A Survey of Machine-learning based detection of Malaware in Android Operative System

María Fernanda Mora Alba
Department of Computer Science
Instituto Tecnológico Autónomo de México
México, Distrito Federal
Email: maria.mora@itam.mx

***Abstract*—In this survey we explore**

***Index Terms*—*xxx*.**

I. Introduction

- Importance and significance of the topic
- Discuss the background and target audience
- Summarize the surveyed research area and explain why the surveyed area has been studied
- Summarize the classification scheme you used to do the survey
- Summarize the surveyed techniques with the above classification scheme

CONTEXTO:

- Número de usuarios de dispositivos móbiles
- Tiempo de uso en dispositivos móviles
- Para qué se usan esos dispositivos (más que servicios de voz)
- Se guarda información sensible: contactos, fotos, banking services, etc. Los usuarios son descuidados porque piensan que es un telefono. Entonces ataques de seguridad son atractivos.
- Android OS: 80% del share.[3]. In September 2015, Android had 1.4 billion monthly active devices Callaham, John (September 29, 2015). "Google says there are now 1.4 billion active Android devices worldwide"
- Apps son esenciales para la experiencia de smartphone: entertainment, productivity, healthcare, online dating, home security, business management. Android Apps: mayores que Apple [3].
- Android is currently the most popular smartphone operating system. % de ataques a Android, pq Android es más propenso además de por el volumen de usuarios. Unlike iOS users, Android OS users do not have to root or jailbreak their devices to install apps from unknown sources.
- qué tipo de cosas te pueden robar: no sólo private information (contact listm text messagesm location) but also cause financial loss of the users by making secretive premium-rate phone call and text messages [3].
- It is prohibitive for app marketplaces such as Google App Store, to throuughly verify if an app is legitimate or maliciuos. Mobile users are left to decide for themselves wheter an app is safe to use [3]. The problem is that users are neither informed nor attentive: in a survey of 308 Android users they showed low attention and comprehension rates as 17% of participants paid attention to permissions during installation, and only 3% of Internet survey respondents could correctly answer all three permission comprehension questions [Felt et al.2012]. These results also showed that Android's permission system, intended to inform users about the risks of installing applications, was not effective in helping the user make correct security decisions.
  therefore we need automatic, scalable and reliable mechanisms.
- On the effectiveness of malware protection on android [58]: mensaje es que no son tan efectivos.
- Específico, based on machine learning, data mining, etc, porque citar [42]
- Se estudiaron los ataques a Android tipo malware porque...
- Explicar la Clasificación que se usó para hacer el survey
- Resumir (brevemente) las surveyed techniques
- Explicar el contenido del survey: Sección overview, discusión (survey), conclusiones y trabajo futuro con los findings.

II. Mobile device security and Android Operative System overview

SEGURIDAD:

A nivel aplicación, red. En cuál nos vamos a enfocar (aplicación). Que tipos de amenanzas hay (Malware)

Objetivo de la seguridad: Integridad, avaialbility, confidenciability. Hay más ataques en donde (paper internet).

Android OS

ANDROID: explicar un poco de la arquitectura relevante para temas de seguridad: kernel, librerías, application framework y on top las aplicaciones. Hablar del AndroidManifest.xml . Sacar todo de [60] (Understanding Android Security) .

The security model of Android heavily depends on the multi-user capabilities of the underlying Linux. Explicar el application security.

A. Android Operative System

Like all operating systems, Android enables applications to make use of the hardware features through abstraction and provide a defined environment for applications [Brahler2010].

Android Operating System is a stack of software components. Its source code is released by Google under open source licenses [Brahler2010]. Its architecture is formed by 4 main layers placed one on top of each other:

1) Linux Kernel at the base
2) Libraries: Android and Android runtime library. Android runtime combines the assets of the Java Virtual Machine and machine Dalvik. Android library consists of C / C ++ language.
3) Application Framework
4) Applications: written in Java

B. Android's Application Structure

C. Android's Security mechanisms

Android applications make use of advanced hardware, software and data with the objective of bringing innovation and value to the consumers. To allow thi value, the platform must offer an environment that guarantees the security of users, data, applications, the device and the network.

III. Survey

- Present the surveyed techniques using the classification scheme in detail
- Identify the trends in the surveyed area. Give evidences for your decision
- Identify some leading research /products/companies/websites
- Identify the unresolved problems /difficulties, and future research issues

IV. Conclusions and future work

Despite the fact that Android is based on Linux, it is not straightforward to take the same desktop analysis approach for Android malware [Yan and Yin2012].

References

[Brahler2010] Stefan Brahler. 2010. Analysis of the android architecture. *Karlsruhe institute for technology*, 7:8.

[Felt et al.2012] Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. 2012. Android permissions: User attention, comprehension, and behavior. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 3. ACM.

[Yan and Yin2012] Lok Kwong Yan and Heng Yin. 2012. Droidscope: seamlessly reconstructing the os and dalvik semantic views for dynamic android malware analysis. In *Presented as part of the 21st USENIX Security Symposium (USENIX Security 12)*, pages 569–584.

- Avoid use of et al. in a bibliography unless list is very long (five or more authors).
- Internet drafts must be marked "work in progress".
- Book citations include publication years, but no ISBN number.
- It is now acceptable to include URLs to material, but ... it is probably bad form to include a URL pointing to the author's web page for papers published in IEEE and ACM publications.
- Leave a space between first names and last name, i.e., "J. P. Doe", not " J.P.Doe

Objetivo del mecanismo Contribution/ Results	Metodología	Scope
Encontrar vulnerabilidades, análisis (comportamiento) y detección (prevención, antes de detectarlo)	Estático o dinámico	Kernel leve