A Survey on Security for Mobile Devices
Mariantonietta La Polla, Fabio Martinelli, and Daniele Sgandurra

Abstract—Nowadays, mobile devices are an important part of our everyday lives since they enable us to access a large variety of ubiquitous services. In recent years, the availability of these ubiquitous and mobile services has significantly increased due to the different form of connectivity provided by mobile devices, such as GSM, GPRS, Bluetooth and Wi-Fi. In the same trend, the number and typologies of vulnerabilities exploiting these services and communication channels have increased as well. Therefore, smartphones may now represent an ideal target for malware writers. As the number of vulnerabilities and, hence, of attacks increase, there has been a corresponding rise of security solutions proposed by researchers. Due to the fact that this research field is immature and still unexplored in depth, with this paper we aim to provide a structured and comprehensive overview of the research on security solutions for mobile devices.

This paper surveys the state of the art on threats, vulnerabilities and security solutions over the period 2004-2011, by focusing on high-level attacks, such those to user applications. We group existing approaches aimed at protecting mobile devices against these classes of attacks into different categories, based upon the detection principles, architectures, collected data and operating systems, especially focusing on IDS-based models and tools. With this categorization we aim to provide an easy and concise view of the underlying model adopted by each approach.

Index Terms—Mobile Security, Intrusion Detection, Mobile Malware, Trusted Mobile.

I. INTRODUCTION

CURRENT mobile devices (henceforth, called smartphones) provide lots of the capabilities of traditional personal computers (PCs) and, in addition, offer a large selection of connectivity options, such as IEEE 802.11, Bluetooth, GSM, GPRS, UMTS, and HSPA. This plethora of appealing features has led to a widespread diffusion of smartphones that, as a result, are now an ideal target for attackers. In the beginning, smartphones came packaged with standardized Operating System (OS): less heterogeneity in OS allowed attackers to exploit just a single vulnerability to attack a large number of different kinds of devices by causing major security outbreaks [1]. Recently, the number of OSes for smartphones (Symbian OS, Windows Mobile, Android and iPhone OS) has increased [2]: as shown in Table I, each mobile OS has now gained a significant market share.

Even if global sales of smartphones will pass 420 million devices in 2011 (according to a recent report by IMS research [3]), the number of mobile malware is still small compared to that of PC malware [4]. Nonetheless, we can expect malware for smartphones to evolve in the same trend as malware for PCs: hence, in the next incoming years we will face a growing number of malware. As an example, as more users download and install third-party applications for smartphones, the chances of installing malicious programs increases as well. Furthermore, since users increasingly exploit smartphones for sensitive transactions, such as online shopping and banking, there are likely to be more threats designed to generate profits for the attackers. As a proof that attackers are starting to focus their efforts on mobile platforms, there has been a sharp rise in the number of reported new mobile OS vulnerabilities [5]: from 115 in 2009 to 163 in 2010 (42% more vulnerabilities). In the same trend, there has been an increase in attention to the security issues from security researchers.

To help understanding the current security problems affecting smartphones, we review threats, vulnerabilities and attacks specific to smartphones and examine several security solutions to protect them. In particular, we survey the literature over the period 2004-2011, by focusing our attention on high-level attacks.

The paper is organized as follows. Section II introduces some background notions on mobile technologies, both for wireless telecommunication and networking standards. Section III describes different types of mobile malware, along with some predictions on future threats, and outlines the differences among security solutions for smartphones and traditional PCs. Section IV discusses current threats targeting smartphones: firstly, it analyzes the different methodologies to perform an attack in a mobile environment; then, it investigates how these methodologies can be exploited to reach different goals. In Sec. V we present security solutions, focusing on those that exploit intrusion detection systems and trusted platform technologies. Finally, Sec. VI draws some conclusions.

II. MOBILE TECHNOLOGIES

In this section, we briefly recall some background notions on wireless and networking technologies that, even if not originally created for a mobile environment, have favored the increasing usage of smartphones.

A. Wireless Telecommunication Technologies

The most important wireless technologies targeted at mobile communications are GSM, GPRS, EDGE and UMTS.

1) GSM: Global System for Mobile communications (GSM) is the first and most popular standard in Europe for mobile telecommunication system and is part of the second-generation (2G) wireless telephone technology. Developed in 1990 by Group Special Mobile, a group created in 1982 by Conférence Européenne des administrations des Postes et des Télécommunications (CEPT), this standard enables the creation of cellular networks where mobile phones (called mobile station in the standard) communicate with each other.
through base stations, networks and switching subsystems. Compared to its predecessor (TACS standard), telecommunications operators can offer new services by exploiting these technologies: for instance, data transmission, digital fax, e-mail, call forwarding, teleconferencing service and Short Message Service (SMS).

2) GPRS and EDGE: These standards stem as an evolution of GSM; General Packet Radio Service (GPRS), also referred as 2.5 generation, was developed to improve performances of GSM network to enable users to achieve higher data rates and lower access time compared with previous GSM standard. GPRS uses packet switching mechanism (as in IP protocol) to enable the exchange of data between users. Moreover, services such as Wireless Application Protocol (WAP) and Multimedia Messaging Service (MMS) are also introduced. In this way, a variety of packet-oriented multimedia applications and services can be offered to mobile users.

Enhanced Data rates for GSM Evolution (EDGE) standard was developed in 2000 to improve the features offered by GPRS by supporting higher transmission rate and higher reliability.

3) UMTS: The Universal Mobile Telecommunications System (UMTS) was introduced in Europe in 2002. This standard represents the third-generation (3G) on cellular system. The transmission rate is higher than 2G and 2.5G by providing a transmission speed up to 2Mbps. Circuit switching connections are supported simultaneously with packet switching connections and users can exploit multiple services and different classes of services, such as conversational, streaming, interactive and background.

B. Networking Technologies

During the last few years, due to ease of installation and the increasing popularity of laptop computers, Wireless Local Area Network (WLAN) has become very popular. This technology enables devices to be linked together through wireless distribution methods and allows users to move in a local coverage area without losing their connection to the network. There are different standards that regulate communications in a WLAN. In the mobile environment, the most popular are Bluetooth and IEEE 802.11.

1) Bluetooth: Bluetooth is a standard that enables devices to exchange data over a small area through short wavelength radio transmissions. Bluetooth is a personal networking technology that enables the creation of Personal Area Networks with high levels of security. This standard, developed by Bluetooth Special Interest Group (SIG) in 1999, is aimed at providing communication between devices having these features:

- lower consumptions;
- short range of communications (1-100 meters);
- small production costs.

As shown in Table II, there are three different classes of Bluetooth devices according to the power consumption and range of communication.

SIG defines several profiles to indicate different services (e.g. Generic Access Profile, GAP, or Headset Profile, HSP) and to describe the service’s implementation.

2) Wireless LAN IEEE 802.11: IEEE 802.11 is a family of standards for WLAN that includes several protocols for communicating at different frequencies (2.4, 3.6 and 5 GHz). These standards can be used in two operation mode:

1) in the infrastructure mode, a device, referred as Access Point (AP), plays the role of the referee: an AP regulates the network access and coordinates the devices that are part of the network;
2) in the infrastructure-less mode (ad hoc mode), no referee exists and devices monitor the spectrum to gain network access.

The most popular protocols included in this standard are defined by the 802.11b and 802.11g protocols. As shown in Table III, the differences between these protocols are related to bandwidth, bit-rate and type of modulation (Complementary Code Keying, CCK, for 802.11b, Orthogonal Frequency-Division Multiplexing, OFDM, for 802.11g).

III. MOBILE MALWARE

This section provides a comprehensive overview of mobile malware and some predictions on future threats. Moreover, it describes the differences among security solutions targeting smartphones and PCs.
**Malware** is any kind of hostile, intrusive, or annoying software or program code (e.g. Trojan, rootkit, backdoor) designed to use a device without the owner’s consent. Malware is often distributed as a spam within a malicious attachment or a link in an infected websites. Malware can be grouped in the following main categories, according to its features (e.g., the vector that is used to carry the payload):

- virus;
- worm;
- Trojan;
- rootkits;
- botnet.

A **virus** is a piece of code that can replicate itself. Different replica of a virus can infect other programs, boot sector, or files by inserting or attaching itself to them.

A **worm** is a program that makes copies of itself, typically from one device to another one, using different transport mechanisms through an existing network without any user intervention. Usually, a worm does not attach to existing programs of the infected host but it may damage and compromise the security of the device or consume network bandwidth.

Malware can also come packaged as a **Trojan**, a software that appears to provide some functionalities but, instead, contains a malicious program.

**Rootkits** achieve their malicious goal by infecting the OS: usually, they hide malicious user-space processes and files or install Trojans, disable firewalls and anti-virus. Rootkits can operate stealthily since they directly apply changes to the OS and, hence, can retain longer control over the infected devices.

Finally, a **botnet** is a set of devices that are infected by a virus that gives an attacker the ability to remotely control them. Botnets represent a serious security threat on the Internet and most of them are developed for organized crime doing attacks to gain money. Example of such attacks are sending spam, Denial-of-Service (DoS) or collecting information that can be exploited for illegal purposes (DoS attacks targeting smartphones are described in detail in Sec. IV-B3).

Mobile malware can spread through several and distinct vectors, such as an SMS containing a link to a site where a user can download the malicious code, an MMS with infected attachments, or infected programs received via Bluetooth. The main goals of malware targeted at smartphones include theft of personal data stored in the phone or the user’s credit.

**Examples**: [6] details a Trojan for Android smartphones, named Trojan-SMS.AndroidOS.FakePlayer.b, which masquerades as a media player and requires the user to manually install it. This fake application is downloaded from an infected webpage in order to view adult content videos. The installation file is very small in size and during installation the application asks the user permissions to send SMS messages. Once the installation has finished, if the user launches the fake application, the Trojan begins sending SMS messages to a premium rate number without the user’s knowledge. These messages result in costly sums being transferred from the user’s account to that of the cybercriminals.

[7] develops a kernel-level Android rootkit in the form of a loadable kernel module that can open a shell for the attacker (using a reverse TCP connection over 3G/Wi-Fi) upon the reception of an incoming call from a trigger number. This results in full root access on the Android device. In this way, an attacker can read all SMS messages on the device, incur the owner with long-distance costs or even potentially pinpoint the mobile device’s exact GPS location.

[8] analyzes three sample rootkits to show how smartphones are as vulnerable as traditional computers to rootkits. In fact, smartphone rootkits can access several distinctive interfaces and information that are unique to smartphones, such as GPS, battery, voice and messaging, which provide rootkits writers with new attack vectors to compromise either the privacy or the security of end users. The first proposed sample rootkit allows a remote attacker to stealthily listen into (or record) confidential GSM conversation using the user’s infected smartphone. The second attack aims at compromising the victim’s location privacy by requiring the infected smartphone to send a text message to the remote attacker including the user’s current GPS location. The final sample attack exploits power-intensive smartphone services, such as those offered by GPS and Bluetooth, to exhaust the battery on the smartphone.

As an example of smart malware, recently a multifarious malware for iOS devices has been designed and implemented by [9] (iSAM). iSAM incorporates six different features of malware:

1) propagation logic;
2) botnet control logic;
3) collect confidential data stealthily;
4) send a large number of malicious SMS;
5) denial of application services;
6) denial of network services.

Moreover, iSAM is able to connect back to its botmaster server to update the programming logic, to implement commands, and to perform a synchronized, distributed, attack.

Table IV reports some notable examples of mobile malware.

### A. Evolution of Mobile Malware

Several papers discuss the evolution of mobile malware: for instance, [10] describes the evolution of malware on smartphones from 2004 to 2006. For an overview on the state of the art of mobiles viruses and worms up to 2006, see Hypponen [11]. In the period 2004-2008, the number of types of mobile malware has increased significantly: as of March 2008, F-Secure has categorized 401 distinct types of mobile malware worldwide, whereas McAfee has counted 457 kinds of mobile malware [12]. In the period 2004-2010, 517 families of mobile viruses, worms and Trojans have been categorized by F-Secure [13]. For a complete list of mobile malware in the period 2000-2008 see [14]; see [15] for mobile malware that spread from January 2009 to June 2011.

The first virus (a Trojan) for mobile phones, developed for Palm devices [16], was discovered in 2000 by F-Secure [17]. In June 2004, the first worm that could spread through mobile phones with Symbian OS appeared: this worm, called Cabir [18], was only a prototype developed by the 29A Eastern European hacker group. Cabir is considered the first example of malicious code that can spread itself exploiting the networking technologies on mobile devices (in this case, Bluetooth) to infect other devices.
Recently, a growing number of viruses, worms, and Trojans that target smartphones have been discovered. As we have already pointed out, the reason for the growing number of mobile malware is due to the widespread use of smartphones. Furthermore, we have to consider that most of the smartphones lack any kind of security mechanisms and are not well prepared against new threats. Within the 2006-2008 period, security issues exploiting several attack vectors have increased [19], and there has been a dramatic escalation of complex attacks targeting lower-layer device functionality: early security threats have turned into sophisticated, profit-oriented, attacks driven by experienced criminals.

A discussion of mobile malware, based on OSes and infection routes, is presented in Töyssy and Helenius [20] that describe and cluster mobile malware with respect to:

- the OS: Symbian, Palm OS, Linux, Windows Mobile;
- the infection routes: MMS, Bluetooth, IP connections via GPRS/EDGE/UMTS, WLAN, copying files, removable media.

The authors propose some prevention solutions and countermeasures, by considering:

- the users, which have to be educated to utilize the device in a secure way;
- the software developer, which can develop security protection targeted at smartphone;
- the network operator, which can enhance the network infrastructure with mechanisms to avoid intrusions;
- the phone manufacturers, which should update the devices automatically so that for attackers it would be harder to exploit security holes;
- new epidemiological models, to forecast if an already detected virus can initiate an epidemic.

Similar solutions are also proposed in [21], where the authors remark that the protection from malicious code should be implemented at every possible entry point to the network. For a comprehensive discussion on evolution of mobile malware, see [22].

**B. Predictions and Future Threats**

Since the first smartphone, discussions about threats targeting these devices have proliferated: the first threat against Symbian and Palm (such as Liberty Crack Trojan [23]) never became widespread and remained just a proof-of-concept. Even if security experts foresee massive attacks to come out at any time, yet they never seem to happen: nonetheless, McAfee Labs [24] predicts that 2011 will be a turning point for threats to smartphones. In fact, in the last months, several new threats to smartphones have emerged: rootkits for the Android platform, remote jail-breaking exploits for the iPhone, and the arrival of ZeuS [25], a largely distributed banking Trojan/botnet. The widespread adoption of smartphones into business environments combined with these attacks is likely to cause the explosion that experts have long anticipated.

According to [26], in the near future cybercriminals will focus their attention on iPhone and Android platforms (this is also confirmed by Cisco [27] and Panda [28]). Symantec [5] explains the state of cybercrime on smartphones by its return on investment: firstly, the installed base of smartphones has grown to an attractive size and they run sophisticated OSes that come with the inevitable vulnerabilities; secondly, Trojans hiding in legitimate applications sold on application stores provides a simple and effective propagation method. Hence, what is currently missing is the ability to turn all this into a profit center.
In March 2011, Google reported [29] that it had removed several malicious Android applications from the Android Market and, in some cases, deleted them from users’ smartphones remotely. The Android marketplace is not very closely monitored, since it adopts the “anything goes” philosophy. This, combined with the current buzz around new smartphones running Android, may make the platform more attractive to cybercriminals [30]. Gostev [26] believes there is also a strong probability that malware may soon be found in products available through Android Market. As an example, many legitimate applications can ask for, and typically be granted, access to a user’s personal data and authorization to send SMSs and make calls. Hence, this places the reliability of the entire Android security concept in doubt. This is confirmed by Juniper Networks [31] that reports that, since Summer 2010, Android malware has increased by 400%.

As [26] points out, no significant malware events occurred that targeted iPhone and which could be compared to the Ikee worm incident of 2009. However, several concept programs were created for this platform in 2010 that demonstrated techniques that could be used by cybercriminals. As an example, SpyPhone allows unauthorized access to information about the user’s iPhone device, such as her location, interests, friends, preferred activities, passwords and web search history. This data can then be sent to a remote server without the user’s knowledge or consent. This functionality can be hidden within an innocuous-looking application.

[27] reveals that in July 2010 the U.S. Library of Congress added jailbreaking to its list of actions that do not violate copyright protections, hence leaving iPhone users free to unlock their devices and download applications not authorized by Apple. Only a week after, JailbreakMe 2.0 appeared, a tool that makes it easier for users to jailbreak their phones. The advent of this tool also revealed a significant security flaw in the iOS 4 that could leave users with jailbroken phones more likely to be attacked by hackers willing to take control of the smartphones.

As pointed out by [32], another feature to consider is the spreading of mobile virus to desktop platforms, e.g. due to devices that are already compromised or tampered with coming off the shelves. As an example, USB devices are responsible for the spread of auto-run malware, while the Conficker/DOWNAD worm contained a propagation capability that used removable drives to increase spread.

It is important to underline that evolution of malware is a continuous race between attackers and defenders: both use the same programming methods, tools and resources either to create a malware or to develop an intelligent malware detection mechanism. A further aspect to be considered concerns the observation of new forms of malware in a testbed environment to predict their behavior: as an example, [33] presents Mobile Agent Malware Simulator (MAISim), a framework that uses the technology of mobile agents for simulation of various types of malicious software (viruses, worms, malicious mobile code) for smartphones.

Furthermore, since the risks at which smartphones are exposed depend also on how they are used, we have to distinguish strictly personal use of a device from uses that also involve business. As suggested in [34], future threats in a mobile environment may affect different assets, such as:

- personal data;
- corporate intellectual property;
- classified information;
- financial assets;
- device and service availability and functionality;
- personal and political reputation.

Finally, Milligan and Hutcheson [35] discuss some risks, threats and countermeasures for smartphones. Some examples of future risks associated with a smartphone include:

- data leakage resulting from device loss or theft;
- unintentional disclosure of data;
- attacks on decommissioned devices;
- phishing attacks;
- spyware attacks;
- network spoofing attacks;
- surveillance attacks;
- diallerware attacks;
- financial malware attacks;
- network congestion.

C. Mobile Security Versus Personal Computer Security

Despite the similarities between smartphones and PCs, there are several notable differences concerning security. Firstly, we have to consider that malware authors can make money from their illicit activities more easily on smartphones than in desktop environments, e.g. due to premium-rate numbers (Felt et al. [36] classify premium-rate calls/SMS as the second most common behavior found in nearly 50% of recent malware1). Secondly, since any event generated by the smartphone has (usually) a cost invoiced by the network operator, from the point of view of the user, the network operator is considered responsible of charging costs even if the event is generated by malware.

We also have to consider the user’s point of view in a mobile environment: to this end, Botha et al. [37] explore the availability of security mechanisms from the perspective of a user who wishes to use desktop-based security mechanisms in a mobile environment. The authors remark that a main difference between smartphones and PCs is that the former is usually a personal device and some issues, such as user authentication, device configuration and content protection, need to be dealt differently. As an example, in the case of user authentication, users face several difficulties when moving from desktop to smartphone. Furthermore, there is a trade-off between security and usability, since many solutions that are used on PCs cannot be applied on smartphones.

Compared to common PCs, the basic security principles of smartphones are quite different. The security problem on smartphones originates particularly from the integration process: nowadays, a single device hosts multiple technologies that allow users to access the Internet from any place at any time. Furthermore, smartphone-specific services often require complex software and infrastructures and expose these devices

---

1Two further common behaviors are stealing/selling users’ information and sending SMS spam, which are found in, respectively, 60% and 17% of the analyzed malware.
to attacks. As described in [38], five key aspects distinguish mobile security from conventional computer security:

- **mobility:** each device comes with us anywhere we go and; therefore, it can be easily stolen or physically tampered;
- **strong personalization:** usually, the owner of device is also its unique user;
- **strong connectivity:** a smartphone enables a user to send e-mails, to check her online banking account, to access lot of Internet services; in this way, malware can infect the device, either through SMS or MMS or by exploiting the Internet connection;
- **technology convergence:** a single device combines different technologies: this may enable an attacker to exploit different routes to perform her attacks;
- **reduced capabilities:** even if smartphones are like pocket PCs, there are some characteristic features that lack on smartphones, e.g. a fully keyboard.

The limited resources of a smartphone are the most obvious difference with a PC. The main limiting factors are CPU and memory. These two factors limit the sophistication of possible security solutions: for example, complex intrusion detection algorithms that work for real-life applications on PCs cannot be easily transferred to smartphones in the near future (see [39] for a distributed solution that tries to circumvent this problem). In addition, a unique characteristic of smartphones is the battery, which severely limits the resources available for a security solution; therefore, it is highly important that a security solution does not constantly drain large portions of available CPU time to avoid battery exhaustion [40].

As an example of the limited resources of a smartphone, according to [41], the ClamAV antivirus engine available for Nokia device requires about one minute of processing time to initialize the signature database and at least 40 MB of memory. To reduce this overhead, the paper proposes a model where mobile antivirus functionalities are moved to an off-device cloud. The Nokia device requires about one minute of processing time to initialize the signature database and at least 40 MB of memory.

As [42] points out, since smartphones have strict resource constraints both in computational capabilities and power consumption, some computationally expensive algorithms for detecting sophisticated threats, such as those implemented by behavioral detection engines, are simply infeasible to be deployed on current smartphones due to their heavy-weight resource requirements. This means that adapting traditional approaches for malware detection might be infeasible for mobile environments as they consume a significant amount of resources and power.

Moreover, a further security threat for smartphones stems from the fact that wireless medium is, by nature, prone to eavesdropping and, therefore, communication confidentiality cannot be taken for granted. Threats to user privacy in a mobile environment are different from those performed on PCs because, on smartphones, sensors (e.g. microphones) are not optional and can be used illicitly to sniff user’s private data. In addition, since mobile applications extensively use, and depend on, sensors and users carry smartphones wherever they go, this increase the opportunities to compromise the privacy of mobile users. Using access control techniques, like those used in PC environments, is not appropriate because it is necessary to take into account also the context in which sensors are used. Furthermore, these attacks work even when the user is not interacting with the mobile phone.

Further discussions on the subject can be found in [43].

IV. ATTACKS ON MOBILE DEVICES

In the following sections, we discuss several kinds of attacks against smartphones. We firstly detail the possible methodologies to perform an attack in a mobile environment and, for each kind of attack, we provide a real example. Secondly, we show how these methodologies can be exploited to reach different goals.

A. Methodologies of the Attacks

The distinct methodologies to perform attacks against smartphones are categorized using the following classes:

- wireless;
- break-in;
- infrastructure-based;
- worm-based;
- botnet;
- user-based.

1) **Wireless Attacks:** There are many different kinds of wireless attacks against smartphones, especially those targeting personal and sensitive data. The most common attack is eavesdropping on wireless transmissions to extract confidential information, such as usernames and passwords. Wireless attacks can also abuse the unique hardware identification (e.g., wireless LAN MAC address) for tracking or profiling the owner of the device. Finally, malware often exploits Bluetooth as a medium to spread its propagation.

[44] discusses security problems in wireless environments and presents the current research activities. A comprehensive review of Bluetooth attacks affecting smartphones can be found in [45]. Some studies for preventing this class of attacks are proposed in [44, 46, 47, 48].

**Example - Cabir:** Cabir is a worm that propagates through Bluetooth. This worm consists of a message containing an application file, *caribe.sis*, that seems like a Security Manager utility. If installed, the worm uses the device’s native Bluetooth functionality to search for other Bluetooth-discoverable devices. Then, the worm attempts to send infected SIS files to the discovered devices as well.

2) **Break-in Attacks:** Break-in attacks enable the attacker to gain control over the targeted device by exploiting either programming errors, e.g. to cause buffer overflows, or format string vulnerabilities. Typically, these attacks are used as a stepping stone for performing further attacks, such as over-billing attacks or data/identity theft.

Some studies for preventing this class of attacks are proposed in [49, 50].

**Example - Doomboot.A:** This Trojan installs corrupted system binaries into the C:\ drive of the device. The corrupted binaries contain further Trojans, as CommWarrior, which are also installed on the device.
3) Infrastructure-based Attacks: Since the services provided by the infrastructure are the basis for essential smartphone functionalities, such as placing/receiving calls, SMS and e-mail services, the economic and social impact of these attacks may be very large, such as the one discussed in [51]. [52] evaluates the security impact of the SMS interface on the availability of the cellular phone network. As an example, if an attacker is able to simultaneously send messages through the several available portals into the SMS network, the resulting aggregate load can saturate the control channels and, therefore, block legitimate voice and SMS communications. The authors demonstrate that an attacker that injects text messages from the Internet can deny voice service in a metropolitan area using hit-lists containing as few as 2,500 targets with little more than a cable modem.

a) GPRS: Since the GPRS architecture is built on the GSM infrastructure, it uses a security architecture based upon the security measures already adopted by GSM (for a review of DoS attacks and confidentiality threats in GSM networks, see [53]). Attacks against GPRS can target the device, the radio access network, the backbone network, and the interfaces connecting GPRS networks with each other or with the Internet. The results of these attacks can be the compromise of end-users security, over bill users, the disclosure or alteration of critical information, the services unavailability, or the network breakdown. We have also to consider that GPRS is more exposed to attackers compared to GSM because it uses the IP technology, which is rather vulnerable.

Attacks against GPRS can be active and passive: active attacks requires a direct intervention of the attacker to listen, modify, and inject data into the communication channel. Furthermore, if the attacker is not part of the GPRS, the attack can be defined external; on the other hand, the attack is defined as internal. A passive attack happens when an attacker taps on a communication channel between two nodes without disturbing the communication to discover some valuable information about the data or control messages.

As described in [54], there are five sensitive area in GPRS security that can be exploited to perform an attack:

1) the mobile station (MS) and the SIM-card: the results of these attacks may be the monitoring of the MS usage, the downloading of unwanted files, the placing of unwanted calls. Attacks on the SIM-card are primary based upon the secret key, which is stored in the SIM-card of the MS. When an attacker retrieves this key, she can intercept data exchanged, or clone the original SIM-card;

2) the interface between the MS and the SGSN (Serving GPRS Support Node): an attacker can perform attacks like DoS or Man-in-the-Middle2. In DoS a malicious third party can jam user data and signaling traffic using special devices called jammer, or induce specific protocol failures, or masquerade as network elements;

3) the GPRS backbone network: they refer to both IP and Signaling System 7 (SS7) technologies, which convey user data and signalling information. Once a malicious MS gets access to the GPRS network may perform various attacks, such as DoS, IP spoofing, compromise of privacy or send large amounts of data to users;

4) the packet network that connects different operators: the Gp Interface provides connectivity between GPRS networks that belong to different operators and supports users roaming. The security targets are the availability of resources and services, the authentication and authorization of users and actions, and the integrity and confidentiality of the data transferred;

5) the Internet: the Gi interface connects the GPRS network to the Internet and service providers that provide services to mobile subscribers. Due to the fact that the Gi interface may carry any type of traffic, the GPRS network elements and the mobile subscribers can be exposed to a variety of threats found on Internet;

Another type of attacks against the SIM-cards of GSM/GPRS is side-channel attacks that allow an attacker to obtain sensitive information from side-channels, as described in [55].

b) UMTS: The UMTS security architecture defines a set of procedures to achieve increased message confidentiality and integrity during their communication. At the kernel of its security architecture lies the user authentication mechanism, known as Authentication and Key Agreement (AKA). Authentication in UMTS is based on a 128-bit symmetric secret key, namely Ki, which is stored in the user’s tamper-resistant Universal Integrated Circuit Card (UICC) and in the corresponding Home Location Register (HLR) of the user’s Home Network (HN).

[56] discusses several vulnerabilities of the UMTS security architecture that can be exploited by malicious attackers to launch DoS attacks. Typically, an attacker tries to access unprotected control messages in order to manipulate specific procedures. The expected results varies from lower Quality of Service to DoS. Some examples of such an attack are:

- dropping ACK signal: an attacker monitors for TMSI Allocation Command messages and then drops any following TMSI Allocation Complete message to repeatedly force the creation of new TMSIs that, eventually, will cause DoS to all the users in that area;

- modification of unprotected Radio Resource Control (RRC) messages: an attacker substitutes a valid RRC Connection Setup Complete with a RRC Connection Reject message to cause a lower Quality of Service or a DoS for the end-users;

- modification of the initial security capabilities of MS: an attacker modifies a RRC Connection Request message to trigger the termination of the connection. For example, she can cause a serious damage by creating a very large number of simultaneous connection requests;

- modification of periodic authentication messages: this happens if the Radio Network Controller (RNC), on receiving a tampered message, releases the connection, disconnecting the MS;

- SQN synchronization: an attacker can ask for a resynchronization procedure to be executed simultaneously for large numbers of users, and repeatedly, to greatly overstress the HLR;
Worms that exploit messaging services (SMS/MMS), as their preferred infection routes, are potentially more virulent, in terms of speed and area of propagation, than Bluetooth ones. In fact, these worms can be easily sent out using just one click and can infect any smartphone in any part of the world with a larger chance of success of propagation.

\( c \) User Mobility Models: Compared with the Internet, mobile phone networks have very different characteristics in terms of topologies, services, provisioning and capacity, devices and communication patterns. These features also characterize the way new types of mobile worms propagate: the most important one is that they do not require Internet connectivity for their propagation and, therefore, can spread without being detected by existing security systems. Hence, mobile worms can infect several devices using proximity attacks against vulnerable devices that are physically nearby.

To model the propagation of these worms, two steps are required:

1. build a model that precisely describes how devices meet each other;
2. understand how malicious code exploits both the mobility of the users and the capacities of the networks.

The dynamics of proximity propagation depend upon the mobility dynamics of a population in a specific geographic region. Unfortunately, an ideal methodology for modeling user mobility does not exist: traces of mobile user’s contacts reflect actual behavior, but they are difficult to generalize and only capture a subset of all contacts due to a lack of geographic coverage.

To model epidemic spreading of malware via proximity-based, point-to-point wireless links, Mickens and Noble [63] introduce a framework called probabilistic queuing to deal with node mobility. To capture the skewed connectivity distributions of mobile networks, the model represents different connectivity levels as distinct queues. Each queue represents a separate epidemiological population. A probabilistic queuing model is proposed to explicitly account for both node velocities and the non-homogeneous connectivity patterns induced by mobility of devices.

The problem of proximity attacks of smartphone is also discussed in [64] where the authors introduce an individual-based model and build analytical expressions for contact-rate calculation and worm transmission. In [65], an event-driven simulator that captures the characteristics and constraints of mobile phone networks is proposed. The simulator models realistic topologies and provisioned capacities of the network infrastructure. The goals of this model are:

- model malware propagation in networks under realistic scenarios to characterize its speed and severity;
- understand how network provisioning impacts propagation and how propagation impacts the network;
- highlight the implications for network-based defenses against malware.

Some further solutions to withstand worm attacks, which are based upon mobility models for Bluetooth worm propagation, are also studied in [66, 67, 68]. See also [69] for analytical models for epidemics in mobile networks.
Examples: [70] and [71] investigate whether a large-scale Bluetooth worm outbreak is viable in practice. The authors use trace-driven simulations to examine the propagation dynamics of a Bluetooth worm in a large population, by showing that the worm outbreak’s start time is very important (e.g., during week-end). Their results suggest that a worm exploiting a Bluetooth could spread very quickly. As defense solution, the authors suggest to locate monitoring points in high-traffic locations.

[72] investigates the effort required to port a smartphone worm for Windows Mobile. The authors were able to build a prototype worm that could spread autonomously if a vulnerability exists in a Windows Mobile service reachable over the network. The authors state that the lower bound of the time required to both build a worm toolkit and find a vulnerability in the network protocol stack of Windows Mobile requires approximately 14 weeks of full time work.

[73] analyzes in detail the first polymorphic worm that affects smartphones running Windows CE platform on ARM processors, known as WinCE.Pmcryptic.A. The worm spreads by generating new polymorphic copies of itself each time and can execute several unwanted actions on a compromised smartphone, including calls to toll numbers.

5) Botnets: Until recently, mobile networks have been relatively isolated from the Internet, so there has been little need for protecting them against attackers trying to create botnets. However, this situation is rapidly changing since mobile networks are now well integrated with the Internet. Hence, threats on the Internet will migrate over the mobile networks (and vice-versa), including botnets, since smartphones can be infected by malware they can be turned into a botclient easily [74].

a) Command-and-Control: The command-and-control (C&C) network, used to remotely propagate messages, tasks, updated payload among the bots and the botmasters (and vice-versa), can be built out using Bluetooth, SMS messages, the Internet (e.g., HTTP), peer-to-peer (P2P) or any combination of them.

Bluetooth C&C: [75] investigates the challenges of constructing and maintaining mobile-based botnets communicating via Bluetooth. By using simulations on publicly available Bluetooth traces, the authors demonstrate that C&C messages can propagate to approximately 66% of infected nodes within 24 hours of being issued by the botmaster. To reduce the amount of traffic observable by the provider to achieve stealthiness, in the developed framework only a small subset of bots (those with the highest degree) communicate directly with the botmaster through cellular channels (e.g., SMS, cellular data). These nodes are selected via their relative frequency of contact with other infected devices: whenever infected smartphones pass within range of each other, they record the identity of the other device. After reaching some threshold set by the botmaster, nodes with a high degree of connectivity send their contact logs to the botmaster which is informed of (i) which devices are under his control (ii) which nodes can help disseminating commands rapidly. The botmaster also disseminates commands and updated payloads through this hierarchical structure by contacting the seed nodes. Due to their high degree of connectivity, these nodes can deliver the payload to the largest number of infected nodes directly and no interaction with the botmaster is required.

SMS C&C: [76] proposes the design of a proof-of-concept mobile botnet resilient even to disruption. The botnet is built out of three components:

1) vectors to spread the bot code to smartphones;
2) a channel to issue commands;
3) a topology to organize the botnet.

Within the testbed mobile botnet, all C&C communications are carried out using SMS messages. To hide the identity of the botmaster, there are no central servers dedicated to command dissemination, since they could be easily identified and removed. Instead, a P2P topology is exploited to allow botmasters and bots to publish and search for commands in a P2P fashion, making their detection and disruption much harder.

Hybrid C&C: [77] shows that it is easy to create a fully functional mobile phone botnet out of Apple’s jailbroken iPhone by discussing the design, implementation and evaluation of an iPhone-based mobile botnet. The authors firstly discuss an SMS-based botnet that is then improved with HTTP to reduce the number of SMS messages that need to be sent for controlling the bots. Finally, the authors show how powerful such a botnet could be if attackers combine P2P (namely, Kademlia) with SMS-HTTP hybrid approach.

Examples: Porras et al. [78] recall how, at the end of 2009, some users of jailbroken iPhones began seeing pop-up windows that redirected the victim to a website where a ransom payment was demanded to remove the malware infection. The vulnerability affected several jailbroken iPhones that have been configured with a SSH service with a known default root password. By scanning some IP addresses from the Internet for vulnerable SSH-enabled iPhones, an attacker could upload a very simple ransomware application to several iPhone users. By exploiting this vulnerability, some weeks after, a second iPhone malware (iKee.A) converted the iPhone into a self-propagating worm to infect other iPhones. This time, the worm succeeded in infecting more than 20,000 victims within a week. Some time later, a new malware (iKee.B), similar to iKee.A, was found: in addition to self-propagation, the iKee.B bot client application introduces a C&C check-in service that enables the botmaster to upload and execute shell commands on all infected iPhone bot clients. This service allows the bot to evolve or to redirect infected iPhones to new C&Cs located anywhere on the Internet. The iKee.B also incorporates a feature to exfiltrate the entire SMS database from the victim’s iPhone.

[79] provides an overview of Yxes, one of the first malware for Symbian OS 9 and a first step towards a mobile botnet. Once installed, using a valid signed certificate, the main tasks of the malware include getting the IMEI and the IMSI of the mobile phone, parsing contacts, killing unwanted applications and propagating. To this end, the malware begins its propagation phase by sending an SMS to each new victim with a link to a malicious server where the victim can download the malware. One of the main issues concerning Yxes is that its code does not exploit any particular Symbian OS vulnerability, but only uses functions of its API in a smart way. For this reason, the author concludes by remarking that the concept
of capabilities, introduced by Symbian, fails to stop malicious
intents, for two reasons: (i) cybercriminals manage to have
their malware signed whatever capability they request; (ii)
capabilities only grant authorizations for given actions but
cannot take into account a context or an intent.

Attackers have also taken a popular legitimate application
and added additional code to it, such as with Pjapps [80].
Several users noticed that something was wrong when the
application was requesting more permissions than should have
been necessary: Pjapps attempted to create a bot network out
of compromised Android devices. This attack clearly demon-
strates that attackers are definitely considering smartphones as
a platform for cybercrime.

6) User as an Attack Vector: User-based attacks contain
every exploit that is not of technical nature. Many of today’s
mobile malware samples are not based on a technical vul-
nerability, but trick the user into overriding technical security
mechanisms [40]. This is an important class of vulnerabilities
and several studies have been performed to evaluate the se-
curity knowledge of the average user; in particular they discuss
the purpose of all the security mechanisms implemented by
the smartphones if the average user does not understand them.
Quite often, if the user clearly understands how to use one
specific mechanism, she might find it difficult to understand
another, possibly new and updated, mechanism.

As [81] points out, very often social engineering attacks
trick a user to override technical security mechanisms, as in
the case of abuse of trust relationships, which might happen
when a malware access the address book of the victim and
send itself to the contacts that trust the infected user. In anoth-
er scenario, a user cannot distinguish if a feature is a legitimate
functionality or an imitated one, e.g. in case of a Bluetooth
message with malicious content.

A survey conducted by Sophos [30] asked users whether
their smartphone was encrypted: 26% percent of users replied
that their data was encrypted, 50% said they were not protected
in the event of theft or loss of the device, and 24% of users
were not sure whether their smartphone was encrypted. These
results show that further education on the security dangers of
smartphones is required.

Examples: Trojan-SMS.AndroidOS.FakePlayer.b is a
Trojan for Android that requires the user to manually install
it [6] and asks the user permissions to send SMS messages.
Hence, the user has to actively participate in the installation
process. Once the installation has finished, if the user launches
the fake application, the Trojan begins sending SMS messages
to a premium rate number without the user’s knowledge.

ZeuS MitMo [25] exploits social engineering to infect
smartphones and performs online banking operations. To this
end, an attacker sends an SMS with a link to a malicious
mobile application to the user: then, if the user installs it,
the attacker can read the SMS message used as a second
authentication factor by the online banking service to access
the user’s account.

B. Goals of the Attacks

In this section, we discuss in detail the goals of the attacker,
which can be:

• privacy;
• sniffing;
• denial of service;
• overbilling.

Moreover, for each class of attack we give some pointers to
related works (which will be discussed in detail in Section V)
that propose either a model or a mechanism to defend against
these attacks.

1) Privacy: Privacy attacks of smartphones concern sit-
uations in which integrity and confidentiality are corrupted,
e.g. when they get lost or stolen. In fact, due to their small
dimensions, smartphones can be stolen or lost more often than
laptop computers. During the period in which the device is not
available for the legal owner, it may be possible that someone
installs a spyware on the phone; furthermore, someone can
read personal data, as contact list or messages.

[82] announces that a researcher has detailed a proof-of-
concept method to steal data from an Android smartphone
using a combination of cross-site scripting and Javascript. [83]
reports how Apple is being sued for allegedly letting mobile
apps on the iPhone and iPad send personal information to
advertising networks without the consent of users. Finally,
[84] presents an overview of recent iPhone privacy issues and
considers the unmodified devices by examining what sensitive
data may be compromised by an application downloaded from
the App Store. Next, it explains how a malicious application
could be crafted to fool Apple’s mandatory reviews in order to
be accepted on the App Store. Finally, it discusses several
attack scenarios, suggesting improvements and basic recom-
mendations.

A further issue, related to Digital Right Manager (DRM)
protection, is discussed in [85]. The authors propose a frame-
work for personal DRMs for Motorola smartphones, where
the user can place DRM protection and keeps control over
her personal content. To this end, the personal DRM system
enables users to define control and generate licenses on custom
content and to securely transfer them to other smartphones. A
user is then able to define and restrict the intended audience
and ensure expiration of the content as desired. Moreover,
compatible DRM devices are able to automatically detect each
other and exchange credentials.

Another fundamental topic related to privacy is location
awareness, i.e. the ability to determine geographical position.
Even if this feature has significant benefits, it raises some
important privacy implications for users of smartphones. A
comprehensive survey on several privacy issues related to
location-aware smartphones is presented in [86].

Some studies for preventing this class of attacks are pro-
posed in [87, 88]. An approach for designing anti-theft solu-
tions for smartphones is described in [89].

Examples: The target of the theft performed by ZeuS
MitMo [25] is information about online banking account.
Exploiting some social engineering techniques and the fact
that many companies use SMS as a second authentication
factor, ZeuS MitMo infects the smartphones and performs
online banking operations in place of the legal owner of the
account. To accomplish its tasks, first of all, an SMS with a
link to the malicious mobile application is sent to an owner of
an online banking account. If the user installs the application,
the attacker gains control of the infected smartphone and can read all the SMS messages that are being delivered. If the user logs into the online banking account, she is asked to enter her mobile number. This number is used by the company to improve verification system: in fact, the customer receives on her number a text message with the transaction details and a code to enter back into the website. Then, the following steps performed by this attack are, in order:

1) the attacker logs in with the stolen credentials using the user’s device and performs a specific banking operation that needs SMS authentication;
2) an SMS with the authentication code is sent to the user’s mobile device from the verification system. The malicious software running in the device forwards the SMS to a system controlled by the attacker;
3) the attacker fills in the authentication code and completes the operation.

[90] discusses a drive-by download attack, i.e. when a user inadvertently downloads a malware (usually a spyware) on visiting a website, against an iPhone 3GS that enables an attacker to steal the SMS database from the phone.

2) Sniffing: Sniffing attacks on smartphones are based upon the use of sensors, e.g. microphone, camera, GPS receiver. These sensors enable a variety of new applications but they can also seriously compromise users’ privacy. If a smartphone is compromised, an attacker can access the data stored in the device and also use the sensors to sniff and record all of the user’s actions.

A defense system against sniffing attacks is proposed in [91].

Examples: [92] describes the design and implementation of Stealthy Video Capturer (SVC), a spyware that can secretly activate the built-in camera on smartphones to compromise the users’ privacy by recording private video information, with little power consumption and is stealth to commercial antivirus.

Soundminer [93] is a proof-of-concept Trojan targeting Android devices that is able to extract private data from the audio sensor. Sensitive data, such as credit card number or PIN number, can be sent out by analyzing both tone and speech-based interaction with the phone menu system. As an example, Soundminer can infer the destination phone number by analyzing audio and reporting this data remotely to a malicious party. Few permissions are requested by the Trojan during installation, specifically that it is granted access to the microphone. Other permissions, in particular network connection or intercepting phone calls, are not requested. For this reason, since Soundminer cannot directly access the Internet, transmissions need to be carried out through a second application, either a legitimate network application or through a program with the networking permission. This allows Soundminer to circumvent mechanisms aimed at mediating explicit communications between two untrusted applications, such as those proposed in [94, 95].

3) Denial-of-Service: With a Denial-of-Service (DoS), an attacker denies availability of a service or a device. DoS attacks against smartphones are mostly due to strong connectivity and reduced capabilities: due to the limited hardware, attacking a smartphone can be accomplished with a small effort; even the traffic generated by only one attacker may be enough to make a device unusable. Specific DoS could quickly drain the batteries, shutdown or dramatically limit the operation time and perform CPU intensive tasks that require a lot of energy or force the device to shut-down. Another type of DoS sends a huge amount of SMS/MMS to the same phone number to either deny users to perform their tasks or degrade the service of an area. [96] shows that by using only SMS communications, i.e. messages sent between smartphones, low-end phones can be forced to shut down. To this purpose, the SMS protocol can be used to transmit small programs that run on a smartphone. Network operators use these files to change the settings on a device remotely. The same approach has been exploited to attack smartphones.

Some studies for preventing this class of attacks are proposed in [97, 59].

Examples: In [98], the authors examine current security mechanisms on smartphones, by identifying some critical vulnerabilities of existing security models. Moreover, they show how such vulnerabilities can be exploited to launch Distributed DoS (DDoS) attacks to public service infrastructures by diverting phone calls. This is achieved by injecting a crafted shell code through a buffer overflow: in fact, in many Linux-based mobile phone the system single-user can easily access root privileges to invoke ptrace() to inject code in any other process. The authors demonstrate that, by only leveraging 1% of Linux-based mobile systems, the service of an emergency-call center, in a region with millions of population, can be disabled.

The battery exhaustion attacks targets a unique resource bottleneck in smartphones, namely the battery power. [99] discusses attacks against smartphones that drains the device’s battery power up to 22 times faster, thus rendering a device useless in short time. To do this, first an attacker has to build a “hit-list” of all the users with an active Internet connection by taking advantage of the insecure MMS protocol, which automatically downloads MMS messages upon receiving notification through HTTP requests. Secondly, the attacker has to periodically send UDP packets to the target smartphone and exploit PDP context retention and the paging channel. The authors show that if a phone is connected to the Internet continuously, its battery life would be completely drained in less than 7 hours.

The water torture attack [100] is another example of battery exhaustion attack that is carried out at the PHY layer. This is achieved by forcing the subscriber station (SS) to drain its battery, or consume computing resources, by sending bogus frames.

An example of attack that targets several Symbian S60 smartphones and that prohibits victims from receiving SMS messages is called Curse of Silence [101]. This attack tries to set the Messages Protocol Identifier to “Internet Electronic Mail” so that an SMS can be used to send e-mails. This attack exploits a vulnerability with some smartphones that cannot handle correctly e-mail address with more than 32 characters: by exploiting this vulnerability, the attacker sends a crafted e-mail in such a way that the device is not able to receive any other SMS message. At the end of the attack, the smartphone displays a warning reporting that the memory is not enough to
receive further messages and that some data should be deleted first.

[102] presents a novel method for vulnerability analysis of SMS-implementations by injecting short messages locally into the smartphone and analyzing and testing all the SMS-based services implemented in the smartphone software stack. The vulnerability analysis is conducted by fuzzing various fields in a standard SMS message including elements such as the sender address, user data and various flags, or by fuzzing the UDH header. Another test concatenates various SMS messages in such a way to force messages to arrive out of order or sends large payloads. The authors state that through the use of the testing tools, they were able to identify several vulnerabilities that can be exploited to launch DoS attacks.

4) Overbilling: The overbilling attacks charge additional fees to the victim’s account and may transfer these extra fees from the victims to the attackers. Since many wireless services are regulated by pay-per-use contracts, these attacks are very specific to wireless smartphones.

Example: A characteristic of GPRS networks is the “always on” mode: users are billed by the amount of traffic instead of the usage time. A typical example of overbilling attack in this network is the one where an attacker, in cooperation with a malicious server located outside of the GPRS network, hijacks the IP address of the target device and starts a download session from the malicious server. Hence, the legitimate user gets charged for traffic that never requests [110].

V. SECURITY SOLUTIONS FOR MOBILE DEVICES

In this section we survey existing mechanisms that are developed to prevent different types of threats for smartphones. We present, first of all, intrusion detection systems for smartphones, then trusted mobile-based solutions. All the solutions are presented in chronological order.

Table V includes some conventional approaches typically implemented by off-the-shelf smartphone applications to provide basic security; instead, table VII lists, in chronological order, the research security solutions (described in the following sections) that provides a prototype. These solutions are classified according to their detection principles, architecture (distributed or local), reaction (active or passive), collected data (OS event, keystrokes), and OS.

A. Intrusion Detection Systems

In this section, we present the state of the art of models and tools that implement Intrusion Detection Systems (IDSes) on smartphones.

IDSes can be based upon two complementary approaches:

1) prevention-based approaches: using cryptographic algorithms, digital signatures, hash functions, important properties such as confidentiality, authentication or integrity can be assured; in this scenario, IDSes have to be running online and in real-time;

2) detection-based approaches: IDSes serve as a first line of defense by effectively identifying malicious activities. Furthermore, there are two main types of detection:

1) anomaly-based (alternative names: anomaly detection, behavior-based), which compares the “normal” behavior with the “real” one;

2) signature-based (alternative names: signature detection, misuse-based, knowledge based, detection by appearance), based upon patterns of well-known attacks.

There exist also hybrid approaches which combine the aforementioned types of detection. With signature-based approaches, the advantage is the false alarm rate that is usually very low. The disadvantage is that they can detect only known attacks. On the other hand, with an anomaly-based IDS we can detect variations of known attacks and even new attacks, but the amount of false alarms is usually quite high. Some of the metrics used to measure their effectiveness are true positive rate, accuracy and response time. [111] provides a general introduction to IDSes in cellular mobile networks.

In the following, we partition existing IDS solutions using these features:

- detection principles:
  - anomaly detection:
    - machine learning;
    - power consumption.
  - signature-based:
    - automatically-defined;
    - manually.

- architecture:
  - distributed;
  - local.

- reaction:
  - active;
  - passive.

- collected data:
  - system calls;
  - CPU, RAM;
  - keystrokes;
  - SMS, MMS.

- OS:
  - Symbian;
  - Android;
  - Windows Mobile;
  - Apple iOS.

First of all, we cluster mobile IDSes based upon the detection principles used to find anomalies: anomaly detection (which includes machine learning and power consumption), signature-based (automatically or manually defined) and runtime policy enforcement. Then, we consider both local and distributed architectures. Next, we distinguish tools that perform any kind of reaction from those that only detect anomalies. We further classify IDSes by considering what kind of data are used as input and by the OS. For each feature, all the solutions discussed are presented in chronological order.

1) Detection Principles: We partition existing IDSes using the following detection principles:

- anomaly detection;
- signature-based;
- run-time policy enforcement.
TABLE V
SECURITY APPLICATIONS FOR SMARTPHONES

<table>
<thead>
<tr>
<th>Product</th>
<th>Features</th>
<th>OS</th>
<th>License</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>WaveSecure</td>
<td>Lock and Wipe, Backup and Restore, Localization and SIM Tracking</td>
<td>Android, BlackBerry, Symbian, Windows Mobile, J2ME</td>
<td>Commercial</td>
<td>[103]</td>
</tr>
<tr>
<td>Norton Mobile Security Lite</td>
<td>Theft of Private Stuff, Unauthorized Access, Mobile Viruses Malware and Threats, Harmful Downloads</td>
<td>Android</td>
<td>Commercial</td>
<td>[104]</td>
</tr>
<tr>
<td>iCareMobile</td>
<td>Parental Control, Automatic Pornographic Content Detection</td>
<td>Symbian, Android</td>
<td>Free</td>
<td>[105]</td>
</tr>
<tr>
<td>ESET Mobile Security</td>
<td>Antivirus, Firewall, SMS/MMS Anti-spam, Anti-theft</td>
<td>Symbian, Windows Mobile</td>
<td>Commercial</td>
<td>[108]</td>
</tr>
<tr>
<td>Lookout Mobile Security</td>
<td>Lock Wipe, Backup, Wipe, Privacy of Data</td>
<td>Android, iPhone</td>
<td>Free</td>
<td>[109]</td>
</tr>
</tbody>
</table>

a) Anomaly Detection: An anomaly detection system compares the “expected” behavior of the smartphone with the “real” behavior (the actions executed at run-time by the device). The solutions included in this section either monitor distinct activities on the mobile, e.g. SMS or MMS services or Bluetooth connections, or analyze the power consumption model of the phone to detect anomalies. Moreover, we detail frameworks that adopt run-time monitoring of the activities.

As discussed in [112], we can split the architecture of a generic smartphone in the following layers:
- user;
- application;
- virtual machine or guest OS;
- hypervisor;
- physical.

For each functional layer, the authors propose several distinct features that should be collected for measuring the phone’s behavior and used by an anomaly detection IDS. Table VI lists some of the capabilities that can be measured at each distinct layer.

Anomaly-based approaches for smartphones are either based upon machine learning techniques or upon monitoring power consumption.

Machine Learning: The paper [57], from the same authors of [59], proposes Proactive Group Behavior Containment (PGBC), a framework aimed at containing malicious software spreading in messaging networks such as IM and SMS/MMS. The primary components of PGBC are service-behavior graphs generated from client messaging patterns and behavior clusters that partition the service-behavior graph into clusters of similar behavior. The authors present an algorithm that, in the presence of a malicious attack, automatically identifies the most vulnerable clients based upon interactions among them. Moreover, they describe a proactive containment framework that applies two commonly-used mechanisms (namely, rate-limiting and quarantine) to the dynamically-generated list of vulnerable clients in a messaging network whenever a worm or virus attack is suspected.

Ho and Heng [46] attempt to identify generic behavioral patterns in mobile malware to build a generic defense model. To slow down the spread of mobile malware, the authors introduce an extended model (which incorporates the works of [47] and [48]), which is a Java-based engine that is independent from the platform. In addition to previous solutions, this model includes a feature for blocking the silent automated transmission attempts of virus installation files from a compromised mobile smartphone via MMS, Bluetooth, Infrared, e-mail and Instant Messaging.

Schmidt et al. [48] exploit the monitoring process of smartphones to extract features that can be used in a machine learning algorithm to detect anomalies. The framework includes a monitoring client, a Remote Anomaly Detection System (RADS) and a visualization component. The monitoring client, which represents the main goal of this paper, is a client that runs on a smartphone and includes three main components:
- user interface;
- communication module, which manages connection with RADS;
- Feature Extractor, which implements measurements and observation of resources and other components.
Features from raw and time-oriented security data. These patterns are integrated with a smartphone security domain knowledge. Using KBTA, continuously measured data (e.g., the number of sent SMSs) and events (e.g., software installation) are monitored to detect suspicious states towards each key system call based on user operational behavior's profiles. Some other solutions use a probabilistic approach to trace malware behavior. For example, [49] devises a behavior-based malware detection system \((pBMDS)\) that adopts a probabilistic approach through the correlation of user’s inputs with system calls to detect anomalous activities. pBMDS observes unique behaviors of the applications and the user’s input and leverages hidden Markov model to learn application and users’ behavior through process state transitions and user behavioral patterns. This statistical approach is used to learn the behavioral difference between applications initiated by user and applications initiated by malware. The most distinguishing feature of the proposed solution is that its malware detection capability focuses on recognizing non-human behavior instead of relying on known attack signatures to identify malware. Therefore, in the training process, pBMDS does not require the number of negative samples to be equivalent to that of the positive ones. The system exploits behavior graphs, which reflect intermediate process states towards each key system call based on user operational patterns, e.g. keystrokes.

In [113], time-stamped security data are continuously monitored within the target smartphone and then processed by the Knowledge-Based Temporal Abstraction (KBTA) methodology. Using KBTA, continuously measured data (e.g., the number of sent SMSs) and events (e.g., software installation) are integrated with a smartphone security domain knowledge-base, which is an ontology for abstracting meaningful patterns from raw and time-oriented security data. These patterns are used to create higher level, time-oriented concepts and patterns, called \textit{temporal abstractions}. Automatically-generated temporal abstractions are then monitored to detect suspicious temporal patterns and to issue an alert. These patterns are compatible with a set of predefined classes of malware as defined by a security expert employing a set of time and value constraints.

Damopoulos et al. [114] use a large dataset of iPhone users’ data log with four machine learning algorithms, namely Bayesian Networks, Radial Basis Function, K-Nearest Neighbors and Random Forest, to detect illegal use of a smartphone. They classify the behavior of users through telephone calls, SMSs and, differently from earlier research, on Web browsing history. To preserve users’ anonymity, each record is hashed through SHA-1. Then, the data are examined either independently or in combination in a Multimodal fashion: in the first case, Random Forest was the most promising classifier with a true positive rate above 99.8\% and accuracy of 98.9\%; in the second case, the best results were given by K-Nearest Neighbors with 99.8\% true positive rate and 99.5\% accuracy.

\textit{Andromaly} [115] is a general and modular framework for detecting malware on Android smartphones using a supervised anomaly detection technique. The framework is based upon a host-based IDS that sample numerous system metrics, such as CPU consumption, number of sent packets, number of running processes. The framework is composed of four main components, namely the feature extractors, processors, main service and the graphical user interface. The classifiers used to evaluate the framework are \(k\)-means, logistic regression, histograms, decision tree, Bayesian networks and Naïve Bayes; furthermore, a filter approach has been used for feature selection. A total of 88 features were collected for each monitored application. Experiment results show that Naïve Bayes and logistic regression were superior over the other classifiers in most of the testbed configurations and Fisher score was the best setting for feature selection.

[116] presents a framework to dynamically analyze applications behavior to detect malware on Android by collecting system traces from several real users through crowdsourcing and a central server. An application client, \textit{Crowdroid}, monitors Linux system calls and, after preprocessing, sends them to a central server, which then parses data and creates a system call vector. Finally, each dataset is clustered through a partitional clustering algorithm, namely 2-means.

\textit{Power Consumption:} An example of monitoring power consumption is presented in [117], where the authors propose a battery-based IDS. This solution performs the sensing of

<table>
<thead>
<tr>
<th>Layer</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypervisor</td>
<td>address spaces, data types, data constructors and data fields, system parameters, virtual registers, system calls, communication protocols</td>
</tr>
<tr>
<td>Application</td>
<td>application manager framework, security framework, messaging framework, multimedia framework</td>
</tr>
<tr>
<td>User</td>
<td>SIM/phone password, key-stroke/T9 usage/spelling analysis, top-n called/texted numbers/contacts based on hour, day, week, month, frequently executed smartphone applications, smartphone application usage analysis, Bluetooth/Wi-Fi usage analysis</td>
</tr>
</tbody>
</table>

**TABLE VI**

ANOMALY DETECTION CAPABILITIES AT DISTINCT LAYER
abnormal battery behavior and energy patterns to detect a variety of attacks and includes two modules:

- **Host Intrusion Detection Engine (HIDE)**, which measures energy consumed over a period of time (established according to an algorithm);
- **Host Analyzed Signature Trace Engine (HASTE)**, which matches frequency patterns of attacks based upon power signature and compares these signatures with a short list of known attack signatures.

Similarly to the previous solution, Kim et al. [97] propose a power-aware malware detection framework that monitors, detects, and analyzes energy-greedy malware. The framework is composed of a power monitor, to collect power samples and to build a power consumption history from the collected samples, and of a data analyzer, to generate a power signature from the constructed history.

Another example of monitoring power consumption is proposed in [118], which bases its observations on the fact that any malware activity on a smartphone consumes battery power. Hence, the proposed solution (VirusMeter) performs malware detection using state machine and power consumption model by comparing the actual measured power consumption with the power that could have been consumed, according to the predefined power consumption model. The user-centric power model characterizes power consumed as a function of common user operations and relevant environmental factors (e.g., calls, SMS, MMS). To collect data, for each user operation a state machine is constructed. This state machine describes the evolution of internal events related to user’s operations.

Even if power consumption is often considered as a limitation, Yan et al. [119] consider its positive impact on preventing and suppressing mobile malware. Even if no specific solutions are implemented, three potential techniques, based upon limitations of smartphone, are proposed:

- **monitoring power consumption**, which concerns the ability to detect an attack depending on battery usage on smartphone;
- **enforcing hardware sandbox**, in which all the hardware modules that are not used frequently and that can be used for malware propagation (e.g. GPS or Bluetooth) are switched off;
- **increasing platform diversity**, in which different APIs are used to develop and execute an application.

**b) Signature-Based:** This paragraph discusses mechanisms that detect anomaly on smartphones using signatures. The signature-based approach checks if each signature derived from an application matches any signature in a malware database. The database of malware signature can be automatically or manually defined.

**Automatically-Defined:** Venugopal et al. [120] apply Bayesian decision theory to the dynamic-link library (DLL) usage of a program to detect viruses. In fact, since most mobile viruses have common functionalities (e.g., deleting system files, sending MMS), these programs need to use DLLs. By exploiting the common patterns of DLL usage among viruses, the proposed approach can detect old and new viruses: a classifier takes as input a binary vector specifying the occurrence (or not) of each DLL function in the feature set.

In [121, 122] the authors propose to adopt a twofold approach to block the spreading of worms on smartphones:

- at the **terminal level**, where a graphic Turing test and identity-based signatures block unauthorized messages from leaving compromised phones;
- at the **network level**, where a push-based automated patching scheme clean compromised phone. This means that network providers automatically push software patches to compromised terminals.

In [123] the signatures to be compared are automatically generated by analyzing the device’s activities. The authors discuss the dynamics of mobile malware that propagate by proximity contact and explore three strategies to detect and mitigate proximity malware, namely:

- **local detection**, in which devices detect when they become infected and disable further propagation;
- **proximity signature dissemination**, in which devices create content-based signatures of malware and disseminate them via proximity communication as well;
- **broadcast signature dissemination**, in which a centralized server aggregates observations from individual devices, detects propagating malware, and broadcasts signatures to smartphones.

Bauckhage et al. [124] present a probabilistic diffusion scheme for detection anomalies indicating malware, which is based upon the device’s usage patterns. The basic idea is to model dependencies of samples and features by means of a bipartite graph, which then serves as the domain of a Markov process. The algorithm is applied to two separate data sets obtained from smartphones during normal daily usage.

**Manually-Defined:** These mechanisms to detect anomaly extract the signatures of mobile malware by manually analyzing the malware and its behavior.

Ellis et al. [125] present a new approach for the automatic detection of worms on smartphones using behavioral signatures, which are manually defined to represent common features across a family of worms. The presented approach focuses on detecting patterns at a higher level of abstraction, where a pattern may be:

- sending similar data between devices;
- tree-like propagation and reconnaissance;
- changing a server device into a client.

Ideally, a pattern is a specific behavior of a spreading worm and should be distinct from normal network traffic. The frequency of and interrelationships between behaviors improve the accuracy of the detection. To evade a behavioral signature requires a fundamental change in the behavior and this task is rather challenging.

In [47], the authors present a behavioral detection framework for viruses, worms and Trojans that extracts key behavior signatures of mobile malware by applying **Temporal Logic of Casual Knowledge** (TLCK) on a set of atomic steps. The authors have generated a database of malware signatures based upon a review of mobile malware: every signature corresponds to the description of the behavior of a family of malware. The run-time monitoring is implemented on the Symbian emulator.
through a proxy DLL to monitor API calls. To distinguish malicious behavior from partial signatures, the framework exploits support vector machines to train a classifier from normal and malicious data.

c) Run-Time Policy Enforcement: The basic idea of these models is that mobile code consumers essentially accept the code “as-is” and exploit a supporting mechanism to enforce the policy associated with the code to detect and stop anomalies.

Nowadays, several smartphones are able to run Java applications, which can also create Internet connections, send SMS messages, and perform other expensive or dangerous operations on the smartphone. Hence, an adequate security support is required to meet the needs of this scenario. To this end, [126] proposes an approach to enhance the security support of Java Micro Edition (J2ME), based upon the monitoring of the usage of the smartphone’s resources performed by MIDlets. A process algebra-based language defines the security policy whereas a reference monitor is exploited to check the resource usage.

Security-by-contract [127, 128] is a run-time policy enforcement solution based upon a digital signature that:
- certifies the origin of the code;
- binds the code with a contract.

A contract contains a description of the relevant features of an application and the relevant interactions with its host platform. A mobile platform can specify platform contractual requirements (a policy), which should be matched by the application’s contract. The authors also propose some algorithms to verify contract-policy matching.

[129] enhances the security-by-contract architecture by adding new modules and configurations for managing contracts. At deploy-time, the proposed system selects the runtime configuration depending upon the credentials of the contract provider; at run-time, the system can both enforce a security policy and monitor the declared contract. According to the actual behavior of the running programs, the architecture can update the trust level associated with the contract provider. The main advantage of the proposed architecture is the automatic management of the level of trust of software and contract releasers.

Finally, [94, 95] propose Kirin security service for Android, which performs lightweight certification of applications to mitigate malware at install time. Kirin certification uses security rules that match undesirable properties in security configuration bundled with applications.

2) Architecture: Local or Distributed: In this section, we partition mechanisms for intrusion detection that use a local architecture from solutions that exploit a distributed architecture.

In a local architecture, both the collecting phase and the analysis phase are locally performed on the device and no interactions with an external server is required. Examples of local solutions are presented, for instance, in [130, 118, 94, 95]. On the other hand, a distributed architecture usually requires a distinct and separated component (i.e., a server) to analyze the activities collected and sent by each device. In this architecture, the external security component can perform all the analysis on the activities monitored on the smartphones without having problems of:
- power consumption;
- small screen size;
- limited resource.

[39] discusses a framework to detect attacks against smartphones using a separate, loosely-synchronized, security server, which hosts one or more exact replicas of the smartphone and applies distinct detection mechanisms. In this way, several expensive detection techniques, which cannot be easily implemented on the smartphone, can be applied on the server to prevent attacks on the phone software. The architecture includes a tracer on the phone itself, to intercept both system calls and signals of a set of protected processes, whereas a replayer on the security server later replays the execution trace and looks for anomalies.

3) Reaction: Passive or Active: In this section, we consider whether existing mechanisms for intrusion detection react or not whenever a new threat is found, e.g. by trying to prevent the attacks to damage the smartphone.

A reaction can be a strategy, to contain the virus or malware propagation, or a mechanism, like alerting the user of the infection. An example of a reaction strategy is presented in [57], where the alerts about potential attacks are collected before starting the reaction strategy. The proposed framework is implemented at the messaging service center where logs of client communication are kept. These logs can be analyzed to generate a service-behavior graph for the messaging network: this graph is then further processed to generate behavior clusters, i.e., groups of clients whose behavior patterns are similar with respect to a set of metrics, namely:
- interaction frequency;
- attachment and message size distributions;
- number of messages;
- number of outgoing connections to other clients;
- list of traced contacts.

When the number of alerts in a particular behavior cluster reaches a threshold, the messages belonging to that cluster are first rate-limited to slow down a potential malware. When the alerts reach a second threshold, the containment algorithm applies proactive quarantine, i.e., it blocks messages from suspicious clients of these behavior clusters. This step essentially enables the behavior clusters to enter into a group defense mode against the spreading malware.

In [130], the proposed system detects viruses running on smartphones using a proxy that performs the analysis of the phone’s behavior and, when a potential virus is detected, sends targeted alerts to both infected devices and a subset of the uninfected devices to prevent the spreading. These devices are chosen based upon the users’ contact list and mobility profiles, to locate those devices they may be in direct contact with an infected device.

A further example of a reaction-strategy is given in [58] where, to limit the spread of MMS- and SMS-based worms, the authors propose a methodology based upon a graph partitioning approach. The problem is tackled by using a social relationship graph where the devices are divided into multiple partitions based upon the social relationship among them. Each
partition includes smartphones that closely interact with each other. Communication patterns are extracted using a network trace and, from these patterns, a social relationship graph of smartphones is built so that an optimal set of phones to be patched can be located firstly. In this way, the system can locate the phones that have the capability to infect the highest number of other phones. The authors propose two patching schemes, namely:

- **balanced partitioning**, where the significance level of each partition is chosen to be as similar as possible so that the worm damage to each partition can be balanced;
- **clustered partitioning**, where edges within each partition have higher weights compared to the edges between the two partitions, so that smartphones that are socially close to each other are in the same partition and nodes that are not close are into different partitions.

Ruitenbeek et al. [131] study the propagation of smartphone viruses based upon MMS and propose several response mechanisms to quantify the effectiveness of virus mitigation techniques. The authors present four MMS virus scenarios: in every scenario the virus on the phone sends MMS messages with an infected attachment file to other phones, which are selected from the contact list of the infected phone or by dialing a random phone number. After receiving this new MMS message, if the user accepts the infected attachment file, the virus is installed, the target phone becomes infected and under control of the attacker. The evaluated response mechanisms for each of the four scenarios are:

- scan of all MMS attachments in MMS gateways to detect viruses;
- user education;
- immunization using software patches;
- monitoring for anomalous behavior;
- blacklist of phones that are suspected of infection.

The experimental results revealed that any response mechanism must be agile enough to quickly react to rapidly propagating viruses and discriminating enough to detect stealthier, slowly propagating, viruses.

4) **Collected Data**: As suggested in [132], monitoring data in a mobile environment can be a challenge due to administrative, technical and conceptual limitations. The visibility of the data can be limited by explicit agreements on exchange of monitored data because, for example, the call records can be under administrative control. Furthermore, we have to consider that due to the integration of several communication interfaces, a smartphone can be connected, at the same time, to different access points (such as Bluetooth, Wi-Fi) and thus the amount of collected data on the communication interfaces can be huge. Finally, we have to take into account that there are some kinds of attacks, e.g. Trojans, not involved in communication activities, so that they are invisible to the network.

Due to the fact that all the solutions based upon intrusion detection need to access several features of a smartphone, we should carefully consider the problem of privacy of the data accessed. By accessing a smartphone, several pieces of information can be retrieved, such as the user’s location, communications, and personal contacts. All these kinds of information are, obviously, private and should not be shared to third parties: hence, any tool for intrusion detection should be run only with the user’s explicit authorization in case it may access private data. Private data should never be carried out of the smartphone by any mechanisms, including system that have to protect the mobile. Therefore, an intrusion detection system should be designed so that:

- the private data used during the intrusion detection process are not transferred out of the device;
- the communication of the intrusion detection functionalities on the smartphone should be restricted to the generated intrusion alarms.

In [91] the authors focus on threats and attacks that violate user’s privacy by sniffing on the sensors on smartphones. The authors develop a threat model based upon the use of sensors and design a general framework for a defense system. This framework consists of three modules: (i) policy engine, (ii) interceptor, (iii) user interaction. The policy engine, based upon the input from user interaction and application monitoring/profiling, determines access: these decisions are based mainly upon application monitoring and profiling without requiring much user intervention. Several policies are considered, such as white-listing, blacklisting and information-flow tracking. The interceptor is interposed between the application and the sensors, and/or between the application and the network, and it enforces the decision of the policy engine. The user interaction is not a mandatory component, since it simply notifies the user by asking her decision. For each module, different mechanisms are explored and discussed but no real implementations are presented.

In the following paragraphs, we discuss solutions that analyze different kinds of data available in a smartphone, namely:

- OS events (e.g., system calls);
- measurements (e.g., CPU, RAM, I/O activities);
- keystrokes;
- communication events (e.g., SMS/MMS).

4) Operating System Events: These events indicate those activities related to the normal functioning of the OS, which can be used to retrieve relevant information about the behavior of the smartphone, and include:

- system calls;
- function calls;
- network operations.

Data from these events can be obtained by exploiting some built-in OS mechanisms. However, in many cases the OS events cannot be monitored because there is no direct access interface. Therefore, in these cases some extensions to the OS are required.

The solution presented in [133] monitors the system calls executed by running processes and labels executing code based upon its access to the network interfaces (e.g., wireless, GSM, Bluetooth). The labels are then transferred between processes and system resources as a consequence of either access or execution. During sensitive operations, the labels collected in this way are compared to a set of rules that allow users to specify fine-grained access control to services and data. The labeling process, which can involve processes and resources, as well as the enforcement of the policies, are performed by a
kernel-level reference monitor. The framework is independent of the OS because it uses a policy language that allows users to express what actions are allowed by specific classes of programs with respect to specific classes of resources.

Another mechanism, MobileSandbox [50], monitors the system calls issued by the software that a user is going to install on its smartphone. MobileSandbox is a background system that samples and translates an operation in a sequence of API calls that the program has issued during its execution. The proposed system can either use the device emulator or an actual device. To perform mobile dynamic malware analysis, the proposed solution includes three steps, namely:

- collecting the software samples as complete as possible;
- analyzing the samples as complete as possible;
- taking certain actions as a response to the analysis.

In [134], the authors discuss in details the implementation of system call interception for Windows CE. Two solutions that analyze these events are proposed: the first one exploits a sandbox approach to analyze malware, whereas the second solution implements the concept of a reference monitor at the OS level.

There are also some solutions to detect anomalies that monitor function calls rather than system calls. One of these solutions is presented in [135] where the authors apply static analysis of function calls to detect malicious applications. Centroid Machine is the name of the light-weight algorithm developed according to common clustering methods. The algorithm can detect Symbian OS malware on the basis of function calls according to the requirements of smartphones, e.g., efficiency, speed and limited resource usage.

A similar solution specific to Android platform is presented by Schmidt et al. [136]: the discussed framework performs static analysis on the executables to extract their functions calls using readelf command. This command returns detailed information on relocation and symbol tables of each Executable and Linking Format (ELF) object file. The output of this analysis is the static list of referenced function calls for each system command. Then, these calls are compared with malware executables for classification.

b) Measurements: A set of measurements includes several performance indicators of a smartphone, such as CPU activity, memory consumption, file I/O activity and network I/O activity. The key idea is that, supposing that changes in the usage of a smartphone are gradual, the normal usage remains constant over time. Therefore, we can extract behavioral profiles and use them for comparison with normal behaviors in order to detect anomalies. Some of these features (e.g., RAM free, user inactivity time, process count, CPU usage, SMS sent count), which are used for anomaly detection, are discussed in [48, 115].

c) Keystrokes: Some solutions exploit keystroke logging (keylogging) techniques to detect anomalies. These techniques track the keys struck on a keyboard to monitor the actions of the user. Typically, the logging is provided in a covert manner so that the user is unaware of the monitoring. The conventional technique of behavior-based anomaly detection focuses on the rhythm of keystroke patterns or transition probability of commands.

In [137], keylogging is applied to smartphones to detect illegal user operations using a scheme that uses a background process to record keystrokes. Keystrokes are divided into long-term and short-time: using the frequency of long-term keystrokes, an anomaly detection algorithm constructs a user profile; then, the short-term keystrokes are compared with the user profile to detect illegal users.

[138] demonstrates that keystroke dynamics of a user can be translated into a set of features for accurately identifying users. To this end, keystroke data of twenty-five smartphone users are collected and analyzed: based upon them, six distinguishing keystroke features are extracted and used for user identification. The results show that the proposed user identification system has an average error rate of 2% after the detection mode and the error rate of rejecting legitimate users drops to zero in the PIN verification mode.

d) Communication Events: Communication events indicate a particular class of events that happen in a device at the application level, such as high-level actions (e.g. sending/receiving of SMS, files). Typically these actions are composed of several elementary actions that cannot be automatically generated by the smartphone’s OS. Communication events include operations such as sending and receiving of messages, or file downloads/uploads.

[139] discusses threats on smartphones mirrored from PCs and proposes a detector application that monitors SMS sent without user authorization. Bose and Shin [59] investigate the propagation of mobile worms and viruses that spread primarily via SMS/MMS messages and short-range radio interfaces (e.g. Bluetooth). Each smartphone is modeled as an autonomous mobileagent capable of sending SMS messages and of discovering other devices equipped with Bluetooth. To identify a set of common behavior vectors, and to develop mobile virus detection and containment algorithms, the existing mobile viruses are investigated. The authors study the vulnerabilities of Bluetooth and SMS/MMS messaging systems in depth and identify the vulnerabilities that may be exploited by future mobile viruses. Finally, they develop the state diagram of a generic mobile virus that can spread via SMS/MMS and Bluetooth. The discovery, infection and replication states of the generic virus are implemented in an agent-based malware modeling framework to study its propagation and containment strategies.

[140] presents a novel approach to the security testing of MMS user agents by taking into account the effects of the infrastructure on the delivery of MMS messages and then by using a virtual infrastructure to speed up the testing phase. As in [102], the paper exploits fuzzing to deliver fuzzed MMS messages to the user agents, by finding several string-length buffer overflows. SmartSiren [130] is a collaborative virus detection and alert system for smartphones that collects communication activity information and performs analysis to detect abnormal behaviors. To halt potential virus outbreaks the authors try to minimize the number of smartphones that can be infected by a new released virus. A light-weight agent runs on each smartphone, while a centralized proxy assists the virus detection and alert process. Each agent keeps track of the communication activities on the device and periodically reports a summary of these activities to the proxy. The proxy
performs joint analysis on the reports and detects any single-device or system-wide viral behaviors.

Finally, [141] proposes a lightweight scheme that can detect anomalous SMS behaviors with high accuracy. The authors start analyzing an SMS trace collected within a five-month period and, according to the analyzed results, four detection schemes are proposed. Each scheme builds normal social behavior profiles for each SMS user and then uses them to detect SMS anomalies in an on-line and streaming fashion. Since a scheme stores only a few states in memory for each SMS user, it imposes very low overhead during on-line anomaly detection. Finally, the authors evaluate these four schemes and also two hybrid approaches with realistic SMS traces by showing that the proposed approach can detect more than 92% of SMS-based attacks with a false alarm rate of 8.5% and 66% of attacks without generating any false alarm.

5) Operating Systems: In this section, existing IDSes are clustered based upon the OS for which they are developed or studied, namely:

- Symbian;
- Android;
- Windows Mobile;
- iPhone OS.

a) Symbian: Symbian is the Nokia’s open source OS and software platform designed for smartphones. The latest Symbian platform includes a user interface component based on S60 5th Edition. Originally developed by Symbian Ltd, Symbian OS is now at the third version, Symbian’3, released in the fourth semester of 2010. Devices based on Symbian accounted for 43.5% of worldwide smartphone sales in the first two semesters of 2010 [142].

Schmidt et al. [48] introduces an approach to monitor Symbian-based smartphones to extract features that can be used by a machine learning algorithm to detect anomalies. In [143], the authors test vulnerabilities on smartphones based upon the Symbian 9.1 OS. Several attacks have been experimented to test the stability of the network stack of the Symbian OS. Some vulnerability has been found that can render the devices unusable.

Further discussions can be found in [144, 145].

b) Android: Android is a mobile OS based upon a modified version of the Linux kernel. The applications are written by a large community of developers to extend the functionality of the devices.

[146] and [147] present a set of results on the evaluation of the security of Android smartphones. Firstly, the authors analyze the Android framework and the Linux Kernel to check security functions, by also surveying some known security tools and mechanisms to increase the smartphones security. Then, the authors analyze the possibilities of applying malware detection mechanisms at the kernel-level, i.e. by monitoring key-kernel events (log file, file system activities). Finally, they apply static function call analysis to detect malware on ELF executables, by exploiting a decision tree for deciding if a new application is suspicious compared to previously analyzed applications (both good and bad ones). [148] presents a collaborative architecture to detect anomalies on Android platforms. [149] presents a review of current security solutions for Android platforms.

In [150], the authors provide a security assessment of the Android framework: firstly, they discuss the current security mechanisms incorporated in Android (namely, Linux mechanisms, environmental features and Android-specific mechanisms); secondly, they propose some security solutions for mitigating threats on Android, using five “threat cluster”:

1) threats that compromise availability, confidentiality, or integrity by maliciously using the permissions granted to an installed application;
2) threats that compromise availability, confidentiality, or integrity that happen when an application exploits a vulnerability in the Linux kernel or system libraries;
3) threats that compromise the availability, confidentiality, or integrity of private or confidential content: e.g., applications that can read the SD card’s contents or attackers eavesdropping on wireless communication remotely;
4) attackers draining a smartphone’s resources: e.g., since applications for Android have neither disk storage nor memory quotas hogging memory or CPU is possible;
5) threats that compromise internal or protected networks: as an example, attackers can use Android devices to compromise other devices, computers, or networks by running network or port scanners, SMS/MMS/e-mail worms, and various other attacks.


TaintDroid [87] is an extension to Android that tracks the flow of sensitive data through third-party applications. TaintDroid assumes that downloaded third-party applications are not trusted, and monitors how these applications access and manipulate users’ personal data. TaintDroid automatically labels data from sensitive sources and transitively applies labels as sensitive data propagates through program variables, files, and interprocess messages. When tainted data are transmitted over the network, TaintDroid logs the data’s labels, the application responsible for transmitting the data, and the data’s destination. The tested performance overhead is 14%.

Enck et al. [152] study 1,100 popular free Android applications using a decompiler to recover Java application source code from its Dalvik installation image and by statically analyzing more than 20 million lines of code. The study shows that several applications misuse privacy sensitive information, in particular phone identifiers (IMES, IMSI and ICC-ID) and geographic location, such as leaking phone identifiers through plaintext requests (such as HTTP GET/POST), tracking users through their phone identifier.

c) Windows Mobile: Windows Mobile is an OS developed by Microsoft for use in smartphones. Based upon the Windows CE 5.2 kernel, Windows Mobile was designed to be similar to desktop versions of Windows and is now superseded by Windows Phone 7. Third-party software development is also available and users can purchase software applications via the Windows Marketplace for Mobile.

Windows Mobile Malware Detection system (WMMD) [153] is a behavior-based malware detection system for Windows Mobile platform. WMMD uses API interception techniques to dynamic analyze application’s behavior and compare it with malicious behavior characteristics library using model
checking. The results show that it can effectively detect the obfuscated or packed malware variants that cannot be detected by other main stream anti-virus products.

An example of anomaly detection system developed for the Advanced RISC Machine (ARM) architecture, is discussed in [154]. The proposed system analyze the relation between system calls and their return address to describe the software behavior on ARM architectures. To this end, during a training phase, the system builds a model using an array of system calls and return addresses and produces a score. Then, at runtime the system freezes the software execution each time a system call is issued, obtains stack information and compares the software behavior with the model using an anomaly score.

d) iPhone OS: iOS (previously known as iPhone OS) is the Apple’s mobile OS that was originally developed for the iPhone, but now it has been extended to support other devices, such as iPad.

To protect its users from malicious applications, Apple has introduced a vetting process, which should ensure that all applications conform to Apple’s rules before they can be offered via the App Store. Unfortunately, this vetting process is not well documented, and there have been cases where malicious applications had to be removed from the App Store after user complaints. To this purpose, Egele et al. [88] study the privacy threats that applications for Apples iOS pose to users. The authors present a novel approach to automatically create comprehensive control-flow graphs from binaries compiled from Objective-C code and perform reachability analysis to identify possible leaks of sensitive information from a smartphone to third parties.

Finally, [155] discusses ten ways to keep user data on iPhone and Android-based devices safe from insecure apps.

B. Trusted Mobile

Trusted Computing Group (TCG) has published a set of specifications to measure, store, and report hardware and software integrity through a hardware root-of-trust, which is the Trusted Platform Module (TPM) and Core-Root-of-Trust-Measurement (CRTM). On a TPM-enabled platform, the CRTM measures the bootloader of the system before it is executed, and then stores the measured value into one of the Platform Configuration Registers (PCRs) inside the TPM. The bootloader then loads OS image, measures it, stores via PCR extension and then executes it [159]. In turn, the OS measures the loaded applications and stores their integrity values in PCRs before executing them. Upon an attestation challenge from a third party, the TPM signs a set of PCR values with an Attestation Identity Key (AIK) and sends back the result. The challenger then can make decisions on the trust status of the platform by verifying the integrity of these values and comparing with the corresponding known-good values.

There are also specifications for mobile phone platforms released by the TCG Mobile Phone Working Group, i.e. the Mobile Trusted Module (MTM) [160]. TCG advocates using MTM to increase the security of smartphones by providing basic cryptographic capabilities, such as random number generation, hashing, protected storage of sensitive data (e.g. secret keys), asymmetric encryption, as well as generation of signatures. These cryptographic primitives can be exploited to implement hardware-based security services, such as device authentication, integrity measurement, secure boot, and remote attestation. The MTM provides a root-of-trust for smartphones in the same way as the TPM does for personal computers. In principle, the MTM is an adaption of the TPM for smartphones and, hence, its specification is similar to that of the TPM, which facilitates interoperability within the existing trusted computing framework for personal computers.

There are two different types of integrity measurement for any application binary: load-time and dynamic measurements. The TCG only specifies load-time integrity measurement, when a piece of code or data is measured or when it is mapped/loaded into main memory. Dynamic measurements refer to the act of measuring the integrity of critical applications at run-time, i.e. when they are executing. In the Integrity Measurement Architecture (IMA) framework [161], measurements are invoked in several system call functions when code or kernel modules are loaded but before they are executed. After a code is mapped into memory and during runtime, it is very difficult to measure the integrity of the process considering very dynamic and nondeterministic behaviors of typical applications, such as loading active code, receiving external inputs, and allocating dynamic memory.

In addition to trusted boot and load-time integrity measurement, integrity protection for mobile phones raises the following extra requirements [158]:

- **secure boot**: a set of mandatory engines [160] reside on a single mobile platform and provide critical and indispensible services that have to be running in known-good states, i.e. their integrity must be verified to assure their trustworthiness. Therefore TCG Mobile Phone Reference Architecture [159] states that secure boot is mandatory for MTM;
- **low booting and run-time overhead**: most smartphones are still limited in computing power. This requires any security solution to be very efficient and integrity measurement during boot and in post-boot state should not degrade the performance and user experience too much;

**run-time integrity assurance**: although run-time integrity measurement is not practical in both PC and mobile platforms, there should be some mechanism to preserve the integrity level of critical applications and resources during run-time, e.g., phone related services (telephony server) and platform management agents. Both TCG and IMA do not propose any mechanism for this purpose.

[156, 158] discuss a framework for mobile integrity measurement and attestation mechanisms, by proposing a secure boot mechanism. The proposed mechanism ensures that a mobile platform can boot into a secure state by exploiting a flow integrity model to achieve high integrity for the system. The solution leverages SELinux MAC mechanisms and adds some SELinux security policy extensions. The framework requires a root-of-trust, such as the MTM. [157] tries to protect the integrity of critical applications from potentially untrusted functionality and develop a small SELinux policy to measure the integrity of a mobile phone using the PRIMA approach [162]. The resulting SELinux policy enables the phone system to be attested to remote parties and protects
critical applications from untrusted code, thus allowing users to install and run trusted applications in a safe way: the policy is 90% smaller than a custom SELinux reference policy.

In [163] the authors propose a practical approach for the design and implementation of trusted mobile platform. The approach, based upon the concept of a trusted platform as a set of trusted engine, defines a method for the take-ownership of a device by the user and the migration (i.e., portability) of user credentials between devices.

[164] identifies three specific problems in the MTM specification and provides some possible solutions. The first one concerns the need of balancing some contrasting goals at the system-level designs, such as performance and power consumption. A suggested solution integrates some TPM features directly into a processor core as opposed to a monolithic implementation of all the functions in a separate module. The second problem considers which cryptographic algorithms a MTM must support: some algorithms, namely RSA and SHA-1, can either have bad performances or security weaknesses. The suggested solution consides elliptic curve cryptography as a viable alternative. Finally, the third problem is related to the implementation of cryptographic primitives: the authors propose a hardware/software solution as opposed to a hardware-based solution, which suffers from poor flexibility. Further solutions can be found in [165].

VI. Conclusions

With the rapid proliferation of smartphones equipped with a lot of features, as multiple connections and sensors, the number of mobile malware is increasing. Differently from PC environment, solutions aimed at preventing the infection and the diffusion of malicious code in smartphone have to consider multiple factors: the limited resources available, including the power and the processing unit, the large number of features that can be exploited by the attackers, such as different kinds of connections, services, sensors and the privacy of the user.

In this work, first of all we have discussed the current scenario of mobile malware, by summarizing its evolution, along with some notable examples; we have also outlined likely future threats and reported some predictions for the near future. Secondly, we have categorized known attacks against smartphones, especially at the application level, focusing on how the attack is carried out and what is the goal of the attacker. Finally, we have reviewed current security solutions for smartphones focusing on existing mechanisms based upon intrusion detection and trusted mobile platforms.
ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions, which have greatly improved the quality of the paper.

Work partially supported by EU-funded project FP7-256980 NESSoS.

REFERENCES


Mariantonietta La Polla received her Bachelor’s and Master’s degrees in Computer Engineering from University of Pisa respectively in 2006 and 2010. Currently she is a Research Fellow at IIT-CNR, Pisa. She is also a Ph.D. student in Information Engineering at the Engineering Ph.D. School “Leonardo da Vinci”, University of Pisa. Her research interests include the study of techniques for Web acquisition and Web data analysis.

Fabio Martinelli (M.Sc. 1994, Ph.D. 1999) is a senior researcher of IIT-CNR, Pisa, where he is the scientific coordinator of the security group. His main research interests involve security and privacy in distributed and mobile systems and foundations of security and trust. He serves as PC-chair/organizer in several international conferences/workshops. He is the co-initiator of the International Workshop series on Formal Aspects in Security and Trust (FAST). He is serving as scientific co-director of the international research school on Foundations of Security Analysis and Design (FOSAD) since 2004 edition. He chairs the WG on security and trust management (STM) of the European Research Consortium in Informatics and Mathematics (ERCIM). He usually manages R&D projects in information and communication security and he is involved in several FP6/7 EU projects.

Daniele Sgandurra received his M.Sc. degree in Computer Science in 2006, and a Ph.D in Computer Science in 2010, both from the University of Pisa. He is currently a PostDoc Researcher at IIT-CNR, Pisa, where his main research fields include virtualization security, intrusion detection systems, critical infrastructures and risk management, cloud security and mobile security.