Assessing privacy risks in Android: A user-centric approach

Alexios Mylonas, Marianthi Theoharidou, and Dimitris Gritzalis

Information Security and Critical Infrastructure Protection Research Laboratory
Dept. of Informatics, Athens University of Economics & Business
76 Patission Ave., Athens GR-10434, Greece
{amylonas,mtheohar,dgrit}@aueb.gr

Abstract. The increasing presence of privacy violating apps in app marketplaces poses a significant privacy risk for smartphone users. Current approaches assessing privacy risk, lack user input, assuming that the value of each smartphone sub-asset (e.g. contact list, usage history) is perceived similarly across users. Thus, per user privacy risk assessment is not achievable. This paper refines our previous work on smartphone risk assessment by proposing an approach for assessing the privacy risk of Android users. Its cornerstone is impact valuation from users, as well as their usage profiles, which enables assessment of per user risk. Threat likelihood is assessed based on the presence of specific permission combinations, which we consider vulnerabilities that enable privacy threat scenarios. These permission combinations correspond to users' app profiles, i.e. to the app categories of Google Play that each user regularly visits. Finally, the proposed method is demonstrated through a case study.

Keywords: Android, Personal Data, Privacy, Risk, Permission, Attacks

1 Introduction

The ubiquity of Android applications ('apps'), Android's openness regarding the sources of available apps, and the variety of available data [30] are some of the reasons that privacy risk is increased in Android [21]. In Android the effort for the implementation and deployment of privacy violating apps is low [22]. Android apps in Google Play were found to leak sensitive information (e.g. device identifiers) [6, 7]. A recent study on Android malware confirms that malware actively harvest data from infected phones (e.g. SMS messages, phone numbers, user accounts) [33]. In response to the increasing malware submissions in its marketplace, Google introduced in February 2012 Bouncer, which performs malware analysis in the marketplace's apps. Also, the latest versions of Android (i.e. v.4.2 and onwards) include a thin client that analyzes all apps on the device - including those installed from alternative sources. Nonetheless, a recent evaluation proves the ineffectiveness of this mechanism (15% detection ratio) [19].

Meanwhile, privacy violations can occur even when a user grants access to protected data (e.g. contact list, exact location, etc.) to a benign app, i.e. one not trying to violate user's privacy. This holds true, since the app may either be used as a confused deputy [4, 11, 16], i.e. accidentally allowing other malicious apps to use its functionality to access the resources, or be bundled with a malicious advertisement library [17, 26, 29], which misuses the shared permissions to violate user privacy. Also, benign Android apps tend to request more permissions than needed for their intended functionality [8].

Currently the assessment of privacy risks in Android apps lacks user input, assuming that the value of each asset is perceived similarly across users. Therefore, a per user privacy risk assessment is not achievable. In this paper, we refine our previous work on smartphone risk assessment [30] and propose a process for accessing privacy risk of Android users. Our contribution includes:

- We provide taxonomy of user data found on a smartphone and their respective Android permissions and discuss ways to assess the impact of their disclosure for a particular user.
- We identify privacy threats applicable to user data. For each threat, we map the permissions required for the threat to occur; we consider that each time a user grants permission to an app, he introduces vulnerabilities and increases threat likelihood [30].
- We crawled apps from Google Play and used this sample to list descriptive statistics for permission combinations that may violate user privacy.
- We assess privacy risk on a per user basis by combining the likelihood of permissions with user input regarding the impact of disclosure. We present the applicability of the method with a case study.

The rest of the paper is organized as follows. Section 2 describes privacy impact assessment. Section 3 introduces the proposed privacy risk assessment method. A demonstration of this method is provided in Section 4 with the use of a case study. The paper concludes in section 5 with discussion and future work.

2 Smartphone Privacy Impact Assessment (PIA)

Privacy essentially refers to the protection of personal data or 'Personally Identifiable Information' (PII), but it can have wider interpretations in different, non-IT contexts, i.e. bodily privacy, privacy of personal behavior or privacy of personal communications [18]. Privacy Impact Assessment (PIA) refers to "a systematic process for identifying and addressing privacy issues in an information system that considers the future consequences for privacy of a current or proposed action" [32]. This process is a risk assessment process, focused on privacy, and it is mainly associated with the collection, use or disclosure of personal information. It is a tool for decision support and it is often a regulatory requirement for public information systems, as it serves as a means to address citizens' concerns over privacy. PIA is currently used in the UK, Canada, USA and Australia for projects of the public sector, e.g. a new technology, pilot, rule,

or other collection. In the smartphone context, PIA could potentially be performed by app developers or by the marketplace, to increase user awareness and to present more clear policies regarding the use of the customers' personal data.

While differences occur between approaches, the underlying principles remain the same (see "fair information principles" [1]). Specifically, herein these principles are defined with respect to the Android marketplace, namely: (i) applications must clearly state why personal information is collected, at or before the time of collection, (ii) user data collected by apps should remain accurate, relevant, up-to-date, and not excessive in relation to the collection purpose, (iii) apps should not retain user data for periods longer than needed for the purpose of the collection, (iv) user data must be collected by apps with the knowledge and consent of the individual, (v) user data must not be communicated to third parties, except under specified conditions that might include consent, (vi) applications must ensure user data are kept under secure conditions, and (vii) user data must be accessible to the individual for amendment or challenge.

These principles indicate that users should be able to review an app's privacy policy prior to downloading it and the policy must clarify what types of data are collected and for which purpose. Nonetheless, in Google Play the inclusion of such a policy is not mandatory for app developers [14]. While Android requires developers to ask users to grant permissions to apps, user studies have proven that Android users ignore permissions, or to not understand them at all [10, 23, 24]. If we consider that permissions are not refined, there are hidden privacy risks for specific permissions, which we identify in Section 3. The main thing missing is that developers do not state why they need the data (for some permissions, the collection purpose is not obvious), how they will use them, how they will be stored or protected and whether they will be transferred to a third party.

2.1 Personal data types

By analyzing the latest Android manifest[1], we identified nine (9) data types that may be the target of a privacy violation, namely: (a) Communication data, such as SMS, MMS, Voice, etc., (b) Sensor data, from the Camera or Microphone,[2] (c) Location data, i.e. GPS data (fine location) and location inferred from networks that the user connects (coarse location) or his social media feeds, (d) External storage data that include app data, documents, emails, etc. (e) Contact data, i.e. the smartphone's contact list or contacts derived by the social media that the user participates in, (f) History/Usage data, which indicate the user's preferences and can be collected by bookmarks, subscriptions to feeds, the call or task logs, social media feeds, the dictionary, (g) Calendar data, which could also be an indicator of contacts and/or the user's location, (h) Identity data, which refers to all the unique identifiers that can be used to identify a user, e.g. his device ID, e-mail addresses, his profile ID, (i) Credentials, the user's authentication tokens, passwords and PINs

[1] At the time of our analysis it is Jelly bean (v. 4.2.2) [15]

[2] These could also be indicators of the user's location. [25]

Each identified data type has been mapped to Android permissions that enable their use or collection by an app (see details in Section 3). These permissions are listed in Table 1 in alphabetical order.

2.2 Access levels to smartphone assets

Regardless of whether we refer to a single user or a category of users, a PIA for smartphones would typical refer to (a). a device (with Android OS in our case) and (b). one or more apps. For each device, the personal information accessed or collected need to be defined, coupled with respective access permissions of apps or controls. For each app, we need to identify the collection and use conditions of PII, privacy risks deriving from either 'dangerous' [12] or unnecessary permission combinations or by the lack of appropriate countermeasures.

Table 1 summarizes the access level of permissions that protect either a data type (e.g. SMS, exact location), or a communication channel (e.g. Internet, SMS). Most of the permissions have a self-explanatory label, specifying the asset or channel that the permission protects (e.g. 'READ_SMS'). The rest of them are briefly described here: i. ACCESS_NETWORK_STATE and ACCESS_WIFI_STATE grant access to information of the carrier and Wi-Fi network respectively, ii. AUTHENTICATE_ACCOUNTS allows an app to perform account authentication, iii. GET_TASKS provides information about apps that are or have been executed in the device, iv. PROCESS_OUTGOING_CALLS allows an app to monitor and manage outgoing calls, v. READ_PHONE_STATE grants access to identifiers such as the IMSI, IMEI, vi. USE_CREDENTIALS allows an app to request an authentication token for an account that has been stored in the device, and vii. WRITE_EXTERNAL_STORAGE provides both read and write access to the external storage.

This table includes the access level, collected by parsing the platforms source code [15]. The analysis focuses on the currently most popular Android versions, namely Jelly Bean (API 17, API 16), Ice Cream Sandwich (API 15) and Gingerbread (API 10), with a distribution among devices 5.6%, 32.3%, 23.3%, and 34.1% respectively [13]. The majority of permissions have a dangerous access level [12], i.e. the user must decide before app installation whether to accept or not the permission request. Five permissions are protected with a normal access level, i.e. they are automatically granted to any app that requests them, namely: ACCESS_NETWORK_STATE, ACCESS_WIFI_STATE, READ_EXTERNAL_STORAGE, GET_ACCOUNTS, and SUBSCRIBED_FEEDS_READ. Finally, in the latest version of Android, READ_LOGS and MOUNT_UNMOUNT_FILESYSTEMS are not available to third-party apps (signatureorsystem access level).

Wang et al. [31] use permission access level as a means to evaluate impact ('harm caused by permission'). They follow the assumption that dangerous permissions are more harmful than normal permissions and appoint impact value to permissions in an ad hoc way. However, among these permissions there are several with 'dangerous' access level that users consider more important than other in terms of privacy [9, 20].

Table 1. Access Level of Permissions

Permission	Access Levels per API			
	17	16	15	10
ACCESS_COARSE_LOCATION	d	d	d	d
ACCESS_FINE_LOCATION	d	d	d	d
ACCESS_NETWORK_STATE	n	n	n	n
ACCESS_WIFI_STATE	n	n	n	n
AUTHENTICATE_ACCOUNTS	d	d	d	d
BLUETOOTH	d	d	d	d
BLUETOOTH_ADMIN	d	d	d	d
CALL_PHONE	d	d	d	d
CAMERA	d	d	d	d
GET_ACCOUNTS	n	n	n	n
GET_TASKS	d	d	d	d
INTERNET	d	d	d	d
MOUNT_UNMOUNT_FILESYSTEMS	s	d	d	d
PROCESS_OUTGOING_CALLS	d	d	d	d
READ_CALENDAR	d	d	d	d
READ_CALL_LOG	d	d	NA	NA
READ_CONTACTS	d	d	d	d
READ_EXTERNAL_STORAGE	n	n	NA	NA
READ_HISTORY_BOOKMARKS	d	d	d	d
READ_LOGS	s/t	s/t	d	d
READ_PHONE_STATE	d	d	d	d
READ_PROFILE	d	d	d	NA
READ_SMS	d	d	d	d
READ_SOCIAL_STREAM	d	d	d	NA
READ_USER_DICTIONARY	d	d	d	d
RECEIVE_MMS	d	d	d	d
RECEIVE_SMS	d	d	d	d
RECEIVE_WAP_PUSH	d	d	d	d
RECORD_AUDIO	d	d	d	d
SEND_SMS	d	d	d	d
SUBSCRIBED_FEEDS_READ	n	n	n	n
USE_CREDENTIALS	d	d	d	d
USE_SIP	d	d	d	d
WRITE_EXTERNAL_STORAGE	d	d	d	d

d:dangerous, s:signatureOrsystem, t:development, -:N/A, n: normal

Each user has a varied risk perception and, thus, perceives the impact of disclosure to a particular type of PII differently. Therefore, he assigns a different impact value to privacy threats compared to other users. He may also consider specific types of personal data as more private or confidential. Such perceptions of impact are affected by

the user's personality, awareness and technological expertise, the type of smartphone use (personal vs. business), as well as the context of the user. Smartphone PIA can provide generic assessment results for the average user or a group of users, or personalized results based on an individual's perceived privacy impact.

3 Smartphone Privacy Risk Assessment

This paper describes an approach for privacy risk assessment which focuses on average smartphone users, i.e. ones that are not savvy in terms of ICT and security. These users are assumed to (a) install apps into their devices, on a regular basis (e.g. daily), (b) install them only from the official app marketplace (i.e. Google Play) and (c) protect their smartphone only with the default security controls of the platform. This means that we do not take into account smartphones with modified operating system, i.e. rooted smartphones, or smartphones protected with third-party security mechanisms (software), which enhance their security. With the former, it is assumed that every app is executed in a sandboxed, permission-based environment [12]. The latter is in accordance with the expectations for an average user - in terms of security and ICT - as well as with the findings of a recent user study [24]. It refers to the poor adoption of third-party software that provides additional layers of security, such as antivirus software, Mobile Device Management (MDM) solutions that isolate corporate from personal data, etc. Finally, it is assumed that the smartphone has Internet connectivity from the mobile carrier.

3.1 Threats and Permissions

Similarly to [30], we consider the permissions of third-party apps, as vulnerability. This happens since in Android access to protected resources happens only if the requesting app is granted the corresponding permission for the resource [12]. Even if the requesting app is benign, its privilege to access private data may be misused by another malicious app performing a deputy attack [4, 11, 16], or by a malicious advertisement library that is bundled with it [17, 26, 29]. Moreover, in Android permission granting happens via the user, who is expected to scrutinize permission requests and decide upon installation whether to authorize them or not [12]. Orthogonally to this expectation, previous work [10, 23, 24] has proven that users tend to ignore permissions, or to not understand them at all. Also, even users that tend to inspect app permission are hampered by deficiencies of Android's permission system [12]. In specific, Android users must either accept permissions in an all-or-nothing approach, i.e. they cannot authorize only a subset of the requested permissions, or cancel the installation of the requesting app. Also, Android permissions are not fine-grained - e.g. the permission SEND SMS allows an app to send SMS messages both to normal and premium numbers - making authorization decisions more difficult. In this context, granting permission combination is herein considered vulnerability and will be used to assess the likelihood of occurrence of various privacy threats.

Among smartphone threats (c.f. [30], [28], [21]) only five pose a potential privacy impact, namely: (a) Tracking/Surveillance (T1), which refers to monitoring users' context3 (e.g. via using the device's sensors [25]), (b) Interception/Eavesdropping (T2), which refers to unlawful interception of communications and is applicable to all communication data types (including external communication data, e.g. call log), (c) Profiling (T3), which refers to user activity monitoring, but for advertising purposes, (d) Phishing (T4), which refers to tricking the user to disclose credentials and (e) Personal Information Disclosure (T5), which refers to the disclosure of all other types of personal information, which do not fall in the other four threat types (e.g. documents, multimedia files, etc.).

Our analysis omits threats that could be used as the attack vector for a privacy breach. For instance, a successful spoofing attack can lead to other privacy attacks, such as phishing, eavesdropping, SSL downgrading etc. [28]. Moreover, threats that exploit vulnerabilities in the OS, which are not related to a specific permission, or refer to vulnerabilities introduced by 'jailbroken' or 'rooted' smartphones (e.g. weak ssh credentials as in IKee [5]) are also out of the scope of the analysis. Also, spyware is not included as a threat, as it would cover several of the above threats depending on the collection purpose. In this paper, every app that has a hidden mechanism to collect personal data without the user's consent is considered to be 'spyware'. For these threats, risk assessment is more complex and needs to take into account several other factors, such as non-privacy threats and vulnerabilities, installed controls, user habits in terms of OS updates and patches, and user security awareness.

In Table 2 we map the personal data of an Android user to permissions that can enable a privacy breach and identify applicable threats to these data types. In order for an app to disclose this information to a third party, a data channel is required. There are three (3) available channels, namely: (a) Internet connection, (b) short messages and (c) Bluetooth. Each of them is protected with a dangerous permission, namely: (a) INTERNET, (b) SEND SMS, and (c) BLUETOOTH.

If the user's app list is available during the privacy assessment, we can then assess the threat likelihood of privacy threats, by identifying 'suspicious' permission combinations that may enable them in the smartphone of a user. Finding an app with a privacy-related permission does not mean, however, that the app is malicious; it means that the app could be potentially used for other purposes than the stated ones. Installing an app that retrieves the app list of a user may be considered too intrusive and may add additional vulnerabilities to the risk profile (of the smartphone/user) we are currently studying. On the other hand, asking the user to 'manually' provide the app list may be cumbersome, especially when the user has limited time and/or many installed apps, whereas asking her device for a manual audit from an analyst may also considered too intrusive for the user. In this context, a less intrusive approach is explained in the sequel.

---

[3] User's location can be retrieved either directly (fine location), or indirectly by using the sensors (e.g. a camera snapshot identifies the user's location), by retrieving in- formation about the networks the user is connected to (coarse location), or calendar entries.

Table 2. Mapping of Data Assets, Permissions and Threats

Data Category	Data Asset Type	Permission	Privacy Threats				
			T1	T2	T3	T4	T5
Communication	SMS	RECEIVE_SMS		✓			
		READ_SMS		✓			
	MMS	RECEIVE_MMS		✓			
	Voice	PROCESS_OUTGOING_CALLS		✓			
	Wap	RECEIVE_WAP_PUSH		✓			
Sensor/Location	Video	CAMERA	✓	✓			
	Audio	RECORD_AUDIO	✓	✓			
	Location	ACCESS_COARSE_LOCATION	✓		✓		
		ACCESS_FINE_LOCATION	✓		✓		
		BLUETOOTH_ADMIN	✓		✓		
		ACCESS_NETWORK_STATE	✓		✓		
		ACCESS_WIFI_STATE	✓		✓		
		READ_SOCIAL_STREAM	✓		✓		
External Storage		WRITE_EXTERNAL_STORAGE		✓	✓		✓
		READ_EXTERNAL_STORAGE		✓	✓		✓
Contacts		READ_CONTACTS		✓			✓
		READ_SOCIAL_STREAM		✓			✓
History/ Usage		READ_CALL_LOG		✓	✓		
		READ_HISTORY_BOOKMARKS			✓		
		GET_TASKS			✓		
		READ_LOGS			✓		
		READ_USER_DICTIONARY			✓		✓
		READ_SOCIAL_STREAM			✓		
		SUBSCRIBED_FEEDS_READ			✓		
Calendar		READ_CALENDAR	✓				✓
Identity		READ_PROFILE			✓	✓	✓
		GET_ACCOUNTS			✓	✓	✓
		READ_PHONE_STATE			✓	✓	✓
Credentials		READ_SMS- RECEIVE_SMS				✓	
		AUTHENTICATE_ACCOUNTS				✓	
		USE_CREDENTIALS				✓	

The user describes the type of smartphone use, by declaring her popular app categories. This way, she indicates what apps she installs and uses often (e.g. on a daily basis). Based on this input, as well as on our statistical analysis on the apps in the Google Play (see Section 4), each category yields varied privacy-sensitive combination of permissions thus, a varied level of privacy vulnerabilities. Since each threat is enabled by a set of permission combinations (c.f. Table 2), per user threat likelihood is the *avg* value of the above frequencies in each category specified by the user. Finally, the per user vulnerability level is assessed based on the *avg* frequency of a particular permission combination in the selected categories, based on the following empiri-

cal semi-quantitative scale[4] : (a) *1–Negligible*: < %10%, (b) *2–Limited*: ≥ 10% and < %40%, (c) *3–Significant*: ≥ 40% and < %70%, and (d) *4–Maximum*: ≥ 70%.

3.2 Threats and Impacts

To assess the impact of a privacy breach to a smartphone, we adjusted the '*Methodology for Privacy Risk Management*' by CNIL [2] to the smartphone context. Initially, permissions in Table 2 are assigned with a level of identification. This parameter refers to the ability to identify an individual solely by accessing the data that the permission protects. This is important because in order for a privacy threat to have an effect, it needs to correspond to a particular individual whose privacy has been breached. The ability to identify individuals must not be confused with profiling (T3), which refers to monitoring users' activity (e.g. among different apps, web sites) for advertising purposes. The Identification Level of permission is assessed by a 4-item scale:

- *1–Negligible*: Identifying a user using this permission appears to be virtually impossible.
- *2–Limited*: Identifying a user using this permission appears to be difficult but is possible in certain cases.
- *3–Significant*: Identifying a user using this permission appears to be relatively easy.
- *4–Maximum*: Identifying a user using this permission appears to be extremely easy.

The proposed levels of identification per Android permission can be found in Appendix A. This static mapping of permissions to levels of identification should be adjusted when changes in permissions occur.

The user is then asked a few questions, in order to assess his individual impact of a privacy breach. Each question describes the effect of disclosure or misuse on various personal data. Therefore, based on Table 2, each question corresponds to specific applicable threats. An example of the short questionnaire, coupled with applicable threat scenarios can be found in Appendix. The questionnaire's answers are also predefined. Each one describes the effect of the privacy threat to the user. In Appendix we map predefined answers to their corresponding Severity Level, based on the following qualitative scale.

- *1–Negligible*: The user is either not affected or may encounter a few inconveniences, which he/she can overcome without any problem.
- *2–Limited*: The user may encounter significant inconveniences, which he/she will be able to overcome despite a few difficulties.
- *3–Significant*: The user may encounter significant consequences, which he/she should be able to overcome albeit with serious difficulties.

[4] The 4 levels of vulnerability [2] are dynamically created by periodically clustering (k-means algorithm) the likelihood values for the top combinations of channel and asset permissions (see Table 3). A 4-item scale was selected to match the impact assessments values [2].

- *4–Maximum*: The user may encounter significant, or even irreversible, consequences, which they may not overcome.

The overall Impact Level (ImL) of a specific threat scenario is then assessed as the sum of Identification Level and Severity Level, based on the following scale: (a) *1–Negligible*: ImL < % 5, (b) *2–Limited*: ImL = 5, (c) *3–Significant*: ImL = 6, and (d) *4–Maximum*: ImL > %6.

3.3 Risk of Privacy Threat

Since each privacy threat requires the presence of specific combinations of permissions on the user's smartphone, the Level of Risk per privacy threat is assessed as a sum of Impact Level and Vulnerability Level, in the following scale: (a) *1–Negligible*: 1 to 3, (b) *2–Limited*: 4 to 5, (c) *3–Significant*: 6 to 7, and (d) *4–Maximum*: 8.

This value corresponds to a user's specific risk profile (personalized PIA), as (a) impact was assessed based on the user's input and reflects his own assumptions regarding the potential effects of a privacy breach and (b) threat likelihood (i.e. vulnerability level) is assessed according to the type of smartphone use the particular user describes, by identifying her common app categories.

4 Case study: Privacy assessment of Android users

This section describes a proof-of-concept case study of our proposed PIA method.

4.1 Statistics for combinations of permissions

To compute the frequency of permission combinations, we crawled apps residing in Google Play from May to June 2013. We collected all the available apps that are indexed in each app category of Google Play, namely 27673 apps. We then analyzed all possible combinations that include pairs of permissions that protect an asset described in Table 2 and a transmission channel. Our sample contains89 such pairs in the permission combinations. Among them, the top 20 most frequent pairs in permission combinations are presented in Table 3.

4.2 Case study

In this proof-of-concept case study we examine two use cases. User A is a teenager who uses his smartphone mainly for leisure, which includes socializing with his classmates and friends, making phone calls and texting, playing games (only sports games) and listening to music. The user specified that his apps fall mainly on the following Google Play categories: (a) *Communication* (e.g. web browser, custom email (e.g. yahoo), chatting apps, etc.), (b) *Sport_Games*, (c) *Social* (e.g. Facebook, Twitter) and (d) *Music_And_Audio*.

Table 3. Top 20 combinations for privacy violating permissions

Channel	*Access to Data*	*Frequency*
INTERNET	ACCESS_NETWORK_STATE	81,13%
INTERNET	WRITE_EXTERNAL_STORAGE	55,14%
INTERNET	READ_EXTERNAL_STORAGE	55,08%
INTERNET	READ_PHONE_STATE	48,92%
INTERNET	ACCESS_WIFI_STATE	31,06%
INTERNET	ACCESS_COARSE_LOCATION	28,67%
INTERNET	ACCESS_FINE_LOCATION	27,75%
INTERNET	GET_ACCOUNTS	18,85%
INTERNET	CAMERA	8,19%
INTERNET	GET_TASKS	7,17%
INTERNET	READ_CONTACTS	6,94%
INTERNET	READ_HISTORY_BOOKMARKS	6,75%
INTERNET	READ_CALL_LOG	5,67%
INTERNET	RECORD_AUDIO	4,63%
INTERNET	READ_LOGS	3,36%
INTERNET	USE_CREDENTIALS	1,92%
SEND_SMS	READ_PHONE_STATE	1,75%
INTERNET	RECEIVE_SMS	1,72%
SEND_SMS	ACCESS_NETWORK_STATE	1,71%
SEND_SMS	WRITE_EXTERNAL_STORAGE	1,68%

User B is a businessman, who uses his smartphone for both business and leisure. His daily use of Android apps includes (a) reading News & Magazines (i.e. apps displaying the content of news web sites (e.g. as CNN)), (b) Socializing (Facebook, Twitter) with colleagues or clients, (c) consulting the Weather and (d) reading maps and navigating by GPS (Travel & Local category), as he regularly travels to visit his clients[5].

Table 4. Questionaire answers

Question	An(A)	An(B)	SL(A)	SL(B)
Q1 (Fine)	N3	L2	1	2
Q2	N/A	N1	N/A	1
Q3	S5	N3	3	1
Q4	N1	N3	1	1
Q5	L4	S1	2	3

An:Answers, SL:Severity Level

[5] Any business data, such as corporate files (e.g. pdf) stored on the external storage, were not included in the case, as they are not considered PII, do not affect privacy, and are under different regulatory requirements. We only examine the effect to a person's reputation, which falls into the scope of privacy and may affect his working conditions.

Table 4 summarizes the responses of the two users in questions Q1-5 (see Appendix for the corresponding text of the questions and answers), as well as their mapping to the severity level of our proposed method for privacy assessment. Each question examines a different threat scenario and, therefore, it requires a different permission combination present, in order to be realized. For space and readability reasons, the subsequent analysis will include only the following permission combinations C_i, which correspond to the threats that are covered with the above questions, namely:

- C_1 `INTERNET, ACCESS_FINE_LOCATION`
- C_2 `INTERNET, READ_PHONE_STATE, READ_CALENDAR`
- C_3 `INTERNET, GET_ACCOUNTS, READ_HISTORY_BOOKMARKS`
- C_{4A} `INTERNET, READ_PHONE_STATE, READ_CONTACTS`
- C_{4B} `INTERNET, READ_PHONE_STATE, READ_CALL_LOG`
- C_5 `INTERNET, READ_EXTERNAL_STORAGE, READ_PHONE_STATE`

Per user threat likelihood is the *avg* value of the frequency of the above combinations in each app category that matches the user's profile (c.f. Section 3). Assuming that the businessman possesses a smartphone with a more modern Android version than the teenager, then the combination for C_4 differs. Access to call history is protected by the permissions `READ_CALL_LOG` and `READ_CONTACTS`, for the businessman and teenager respectively [15]. Hence, the threat that is realized with the occurrence of C_4 (i.e. T3) gets different scores in the vulnerability level from the common category (i.e. Social), as a result of the two different combinations C_{4A} , C_{4B} (c.f. Appendix, Table 7). Table 5 summarizes the vulnerability levels for per user threat, as well as the identification level, impact level and risk level of this case study.

Table 5. Case study of privacy risk assessment

PC	Th	IdL(A)	ImL(A)	VL(A)	RL(A)	IdL(B)	ImL(B)	VL(B)	RL(B)
C1	T1	1	1	2	1	1	1	3	2
C2	T1	N/A	N/A	N/A	N/A	4	2	1	1
C3	T5	3	3	1	2	3	1	1	1
C4	T3	4	2	2	2	4	2	1	1
C5	T5	4	3	2	2	4	4	2	3

PC: Permission Combination, Th: Threats, IdL: Identification level, ImL: Impact Level, VL: Vulnerability Level, RL: Risk Level

The case study includes permission combinations with different identification levels (c.f. Table 5). For instance, C_1 is assessed with negligible identification level (the device location provides only a weak correlation with the user's identity), whereas C_5 is assessed with maximum identification level, due to the unique identifiers (e.g. IMSI, IMEI) accessible via the `READ_PHONE_STATE` permission.

For threats involving access to the device location, the businessman's vulnerability level is greater due to his preference to navigation apps. The teenager has a greater vulnerability level for the threats that involve access to calling history, while the rest

threats have a similar vulnerability level for the combination of permissions that are included in the case study's scope (c.f. Table 5).

The businessman responded that he is more upset about unauthorized access to his data in the external storage and the teenager about disclosure of his browsing history to his friends. For the former, this is the threat with the highest risk level (RL=3, c.f. Table 5). For the latter, the three threats that are realized with the combinations $C_3 - C_5$ were assessed as highest with our method (RL=2). Finally, it is assumed that the teenager responded that he is not using his calendar. Thus, the threat that is realized with C_2 is not applicable to him and no risk score is assigned to it.

5 Related work

This work refines the risk assessment method proposed in [30]. Its focus is on a subset of smartphone threats that were presented in [30], namely privacy threats. The privacy threats that are applicable to user data are identified and mapped to the permission combinations that are required for the threat to occur. Threat likelihood is computed from the frequency of these permission combinations on Google Play. Our work relates to [27], which studies permission combinations as risk signals. The analysis of risk signals, however, is based on an outdated app sample (collected in 2011, before the introduction of Bouncer that changed the frequency of permission combinations by filtering out apps from Google Play). Also, the analysis focuses only to a subset of the available permission combinations that may violate user privacy. DroidRisk [31] is, to the best of our knowledge, the first attempt to quantitatively assess the risk levels of both Android permissions and apps. Its assessment is based on patterns of permission requests in malicious and benign apps. However, DroidRisk's analysis is limited only to statistics on individual permissions and not on their combinations.

Our method can benefit from a generic impact valuation such as [9], which includes a ranking of risks according to user upset. This generic impact valuation can be facilitated to create static, generic risk profiles. Finally, previous works (e.g. [3, 33]) often include statistics about the popularity of individual permission in Google Play. Our work provides up to date popularity of permission combinations that can violate user privacy when they are misused by apps.

6 Conclusions

This paper extends our previous work on smartphone risk assessment by describing a method for Privacy Impact Assessment (PIA) for Android apps. As opposed to previous works that delve into privacy violating apps, our approach is user-centered. The cornerstone of our assessment is impact valuation from the user, as well as her usage profile, which enables per user risk assessment. Threat likelihood is assessed based on the presence of specific permission combinations, which we consider to be vulnerabilities that enable privacy threat scenarios. For the demonstration of the method, a case study is presented with input from two hypothetical users and actual app data, i.e. permission combinations from apps in Google Play. Our proposed method is envi-

sioned as a complement to the existing protection from privacy violating apps. For instance, the privacy risk level of a user can be used to generate app analysis policies, which would filter out Android apps based on user privacy requirements, or be used to provide security awareness that is tailored to the user's risk profile.

The dynamic computation of permission combinations, which are used as input by our method, is limited only to apps that are available from Google Play. The frequency of permission combinations in other app marketplaces may be different, implying different threat likelihood. Our method assumes the participation of the user and is prone from the subjectivity of his/her impact perceptions. Also, our analysis could be extended to include the existence of safeguards that may decrease threat likelihood – even though past literature has proven that currently smartphone users poorly adopt them. We leave this task for future work, with an eye towards building upon our publications in the area of security-critical applications/infrastructures [34-37].

Acknowledgements. This research has been co-funded by the European Union (ESF) and Greek national funds, through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (Program HERACLEITUS II: Investing in knowledge society through the ESF).

References

1. Office of the privacy commissioner of canada privacy impact assessments (2007), http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18308
2. Methodology for privacy risk management (2012), http://www.cnil.fr/fileadmin/documents/en/CNIL-ManagingPrivacyRisks-Methodology.pdf
3. Barrera, D., Kayacik, H., van Oorschot, P., Somayaji, A.: A methodology for empirical analysis of permission-based security models and its application to Android. In: Proc. of the 17th ACM Conference on Computer and Communications Security. pp. 73–84. ACM (2010)
4. Chin, E., Felt, A.P., Greenwood, K., Wagner, D.: Analyzing inter-application communication in Android. In: Proc. of the 9th International Conference on Mobile Systems, Applications and Services. pp. 239–252. ACM, USA (2011)
5. Cluley, G.: First iPhone worm discovered - ikee changes wallpaper to Rick Astley photo (November 2009), http://nakedsecurity.sophos.com/2009/11/08/iphone-worm-discovered-wallpaper-rick-astley-photo/
6. Enck, W., Gilbert, P., Chun, B., Cox, L., Jung, J., McDaniel, P., Sheth, A.: Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones. In: Proc. of the 9th USENIX Conference on Operating Systems Design and Implementation. pp. 1–6. USENIX Association, USA (2010)
7. Enck, W., Octeau, D., McDaniel, P., Chaudhuri, S.: A study of android application security. In: Proc. of the 20th USENIX Conference on Security. pp. 21–21. USENIX Association, USA (2011)
8. Felt, A., Chin, E., Hanna, S., Song, D., Wagner, D.: Android permissions demystified. In: Proc. of the 18th ACM Conference on Computer and Communications Security. pp. 627–638. ACM (2011)

9. Felt, A., Egelman, S., Wagner, D.: I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns. In: Proc. of the 2nd ACM Workshop on Security and Privacy in Smartphones and Mobile Devices. pp. 33–44. ACM, USA (2012)
10. Felt, A., Ha, E., Egelman, S., Haney, A., Chin, E., Wagner, D.: Android permissions: User attention, comprehension, and behavior. In: Proc. of the 8th Symposium on Usable Privacy and Security. ACM (2012)
11. Felt, A., Hanna, S., Chin, E., Wang, H.J., Moshchuk, E.: Permission redelegation: Attacks and defenses. In: In 20th Usenix Security Symposium (2011)
12. Google: Android security overview (July 2013), http://source.android.com/devices/tech/security/index.html
13. Google: Dashboards - Android developers (July 2013), http://developer.android.com/about/dashboards/index.html
14. Google: Privacy policies for android apps developed by third parties (2013), https:// support.google.com/googleplay/answer/2666094?hl=en
15. Google: Refs - platform/frameworks/base - git at google (2013), https://android.googlesource.com/platform/frameworks/base/+refs
16. Grace, M., Zhou, Y., Wang, Z., Jiang, X.: Systematic detection of capability leaks in stock Android smartphones. In: Proc. of the 19th Network and Distributed System Security Symposium (2012)
17. Grace, M., Zhou, W., Jiang, X., Sadeghi, A.: Unsafe exposure analysis of mobile in-app advertisements. In: Proc. of the 5th ACM Conference on Security and Privacy in Wireless and Mobile Networks. pp. 101–112. ACM (2012)
18. ICO: Privacy impact assessment handbook, v2.0, Information Commissioner's Office, United Kingdom
19. Jiang, X.: An evaluation of the application ("app") verification service in android 4.2 (December 2012), http://www.cs.ncsu.edu/faculty/jiang/appverify/
20. Lin, J., Sadeh, N., Amini, S., Lindqvist, J., Hong, J., Zhang, J.: Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing. In: Proc. of the 2012 ACM Conference on Ubiquitous Computing. pp. 501–510. ACM (2012)
21. Marinos, L., Sfakianakis, A.: Enisa threat landscape. Technical report, ENISA (2012)
22. Mylonas, A., Dritsas, S., Tsoumas, B., Gritzalis, D.: Smartphone security evaluation - the malware attack case. In: Proc. of International Conference of Security and Cryptography. pp. 25–36 (2011)
23. Mylonas, A., Gritzalis, D., Tsoumas, B., Apostolopoulos, T.: A qualitative metrics vector for the awareness of smartphone security users. In: 10th International Conference on Trust, Privacy & Security in Digital Business. pp. 173–184 (2013)
24. Mylonas, A., Kastania, A., Gritzalis, D.: Delegate the smartphone user? Security awareness in smartphone platforms. Computers & Security 34(0), 47–66 (2013)
25. Mylonas, A., Meletiadis, V., Mitrou, L., Gritzalis, D.: Smartphone sensor data as digital evidence. Computers & Security 38(0), 51–75 (2013)
26. Pearce, P., Felt, A.P., Nunez, G., Wagner, D.: Android: Privilege separation for applications and advertisers in android. In: Proc. of the 7th ACM Symposium on Information, Computer and Communications Security. pp. 71–72. ACM (2012)
27. Sarma, B.P., Li, N., Gates, C., Potharaju, R., Nita-Rotaru, C., Molloy, I.: Android permissions: a perspective combining risks and benefits. In: Proc. of the 17th ACM Symposium on Access Control Models and Technologies. pp. 13–22. ACM (2012)
28. Souppaya, M., Scarfone, K.: Guidelines for Managing the Security of Mobile Devices in the Enterprise. NIST (June 2013), NIST Special Publication 800-124, rev. 1

29. Stevens, R., Gibler, C., Crussell, J., Erickson, J., Chen, H.: Investigating user privacy in Android ad libraries. In: Workshop on Mobile Security Technologies (2012)
30. Theoharidou, M., Mylonas, A., Gritzalis, D.: A risk assessment method for smartphones. In: Proc. of the 27th International Information Security and Privacy Conference, pp. 443–456. Springer (2012)
31. Wang, Y., Zheng, J., Sun, C., Mukkamala, S.: Quantitative security risk assessment of android permissions and applications. In: Data and Applications Security and Privacy XXV II, pp. 226–241. Springer (2013)
32. Warren, A., Bayley, R., Bennett, C., Charlesworth, A., Clarke, R., Oppenheim, C.: Privacy Impact Assessments: International experience as a basis for UK Guidance. Computer Law & Security Review 24(3), 233–242 (2008)
33. Zhou, Y., Jiang, X.: Dissecting android malware: Characterization and evolution. In: Proc. of the 2012 IEEE Symposium on Security and Privacy. pp. 95–109. IEEE Computer Society (2012)
34. Gritzalis, D.: Embedding privacy in IT applications development. In: Information Management and Computer Security, 12(1), 8-26 (2004)
35. Gritzalis, D.: Enhancing security and improving interoperability in healthcare information systems. In: Informatics for Health and Social Care, 23(4), 309-324 (1998)
36. Theoharidou, M., Kotzanikolaou, P., Gritzalis, D.: Risk assessment methodology for interdependent critical infrastructures. In: International Journal of Risk Assessment and Management, 15(2-3), 128-148 (2011)
37. Theoharidou, M., Kotzanikolaou, P., Gritzalis, D.: A multi-layer criticality assessment methodology based on interdependencies. In: Computers & Security, 29(6), 643-658 (2010)

Appendix

Table 6. Level of Identification per Permission

Identification Level	**Permission**
1	ACCESS_COARSE_LOCATION,ACCESS_FINE_LOCATION, ACCESS_NETWORK_STATE,ACCESS_WIFI_STATE,BLUETOOTH_ADMIN, GET_TASKS,READ_CALENDAR,READ_HISTORY_BOOKMARKS, READ_LOGS,READ_USER_DICTIONARY,RECEIVE_WAP_PUSH, SUBSCRIBED_FEEDS_READ
2	CAMERA,PROCESS_OUTGOING_CALLS,READ_CALL_LOG,READ_CONTACTS,READ_EXTERNAL_STORAGE,READ_SMS,READ_SOCIAL_STREAM,RECEIVE_MMS,RECEIVE_SMS,RECORD_AUDIO,WRITE_EXTERNAL_STORAGE
3	AUTHENTICATE_ACCOUNTS,GET_ACCOUNTS,USE_CREDENTIALS
4	READPHONE_STATE,READ_PROFILE

Level of Identification. Table 6 presents our proposed static mapping of the ability of a permission to identify an individual. Permissions for assets that provide a weak correlation with the identity of the Android user (e.g. under specific assumptions the

location of the device may be used to infer the user's identify) are assigned to score 1. For the assets where these assumptions are more likely to occur in certain cases (for instance a camera or audio snapshot identifies the user, the user refers to his identity in a text message), their permissions are mapped to level 2. Finally, the permissions that provide access to data that relatively easy identify users (e.g. emails, Facebook accounts, etc.) are assigned to level 3 and the ones that provide strong correlations with the user identity (e.g. users name or phone number) are assigned to level 4.

Severity Assessment. This is an indicative questionnaire used to address the severity of impact of applicable threat scenarios.

- **Q1**. What will happen if someone tracks your *GPS location*? (**T1,T3**)
- **Q2**. What will happen if someone reads your *calendar*? (**T1**,**T5**)
- **Q3**. What will happen if your *bookmarks* and *browsing history* is sent to your friends? (**T5**)
- **Q4**. What will happen if your *calling history* is sent to advertisers? (**T3**)
- **Q5**. What will happen if someone reads without your permission your *documents or media from the external storage*? (**T5**)

The predefined answers that the user can select are given in the following sets, namely: (a) ***Negligible***: {Nothing (*N1*); This would be annoying (*N2*); I would be irritated (*N3*); I would have to reenter or modify it/them (*N4*)}, (b) ***Limited***: {This would cost me money (*L1*); I would be a bit afraid or confused (*L2*); I would be stressed (*L3*); I would be embarrassed (*L4*)}, (c) ***Significant***: {I may lose my job (*S1*); This may affect my health (*S2*); This would cost me lots of money (*S3*); I may face legal problems (*S4*); I would be humiliated (*S5*)}, (d) ***Maximum***: {I would not be able to work again (*M1*); I would get ruined financially (*M2*); My health would be damaged (*M3*); I could lose my life (*M4*); Nobody would speak to me again (*M5*)}.

Case study supporting data. The table below presents the frequency of the 6 combinations of permission that are studied in section 4. Due to space limitations the following notation is used: SG:`SPORTS_GAMES`, SO:`SOCIAL`,CO:`COMMUNICATION`, MA:`MUSIC_AND_AUDIO`, WE: `WEATHER`, TL: `TRAVEL_AND_LOCAL`, `NEWS_AND_MAGAZINES`.

Table 7. Percentages for permission combinations

PC	SG	SO	CO	MA	WE	TL	NM
c1	21,5	36,3	22	6,5	51,7	72	25,2
c2	0,3	1,6	1,6	1,1	0,6	1,1	0,5
c3	4,3	2,5	2,1	3,5	1,9	1,5	0,7
$c4_{A,B}$	5	(12,4; 9,2)*	27,7	12,2	2,1	5,2	1
c5	43,4	33,4	37,4	45,1	18,6	27,5	36,5

* values for (C_{4A}%; C_{4B}%)