction of a closed formula by a state.

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\neg$</th>
<th>$\perp$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>$\neg$</td>
<td>$\perp$</td>
</tr>
<tr>
<td>$p \models \varphi \land \psi$</td>
<td>iff</td>
<td>$p \models \varphi$ and $p \models \psi$</td>
</tr>
<tr>
<td>$p \models \varphi \lor \psi$</td>
<td>iff</td>
<td>$p \models \varphi$ or $p \models \psi$</td>
</tr>
<tr>
<td>$p \models \langle K \rangle \varphi$</td>
<td>iff</td>
<td>$\forall p'. \forall \alpha \in K. p \overset{\alpha}{\rightarrow} K \cup R p' ; \text{implies} ; p' \models \varphi$</td>
</tr>
<tr>
<td>$p \models \nu Z \varphi$</td>
<td>iff</td>
<td>$p \models \nu Z^n \varphi$ for all $n$</td>
</tr>
<tr>
<td>$p \models \mu Z \varphi$</td>
<td>iff</td>
<td>$p \models \mu Z^n \varphi$ for some $n$</td>
</tr>
</tbody>
</table>

where:
- for each $n$, $\nu Z^n \varphi$ and $\mu Z^n \varphi$ are defined as:
  - $\nu Z^0 \varphi = \perp$
  - $\nu Z^{n+1} \varphi = \varphi[\nu Z^n \varphi/Z]
  - $\mu Z^0 \varphi = \top$
  - $\mu Z^{n+1} \varphi = \varphi[\mu Z^n \varphi/Z]

We can then add a predicate $p$, and obtain the formula:

$vY. p \land \langle a \rangle Y$

saying that “there is an infinite sequence of $a$-transitions, and all states in this sequence satisfy $p$”.

With two fixpoints, we can write fairness formulae, such as:

$vY. \mu X. (p \land \langle a \rangle Y) \lor \langle a \rangle X$

meaning that “on some $a$-path there are infinitely many states where $p$ holds”.

Changing the order of fixpoints we obtain:

$\mu X. vY. (p \land \langle a \rangle Y) \lor \langle a \rangle X$

saying “on some $a$-path almost always $p$ holds”.

In this paper we use CAAL (Concurrency Workbench, Aalborg Edition) (Andersen et al., 2015) as formal verification environment. It is one of the most popular environments for verifying systems. In the CAAL the verification of temporal logic formulae is based on model checking (Clarke et al., 2001).

3 A FORMAL FRAMEWORK FOR SPYWARE DETECTION

In this section we describe our approach aimed to detect spyware Android applications. The approach models the Android application under analysis as a labelled transition system capturing the behaviour of...
3.1 Spyware Characterization through Temporal Logic Formulae

Temporal logic allows us to reason about changes in the behavior of a system over time, without explicitly mentioning specific instances of time. In particular, a formula may specify that some property eventually turns true, or always holds, or never turns true. In this section we use the mu-calculus logic to specify the spyware behaviour occurring in Android applications.

We consider the model checking technique to detect spyware application for the following main reasons:

- The checking process is automatic. There is no need to construct a correctness proof.
- The possibility of using the diagnostic counterexamples. If the specification is not satisfied, the model checker will produce a counterexample execution trace that shows why the specification does not hold. The counterexamples are invaluable in analyzing an application, since they can be used to understand where the spyware behaviour is in the application under analysis.
- Temporal logic can easily and correctly express the behaviour of a spyware application.
- There is no problem with partial specifications. It is unnecessary to completely specify all the application before beginning to model check properties. Thus, model checking can be used only to verify part (methods) of the application.
- Formal verification allows evaluating all possible scenarios, the entire state space all at once. Model checking allows checking if, in each state, the system obeys certain properties. In particular, it allows verifying if the system under analysis exposes a certain behaviour expressed using a temporal logic formula. Spyware is a malware able to perform harmful actions in order to steal sensitive information. Basically, it is a software exposing in its code some malicious behaviours. Roughly speaking, in its code, there are some instructions performing these actions. We can imagine this like a software specification: the software is designed to do something malicious. Now, applying formal verification we investigate whether the software exhibits this malicious behaviour.

Table 2 shows an example of temporal logic formula written in mu-calculus logic. It catches the reading phone contacts suspicious behaviour. In Android environment the ContentProvider allows reading phone contacts. In order to access to all contact information a ContentResolver object
must be used. In our logic formula this operation is specified by the action `invokegetContentResolver`. After that it is necessary to communicate with the contacts applications performing a query to the URL of the contacts table (URI: `ContactsContract.Contacts.CONTENT_URI`). This step is specified in our logic formula by the sequence of actions:
`getstatic android.provider.ContactsContract.Contacts and invokequery`. Finally the action `invokegetString` returns the contacts information as contact name, contact number, etc.

In order to better understand the behaviour specified in our logic formula, we report the corresponding Java code snippet in Figure 2. In particular, the line highlighted in yellow shows the query to Content Provider and the lines corresponding to get the contact information (i.e., invocation of the `getString` method in Figure 2). Our logic formula specifies in mu-calculus logic the instructions show in Figure 2.

It should be underlined that we have formulated also the formula able to catch read phone contact for Android application with an API level less than or equal to 5. We have specify also the formula considering the URI: Contacts.Phones, deprecated in API level 5. The formula verified on the applications is $\kappa$. It is the logical disjunction between the formula considering the API levels greater than the API level 5 ($\xi$) and the formula considering the other ones less than or equal to API level 5 ($\gamma$). In the following manner the formula covers all the Android API levels.

4 EXPERIMENTAL EVALUATION AND ASSESSMENT

In the following section, we detail how we generated the experimental dataset and we discuss the performances obtained by the proposed framework. In order to evaluate the effectiveness of the proposed method, we generated a set of Android spyware exploiting a framework able to automatically generate malicious samples: the Android Framework for Exploitation.

4.1 Android Framework for Exploitation

The Android Framework for Exploitation (i.e., AFE) is an open-source python-based project aimed to evaluate Android vulnerabilities. It is composed by several modules, we exploit the Malware Creator and the Stealer (able to inject code with the ability to steal information from the attacked device including contacts, call logs, text messages and files from SD card).

Basically the Malware Creator module in order to inject the malicious behaviour implemented in the Steal module, it considers a pre-defined template able to embed the malicious payload (provided by the Steal module) and call it from a Service (declared in the Android Manifest file): the Service will be called when the Main activity is called (i.e., when the application is launched on the infected mobile device).

Basically, AFE considers following steps to automatically inject the malicious code into a legitimate applications: (i) it decompiles it into the small language, (ii) the malicious payload is added and (iii) the app with the spyware behaviour is rebuilt.

Figure 3 depicts the difference between an Android application before and after the AFE injection.

As shown in Figure 3 in the injected version there is the xybot package added by AFE containing the spyware malicious payload.

Figure 4 shows the classes included in the xybot package.

The main class responsible for the malicious behaviours is `com.xybox.infect.class` (highlighted from a red circle in Figure 4): a java byte-code snippet belonging to this class is shown in Figure 5.

From the snippet in Figure 5 it is possible to see the device contact gathering malicious action: as a matter of fact, basic contact information in Android are stored in Contacts table with detailed information stored in individual tables. The snippet shows a query to retrieve the records stored in `ContactsContract.Contacts.CONTENT_URI` (the instruction is highlighted by the red arrow).

4.2 Dataset Building

In order to evaluate the effectiveness of the proposed framework, a dataset composed by legitimate and spyware Android applications is considered. We collected 80 freely applications belonging to 26 different categories from Google Play Store (i.e., Books and Reference, Lifestyle, Business, Live Wall- paper, Comics, Media and Video, Communication, Medical, Education, Music and Audio, Finance and News, Magazines, Games, Personalization, Health and Fitness, Photography, Libraries and Demo, Productivity, Shopping, Social, Sport, Tools, Travel, Local and Transportation, Weather, Widgets). Their dimensions are ranging from 24 kB to 37 MB. We have selected an equal number of applications belonging to each

---

2https://github.com/appkno/AFE
category. The applications were downloaded in the time-window between March 2018 and April 2018.

We submitted the Play Store apps to the VirusTotal service: whether the 59 antimalware provided by VirusTotal marked as clean the application, we label the application as trusted.

To embed into the legitimate applications the spyware malicious behaviour we considered the AFE framework. For each applications downloaded from Play Store, through AFE a spyware version of the application was generated. We labeled the applications generated by AFE as spyware.

Furthermore, we generated an obfuscated version for each application submitted to the AFE framework using DroidChameleon tool (Rastogi et al., 2013). DroidChameleon applies code transformations to the smali code of the application under analysis. We consider obfuscated spyware to demonstrate that the proposed framework is resilient to the most widespread code obfuscation techniques implemented by malware writers in order to elude the current signature based detection provided by antimalware technologies (usually ineffective against trivial code transformations (Canfora et al., 2015a; Rastogi et al., 2014; Zheng et al., 2012)). As a matter of fact, antimalware soft-

4https://www.virustotal.com/#/home/upload
The samples generated with the AFE framework were injected with the following obfuscation techniques: (i) changing package name; (ii) identifier renaming; (iii) data encoding; (iv) call indirection; (v) code reordering; (vi) junk code insertion.

At the end of this transformation process, we have collected 60 obfuscated applications which are a morphed version of spyware samples. It should be underlined that only in one sample the read contacts suspicious behaviour is defined in the run method of a thread. In this case we can consider the sample under analysis suspicious. In the other three samples the identified behaviour is located in parts of code that seem harmless. Thus, in these cases we have to consider the identified samples as False Positive since our method classified them as spyware but they seem to be trusted.

Furthermore, the proposed method is able to locate the code snippet where the logic formula results true. In particular, our framework provides as output both the label (spyware or not spyware) and, if the formula is true, the exact location in the code in terms of the method name, class name and packages where the formula is resulted verified. In fact, from the localization results, it has emerged that all the spyware samples contain the malicious payload in the `com.xybot.infect.class` class (i.e., the class injected by the AFE framework).

It is worthy of note that for the 4 trusted samples the logic formula turned out to be true in another class belonging to another package different from `com.xybot`. In particular, during the analysis of spyware samples, the logic formula results verified in two different classes. Only in one application, it results verified on three classes.

With regard to the obfuscated versions of the spyware applications, the proposed framework was able to correctly identify as spyware all 60 morphed samples.

In order to evaluate the obtained results we compute following metrics: Precision, Recall and F-Measure.

The precision has been computed as the proportion of the examples that truly belong to class X among all those which were assigned to the class. It is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved:

\[
\text{Precision} = \frac{tp}{tp + fp}
\]

where \(tp\) indicates the number of true positives and \(fp\) indicates the number of false positives.

The recall has been computed as the proportion...
of examples that were assigned to class X, among all the examples that truly belong to the class, i.e., how much part of the class was captured. It is the ratio of the number of relevant records retrieved to the total number of relevant records:

\[
\text{Recall} = \frac{tp}{tp + fn}
\]

where \( tp \) indicates the number of true positives and \( fn \) indicates the number of false negatives.

The F-Measure is a measure of a test’s accuracy. This score can be interpreted as a weighted average of the precision and recall:

\[
\text{F-Measure} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

Table 4 shows the performances in terms of the metrics we defined.

<table>
<thead>
<tr>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.98</td>
<td>1.00</td>
<td>0.98</td>
</tr>
</tbody>
</table>

As shown in Table 4 the proposed framework is able to reach a precision value equal to 0.98, a recall value equal to 1 and an F-Measure of 0.98.

5 RELATED WORK

Several studies in current state of the art literature are mainly focused on generic mobile malware detection (Chen et al., 2016; Suarez-Tangil et al., 2017; Nix and Zhang, 2017; Duc and Giang, 2018). These works are mainly exploiting machine learning techniques by extracting distinctive features from samples under analysis to discriminate between malicious applications and trusted ones. Contrarily, in this paper we investigate for a specific threat (i.e., the mobile spyware). Another difference with the these methods is that the proposed model checking based approach is behavioural: it models the code behaviour and then, it checks against it the temporal logic formulae by specifying the malicious behaviour.

Shan et al. in (Shan et al., 2018) investigate about self-hiding behaviours (SHB), e.g. hiding the app, hiding app resources, blocking calls, deleting call records, or blocking and deleting text messages. First of all the authors provide an in-deep characterization of SHB, then they present a suite of static analyses to detect such behaviour. They define a set of detection rules able to catch SHB. They test their approach against more than 9,000 Android applications. Differently from the method we propose, authors are...
not mainly focused on spyware detection even if they define a set of rules able to detect specific behaviours.

At the best of our knowledge the only work focusing on Android spyware detection is the one proposed in (Chatterjee et al., 2018). Authors are focused in spyware used as intimate partner surveillance (IPS). The authors crawled apps from Google Play Store and using a combination of manual inspection and machine learning based approach discovered a large number of apps which are designed for legitimate use but also repurposed for IPS. Differently from this method we consider the model checking technique in order to identify spyware apps. Authors extract distinctive features from applications in order to apply machine learning based approach, instead, we define temporal logic formulae, which are behavioural based, to recognize Android spyware. Furthermore, we are focused about spyware with information gathering ability (i.e., the most widespread spyware in mobile environment (Wei et al., 2012)).

Zhang et al. in (Zhang et al., 2018) demonstrate that Google Assistant can be targeted since it suffers from some vulnerabilities. They develop an attacking framework able to record the voice of the user. This framework launches the attack using the recorded voice. This is a very dangerous vulnerability since the built-in voice assistant is able to access system resources and private information. Thus, hacking this assistant can lead to the leak of private and sensitive information. Differently, the proposed framework is able to recognize spyware applications in mobile environment to stem these types of attacks.

6 CONCLUSION AND FUTURE WORK

Nowadays smartphones collect a large amount of personal information. This is the reason why malware writers target these devices. More specifically, there is a kind of malicious software aiming to steal and collect these sensitive information and it is known as spyware.

Thus, in this paper we described a spyware detection framework. We exploit model checking technique and we use temporal logic formulae to detect Android spyware. We generated a synthetic dataset injected by spyware malicious payload in order to evaluate the effectiveness of the proposed method.

As future work, we plan to extend the experimental dataset including applications belonging from third-party marketplaces. We want also largely investigate for many other applications belonging to the Android official market. Thus, we want to perform an in-deep analysis of the applications available in the stores. Furthermore, also secure information analysis will be investigated (Avvenuti et al., 2012).

Furthermore, we intend to compare our approach with other solutions proposed in literature, for example the approach proposed by (Chatterjee et al., 2018).

ACKNOWLEDGMENT

This work was partially supported by the H2020 EU funded project NeCS [GA #675320], by the H2020 EU funded project C3ISP [GA #700294].

REFERENCES


