Survey: Data Protection in Smartphones Against Physical Threats

Ildar Muslukhov
The University of British Columbia
Department of Electrical and Computer Engineering
2332 Main Mall,Vancouver, Canada
ildarm@ece.ubc.ca

ABSTRACT

In this paper we present a survey of the data protection systems in the presence of physical threats, such as theft, loss, damage and malicious use. Recent studies suggests that the aforementioned threats are real and happen with very high rates, e.g., it has been reported that more than 50% of Maimi city had experience with theft and loss of their device in 2011. In this paper we also present adversarial model and users' requirements to data protection systems which were based on the result from the interview and survey studies we conducted. Analysis of the existing systems in the light of the aforementioned adversarial model and users' requirements revealed that existing system (a) do not consider users as a source of threats, (b) do not provide adequate protection against an adversary with physical access, and (c) do not meet users' requirements on usability and privacy. Finally, we suggest areas of further research, which would help to address the problem of data protection against an adversary with physical access to the smartphone.

1. INTRODUCTION

Smartphones have become truly ubiquitous devices of today. Diversity of services and functionalities provided by modern smartphones include gaming, web browsing and emails to GPS navigation, voice search and high definition video. Such functionalities attracted many users, which allowed smartphones to overtake laptops and desktops in the number of annually sold items [5, 29]. Smith reports that 46% of adults in the US have smartphones [53] (an increase by 11

Such spread of smartphone use, however, is accompanied with a rise of adversaries who use different ways to monetize attacks on smartphones. As a result, the number of different types of malware grew [20, 57] and physical threats, such as loss, misuse, theft or damage of the smartphone grew too [6, 15]. Moreover, adoption of smartphones in businesses creates new attack vectors on corporate data records, where sensitive and confidential information is at a greater risk due to higher mobility of smartphones. Ponemon Institute [9] shows that in 2010 cases which involved lost, stolen, misused data-bearing devices, such as smartphones, were ranked among the most expensive ones to recover from and among the hardest to detect, with an average cost of $156 per breached record or file. Such cost includes both (a) expenses on the investigation, and (b) any types of fees for data breaches. Lookout, an antivirus vendor, reports that loss and theft were among the biggest threats to smartphones in the middle of 2011. For example, in some of the US cities proportion of the population, which had experienced loss or theft of their mobile devices was up to 52% [6]. Finally, Symantec reports that in 96% of cases when a smartphone is lost a stranger that finds it will try to access sensitive information, such as social networking applications, emails, pictures, passwords, and banking applications [13].

One way to explain this it the small size of the modern smartphones, that might make it easier to lose or to steal it. For instance, it was reported that 31,544 smartphones have been left in taxi cabs of New York city during the first six months of 2010 [15]. Landman reports that over 60,000 mobile phones and 5,500 Personal Digital Assistants were left in the cabs around London, UK in the six months of 2005, in comparison to 4,500 laptops during the same period [33].

Given the popularity of smartphones, their intrinsic mobility and the rise of the threats to smartphone data, we believe that it is very important to address the problem of smartphones' data security, especially in cases when a smartphone is being lost, stolen, misused or damaged, i.e., in the presence of physical threats. Our exploratory study showed that users perceive loss, theft, misuse and damage of their smartphone as a serious threat and that they want to protect their data from consequences of physical threats being realized. Users, however, do not protect such data, because they either find that existing tools and systems do not comply with the sets of their requirements or inconvenient to use. Lets consider PIN-lock. The participants, who did not use PIN-code lock for the phone or stopped using it, told us that such lock is too coarse and does not allow them to protect particular applications or particular data records. Moreover, those who did use such lock told us that they would like to have different PIN-codes for different sets of applications and data items and to have the ability to switch it off sometimes.

We also found that, participants do not backup smartphone data regularly, because they do not keep track of which data item was changed on the smartphone and should be backed up. They also found this procedure to take a lot of their efforts. Interestingly, most of the participants agreed that they prefer local storage to a cloud storage, such as DropBox [24]. Among the reasons for that were local storage is faster, cheaper, users have more physical control over it, and they have privacy concerns with the cloud storage. The last reason, however, is not surprising and is coherent with the results shown by Ion et al. [31], where they studied Internet users' attitudes and preferences towards cloud storage. Given that there is a gap between what users need from smartphone data protection systems and what current systems have to offer, it is an important problem to identify which required functionality is missing or does not meet users' requirements fully and need to have a better support.

In what follows we first define risks associated with physical threats and present adversarial model, where we discuss adversaries' capabilities. We then discuss users' requirements to data protection system (DPS) based on related literature and interviews and survey we conducted. After we discuss each property of data security (i.e., availability, integrity and confidentiality) in the light of the existing systems, adversarial model. Finally, we conclude with discussion of the limitations of the existing solutions and propose areas for further research.

2. PHYSICAL THREATS, RISKS AND ADVERSARIAL MODEL

Data in smartphones, potentially, could be sensitive or valuable. We define data as being *sensitive* if such data is either confidential or users would have concerns with revealing such data. We define data as being *valuable* if there is financial, nostalgic or personal value for a user in such data. We divide risks to data into two groups (1) an unauthorized access to data, and (2) data loss. Indeed, recent report show that there is an increase of unauthorized access to sensitive data on the phone [9]. Furthermore, McAfee reports that "Four in 10 organizations have had mobile devices lost or stolen and half of lost/stolen devices contained business critical data" [10].

Not all physical threats are caused by an adversary. For instance, loss or damage of the device might lead to data loss, however, these threats do not need an adversary in order to happen. This, however, does not mean that we should not pay attention to these threats, since they do impose risk to valuable data and sensitive data.

In the presence of an adversary, both risks (unauthorized data access and data loss) are possible. Our adversarial model assumes that an adversary would have at least one of the following objectives:

- Read sensitive data, e.g., personal messages in online social networking application
- Modify valuable data, e.g., replace event date and time in the calendar or change contact phone number
- Get the smartphone itself as a valuable device

Physical threats by definition imply that an adversary is able to get a physical access to the device. Moreover, we assume that an adversary before stealing or borrowing the device is able to observe how victims use their smartphones. The aforementioned capability of observation in conjunction with the results from recent studies suggest that such an adversary would be able to capture authentication secret. In particular, De Luca et al. [23] showed that most of the participants do not protect their PIN codes from credit cards while using ATM machines. Zakaria et al. [56] studied how easy it is to capture Draw-a-Secret authentication secret and showed that almost all participants were able to capture authentication secret for DAS in 5-by-5 grid configuration. Deployed DAS authentication configurations, however, are 3-by-3, which would be even easier to captures due to small and easier pattern. Finally, Raguram et al. [50] presented a system that allows to restore users' input into smartphone from reflections of the display from other objects in the environment, such as glasses or windows.

Different objectives breach different properties of data security. An adversary that gets unauthorized read access to data might compromise data confidentiality. An adversary that gets unauthorized write access to data might compromise data integrity or availability. Integrity might be at stake if an adversary modifies data with malicious intent, e.g., change the phone numbers in the address book. If such activity goes unnoticed for a while or no additional copies of such data are stored anywhere else, availability of such data will be compromised as well. Finally, an adversary that aims at financial value of smartphone might compromise data availability.

We assume that an adversary is not able to replace the core software of the smartphone, and thus gain full control over the operating system. This assumption, however, does not mean that an adversary cannot perform usual users' actions, such as application installation, control network connectivity and resetting device to factory settings, i.e., wiping all data and applications in smartphone.

3. USERS' REQUIREMENTS TO DPS

In the interview (22 participants) and survey (58 participants) studies we found that users do have usability and privacy related requirements to the DPS, such as storage location and ease of use of locking system.

For availability property of data security most of the participants stated that they would like to have backup of their data to be stored on the local storage (e.g., desktop/laptop, external hard drive). This was particularly true for sensitive data, however, it also was true for valuable data. Users saw local storage as a better solution for them because of the following reasons (1) they have more control over local storage (e.g., they can physically unplug the storage from the Internet, or control who has access to it), (2) local storage is faster, and (3) local storage is cheaper. Indeed, recent study study by Ion et al. [31] shows that users do have privacy concerns when storing sensitive data in the cloud based storage. Moreover, recent incidents with systems based on online storage only make this matter worse. For instance, in the middle of 2011 DropBox made all users' files accessible without authentication [25]. Mulazzani et al. [40] showed an attack on DropBox protocol, where an attacker was able to test whether a user stores specific file.

On the other hand, for confidentiality property of the data security users wanted to have more control over what should be protected. For instance, more than 20% of users found the simplest authentication methods, such as PIN-codes and DAS, very unusable in the current implementation, when users need to authentication each time they access phone. If a locking system of a smartphone could be more fine grained, i.e., provide different levels of protection for different data items and functionalities, this would allow users to use a stronger authentication methods for more sensitive data items and weaker authentication methods for less sensitive items. Moreover, making access to non sensitive data without authentication and using various authentication methods for sensitive data might increase efforts need by an adversary in order to capture authentication secret, since he/she would need (1) to wait more for the moment when users access sensitive data, and (2) to consider stronger authentication methods being used with different authentication secrets for different data types or items.

As for those who did use locking system they render themselves vulnerable to the aforementioned adversary, because most of them (90%) used weak authentication methods, such a PIN or DAS. The participants stated that the main reason for such decision is that they need to authenticate themselves too often and this make it impossible for them to use stronger authentication methods. Furthermore, they stated that for some data types or items they would like to have several levels of defense, i.e., they would like to keep the PIN code for the whole phone, but also have an additional lock for some pictures or some messages. In related literature it has been shown that they way users use their smartphones is different from the way we use desktops/laptops, where with smartphones users tend to work in short bursts of interactions, with very frequent distractions [47]. This suggests that a more flexible way, such as fine grained access control should be introduces in order to enable users to access non important data without need of authentication.

4. EXISTING DATA PROTECTION SYSTEMS

In this section we review existing systems in the light of the aforementioned adversarial model and users requirements. We divide our discussion into sub sections related to each property of data security, i.e., availability, integrity and confidentiality. We combine confidentiality and integrity, because system that provide confidentiality protection also could be used for integrity protections. For instance, by limiting access data modification functionality. Because of the same reason we discuss together data availability and integrity, where reservation systems with support of versioning could be used to overcome data corruption by rolling back to previous version.

4.1 Confidentiality and Integrity

There are two approaches that are used to reduce the risk of an unauthorized access to data. The first one is based on access control, which could use privilege separation or domain isolation. The second one is based on malicious use detection, for instance by using intrusion detection systems (IDS).

4.1.1 Access Control

Data isolation approach was explored by many authors. For instance, Ni et al. [44] proposed a smartphone sharing system based on privilege separation, called DiffUser, which introduces three user accounts: Administrator, Normal and Guest. Each account has different permissions on which applications or phone services could be accessed. In particular, a Guest user account has access to voice calling functionality, can view contacts, but cannot change them and has no access to other applications. Normal user account, additionally to what a Guest account has, would also have ability to run most of the installed applications. An Administrator user account, additionally to Normal user account, would also be able to install and remove applications and configure general phone settings, such as network connectivity. In order to switch between different accounts user need to authenticate himself with a password, which is different for Normal user and Administrator. Even though, it addresses the problem partially it is still far away from meeting users' requirements. First of all, it considers all data items and functionalities within applications as being of the same sensitivity level. Second, DiffUser does not allow custom settings for access scope for the introduced accounts.

Somewhat similar and simpler applications exists on the mobile applications markets already. For example, the Smart AppLock Pro and Ultimate App Guard [7, 8]. These applications introduce privilege elevation to mobile platform and divide installed applications into two parts. One part will always require authentication, while another would be accessible without it. These applications, however, share the same limitations of the DiffUser system. Moreover, reference monitor in the aforementioned applications is prone to terminations by both OS, due to resource allocation optimization [12], or by third party applications, such as Advanced Task Killer [1].

Barr et al. [16] proposed data isolation by virtualization based on hypervisor. This approach implies use of separate instances of mobile OS in virtual machines, where separation for different types of data is enforced, e.g., personal data in one virtual machine and corporate data is in another. By separating data into different virtual machines, the aforementioned system reduces risks to data in case if another instance of virtual machine got corrupted by a malware. Virtualization, however, has its own drawbacks when applied in mobile platform, such as performance and energy overheads, which are highly important for mobile devices, especially for smartphones, since users rely on them. This also makes such approach less attractive since current smartphones would not be able to host more than two of three virtual machines, which could be insufficient for users.

Finally, Bugiel et al. [18] introduced framework that allows applications to be isolated into different domains, such as personal or business. Even though, their work targets malware threats, one can still use such approach in combination with authentication in order to separate domains. Such approach would introduce less computational and memory overheads than virtualization, since it does not require several instances of an OS to be run on the smartphone. However, this approach does not provide users with means to differentiate data within an application and tools to provide different levels of protection.

All data confidentiality protection systems that rely on access control as a way to protect data confidentiality (and integrity) need also to rely on some type of authentication, in order to tell legitimate and illegitimate users apart. The popularity of weak authentication methods in smartphones, such as PIN and DAS [21, 32], makes such approach highly vulnerable to an adversary, described in section 2. Moreover, findings from our survey study show that 90% of smartphone users use either PIN-code or DAS authentication methods.

One way to address weakness of PIN/DAS based authentication methods is to use stronger authentication method or password policy, such that provides better resistance to shoulder surfing attacks. Substitution candidates might include a stronger alpha-numeric passwords, gesture based authentication [37], face or voice recognition, near field communication, RFID or bluetooth tokens, or finger print scanner. Some of the aforementioned methods have usability problems, for instance Kurkovsky and Syta [32] showed that striker password policies impact memorability and users often forget them.

Other authentication methods could be used in particular environments. For example, gesture authentication could be used in the environment where nothing but user interacts with accelerometer. Moreover, users should be able to move their hand freely while reproducing gesture. All this make such method hardly applicable in such cases when users are on the bus, or in very crowded environments, which are often the cases where users use their smartphone [47]. Face and voice recognition falls to the same problem, they require clear enough from noise environments, so that users' voice could be clearly distinguished. Moreover, they are easy to overcome, voice authentication could be easily recorded by an adversary and current system based on authentication by face recognition could be overcome by presenting static picture of the user [4].

Hardware tokens, on the other hand, are highly usable, since they do not require users to remember any authentication secret. Moreover, the authentication process could be done in the background, by constantly evaluating proximity of the token, which would provide users with seamless user experience on the smartphones. Even though, such approach reduces mental task of remembering authentication secret, it introduces another mental and physical task of remembering to not to lose the hardware token and taking care of its availability (i.e., batter power), in case of such bluetooth or NFC tokens. Moreover, an adversary that observes such protection might choose to steal the device with token, or to "borrow" the device and use it within proximity of the token, e.g., in the next room.

One can also use several authentication modalities in order to address the weakness of PIN/DAS based authentication methods. For example, one can add voice recognition or use keystrokes dynamics [17, 28, 39] in parallel with PIN or DAS. Such approaches, however, would be either very intrusive (i.e., require much more data to be provided by users through input interfaces), or extremely limited by small input interface of the smartphones (i.e., such interfaces do not allow to collect enough data for classification algorithm to work). For example, an unauthorized data read access would not necessarily generate enough keystrokes for classification algorithm to work [17].

4.1.2 Intrusion Detection Systems

Another way to provide confidentiality protection to sensitive data is use IDS and to detect a malicious smartphone use, i.e., an intrusion. Such an IDS, relatively to the smartphone, could be either passive or active.

Passive IDS, such as Network based IDS (NIDS) use network traffic of the smartphone in order to detect whether a smartphone is being abused and should be disconnected from services, in order not to disrupt others users access to such serves. NIDS, however, has conceptual limitation in the presence of a physical adversary, because it does not consider data stored in the mobile device as assets that need protection. Moreover, it is doubtful that an adversary with physical access to the device and interested in the data inside the device would generate enough network traffic. Moreover, the adversary, described in section 2, most probably will disable all network connections in order to go unnoticed and to disable remote phone management, such as Find My iPhone [3]. We will exclude NIDS from further discussion because of the aforementioned limitations.

Active IDS, or Host based IDS (HIDS), which are meant to be run on the system they protect, are more suitable for the physical threats and could be based either on signatures detection or anomaly detection approaches. Signature and anomaly based HIDS differs in the way they establish "ground truth" about malicious activities. Signature based approach requires a set of malicious behavior patterns to be defined before such system to be used. Anomaly based approach, on the other hand, does not need any previous knowledge and build users' behavioral model during the system use. It then uses this model to detect cases when users' behavior differs significantly from what is predicted by model. In what follows we first discuss existing signature based IDS and follow with discussion of anomaly based IDS.

4.1.3 Signature based Intrusion Detection Systems

Similar signature based HIDS were proposed by many authors. For example, Nauman et al. [43, 42], Conti et. al [22], Enck et al. [27], and Ongtang et al. [46].

A fine-grained applications' permissions model that was proposed in the Apex system by Nauman et al. [43], gave users control over which data could be used by installed applications. This was achieved by rules that define scenarios and contexts when a particular data type could be accessed. Context was defined by the network a smartphone is connected to, e.g., WiFi network at work, at home or cellular network. This system, however, focuses on malware threats, and on inter process communications (IPC) in particular, and does not consider users as a source of threats.

Another system, the CRePE [22] added to the Apex functionality ability to block some applications from being run in specific contexts. For instance, if a smartphone is in roaming or hostile WiFi network prohibit running email clients or banking applications. Even though, this system improved security of applications in hostile network environment, it still focuses only on data leaks through network communications and does not provide protection against an adversary with physical access.

The Kirin system [27], proposed by Enck et al., introduced a notion of "dangerous" permission sets, which prevents application installation when requested permissions were "dangerous". If such system is installed and "dangerous" permission sets are defined, then users receives protections against applications which might abuse received permissions and leak data through network or IPC. This system, however, does not consider malicious users as a source of threat to data confidentiality and integrity.

The Saint system [46] extends existing Android's permission model in order to allow third party applications to assert permissions granted to its application program interface (API). Main goal of this system was to prevent abuse of third party applications' API by malware. The Saint system shares limitations of the Kirin system in the presence of the adversary with physical access.

Another signature based HIDS was proposed by Chaugule et al. [19], which was focused on malware. In particular, authors make assumption that any data access that originates from users, i.e., data read or write request that is a sequence of users opening a file for view or saving it, is benign, and, thus, should be excluded from being monitored. This is crucial point limitation for us, since we assume the opposite, not all users interactions are benign and an adversary will access data as a usual user, through user interface.

Data monitoring techniques, such dynamic data tainting, were also employed in several HIDS for smartphones. For example, the TaintDroid system, proposed by Enck et al. [26], tracks data leakage through IPC with the help of dynamic data tainting. Collected data routes then were used as an input for a rule-based security policy system, which detected whether a process is trying to get an unauthorized access to sensitive data. In case if such access attempt was detected policy enforcement point prohibited further access to such data for the process which requested access.

Another example of dynamic data tainting is similar systems, proposed by Ongtang et al. [45] and by Hornyack et al. [30]. The aforementioned systems focused on data being sent off the device in some contexts, such as being in a roaming or in unprotected WiFi. They are based on rule based system that define appropriate actions (e.g., prohibit the connection) for cases when sensitive data are being sent off the device. Such system, however, do not consider users as a source of threats to sensitive data, render themselves inefficient against an adversary who uses physical access to the mobile device.

Common limitation of all aforementioned systems is the fact that they target malware threats and do not consider users as a source of any threats. Most of the time they rely on authentication only, which, if used with weak authentication methods, could be easily overcome. Furthermore, it is not clear how effective such systems could be if applied to physical threats, and how feasible it is to codify malicious activities of a physical adversary, given that such codification is user specific. Finally, an inherited limitation of all signature detection based HIDS is ability to detect only known attacks. This relates to physical threats as follows, if a sensitive data item has not been considered as sensitive before (e.g., users have not thought about it yet) then such data would be left unprotected.

4.1.4 Anomaly based Intrusion Detection Systems

Anomaly detection based HIDS against physical threats have also received some attention from research community, although very limited. Such approach is usually based on building users behavior model, which is then used as a predictor of the next user's action. If a current actions has a lower probability than some threshold an intrusion is detected. In the case of an adversary who has physical access to the device the goal of such system would be to discriminate different users, when they are using the same device. By the same device we mean a smartphone that has the same set of applications, data items and configuration. This does not include such cases when a factory reset function is used in order to wipe out all data and applications on the device before such device is to another user for permanent (e.g., sold smartphone).

To the best of our knowledge, work done by Li et al. [34, 35, 36] is the the only example of anomaly based detection IDS application for smartphone. In their work Li et al. proposed to use users' behavioral models to tell different users apart. In their adversarial model they assume that the sole objective of the adversary is to get the mobile device as a valuable asset, thus, an adversary would use factory reset function and wipe out the stolen device. In order to evaluate how easy it is to discriminate such users authors conducted several laboratory experiments, where they evaluated different sets of features for the aforementioned purpose. In particular, they investigated use patterns of voice calling, SMS messaging, Bluetooth connection functionalities, as well as third party applications being used. The results of the experiments suggest that phone numbers which users had dialed and sent SMS messages to allows to discriminate users with equal error rates (ERR) up to 10-15%. On the other hand, applications being run on the smartphone did not show the same performance, and on average had 35% ERR. Author admit, though, that such bad performance was due to the fact that users tend to use the same set of applications, moreover, the data set which was used in their experiment was very limited. In particular, their data set did not have information about which data items were accessed during application run, in opposite, it only contained information when an application was launched, that did not allow authors to differentiate users which accessed only one data item (e.g., one email messages) apart from users that access multiple data items (e.g., all email messages).

Even though authors tried to address physical threats such as theft or loss, they had completely different objective. Main objective of this study was to unveil cases when a smartphone is stolen from service provide perspective, i.e., recognize that this smartphone was used by someone else before, on the basis of used services (phone calls, SMS messages). Authors also assume that an adversary is not interested the data stored on the smartphone, and, thus, would delete all data immediately. We assume the opposite, an adversary with physical access will try to access victim's data on the smartphone. It is safe to make such assumption since recent study by Symantec shows that in 96% of the cases when a smartphone is lost users that found it would try to access sensitive data [13]. Furthermore, our survey study revealed that more than 10% of smartphone users have tried to sneak into someone mobile device before. This assumption renders the approach proposed by Li et al. ineffective against an adversary with physical access.

Finally, some authors explored ways of using desktop level anti-malware systems and IDS in smartphones with the help of virtual machines. For example, Portokalidis et al. [49] proposed the Paranoid Android system, which used a complete replica of a smartphone on a dedicated server. For such system to work a smartphone would require to have a special mobile agent installed which records all call traces and passes them to the replica server. After replica server receives call traces updates it reproduces them on the smartphone replica by executing the same functions calls in the same sequence. This helps replica server to maintain a copy of the smartphone up to date, which is later used for detection of malware and data leakage with application of desktop level anti-malware software. Even though, the Paranoid Android system was tailored against malware, it could be relatively easily extended to include users interactions, so that malicious use could be detected on the server either by using Signature or Anomaly based detection. Such approach, however, has crucial limitation in the presence of the adversary who use physical access to the device. In particular, we assume that an adversary with physical access is able to circumvent all communication channels, which will render such approach ineffective.

4.2 Availability and Integrity

Another risk to data is data loss, which could be addressed by various data replication techniques, such as backup or synchronization. Existing solution could be divided into three groups (1) cloud storage based, (2) local storage based, and (3) mixed. Moreover, a support of versioning could be used to recover from data corruption, which could be supported in all three aforementioned groups.

Most popular system for synchronization and backup iCloud and DropBox [14, 24] have many advantages. For example, such systems have higher accessibility and availability than local based systems, e.g., home desktop or external hard drive. They also provide hardware protection, by storing several replicas of data for redundancy purposes where in the case of local solutions, users often use doomed to lose their data if hardware fails. Furthermore, some of them, e.g. the DropBox system, support file versioning, thus provide users with ability to recover from data corruption. Systems that are based solely on cloud storage, however, do not meet users' requirements, discussed in section 3.

Local storage based system, such as iOmega Network Storage System [11], or iTunes [2] are often bounded to home networks, i.e., render themselves unaccessible when users are not within home network. Even if they have support functionality of remote access, dynamic nature of IP addresses for consumer routers make them inefficient or very hard to configure and use for average users.

Research community also tried to address data loss risk by investigating different methods of data synchronization. For instance, Peek et al. [48] proposed the EnsemBlue system that facilitate integration of storage from multiple consumer electronic devices (CED) and provides tools for synchronization such CEDs. The system architecture is based on the central sever (a PC) that stores all data and distributes data to CEDs according to predefined query filters. Such query filters is used to define date items or types which need to be transfered to a CED, e.g., all pictures that are no more than one year old should be transfered to digital frame.

Salom et al. [52] proposed similar to EnsemBlue system which used decentralized storage system, so called the Perspective. It gave users ability to configure data redundancy and partitioning, in order to increase data storage reliability and improve accessibility on CEDs, such as Personal Video Recorders (PVR).

Another system, which was targeted at synchronization in databases (DB) was proposed by Ramasubramanian et al., so called the Cimbiosys system [51]. Even though authors considered data base types system, such system could be easily deployed for smartphones, considering that (1) data IO operations could be relatively easy converted to SQL queries, and (2) many data structures in smartphones are already implemented in the format of lightweight SQL DBs. In particular, the Cimbiosys system introduced the notion of synchronization rules which defined data eventual storage locations. It also used data versioning as a way to propagate data changes to the storage locations. Later, the Cimbiosys system was enhanced by Mahajan et al. [38], where authors added ability to recover from data corruption in an efficient way, i.e., with minimum IO operations needed. It remains an open question, however, how efficient and effective such system would be in smartphone, considering limited resources (CPU, and storage) which are needed for this system to compute hashes of versions and store several versions of the same data.

Systems which use both local and cloud based storage try to address limitations of the aforementioned systems. For example, the Wuala system [55] allows its users to share their space between each other. If a user shares his computer storage he/she will receive in return storage for their needs on someone else computer. Users are also able to buy additional space on the cloud storage and combine both types of storages in order to increase availability of the data. Somewhat similar the FriendStore system, which was limited to the network of friends on Facebook, was proposed by Tran et al. [54]. In this system friends from online social networks negotiated directly storage location share between each other. Data were encrypted before left owner computer.

All the aforementioned systems rely on backup schedule or on an explicit request from a user to perform backup or synchronization operation. Furthermore, non of the existing systems address data loss risk in such scenarios when data have not been backed up yet from a smartphone (reside only on a smartphone), but have been modified by an adversary with a malicious intent.

5. DISCUSSION

In what follows we discuss limitations of the existing solution and highlight possible further research.

5.1 Confidentiality and Integrity

Majority of the existing systems target malware threats and do not consider users as a source of any threats to the data in smartphone. In particular, access control systems rely on authentication methods as the way to detect illegitimate users. Unfortunately, existing systems do not consider usability of the authentication method and assume that users will adopt strong enough. However, results of the recent studies, including [41], suggest that users tend to adopt weak authentication methods due to limited user interface on the smartphones.

One way to reduce authentication burden on users is by decreasing frequency of authentication prompts. One way is to use timeout, which, however, introduce a very easy way to bypass locking system for an adversary. Another way to introduce fine grained access control system which would allow to lock only part of data items, applications and applications' functionalities. However, existing system allow only to lock at application level, which implies the same level of sensitivity for all data and functionalities with in an application. It is an open research question how a fine grained locking system could implemented so that it is secure, effective, efficient and usable. Moreover, a usable configuration interface has to be provided to such system, otherwise such system might deem itself to be abandoned by users.

It also not clear whether the only way of locking data is by authentication mechanism. We believe that other mechanisms, which could striker or softer could be used as well. For example, a deterrence mechanism of journaled access could be put in place, where each access to sensitive data is recorded. Moreover, such audit record could also include a photo taken from front facing camera, which is a standard in modern smartphones. This, however, requires study not only efficiency and effectiveness of such mechanisms, but also on study on their usability and whether users would like to adopt such tools.

Yet another perspective approach to data confidentiality protection in smartphone is to try to remove users from security loop as much as possible, i.e., automate most of the decisions. This could be accomplished by use of IDSes based on human behavior. It is an open research questions how well an average user can define what would constitute a malicious use of his smartphone for signature based approach and how effective such approach would be against an adversary with physical access. However, given that the number of such signatures, would be in order of several magnitudes smaller than the number of signatures for malware, this approach could be easily deployed on smartphones. For instance, recent malware DBs contains hundreds of thousands of malware signatures, where it is highly doubtful that users would define more than a thousand or ten thousands of signatures.

On the other hands, one can use anomaly based IDSes in order to detect when use of smartphone is not consistent with a normal use. The task of building and evaluating of such system, however, is very challenging. First of all, there is one very important difference between the problem of anomaly detection in user behavior and the problem of anomaly based detection of malware, that is a malware on one smartphone is a malware on the other smartphone. However, with user behavior it is not the case, because sensitivity of data is user specific. For instance, a data access pattern for one user could be identified as being malicious, and, in opposite, the same pattern would be identified as non malicious for another user. Second, one can easily capture malware and replay it in order to reproduce attack. Moreover, many repositories of different classes of malware are available for use. This is not the case with malicious users' activity, and to the best of our knowledge, no such data has been collected and made public. That is why it is an open research problems of how such system could be implemented, how effective and efficient such system could be and, most importantly, how such system should be evaluated.

5.2 Availability and Integrity

Most important limitation of the systems that are widely deployed today, which are based on cloud based storage, is privacy concerns users have with such storage. This however should be address by adoption of the local based or hybrid based solutions. However, such systems should use cloud storage carefully, providing full secrecy for the stored data.

Another aspect of data protection system is to provide means for users to recover from data corruptions. Even though, some of the systems support versioning of the data, e.g. DropBox, they do not provide such functionality for data which are stored only on the smartphone. Data could be stored only on the smartphone because of multiple reasons, e.g., have not been backup yet, user do not want to backup it because of sensitivity concerns.

Finally, because of the resource limitations in smartphones (network connectivity, batter power, CPU) all existing systems require either explicit users actions to start synchronization process or do it only in some specific contexts, such as home WiFi or when connected to power adapter. This could cause problems, since users might charge their phones outside of the home networks, thus, local solutions might be not accessible or they might forget to initiate such action. Furthermore, future research should focus on the designing of a system that would provide users with usable tools for controlling and managing their local storages where they would be able to backup data from theirs smartphones. Additionally, such system should overcome the limitation of existing systems, which are bounded to home networks.

6. CONCLUSION

In this paper we present a problem of data protection against physical threats, where an adversary would be able to get a physical access to the device, and would be able to overcome weak authentication methods, such as PIN-code or DAS. We also presented users' requirements to data protection systems, which were elicited in the interview and survey studies. In particular, we highlight what impede adoption of DPS.

We then reviewed existing systems in the light of the proposed adversarial model and users' requirements. Our findings suggest that existing systems are faraway from providing adequate protection against an adversary with physical access to the mobile device and futher research should be done in order to address the problem of data protection against physical threats.

For data confidentiality and integrity we propose further research in access control area where a new fine-grained system should be designed. Moreover, we suggest that other defense mechanisms should be studied, such as access journaling and decoys. We also proposed an approach where users participation in security loop is limited as an alternative way of detecting malicious data access, which could be based on either signature or anomaly detection based IDS. Finally, we highlight the necessity of investigating of a usable controlling interface for all the aforementioned systems.

For data availability and integrity we propose research direction that will investigate an efficient and effective design of a backup system which would be based on local storage and which would overcome the limitation of home network boundaries. Additionally, such system would provide efficient and effective tools for data corruption recovery for data which are stored only on smartphones.

7. REFERENCES

[1] Advanced task killer. `https://market.android.com/details?id=com.rechild.advancedtaskkiller`. last accessed February 4, 2012.

[2] Apple - itunes - everything you need to be entertained. http://www.apple.com/itunes/. last accessed February 4, 2012.

[3] Find my iphone. `http://itunes.apple.com/ca/app/find-my-iphone/id376101648`. last accessed February 4, 2012.

[4] Galaxy nexus android 4.0 face unlock broken by picture. `http://www.neowin.net/news/galaxy-nexus-android-40-face-unlock-broken-by-picture`. last accessed March 4, 2012.

[5] 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.

[6] 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.

[7] Smart applock pro. `https://market.android.com/details?id=com.thinkyeah.smartlock`. 2012.

[8] Ultimate app guard. `https://market.android.com/details?id=com.anttek.appguard`. last accessed February 4, 2012.

[9] 2010: Anual study: Global cost of a data breach. http://www.symantec.com/about/news/resources/press_kits/detail.jsp?pkid=ponemon, 2011.

[10] Mobility and security: Dazzling opportunities, profound challenges. `http://www.mcafee.com/mobilesecurityreport`, 2011.

[11] Nas network storage, online storage network devices and data backup software by iomega. http://iomega.com/nas/us-nas-comp.html, 2012.

[12] Processes and threads. `http://developer.android.com/guide/topics/fundamentals/processes-and-threads.html`, 2012.

[13] Webview. `http://www.symantec.com/about/news/resources/presskits/detail.jsp?pkid=symantec-smartphone-honey-stick-project`, 2012.

[14] Apple. icloud. https://www.icloud.com/, 2011.

[15] L. Banks. Mobile devices pose security dilemma for CIOs. http://www.cio.com.au/article/346474/mobile_devices_pose_security_dilemma_cios/, 2010.

[16] K. Barr, P. Bungale, S. Deasy, V. Gyuris, P. Hung, C. Newell, H. Tuch, and B. Zoppis. The vmware mobile virtualization platform: is that a hypervisor in your pocket? *SIGOPS Oper. Syst. Rev.*, 44:124–135, December 2010.

[17] A. Buchoux and N. Clarke. Deployment of keystroke analysis on a smartphone. In *Australian Information Security Management Conference*, 2008.

[18] S. Bugiel, L. Davi, A. Dmitrienko, S. Heuser, A.-R. Sadeghi, and B. Shastry. Practical and lightweight domain isolation on android. In *Proceedings of the 1st ACM workshop on Security and privacy in smartphones and mobile devices*, SPSM '11, pages 51–62, New York, NY, USA, 2011. ACM.

[19] A. Chaugule, Z. Xu, and S. Zhu. A specification based intrusion detection framework for mobile phones. In *Proceedings of the 9th international conference on Applied cryptography and network security*, ACNS'11, pages 19–37, Berlin, Heidelberg, 2011. Springer-Verlag.

[20] E. Chien. The motivations of recent android malware. http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/motivations_of_recent_android_malware.pdf, 2011.

[21] N. Clarke and S. Furnell. *Computers & Security*, 24(7):519 – 527, 2005.

[22] M. Conti, V. T. N. Nguyen, and B. Crispo. Crepe: context-related policy enforcement for android. In *Proceedings of the 13th international conference on Information security*, ISC'10, pages 331–345, Berlin, Heidelberg, 2011. Springer-Verlag.

[23] A. De Luca, M. Langheinrich, and H. Hussmann. Towards understanding atm security: a field study of real world atm use. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 16:1–16:10, New York, NY, USA, 2010. ACM.

[24] Dropbox Corporation. Sync your files online and across computers. http://www.getdropbox.com/, 2009.

[25] P. Ducklin. Dropbox lets anyone log in as anyone - so check your files now! `http://nakedsecurity.sophos.com/2011/06/21/dropbox-lets-anyone-log-in-as-anyone/`, 2011.

[26] 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*, OSDI'10, pages 1–6, Berkeley, CA, USA, 2010. USENIX Association.

[27] 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*, CCS '09, pages 235–245, New York, NY, USA, 2009. ACM.

[28] S. Furnell, N. Clarke, and S. Karatzouni. Beyond the pin: Enhancing user authentication for mobile devices. *Computer Fraud & Security*, 2008(8):12 – 17, 2008.

[29] M. Hamblen. Spending on chips for mobile devices outpaced computers in '11. http://www.computerworld.com/s/article/9223886/, 2011.

[30] P. Hornyack, S. Han, J. Jung, S. Schechter, and D. Wetherall. These aren't the droids you're looking for: retrofitting android to protect data from imperious applications. In *Proceedings of the 18th ACM conference on Computer and communications security*, CCS '11, pages 639–652, New York, NY, USA, 2011. ACM.

[31] I. Ion, N. Sachdeva, P. Kumaraguru, and S. Capkun. Home is Safer than the Clould! Privacy Concerns for Consumer Cloud Storage. In *Proceedings of Symposium on Usable Privacy and Security*, pages 1–20, Pittsburgh, PA, USA, July 2011.

[32] S. Kurkovsky and E. Syta. Digital natives and mobile phones: A survey of practices and attitudes about privacy and security. In *Technology and Society (ISTAS), 2010 IEEE International Symposium on*, pages 441 –449, june 2010.

[33] M. Landman. Managing smart phone security risks. In *2010 Information Security Curriculum Development Conference*, InfoSecCD '10, pages 145–155, New York, NY, USA, 2010. ACM.

[34] F. Li, N. Clarke, M. Papadaki, and P. Dowland. Behaviour profiling on mobile devices. *Emerging Security Technologies, International Conference on*, 0:77–82, 2010.

[35] F. Li, N. Clarke, M. Papadaki, and P. Dowland. Behaviour profiling for transparent authentication for mobile devices. In *Proceedings of the 10th European Conference on Information Warfare and Security (ECIW), Tallinn, Estonia*, ECIW '11, pages 307–314, 2011.

[36] F. Li, N. Clarke, M. Papadaki, and P. Dowland. Misuse detection for mobile devices using behaviour profiling. *International Journal of Cyber Warfare and Terrorism (IJCWT)*, 1(1):41–53, 2011.

[37] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan. uwave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive Mob. Comput.*, 5:657–675, December 2009.

[38] P. Mahajan, R. Kotla, C. C. Marshall, V. Ramasubramanian, T. L. Rodeheffer, D. B. Terry, and T. Wobber. Effective and efficient compromise recovery for weakly consistent replication. In *Proceedings of the 4th ACM European conference on Computer systems*, EuroSys '09, pages 131–144, New York, NY, USA, 2009. ACM.

[39] A. C. Morris, S. Jassims, H. Sellahewa, and J. Koreman. Multimodal person authentication on a smartphone under realistic conditions. In *SPIE Conference on Mobile Multimedia/Image Processing for Military and Security Applications*, 2006.

[40] M. Mulazzani, S. Schrittwieser, M. Leithner, M. Huber, and E. Weippl. Dark clouds on the horizon: using cloud storage as attack vector and online slack space. In *Proceedings of the 20th USENIX conference on Security*, SEC'11, pages 5–5, Berkeley, CA, USA, 2011. USENIX Association.

[41] I. Muslukhov, Y. Boshmaf, C. Kuo, J. Lester, and K. Beznosov. Understanding users' requirements for data protection in smartphones. In *ICDE Workshops, Washington DC, USA*. IEEE, 2012.

[42] M. Nauman and S. Khan. Design and implementation of a fine-grained resource usage model for the android platform. *Int. Arab J. Inf. Technol.*, 8(4):440–448, 2011.

[43] M. Nauman, S. Khan, and X. Zhang. Apex: extending android permission model and enforcement with user-defined runtime constraints. In *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, ASIACCS '10, pages 328–332, New York, NY, USA, 2010. ACM.

[44] X. Ni, Z. Yang, X. Bai, A. C. Champion, and D. Xuan. DiffUser: Differentiated user access control on smartphones. In *Mobile Adhoc and Sensor Systems, 2009. MASS '09. IEEE 6th International Conference on*, pages 1012 –1017, Oct 2009.

[45] M. Ongtang, K. Butler, and P. McDaniel. Porscha: policy oriented secure content handling in android. In *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, pages 221–230, New York, NY, USA, 2010. ACM.

[46] 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*, ACSAC '09, pages 340–349, Washington, DC, USA, 2009. IEEE Computer Society.

[47] A. Oulasvirta, S. Tamminen, V. Roto, and J. Kuorelahti. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile hci. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '05, pages 919–928, New York, NY, USA, 2005. ACM.

[48] D. Peek and J. Flinn. Ensemblue: integrating distributed storage and consumer electronics. In *Proceedings of the 7th symposium on Operating systems design and implementation*, OSDI '06, pages 219–232, Berkeley, CA, USA, 2006. USENIX Association.

[49] 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*, ACSAC '10, pages 347–356, New York, NY, USA, 2010. ACM.

[50] R. Raguram, A. M. White, D. Goswami, F. Monrose, and J.-M. Frahm. ispy: automatic reconstruction of typed input from compromising reflections. In *Proceedings of the 18th ACM conference on Computer and communications security*, CCS '11, pages 527–536, New York, NY, USA, 2011. ACM.

[51] V. Ramasubramanian, T. L. Rodeheffer, D. B. Terry, M. Walraed-Sullivan, T. Wobber, C. C. Marshall, and A. Vahdat. Cimbiosys: a platform for content-based partial replication. In *Proceedings of the 6th USENIX symposium on Networked systems design and implementation*, NSDI'09, pages 261–276, Berkeley, CA, USA, 2009. USENIX Association.

[52] B. Salmon, S. W. Schlosser, L. F. Cranor, and G. R. Ganger. Perspective: semantic data management for the home. In *Proccedings of the 7th conference on File and storage technologies*, pages 167–182, Berkeley, CA, USA, 2009. USENIX Association.

[53] A. Smith. Nearly half of american adults are smartphone owners. `http://pewinternet.org/Reports/2012/Smartphone-Update-2012.aspx`. last accessed March 5, 2011.

[54] S. Tran and M. Mohan. Security information management challenges and solutions. `http://www.ibm.com/developerworks/db2/library/techarticle/dm-0607tran/index.html`, 2006.

[55] Wuala - Secure Online Storage - Backup. Sync. Share. Access Everywhere. Sync your files online and across computers. http://www.wuala.com/, 2010.

[56] N. H. Zakaria, D. Griffiths, S. Brostoff, and J. Yan. Shoulder surfing defence for recall-based graphical passwords. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, SOUPS '11, pages 6:1–6:12, New York, NY, USA, 2011. ACM.

[57] V. Zakorzhevsky. Monthly malware statistics, march 2011. http://www.securelist.com/en/analysis/204792170/Monthly_Malware_Statistics_March_2011. last accessed August 18, 2011.