

# Understanding Users' Requirements for Data Protection in Smartphones

Ildar Muslukhov <sup>#1</sup>, Yazan Boshmaf <sup>#2</sup>, Cynthia Kuo <sup>\*3</sup>, Jonathan Lester <sup>\*4</sup>, Konstantin Beznosov <sup>#5</sup>

*#Electrical and Computer Engineering Department, The University of British Columbia  
Vancouver, Canada*

<sup>1</sup>ildarm@ece.ubc.ca

<sup>2</sup>boshmaf@ece.ubc.ca

<sup>5</sup>beznosov@ece.ubc.ca

*\*Nokia Research Center  
Palo Alto, USA*

<sup>3</sup>cynthia.kuo@nokia.com

<sup>4</sup>jonathan.lester@nokia.com

**Abstract**—Securing smartphones' data is a new and growing concern, especially when this data represents valuable or sensitive information. Even though there are many data protection solutions for smartphones, there are no studies that investigate users' requirements for such solutions. In this paper, we approach smartphones' data protection problem in a user-centric way, and analyze the requirements of data protection systems from users' perspectives. We elicit the data types that users desire to protect, investigate current users' practices in protecting such data, and show how security requirements vary for different data types. We report the results of an exploratory user study, where we interviewed 22 participants. Overall, we found that users would like to secure their smartphone data, but find it inconvenient to do so in practice using solutions available today.

## I. INTRODUCTION

Recent market research shows that users are gradually shifting from Personal Computers (PC) to smartphones. For example, it is predicted that the number of sold smartphones will surpass the number of sold laptops by the end of 2011 [1], and that smartphones will become the main access device for the Web [2]. This shift is accompanied by an increased interest in mobile applications. In fact, users are spending more time per day using applications on the smartphones than browsing the Web [3]. Recent report shows that 96% of the U.S. population have a mobile phone, 35% of these mobile phone owners have smartphones, and 25% of the smartphone owners use their smartphone as a primary device for accessing the Internet [4].

Today's smartphones employ sophisticated Operating Systems (OSs) that enable a wide set of functionality such as personal data storage, PC-like Web browsing, music and video streaming, GPS navigation, video and voice recording, and third-party applications. Thanks to the large internal storage of modern smartphones, users can now store gigabytes of data, which could be valuable or sensitive such as personal photos, contact details, personal and work related documents, calls history, private messages, etc.

Unlike PCs, smartphones are highly portable and are more likely to be stolen, lost or damaged. Recent report shows that 52% of Miami city's population have experienced cell phone loss or theft [5]. In addition, the same report shows that 54%

of cell phone owners do not use any locks, such as passwords or PIN-codes, on their phones, which makes it easier for an adversary to get full access to users' data on the smartphone once it is stolen.

Other than physical threats, a new generation of malware, such as Geinimi Trojan and Ikee Worm, started to target popular smartphone platforms with an objective to steal or damage users' data [6], [7]. Moreover, mobile Spyware is on the rise too, and can be purchased from smartphones "App" markets. For example, if a user wants to spy on someone and is able to gain physical access to the phone, he can install the *Android.Tapsnake* application, an App that passes itself off as a game, but includes a complete set of tools for surveillance [7].

Clearly, the threat of mobile phones being lost, stolen, or infected makes the problem of data protection in smartphones a very important and challenging one. In this paper, we approach this problem from users' perspective by employing a user-centric design approach. In particular, we investigate users' requirements of data protection in smartphones, and seek to understand how users classify their data in smartphones based on its value and sensitivity. We also show what practices the users employ today in order to protect their data on smartphones. Overall, we show how security requirements vary for different data types. In particular, our investigation shows that users consider data security as a serious concern, but they tend not to take any actions to ensure it.

## II. RELATED WORK

In response to the increased number of malware attacks on smartphones, malware detection and mitigation on these devices has received much attention from the research community over the past few years [8], [9], [10], [11]. Most of the proposed solutions, however, are application or OS specific and target the confidentiality and the integrity aspects of smartphone data protection, but not its availability. Moreover, most of them do not take into account different users' requirements to security for different data types.

Ongtang et al. [8] proposed an application-specific Access Control List (ACL) design, where an application developer can

define an ACL for his application and the functionalities that application exposes. This approach, however, does not address the integrity and the availability of applications' data. It also requires the application developer to define an ACL for each of the exposed functionality of the application, which is not necessarily in compliance with the end users' requirements and preferences. Moreover, it does not protect the data from unauthorized accesses in case the mobile device is stolen.

Enck et al. [9], [10] proposed the Lightweight Application Certification (LAC) system and the TaintDroid system. The main goal of the LAC system is to prevent an application from running in case it has a dangerous set of permissions, as defined by the user. This, however, does not address the problem of physical threats, where an adversary can access the smartphone's data directly. Likewise, it does not address data leakage through Inter Process Communication (IPC). The TaintDroid system, on the other hand, tackles the problem of data leakage via IPC, where data flows are defined and used to prevent data theft. This system, however, fails to detect data leakage, which happens when the data representation is changed during the transfer from one process to another by, for instance, encryption. Moreover, the TaintDroid system does not address threats to smartphones' data that arise from loss, theft, or damage.

Another system, called Paranoid Android (PA), was proposed by Portokalidis et al. [11] is aimed at malware detection in smartphones. This system replicates the state of the smartphone in a virtual machine on a dedicated server and runs PC level malware detection tools, such as antiviruses. This system, however, relies on persistent connection of the mobile device to replication server and fails if an adversary is able to subvert communication channels between smartphone and server. It also does not address problems that arise from mobile phone being lost or stolen.

Most of today's smartphone OSs provide synchronization and backup services that are facilitated through online storage, such as Android's Gmail Sync and iOS's iCloud [12]. Third-party applications are also available and can be used to backup users' data. For example, DropBox [13] and Wuala [14] are two popular applications that are used as a cloud-based file backup solution for smartphones. Ion et al. [15], however, shows that most users do not trust the cloud to store their private data, and would prefer home-based solutions instead. There seems to be a gap between what users need from a data protection system in smartphones, and what the current solutions have to offer.

### III. RESEARCH QUESTIONS

Smartphones' data security is a relatively new research field, which is primarily because smartphones are a new phenomenon. Moreover, to the best of our knowledge, there has been no academic studies on users' requirements for data protection systems in smartphones. To get a better understanding of the problem domain, we ask qualitative types of questions in this study, which is an approach better suited for cases when a better understanding of the problem

space is required. In our exploratory study, we aim to answer the following Research Questions (RQs):

- **RQ1:** What types of data do users usually store on their smartphones?
- **RQ2:** How *sensitive* or *valuable* is each data type?
- **RQ3:** What practices do users employ in order to ensure the confidentiality, integrity, and availability of their data?

We define *sensitive* data as data that represents private or confidential information such as financial transactions, work related documents, and personal messages. We define data as being *valuable*, on the other hand, if the loss of such data is considered to be an issue, where a user would desire to recover a copy of the lost data and the user thinks that this activity will require a great amount of effort from him. Note that we classify data as not being valuable if a user does care about such data, but have a copy somewhere else, and it is not hard for him to recover this data on his mobile phone.

Answering RQ1 would give us a better understanding of what kind of data types a data protection system should support on today's smartphones. Answering RQ2, on the other hand, gives us an insight into how security requirements vary for different data types. This would allow data protection systems to focus on what is needed to be protected and prioritized according to users' demands. Finally, answering RQ3 will give us and the wider research community a clearer picture of everyday issues that are faced by users when they try to protect their smartphone data.

### IV. METHODOLOGY

An exploratory study was designed in a form of semi-structured interview with an objective of gathering qualitative types of data. We decided to begin with a qualitative rather than quantitative method in order to explore the problem domain and get a better understanding of users' requirements for data protection systems in smartphones. One of the advantages of semi-structured interviews is that an interviewer can easily deviate from the initial interview structure and plan and focus on the questions that are highly related to the topic, especially for questions that were not expected by researchers.

We use theoretical sampling rather than random sampling during the selection process of participants, as the demographic diversity of participants was more important for this study than having a uniform random sample of smartphone users [16]. Diversity is often more important during qualitative study phases, especially when questions on variability of some parameters need to be answered. Accordingly, we based our selection of participants on the demographics data we obtained via a pre-interview questionnaire, which was sent to a candidate participant, before scheduling an appointment for the interview with him. In the questionnaire, we asked seven questions about age, gender, completed education, job(s) position and area of work, hobbies, annual household income, and native language. Each interviewed participant was paid \$25 CAD for a one-hour long interview.

All interviews began with a set of simple questions, such as “what applications did you use during the last few days?” or “what was the first thing you have done with your smartphone today?”. We then proceeded by asking the participants about the applications they were using on their smartphones and recorded all applications mentioned by users. For business types of applications, such as calendar, emails, and messengers, we also asked whether the participants used them for work or for personal cases. After that, we asked the participants how these application are being used, in particular we were interested what data types are stored with each application and whether or not they save an account’s password for each application. For instance, if participant used an email client we asked what kind of email accounts did he use and what kind of emails did he receive on each of the accounts, we also asked whether he saved the password of any of the email accounts.

During a pilot study of six participant, we found that it is difficult for participants to provide a clear answers about the sensitivity and the value of their data. To address this issue in the our study, we gave several scenarios to participants, that were aimed to communicate probable risks clearer. For sensitivity of the each data type we asked participants about consequences of data being viewed by a stranger or thief. We considered two types of such persons, a complete stranger and someone from user’s social circle who knows the victim. For the value of the data we asked participants about to think about consequences if they lose their smartphone at the time of the interview. Additionally, with those who had lost their phones in the past, we asked about their experience with data loss and recovery.

Finally, by the end of the interview, we asked the participant about the practices they use today to ensure that their valuable or sensitive data are kept confidential and promptly available. During this interview phase we tried not only capture what type of tools are being used, but also why they use or do not use some of the tools.

We conducted all of the interviews in the presence of two interviewers in order to ensure that all important questions were asked. We audio recorded and transcribed all interviews verbatim. Later transcriptions of the interviews were coded, analyzed and checked by both interviewers. To ensure sufficient number of participants in our study we used information saturation analysis, where we categorized each additional unique piece of evidence obtained from each additional participant. We stopped recruitment when we observed that we are not getting any new data from additional participants.

## V. RESULTS

We conducted 22 interviews during the month of October, 2011. Half of them were conducted at the University of British Columbia (UBC) Point Grey campus, and the rest at UBC’s Robson Square campus. As it shown in Table I the demographics of the recruited participants are diverse and include high range of different occupations and age. Note, that some of the participants had more than one job, that is

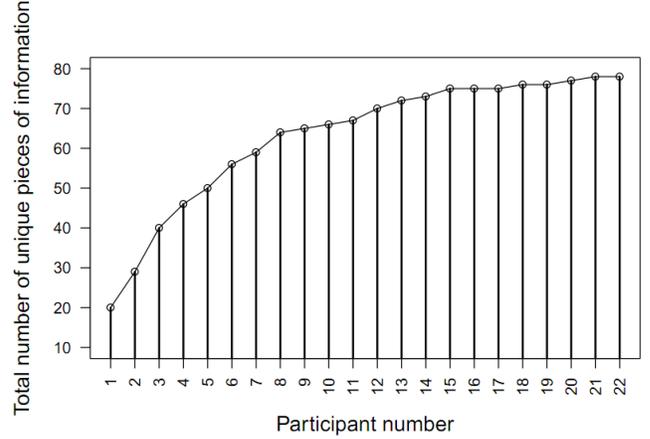


Fig. 1. CDF of new collected information across the interviewed participants.

why sum of the number of participants per occupation will not be equal to the total number of participants. After the 18<sup>th</sup> participant we observed no new information arising in the interviews, that is why we decided to stop recruiting new participants, according to the theoretical sampling approach. The graph, shown in Figure 1, supports this decision and shows that saturation in data collection was reached.

We define two classes of user data based on its protection requirements: valuable data and sensitive data, as discussed in Section III. Moreover, we categorize the smartphone’s data into 11 data types as shown in Table II and security practices and experience with loss/damage of their mobile phones, which we compiled from the transcribed interviews. Similarly, we show in Table II what data class each data type has according to the collected data. In Table II fully filled circle in sensitive columns shows that a data type most of time was considered to be sensitive. On the other hand, empty circle shows that a data type was not considered sensitive by all participants. Data types that were considered sensitive only by a minority, i.e. at least one participant, are shown as half filled circles. Same logic is used for valuable class of data. We do not provide descriptive statistics about the data types or their classes as this was not the goal of the study and requires different, quantitative, methods to be used. In this paper we provide our participants’ reasoning and explanations for why each data type was assigned to its corresponding class. In the following, we provide further explanation for selected data types that we found interesting.

**Passwords:** Some of our participants stored their passwords on their smartphones using different means. One participant stored passwords for online banking as a contacts in the address book in clear text, another created notes for PIN-codes from doors at her work. Sub group of the participants used special applications, such as password managers, to store their list of passwords. Most of participants opted to let applications to save the associated passwords, so that they did not have to enter a password every time they

TABLE I  
DEMOGRAPHICS OF INTERVIEWED PARTICIPANTS

Parameter	Property	Participants
Gender	Males	10
	Females	12
Age	under 18	1
	19-24	7
	25-30	5
	31-35	2
	36-40	3
	41-45	3
Education	46-50	1
	Still in High School	1
	High School	6
	Professional School or College Degree	4
	University (Bachelor's)	6
Household income	Graduate School (Master's or PhD)	5
	under 15K	6
	15K-30K	3
	30K-50K	3
	50K-80K	7
Smartphone OS	more than 80K	3
	iOS	11
	Android	4
	Symbian	2
	BlackBerry OS	4
Occupation	WebOS	1
	Caregiver	1
	Curator Assistant	1
	Entrepreneur	1
	Graduate Student	2
	High-school Student	1
	Language Teacher	1
	Marketing	2
	Municipal Worker	2
	Network Administrator	1
	Nurse	1
	Librarian	1
	Pilot Instructor	1
	Proof-reader	1
	Sales Person	4
	Security Guard	1
	Software Engineer	1
Tailor	1	
Undergraduate Student	1	
Unemployed	1	
Data Stored	Work Related	9
	Personal	22
Phone Ownership	Personal	19
	Company	3

ran such an application (e.g., the email clients, Facebook application). All participants considered these passwords to be sensitive data. Those participants, who used password managers, were less worried, since such application required additional password, although they admitted that the password they used was a name of a person or a dictionary word. Interestingly, some of the participants considered password lists as being highly valuable, as these passwords lists were stored only on their smartphones and loss of the list will incur

TABLE II  
TYPES OF DATA AND THEIR SENSITIVITY AND VALUE FROM USERS' PERSPECTIVE

Data Type	Sensitive	Valuable
SMS Messages		
Photos/Videos		
Voice Recordings		
Notes		
Contacts		
Music		
Passwords		
Emails		
Documents		
Events in Calendar		
Recorded GPS Tracks		

a lot of work for the restoration of the accounts and passwords.

**Music and Events in the Calendar:** Music and events, from another end, were never mentioned as being sensitive or valuable. The most common explanation from the participants for that was that they have a copy of such data on their computer, or online, or they remember it all. For the appointments, in case if they had a lot of them, they kept a copy in physical agenda book.

**Voice Recordings:** We found that users sometimes use their smartphones to audio tape particular conversations, which could be confidential, and thus, sensitive. For example, one of the participant was a quality assurance specialist, and has recorded a conversation with an employees without their consent. From another side, participants used their voice recorders for notes and memos, which they did not consider as being sensitive. Participants generally considered such recordings valuable as they might not be able to recover them if lost.

**Photos and Videos:** Some of the participants defined some of the photos and videos in smartphone as both sensitive and valuable. Others considered their pictures and videos sensitive for cultural reasons. For example, one of the participants stated that photos of his family are sensitive, as women in his culture wear Hijab in public. Interestingly, most of the participants, who took pictures and videos on their smartphones, kept them there for some period of time in order to accumulate worthy amount of them, before transferring them to a PC. Several participants who lost or damaged their phones recently admitted that they lost valuable pictures too. Moreover, it was hard for participants to tell at the spot

whether they have valuable or sensitive pictures, without looking into their image and video gallery.

**SMS Messages:** We found that SMS messages have a short temporal value, which is lost once participants read them. Most of the participants stated that they do not use SMS for highly meaningless conversations, rather than friendly chatting and a way of poking their friends. Moreover, we found that participants considered some messages as being sensitive if they are read by particular people, such as their parents “I do not like the idea of someone, especially my parents, going through my messages...”, but they were comfortable sharing them with others, such as their friends.

**Contacts:** The participants sometimes considered the contacts as being sensitive, mostly because they were not comfortable sharing such data with others. Reasons varied from reputation consequences “*If someone got hold on of my contacts, I would feel uncomfortable, because I feel like those people trusted me to keep their phone numbers private*” to expected threats to people “*I am not sure what those who got my contacts numbers will do with them, they could call them or send them spam*”. The value of contacts details was justified mainly by the lack of synchronization with a PC or an online account. Interestingly, some of the participants stated that they had a copy of their contacts details in a small paper book, which they carry around with them, because they had experienced loss of their phone or experience problems with batter lasting on their smartphones.

**Email Messages:** Participants classified their emails as being not valuable, because all of them were able access emails either online or on a PCs. Most of the participants had more than one email accounts, and setup several of them on their phones. They classified their emails as “junk-collecting” or “sign-up” email, personal and work related. Nine of the participants used work email accounts on their phones and received confidential emails, such that contained new products details, marketing companies budgets, business proposals. The mix of unimportant and important emails defined Emails sensitivity as “could be sensitive”.

**Documents:** Some participants uploaded work-related documents to their smartphones. These documents often contained confidential information, such as description of a new product or sales figures. Reasons for having such data on their mobile phone were explained by necessity to have some vital numbers on the go during travels. Such documents were not considered valuable, because were easily recoverable either from participants’ PCs or from network storage at work.

**GPS Tracks:** GPS tracks are usually saved by training assistant programs, such as miCoach for iOS, which are used by people to do outdoor exercise and track their performance in terms of energy and mileage. Also these applications used to store tracks on the maps, where user was during his exercise.

TABLE III  
SECURITY PRACTICES AND EXPERIENCE OF INTERVIEWED PARTICIPANTS

Parameter	Property	Participants
Use pin-lock	Yes	7
	No	11
	No, but used to	4
Had experience of	Losing phone	5
	Breaking phone	4
	Losing and	
	Breaking phone	1

Such tracks were considered to be very sensitive, mainly because such tracks most of the time lead to the users’ homes.

In our interviews we tried to understand whenever users use some tools to ensure protection of their data, such as backup or PIN-code locks. We also tried to understand what were the reasons for and against using specific tools. In what follows we present results on what practices were used by our participants and their justification. Short summary of the results is shown in Table III.

We found that most of the participants, but not all of them, backup valuable data whenever they “feel” that they have to, which varied from *once a week* to *once in six months*. Those who lost their device and valuable data too admitted that they are paying more attention now to this practice. Another interesting observation was the main reasons for such infrequent and irregular backups. Among them we highlight (1) inconvenience of current systems, (2) lack of time, and (3) lack of information on what data needs to be backed up.

Several participants stated that they do not trust the security of their smartphones at all, and, thus, decided not to store any sensitive or valuable data on such devices. Their decision was based on concerns they have in regard of risks to smartphones, such as loss, theft or infection by a “virus”. Interestingly, 20 participants told that they think that smartphone is less secure device in terms of storing their data than a PC. As the main reason for this participants mentioned high mobility of the smartphones, and, thus, higher chances for it to be lost or stolen.

When we asked what they would do if they lose their mobile device, most of the participants told that the first action they would take is try to recover their device, by going through places they visited in last couple of hours. Four participants told that they are using application to track their device, such as “Find my iPhone” [17] and will try to find their phone through this application first. Answers of those who did lost their phones before were the same, moreover, those who did lose their phones in the past tried to recover their devices first. In the case when users were not able to recover their smartphone within couple of hours all participants told that they would call their service providers and block their line, as to avoid paying for someone using their high cost services. All participants who stored passwords of the phone in any form told that they would change their passwords in a day or

two after the loss. Not surprisingly, the phone itself was also mentioned as a financially valuable asset to lose.

In the scenario when their phone been stolen or used by someone else, participants showed different perception of risks from different types of persons that used their smartphone. Threats expectations were higher for 17 participants when a person who used or stolen their phone was someone who knows them. Our participants also stated that if they lend their phone to their friends they would like to keep an eye on them, mainly because they had concerns about this person looking into their personal data, such as messages and pictures. Most of the time they did not care about showing some data, such as messages and emails, to complete strangers, but did care if such data were seen by someone from their social circle.

Not surprisingly, 21 participants stated that they would like to store backups of sensitive data at home on their PCs or external hard drives rather than having them online. Two Android users decided to disable synchronization with Gmail account completely only because of privacy concerns. Moreover, half of the participants used external hard drives already as a backup solution for their home media files, such as videos, pictures and documents. Although, most of them did use some form of “cloud” storage, such as Gmail, Facebook or Dropbox, they preferred to store only “shareable” content at such services. This is, also, consistent with Ion et al. [15] findings, where they studies users’ attitudes toward cloud storage in general.

Finally, we observed only seven participants who used PIN-locks on their smartphones. One of the participants used it only because of the company enforcing policy, and told that he would not use it if that would be possible. Another participant told that she used PIN-lock only to protect her SMS messages from her parents, and found it annoying that she were not able to protect only these messages. All participants that used PIN-codes stated that they type PIN-code very often for data that is both not sensitive and not valuable to them, such as weather forecast or games. Those who did use lock but gave up explained this by necessity to have fast access to some data and functions of the phone on the go or in some specific scenarios. For instance, one of the participants told that she gave up on using PIN-codes when she was at the party and needed constant access to the Internet, to check so information. She found it inconvenient to type her PIN-code all the time so she decided to switch it off completely. Moreover, cases when they need to type PIN-code or password on the go render them unusable, especially when users are in rush. Likewise, users who did not use any type of the PIN-codes agreed that typing a PIN-code for every application or data on their phone does not make sense. For those who did not use such locks, they main reasons for not using it were (1) they do not have any sensitive data on their smartphones, or (2) it is too inconvenient for them to type a PIN code or a password, or (3) they felt “socially-awkward” to type a password in front of their friends or family.

## VI. DISCUSSION

In this section, we discuss the implications of our results on smartphone data security, focusing on data availability, integrity and confidentiality.

**Confidentiality** is at risk when sensitive information is exposed. This potentially might include, but is not limited to, pictures, bank accounts statements, contracts, voice memos and emails. Users’ tendency to save passwords for applications on their smartphones make this problem even worse, especially when 54% of smartphones users do not use PIN-locks on their phones [4]. Systems that aim to address data confidentiality, such as data encryption, could be improved based on our results by selecting and prioritizing what to encrypt, and thus, consume less energy.

Different requirements to security of different data types suggest that smartphones’ PIN-locks should introduce fine-grained control, where a user can specify which application and data should be locked and which one should be accessible instantly. This also will reduce burden on users of constantly typing their PIN-codes for unimportant applications and data. Moreover, similar applications already available for Android OS, for instance Smart AppLock [18], which was downloaded more than 500,000 times. Although such applications does not allow to separate between different data in the same application, what could be needed when one application has several accounts of different value, for instance email clients that has accounts for “junk” emails and work emails.

Another problem with such applications is process recycling in mobile OSs, where a process that is not part of the OS could be killed by OS in order to released resources needed for another application, or an application, such as Advanced Task Killer [19] could be used to kill such process. That is why they should be integrated into operating system. Another possible approach to address this problem is it to privilege elevation, similar to what is used in Mac OS and Windows, that could used when a users wants to access sensitive or valuable data. Confidentiality problem is also important in the light of this study results, where we found that users tend to spend sometime trying to track their phone before blocking it, which might give enough time for an adversary to get sensitive data out of the phone.

One can argue that if users are to protect their emails then they should not save their password on the email client. We believe, however, that this does not solve the problem because (1) email clients will have a local copy of several previous emails, which can be accessed without typing in a password, and (2) these email clients usually require much stronger passwords, which are more difficult to memorize than PIN-codes and are harder type on small devices such as smartphones.

**Availability** of the valuable data is at risk when a malware or another person deletes or corrupts data on the smartphone or the mobile device is lost or damaged and, at the same time, user do not have a copy of such data anywhere else. Our study showed that smartphone users tend not to do frequent

and regular data backups. Moreover, participants felt that this activity takes a lot of their time, is inconvenient and it is also hard for them to know what exactly need to be backed up. At the same time, participants did store valuable data on their phones, which they did not store anywhere else, such as contact details, passwords lists and photos. Two participants who have lost their mobile phone before told that they also lost valuable photos, which they did not saved on their PC. Finally, 21 participants stated that they would like to have a “local” backup storage, rather than an online storage, mainly because (1) they do not trust their sensitive data to “someone” who can easily see them, (2) they have concerns about cost and availability of such services, and (3) they are not satisfied with access speeds. This suggest that data protection solutions that are aimed at the problem of data availability have to consider automatic backup and synchronization with a local storages, in a such way that requires minimal users intervention.

**Integrity** of data is at risk when a smartphone is used without the user’s consent or knowledge or data are corrupted by a malware. For instance, an adversary could change or delete some data, such as phone numbers or pictures, or a malware could send SMS messages to premium numbers. One way to reduce such risks to data integrity is to use authentication/authorization for all write operations on valuable data types or use of costly services. Although, asking to make a decision on each write could be overwhelming for users and additional user studies should be conducted to advise a better design for such systems. Another option for data integrity is to support data versioning, so that data could be restored to a previous state(s) when integrity is corrupted. This however requires users to be vigilant and detect when corruption case occurs.

## VII. LIMITATIONS

In what follows we discuss limitations of this work.

First of all, in qualitative research where researcher is an interviewer *researcher bias* could rise from researcher’s expectations on the outcome of the study. In order to reduce this threat we had two interviewers and we did member checking with participants to test whether emerging ideas and concepts make sense. Likewise, interviews were transcribed, checked and coded by two researches and all conflicts on coding were discussed until an agreement is reached.

In this study we did not have generalizable population sample, which is due to participants selection method, where we aimed at a diverse population rather than a representative one. For instance, we rejected more than 20 participants only because all of them were students and by that time we already had four students in our sample. Thus a follow up study is required which will use quantitative data collection methods.

Participants’ answers might not reflect their real opinions, since all interviews were conducted outside of the usual contexts, where users usually use their smartphones. We tried to address this issue by providing users with use scenarios, although this does not address problem fully, we do not see any other ways to address this issue, rather than inquiring

participants in the context. This, however, practically impossible, because people tend to use their smartphones sporadically during the day and “on the go”. Moreover, some important applications, such as e-banking, are used only once a month.

We admit that some of the questions might have been misinterpreted by our participants, thus they gave incorrect or not precise answers. We tried to mitigate such risk by (1) conducting a pilot study with six participants and testing questions, where we asked pilot study participants whenever a questions was clear or not, and (2) by constant analysis of arriving data and checking our understanding with participants. Where appropriate, we provided users with scenarios, which meant to help them with possible risks assessment. Most of the emerging concepts were checked with participants at the end of the interview. We also asked participants to show us the matter in question on their device. For instance when a participant told that he uses a PIN-lock we asked him to show how it works, thus we were able to classify what type of PIN-lock he used.

Conclusions made in this work are not final and will be adjusted after a follow up study. In particular, the list of data in Table II could be extended with new types of data, unseen during this study. We also considered data only from application point of view, and we did not consider semantic link between these data, for instance when different data types are linked by a project, which we plan to address in future research. Additional reasons could appear for using or not using specific data protection tools, such as backup synchronization and PIN-codes. Likewise, new insights on the design of such systems might appear, such as a clear prioritization of the aforementioned problems that users face today.

## VIII. CONCLUSIONS

In this paper we presented results of an exploratory qualitative user study, which aimed to get a better understanding of users’ requirements to data protection systems in smartphones. We compiled a list of data types that we found being used or stored by our participants and justifications why each data type is considered to be sensitive or valuable. We showed what security practices are used by our participants and their preferences when using such systems. Likewise, we presented reasons why some of the tools are not being used, such as PIN-codes and backup systems.

Overall, we found that users do store sensitive and valuable data on their smartphones, and consider the security of their data as a concern. They, however, tend to not take any actions in order to ensure confidentiality, integrity and availability of this data. We also found that PIN-codes and passwords locks for smartphones to be unusable for most of the users, which is mainly due to their demand of an instant access to data types and applications that are not sensitive or valuable such as games, weather forecast applications, and Internet browsers.

We also showed that users are reluctant to store sensitive data in an online storage and prefer to use local storage solutions, such as external hard drives. Moreover, users felt

that existing solutions require to much efforts from users to do a frequent and regular backups. In case of those participants who had lost their phones before, this resulted in loss of valuable data, such as photos and contact details.

## IX. FUTURE WORK

In this paper, we presented qualitative results of an exploratory user study on users' requirements for data protection in smartphones. We acknowledge that with the current study design we cannot generalize our results, but we were able to better understand the problem domain and capture users' requirements. We plan to conduct a confirmatory study on a larger population sample, where we will collect descriptive statistics on users' preferences and security practices. We are forming a set of hypotheses, which we will test using an online survey. Based on these findings we plan to build a prototype of data protection system, that will offer tools for Confidentiality, Integrity and Availability of users' data, which we plan to evaluate thoroughly on efficiency, efficacy and usability.

## ACKNOWLEDGMENT

We would like to thank San-Tsai Sun, Patrick Conroy and the members of the Laboratory for Education and Research in Secure Systems Engineering (LERSSE) for their kind feedback on an early draft on this paper. This research is sponsored by a grant from NSERC Internetworked Systems Security Network (ISSNet).

## REFERENCES

- [1] "Smartphone sales to surpass laptop sales by 2012," <http://www.switchfast.com/switchfast-blog/2009/10/28/smartphone-sales-to-surpass-laptop-sales-by-2012.aspx>, last accessed August 18, 2011.
- [2] "Gartner highlights key predictions for it organizations and users in 2010 and beyond." <http://www.gartner.com/it/page.jsp?id=1278413>, last accessed August 18, 2011.
- [3] "Flurry: Time spent on mobile apps has surpassed web browsing," <http://techcrunch.com/2011/06/20/flurry-time-spent-on-mobile-apps-has-surpassed-web-browsing/>, last accessed August 18, 2011.
- [4] A. Smith, "Smartphone adoption and usage," <http://www.pewinternet.org/Reports/2011/Smartphones.aspx>, last accessed August 18, 2011.
- [5] "Lost and found: The challenges of finding your lost or stolen phone," <http://blog.mylookout.com/2011/07/lost-and-found-the-challenges-of-finding-your-lost-or-stolen-phone/>, last accessed August 18, 2011.
- [6] V. Zakorzhnevsky, "Monthly malware statistics, march 2011," [http://www.securelist.com/en/analysis/204792170/Monthly\\_Malware\\_Statistics\\_March\\_2011](http://www.securelist.com/en/analysis/204792170/Monthly_Malware_Statistics_March_2011), last accessed August 18, 2011.
- [7] C. Eric, "The motivations of recent android malware." [http://www.symantec.com/content/en/us/enterprise/media/security\\_response/whitepapers/motivations\\_of\\_recent\\_android\\_malware.pdf](http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/motivations_of_recent_android_malware.pdf), 2011.
- [8] M. Ongtang, S. McLaughlin, W. Enck, and P. McDaniel, "Semantically rich application-centric security in android," in *Proceedings of the 2009 Annual Computer Security Applications Conference*, ser. ACSAC '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 340–349. [Online]. Available: <http://dx.doi.org/10.1109/ACSAC.2009.39>
- [9] W. Enck, M. Ongtang, and P. McDaniel, "On lightweight mobile phone application certification," in *Proceedings of the 16th ACM conference on Computer and communications security*, ser. CCS '09. New York, NY, USA: ACM, 2009, pp. 235–245. [Online]. Available: <http://doi.acm.org/10.1145/1653662.1653691>
- [10] W. Enck, P. Gilbert, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth, "Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones," in *Proceedings of the 9th USENIX conference on Operating systems design and implementation*, ser. OSDI'10. Berkeley, CA, USA: USENIX Association, 2010, pp. 1–6. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1924943.1924971>
- [11] G. Portokalidis, P. Homburg, K. Anagnostakis, and H. Bos, "Paranoid android: versatile protection for smartphones," in *Proceedings of the 26th Annual Computer Security Applications Conference*, ser. ACSAC '10. New York, NY, USA: ACM, 2010, pp. 347–356. [Online]. Available: <http://doi.acm.org/10.1145/1920261.1920313>
- [12] Apple, "iCloud," <https://www.icloud.com/>, 2011.
- [13] Dropbox Corporation, "Sync your files online and across computers," <http://www.getdropbox.com/>, 2009.
- [14] Wuala - Secure Online Storage - Backup. Sync. Share. Access Everywhere., "Sync your files online and across computers," <http://www.wuala.com/>, 2010.
- [15] I. Ion, N. Sachdeva, P. Kumaraguru, and S. Capkun, "Home is safer than the cloud! privacy concernc for consumer cloud storage," in *Proceedings of Symposium on Usable Privacy and Security*, Pittsburgh, PA, USA, July 2011, pp. 1–20. [Online]. Available: <http://cups.cs.cmu.edu/soups/2011/proceedings/a13-Sachdeva.pdf>
- [16] B. G. Glaser, *Theoretical sensitivity : advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press, 1978.
- [17] Apple, "App store - find my iphone," <http://itunes.apple.com/ca/app/find-my-iphone/id376101648?mt=8>, 2011.
- [18] ThinkYeah, "Smart AppLock Pro," <https://market.android.com/details?id=com.thinkyeah.smartlock>, 2011.
- [19] ReChild, "Advanced Task Killer," <https://market.android.com/details?id=com.rechild.advancedtaskkiller>, 2011.