Locked Your Phone? Buy a New One? From Tales of Fallback Authentication on Smartphones to Actual Concepts

Alina Hang[1], **Alexander De Luca**[1, 2], **Emanuel von Zezschwitz**[1], **Manuel Demmler**[1], **Heinrich Hussmann**[1]
[1]Media Informatics Group, University of Munich (LMU), Munich, Germany
[2]DFKI GmbH, Saarbrücken, Germany
(alina.hang, alexander.de.luca, emanuel.von.zezschwitz, hussmann)@ifi.lmu.de,
demmler@cip.ifi.lmu.de

ABSTRACT
We describe three scenarios in which fallback authentication on smartphones can occur and evaluate their real-life occurrences in an online survey (n=244) and complementing interviews (n=12). The results provide first insights into frequencies, reasons, countermeasures taken and problems of lockout experiences. Overall, study participants were satisfied with current fallback schemes, but at the same time, fallback authentication was aggravated when special circumstances applied and thus, leave room for improvements. Based on this, we propose an alternative concept for fallback authentication that quizzes users about installed and not installed apps on their device. Authentication succeeds, when users identify a certain number of apps correctly. Our evaluation showed that the concept yields an overall accuracy of 95%.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Fallback Authentication; Smartphones; Survey; Apps

INTRODUCTION
Smartphone owners spend a lot of their interaction time with authentication tasks (e.g. [6, 15]). While it is indisputable that users make errors during these tasks and that users forget passwords [1], there is little knowledge about how often (if at all) these tasks fail so that complete lockouts occur. Hereby, the term complete lockout refers to situations in which a user's account or device becomes inaccessible, for example, due to consecutive failed authentication attempts.

MobileHCI'15, August 24–27, 2015, Copenhagen, Denmark.

ACM 978-1-4503-3652-9/15/08...$15.00.
DOI: http://dx.doi.org/10.1145/2785830.2785839

In order to gain first insights into this matter, we ran an online survey with 244 participants and 12 complementing interviews, covering different lockout scenarios as well as the circumstances in which fallback authentication becomes necessary, the countermeasures taken and the problems encountered during the fallback experiences.

The majority of smartphone owners was unaware of the possibility to protect their SIM card with a PIN or even mixed it up with other authentication schemes available on smartphones (e.g. PIN protected lock screens). But overall, our study participants were satisfied with current fallback schemes. Nonetheless, there were also aggravating circumstances that made fallback authentication difficult, for example, when users were out of their usual context or ignorant of the available options they have. Based on this, we explored an alternative concept for fallback authentication that is targeted at these kinds of users. The basic idea is to create security questions about a user's installed apps (i.e. whether a particular app is or is not installed on the respective device).

The contribution of this paper is threefold: It is the first paper to provide a detailed overview of fallback authentication on smartphones. Furthermore, it makes recommendations for the improvement of current and future solutions through design, and last but not least, it presents an exemplary implementation as well as evaluation of an alternative fallback scheme which can ease the act of re-authentication.

FALLBACK AUTHENTICATION ON SMARTPHONES
There are three possible scenarios in which fallback authentication may take place on smartphones.

SIM card lockout. SIM cards are used to identify the smartphone with the mobile network. Not all smartphones have a SIM card, but the ones that do, can be protected using a four-digit PIN. This protection is either activated automatically or must be activated using the smartphone's menu. Restarting the phone is an exemplary situation in which this PIN needs to be entered. In case users fail to provide the correct PIN within x trials, the SIM card gets blocked and a so-called personal unblocking code (PUC) is required. The PUC is usually sent to the user by mail or can be looked up online.

Lock screen lockout. In order to restrict access to their smartphones, users can select from a variety of authentication schemes, ranging from knowledge-based schemes to biometric approaches. In case they fail to authenticate multiple times in a row, the device usually gets disabled and users have to switch to alternative authentication schemes.

The availability of these schemes often depends on the operating system (e.g. Android, iOS, Blackberry OS or Windows Phone) and the type of authentication scheme used. For example, Android users who use Face Unlock for authentication have to provide a graphical pattern as fallback option. In case the pattern is forgotten users have the possibility to wait a certain amount of time or to conduct email-based reset using their Google account. In the last resort, when all means fail, the device can be restored to factory settings.

In turn, iOS users who authenticate with TouchID have to provide a four-digit PIN as an alternative. Similar to Android, they have to wait a certain amount of time when the PIN has been entered incorrectly multiple times in a row. The waiting times can be circumvented when connecting the device with Apple's iTunes software. However, this method for fallback is coupled with data loss.

Documentation on fallback authentication for the operating systems Blackberry OS or Windows Phone is rare. In order to regain access to their device, users have to reset their device (and thus, lose all their data). There is an exception for Blackberry users that have an enterprise account as they can contact their technical support to perform password reset.

Account lockout. Many web services provide app versions for their websites. Users can access their accounts either using the corresponding apps or visiting the websites in a mobile browser. Most of the times a combination of username and password is required to login. In case users forget their passwords, a variety of solutions are available for account recovery (e.g. email-based resets or security questions [3]). The latter approach comes in different variations and is known for its shortcomings [9, 13]. Thus, alternative solutions like preference-based authentication were proposed [8, 14].

ONLINE SURVEY DESIGN AND INTERVIEW DESIGN

We conducted an online survey using Amazon's Mechanical Turk (MTurk), a method that was proven to be effective as long as certain preventive measures (e.g. control questions) are taken [10]. The survey was further complemented by anecdotes from semi-structured interviews that was proven to be useful to back up formal research [2].

Online Survey

An MTurk task was created for the survey, which was open to U.S residents only and required participants to have a HIT approval rate of 90% or higher. The rate depicts the percentage of approved assignments for each worker. The higher the approval rate, the better.

As a prerequisite, participants had to own a smartphone. We checked this by asking users to visit an URL with their mobile browser to enter their worker ID. The website used the *Mobile Detect Library*[1] for device detection. The survey lasted about 15 minutes and Workers received US$ 1 for their effort. The usual rate for MTurkers is US$ 4 per hour.

The general structure of the survey was divided into four parts. The first part covered questions about demographics (e.g. gender, age, smartphone experience). The remaining parts were dedicated to the three previously described fallback scenarios. Each of them started with a brief introduction to the scenario, and was followed by the question if the participants use or have used the corresponding protective measure on their smartphones (e.g. *"Do you use a PIN code to protect your SIM card?"*). In case they affirmed the question, they were asked if they ever had experienced [*SIM card|lock screen|account*] lockout. This was then followed by more specific questions about their last fallback authentication experience, covering questions about the reasons for lockout, their emotional state and the taken countermeasures to solve the problem. In case the first or second question was negated, we asked questions like *"What do you do to not forget the PIN of your SIM card?"*. This was done to create equal amounts of questions for each branch to prevent participants answering untruthfully (i.e. always with *no*) to finish the survey faster.

Interviews

The interviews were based on a between-subjects design, meaning that each interviewee was assigned to one of the three lockout scenarios, depending on their previous lockout experiences (i.e. four interviewees per scenario). In order to participate, interviewees were required to own a smartphone and to have experienced at least one of the three lockout scenarios. The interviews lasted for about 30 minutes. Interviewees received 5€ gift vouchers for their participation.

The structure of the interviews was similar to the online survey, but allowed more flexibility when interesting aspects were mentioned as interviewees were encouraged to elaborate on their answers.

Participants and Interviewees

In order to avoid confusion, the remainder of this paper will use the term participants for those who took part in the online survey, while the term interviewee will refer to those who took part in the interviews.

The survey was open for a period of one week, during which we received 272 submissions. We removed 28 (10%) submissions from the final data set after checking the control questions and the provided answers (e.g. when participants did not own a smartphone). The remaining 244 MTurkers (104 female) were aged between 18-66 years (average: 32 years).

The majority of participants (191; 78%) was employed and came from diverse professional backgrounds (e.g. IT, Design, Healthcare, etc.). Twenty-nine (12%) were students and the remaining 24 (10%) were not employed. The majority of MTurkers owned an Android device (139; 57%), 99 owned an Apple device (41%) and 6 owned a Nokia device (2%).

[1]Mobile Detect. http://mobiledetect.net/ (last accessed:30/01/2015)

The 12 interviewees (6 female) were aged between 17-55 years (average: 27 years). They were mostly students with backgrounds like computer science, teaching or law. The other interviewees were employed (2) or still attending school (1). Ten of them owned an Apple iPhone, the remaining ones owned an Android device (1) or a Blackberry (1).

RESULTS

We used an inductive coding approach to analyze the open-ended questions. Hereby, two researchers worked independently from each other to develop the codes. They then met to create the final code plan, which was used by a third researcher to code the answers of the questions.

Table 1 provides an overview of the number of participants that have and have not experienced the three lockout scenarios. Details for each scenario are provided in the following. We first report the results from the online survey and then complement them with anecdotes from the interviews.

SIM Card Lockout

In general, 133 participants (55%) stated to have never used a PIN for their SIM card before. The main reason for not using the SIM PIN at all was ignorance. Seventy-four of 133 participants (56%) did not know about the possibility of protecting their SIM card or mixed it up with the PIN code often used on lock screens. Forty-five participants (34%) were unconcerned about the security on their phone as they do not see any serious threats for their SIM card. The remaining participants gave reasons like inconvenience or fear of lockout.

In turn, 111 participants (45%) use or have used a PIN for their SIM card. Only four of them (4%) stated to have experienced lockout. The reasons were forgetting the PIN, mistyping it, mixing it up with another PIN or involving another person (e.g. mother tried to access the phone). In order to regain access, participants had to enter the PUC or call the service provider who sent a new PUC. Another one failed completely and had to buy a new SIM card.

Similar reasons for lockout where given by our interviewees, who all were not capable of unblocking their SIM card without additional help. For example, one interviewee reported that he was on vacation when lockout occurred and thus, had to rely on his accompanying friend (to borrow her phone) and family members (to look up the PUC at his home). This procedure was coupled with long waiting times and similar complaints were made by the remaining interviewees who had to wait several days before the SIM card was unblocked with the help of the service provider.

Lock Screen Lockout

Altogether, 79 participants (32%) stated that they had never used any kind of security for their lock screen, while 165 participants (68%) stated to protect or to have protected their lock screen in the past. With respect to the authentication scheme, 106 (64%) used PIN-codes, 52 (32%) used graphical patterns, 13 (8%) used TouchID, 2 (1%) used other authentication schemes. Lock screen lockouts were mentioned by 56 participants (34%).

		SIM	Lockscreen	Account
Used	**Yes**	111 (45%)	165 (68%)	237 (97%)
	No	133 (55%)	79 (32%)	7 (3%)
Locked	**Yes**	4 (4%)	56 (34%)	75 (32%)
	No	107 (96%)	109 (66%)	162 (68%)

Table 1. Overview of the number of participants that use a PIN for their SIM card, protect their lock screen (e.g. with PIN) or use web accounts on their smartphones. The table also shows the number of participants that have experienced lockouts.

The main reason for the need of fallback authentication was mistyping the passwords. This was mentioned by 39 participants (70%) and happened, for example, when participants were in a hurry (stated by 14 participants; 25%) or inattentive (stated by 6 participants; 11%). Two participants (4%) mentioned that they were *"too drunk or otherwise intoxicated to correctly input the pattern"* (this refers to the password used for the Android unlock pattern authentication scheme). Eight participants mentioned that they had just recently changed their password, so that they accidentally entered the old one (14%). There were six participants (11%) who forgot their passwords momentarily. In three cases (5%) the lockout occurred due to a third person (e.g. child playing with the phone). Table 2 gives an overview of all lockout reasons and the countermeasures taken. Please note that the numbers do not add up to 100% since multiple reasons could be provided.

Almost all participants (55, 98%) had no problems during fallback authentication, but described the experiences as annoying. However, some of them (13, 23%) remarked that the experience gave them the feeling of security as the lockout was a proof that the security of their device works. Only in one case, major problems were reported (e.g. *"I couldn't unlock the phone because my daughter reset the code and couldn't remember it. Had to buy a new phone"*).

In order to regain access to their device, 48 participants (86%) had to wait a certain amount of time and then were able to reenter their PIN or password. The waiting times ranged from 30 seconds to 15 minutes. The participant who had to buy a new phone, reported a waiting time of several years. In general, waiting times were rated ambiguously. While 21 participants (38%) considered them as quick and easy, another 21 participants (38%) found them too long. These observations are supported by the comments from the interviews. For example, one interviewee told us that his friends had tried to access his phone multiple times to play him a prank and as a consequence, locked his device for up to 48 hours. Since the interviewee did not want to wait that long, he took the phone to a store to get the lock removed.

Account Lockout

Only seven participants (3%) have never used any apps or websites on their smartphones that require login, while 237 participants (97%) stated to use or have used such apps/websites. In this context, 75 participants (32%) had experiences with account lockout.

The main reason for this was forgetting the password. This was mentioned by 50 participants (67%). Another five participants (7%) noted that they had mixed up their passwords

from other accounts. Mistyped passwords were identified as a problem by 21 participants (28%), 6 (8%) mentioned device constraints (e.g. small keyboard and screen) and 5 participants (7%) named technical reasons (e.g. automatic account logout). One participant did not even know what happened.

Most participants had no problems during fallback authentication (73; 97%) and for 23 participants (31%) the availability of fallback schemes gave them, similar to lock screen lockouts, peace of mind (as their accounts seem safe against attackers). Nonetheless, 45 participants (60%) described the situation as annoying. In order to regain access to their device, 63 participants (84%) stated to have used common fallback schemes like email-based resets (13; 17%) or security questions (5; 17%). Five participants (7%) used password managers or written down notes to retrieve their credentials. Two participants (3%) remarked that they postponed the fallback until they had access to their computer. Table 3 lists all lockout reasons and the corresponding countermeasures.

These insights are supported by the interviews. Only one interviewee reported problems with account recovery while being on vacation. She tried to access her Facebook account on her phone, but failed though she was sure to have entered the correct passwords. Even trying it on other smartphones did not solve the problem. Only login on a desktop computer revealed that the problem could not be solved on a mobile.

Device and System Impact On Lockout Frequencies

We used Cramer's V to analyze the correlation strength between the lockout frequencies and the device type (e.g. Android, iOS, etc.) as well as the type of authentication scheme used (e.g. PIN, Pattern, etc.).

The first analysis tested whether the use of different smartphone models (i.e. *device type*) had an impact on the frequency of lockouts (e.g. did Android users get locked out more frequently than users of other operating systems?). The results showed that there was a weak relationship between device type and account lockouts (Cramer's V = 0.08). The relationship was weak for lock screen lockouts (Cramer's V = 0.16) and SIM card lockouts (Cramer's V = 0.21).

We further investigated the relationship between the different types of lock screens and the lockout frequencies. The correlation was weak as well (Cramer's V = 0.18).

DISCUSSION

The results suggest that fallback authentication happens infrequently. Nonetheless, the individual reports depict interesting aspects of current fallback schemes that should be improved through design. They are also helpful for those who aim at developing alternative fallback schemes to current solutions.

Is It Worth the Effort?

With the small number of users who had difficulties with current fallback schemes, it is reasonable to wonder why one should care for designing alternatives. Though the group of troubled users was a minority, these users are exactly the ones that need the most help and thus, should not be neglected. As researchers, we should always try to see if we can do any better than the status quo.

Lock screen lockout			
Reasons		*Countermeasures*	
Category	*Count*	*Category*	*Count*
Mistypes	39 (70%)	Wait	48 (86%)
Hurry	14 (25%)	Code	4 (7%)
Password Mixup	8 (14%)	Email	2 (4%)
Inattentiveness	6 (11%)	No Memory	2 (4%)
Forgotten PIN	6 (11%)	New Phone	1 (2%)
Ext. Factors	6 (11%)		
Techn. Constraint	3 (5%)		
Third Person	3 (5%)		
Intoxication	2 (4%)		
Habit	1 (2%)		

Table 2. Overview of the number of mentions by the 56 participants for the reasons for lock screen lockout and the countermeasures taken.

Account lockout			
Reasons		*Countermeasures*	
Category	*Count*	*Category*	*Count*
Forgotten PW	50 (67%)	Common FA	63 (84%)
Mistypes	21 (28%)	Email	13 (17%)
Dev. Constraint	6 (8%)	Questions	5 (7%)
Password Mixup	5 (7%)	Helpdesk	4 (5%)
Techn. Constraint	5 (7%)	Captcha	2 (3%)
Unknown	1 (1%)	Phone Code	1 (1%)
		Lookup	5 (7%)
		Postpone	2 (3%)

Table 3. Overview of the number of mentions by the 75 participants for the reasons for account lockout and the countermeasures taken for fallback authentication (FA).

Aggravating Circumstances

Though most participants were satisfied with current fallback schemes, special circumstances made fallback authentication on smartphones more difficult. These situations are mainly influenced by two factors: First, ignorance and second, the unavailability of alternative options. This indicates that when designing alternative fallback schemes, these two special target groups should be taken into consideration.

Lack of Communication

The results of the online survey showed that the majority of users were not aware of the availability of a SIM PIN and with respect to lock screen lockouts, many participants were not familiar with alternatives other than waiting times. This observation is critical, as users may be caught by surprise when lockout situations occur. For example, they may be prompted to provide things that they have never heard of or they may be helpless when the offered fallback scheme fails.

Thus, communication should be an essential part in the design process of fallback schemes. Based on our data, we see three potential stages during smartphone interaction where communication should take place. First, before the lockout even happens to prepare users for potential lockouts. This could happen, for example, when users set up their phone for the first time. Currently, PUCs are often sent to the user by mail without further explanation about when it is needed or how it is used. Second, when lockouts are about to happen. For example, participants of our study mentioned inat-

tentiveness or habitual inputs as reasons for lockouts. Interrupting users through warnings may prevent lockouts to happen. Third, communication should also take place when the actual lockout has occurred to inform users about the possibilities that they have and guide them through the fallback process. This would prevent some users from disposing their phone and buying a new one.

Lack of Alternative Options

Most problems were reported when users were out of their usual context (e.g. vacations) and thus, could not fulfill the needed requirements of current fallback schemes. This does not mean that current solutions should be replaced, but instead, we propose to complement them with additional alternatives, so that users can choose the option that is most suitable for their situation.

The availability of alternatives is also important to offer users a safety net in case some of the fallback options fail (e.g. unacceptable waiting times). This implies that fallback authentication should be considered as a chain of multiple fallback schemes that are selected based on the users current context. However, with each component in the chain, the system's vulnerability increases through its weakest link. Thus, the design of alternative schemes must ensure that it is at least as secure as the primary authentication scheme (e.g. PIN).

Annoyed, but Secure

The experience of lockouts gave users a feeling of security in two ways: a) they felt appeased that their data is not lost in case they forget their password or enter it incorrectly for too many times; b) they felt that their data is safe from potential adversaries who might try to hack their accounts/phone. Nonetheless, lockouts were also described as annoying and were often mentioned in conjunction with waiting times.

This should be taken into account when designing alternative fallback schemes. For example, by trying to engage users more actively in the overall process in the form of small security tasks. Though this might take longer than waiting, the perceived duration may be shorter. Similar observations have been made by previous work, where perceived duration was influenced by factors, such as likeability [16].

Third Parties

Though most reported fallback experiences were self-inflicted, some of them were also caused by third parties who were close to the user (e.g. family and friends). The lockouts were either caused unintentionally or with malicious intentions. In particular the latter reasons are critical and must be taken into consideration as previous work has shown that these kinds of attacks are very likely [11]. For the design of fallback schemes this means, that insider threats must be taken into account. For example by evaluating new systems with persons that have advanced knowledge about the user to ensure the system's security (e.g. [5]).

Desktop vs. Mobile

Fallback situations that may happen on desktop computers as well as on mobile devices (e.g. account lockouts) often use the same fallback schemes for both environments. On the one hand, this has the advantage that users are already familiar with the fallback schemes, on the other hand, these schemes do not take into account mobile factors that may impact the usability of mobile fallback authentication, such as the small form factor of mobile devices or the difficult text input.

Some participants reported to postpone fallback authentication until a desktop computer is available. This means that when designing fallback schemes, the device's limitations should be taken into account. This can either be done by adapting existing solutions or by offering alternative options that are specifically designed for mobile devices.

IMPLICATIONS FOR DESIGN

In the preceding discussion we made suggestions for improvements. The main design implications can be summarized as follows. Fallback authentication should be ...

... independent from third parties or additional tokens.

... immediate so that users can authenticate anytime and anywhere (e.g. when Internet access is not available).

... engaging users during the authentication process to reduce perceived authentication time.

... guiding users and showing them alternative options in case they fail authentication.

... evaluated with persons close to the user as they are very likely attackers and thus, a good security indicator.

... at least as secure as the primary scheme (e.g. PIN), since the security of a system is as strong as its weakest link.

EXEMPLARY CONCEPT

Taking these implications into account, we designed an alternative concept for fallback authentication on smartphones.

Although researchers have proposed alternative approaches for fallback authentication in the desktop environment that could be employed in the mobile world as well, they often suffer from outdated data (e.g. [7]) or memorability issues (e.g. [14]). Thus, Hang et al. explored the potential of dynamic approaches by using icon arrangements on home screens [4] or smartphone activities [5] to design dynamic security questions. While the former concept still needs much more refinement to be usable and secure, the latter showed very promising results. However, the authors noted that more app-related questions need to be researched to create a sufficiently secure system, since they only tested two types of app-related questions (e.g.*Which app did you use last week? or Which app did you install yesterday?*).

We tie in with this research to extend the design space for app-related question. We explore the ability of users to identify apps that are or are not installed on their devices for authentication. Since users are likely to possess more apps than they actually use [12], our approach may also overcome the problem of data availability that was often mentioned when users utilize very few apps on a regular basis or when the designs are limited to a smartphone's home screen [4, 5].

Our main assumption is that smartphone owners are better than potential adversaries in distinguishing between apps that they actually own and do not own. Since some apps may be easier to guess for adversaries, one correct answer is not sufficient to authenticate successfully (similar to [7]). Instead, the authentication procedure consists of multiple yes/no questions. Users are consecutively prompted whether a particular app is or is not installed on their device (see figure 1, left). Depending on the number of apps that users identify correctly, they are either authenticated or rejected.

STUDY PROTOTYPE

The study application was implemented for Android smartphones with Jelly Bean (version 4.1) or higher.

Device Apps

In order to retrieve a list of installed apps, the study application scans for all installed packages in the storage of the user's device. In addition to this, it filters apps with a system flag to exclude them from the list. This was done for two reasons. First, many system apps are (pre-) installed without the knowledge of the user. Second, there are hundreds of system apps per device that would overwhelm users during the study.

Library Apps

In addition to device apps, users were also quizzed about apps that were not installed on their device. For this, we wanted to create a diverse app library by selecting apps from different app categories to avoid that library apps are too similar (and maybe easier to be guessed correctly). Altogether, the library consisted of 49 apps from the Google Play Store. Thirty of them (15 paid and 15 free) were taken from the list of top apps by the Google Play Store (as of May 2014). We further included less popular apps by randomly selecting them from 19 different app categories (e.g. Sports, Tools or Education).

Question App

Altogether, users were quizzed about all their device apps plus a random number of library apps. This number was selected randomly and ranged from zero to the maximum number of library apps (after cleaning it from apps that were also available on the user's device). Please note that we quizzed all device apps for evaluation purposes. In a real world deployment, usually a subset of those apps would be used for security reasons.

For each question, an app icon and its name was shown to the user who then had to answer whether the app is or is not installed on their device (see figure 1, left). We opted for yes/no questions to increase the usability of our approach. Previous work has shown that the availability of multiple answer options can make it more difficult for users to answer a question [4]. Although there is a 50% chance that a question is answered correctly, we have to keep in mind that the chances are reduce with each additional question (see threat model for additional details).

Users did not see whether they answered a question correctly or not. Only after submitting the answer to the last question, an overview of their performance was revealed (see figure 1, right). Please note that we displayed the results for discussion purposes. In a real world deployment such a list must not be shown to prevent adversaries from learning the answers.

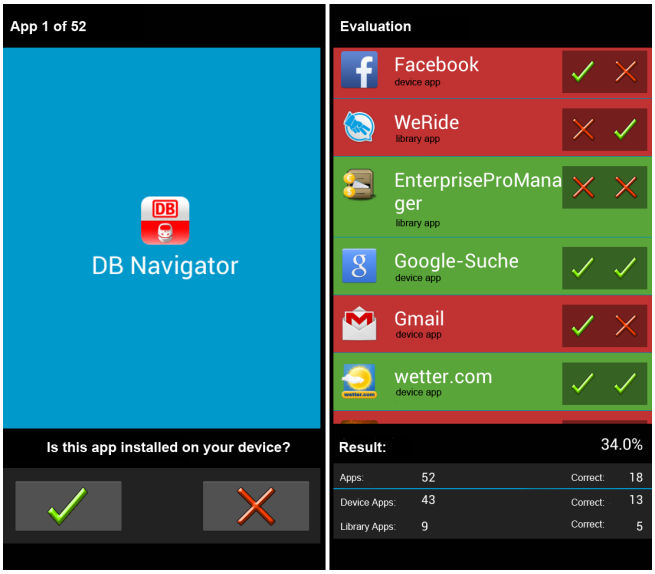

Figure 1. Screenshots of the study application. The left one shows an exemplary question that users were quizzed during the study. The right one is an overview of the performance of a participant during the study. Original language: German.

THREAT MODEL

To test the proposed concept with a worst-case scenario, we assume an adversary with advanced knowledge about the user. This way, guessing the correct answer is not plain luck anymore. Furthermore, these kinds of adversaries have shown to be very likely [11].

Since our approach is based on yes/no questions, there is a chance of $(\frac{1}{2})^x$ for a random adversary to guess the correct answers to a set of x questions, when equal answer distributions are assumed. In order to achieve a comparable security level to PIN authentication, where the chances are $\frac{1}{10000} = 0.0001$, we need to ask at least 13 questions as $(\frac{1}{2})^{13} = 0.0001$.

USER STUDY

We conducted a user study to test the feasibility of our approach in terms of usability (i.e. recognition of (not) installed apps) and security (i.e. insider threats) for fallback authentication on smartphones. We did not opt for a baseline comparison with similar work (e.g. [4, 5]) because our work is of exploratory nature. The main goal is to see if our concept could potentially work to complement existing app-related ideas (and not to replace them).

Study Design

All participants completed the same tasks. There were two types of participants: participants who owned a smartphone (that was used to install the study application) and participants who acted as adversaries. In order to avoid confusion, the remainder of this paper will use the terms participants for smartphone owners and adversaries for the other persons.

Participants were required to own a smartphone to take part in the study. Furthermore, they were required to bring a person that they are close with to act as adversary. During recruitment, we gave participants examples for close persons (e.g. partner, best friend). Though we told participants that the study is about fallback authentication, we did not reveal additional information to prevent them from learning the apps that they have installed on their devices.

Study Procedure

Participants and adversaries were invited to our lab for the experiment. After a brief introduction to fallback authentication, we asked adversaries to leave the room and wait outside.

In the meantime, the study application was installed on the participant's smartphone. Once installation was done, the actual study began. Participants were asked to complete the study task on their smartphones (i.e. to mark apps as installed or not installed on their device), which was followed by a demographic questionnaire and a post-task interview, in which we showed participants the results of their performance and a list of all apps quizzed during the task (see figure 1, right). For each app, we encouraged participants to comment on the strategies and problems they had during identification. The same procedure was repeated for adversaries.

The study lasted for about an hour. Participants and adversaries received 10€ gift vouchers each.

PARTICIPANTS

Altogether, 15 participants (5 female) and 15 adversaries (12 female) took part in the experiment. Participants were aged between 20 and 34 years (average: 24 years). Adversaries were aged between 21 and 49 years (average: 26 years).

Nine participants were students with a technical background, five were employed in different areas and one was a high school student. Similar demographics were found for adversaries. Nine of them were students, five were employed and one stated to be unemployed. The relationships between participants and adversaries were as follows. Seven brought their significant other, another seven brought a close friend and one brought a family member. There were no differences between the relationships mentioned by participants and adversaries.

Eleven participants reported that they had shared their device with the other person in the past (e.g. to make a call or to view photos). However, they also noted that sharing was limited to short time spans (between 2-30 minutes) and had happened altogether only between 1-4 times.

RESULTS

In total, participants had 646 apps installed on their devices, with an average of 43 apps per user (min=14; max=91).

Number of Correctly Identified Apps

An app was considered as correctly identified when a device app was marked as installed or when a library app was marked as not installed. In the following, we will report the quota of correctly identified apps and the number of quizzed apps (in percent). Table 4 gives an overview of the total number of correct answers provided by users and adversaries.

	Device	Library	Total
Users	95.4%	95%	95.2%
Adversaries	60.5%	82%	68 %

Table 4. Overview of the number of correctly identified apps (device apps, library apps and both) for users and adversaries.

In general, all participants performed better than their corresponding adversary. While participants identified 95% of apps correctly (min=89%, max=100%), adversaries reached only 69% (min=45%, max=82%). Participants answered 77% of device apps and 68% of library apps correctly. In turn, adversaries identified 61% of device apps and 82% of library apps correctly.

A mixed ANOVA with the between-groups factor *user type* (i.e. participant and adversary) and the within factor *app type* (i.e. device app and library app) showed significant differences in the performance between participants and adversaries ($F(1,28) = 66.41, p < 0.01$). However, no main effects were found for *app type*. There were also no interaction effects between *user type* and *app type*.

Time

We also measured the time participants and adversaries needed to provide the answer for an app. The measurement started as soon as the app icon appeared and ended with the submission of the answer. On average, participants needed 2.4s per app (min=1.8s; max=3.3s), while adversaries took on average 6.5s (min=5s; max=8.6s). A mixed ANOVA (using the time average of each participant) with the between-groups factor *user type* (i.e. participant and adversary) and the within factor *app type* (i.e. device apps and library apps) showed that participants were significantly faster than adversaries in identifying apps ($F(1,28) = 154.9, p < 0.01$). However, no differences were found for *app type*. Also, no interaction effects between *user type* and *app type* were found.

Strategies of Adversaries

During the post-task interviews, adversaries were encouraged to reveal the strategies they used to decide whether an app was or was not installed on the corresponding device.

Adversaries reported that the popularity of apps played an important role in decision making. In particular social apps, communication apps as well as local apps were considered as popular and thus, often assumed to be installed. This included Facebook, WhatsApp or apps for local public transport.

Another approach was to estimate how well an app would fit to the victim's characteristics and preferences. For example, one participant told us about her boyfriend's special taste in games. She marked all apps as installed that met this criterion. Other participants assumed that their victim would not be willing to pay for apps and thus, marked all apps labeled with the terms *premium*, *plus* or *full version* as not installed.

Adversaries also based their decisions on previous selections. For example, in case they had previously marked an app from a particular category as installed (e.g. weather apps), they were reluctant to do so again, when they were quizzed about similar apps (e.g. another weather app).

40 apps, 5 allowed errors		
Accuracy	Frequency	Frequency (in %)
79.3	1	0.1
82.8	3	0.3
86.2	36	3.6
89.7	123	12.3
93.1	283	28.3
96.6	369	36.9
100	185	18.5

45 apps, 6 allowed errors		
Accuracy	Frequency	Frequency (in %)
78.6	2	0.2
82.1	8	0.8
85.7	20	2
89.3	102	10.2
92.9	272	27.2
96.4	371	37.1
100	225	22.5

Table 5. Overview of the occurring frequency (after 1000 calculations) of accuracy values (in %) for 40 (left) and 45 (right) quizzed apps under consideration of 5 allowed errors.

Problems by Users

We asked participants about the problems that they encountered during decision making. Participants reported that, at times, it was difficult to distinguish between similar apps and thus, to decide whether the corresponding app was installed or not. For example, there exist multiple versions for the puzzle game 2048 from different companies. Furthermore, quizzing users about apps that they use infrequently made it difficult for them to decide whether an app was still installed on the device or whether it had already been removed.

Accuracy

So far, we only considered the number of apps that participants and adversaries were able to identify correctly. Based on these numbers, we can say that if we quizzed all apps on a device and required users to answer more than 82% of the questions correctly, all participants would be able to authenticate successfully, while all adversaries failed. Although this is an interesting anecdote, it is not sufficient to assess the security of the presented approach.

Thus, the following section will use accuracy calculations for a more thorough security analysis. Accuracy is a good indicator to describe how well a system works. It takes into account the number of succeeded and failed authentication attempts by legit users (i.e. true positives (TP) and false negatives (FN)) as well as the number of succeeded and failed attacks by adversaries (i.e. false positives (FP) and true negatives (TN)). The formula for calculation is as follows:

$$Accuracy = \frac{\sum TP + \sum TN}{\sum TP + \sum FP + \sum TN + \sum FN}$$

It returns a value between 0 and 1 (or 100 in %). While a value of 0 is bad (as all attacks succeed), a value of 1 (or 100) is very desirable (as all attacks fail and users always succeed).

In addition to this, we further consider two different parameters for accuracy calculation. First, the number of apps to be quizzed. Second, the number of errors allowed (i.e. the number of apps that have to be identified correctly in order to be authenticated successfully). This is important to estimate how strict/usable the concept is and how it could look like in a real world deployment. The parameters must be chosen so that authentication is as easy as possible for the user, but still secure enough against potential adversaries.

For each participant, we successively selected $x = \{15, 20, 25, 30, 35, 40, 45\}$ of their quizzed apps and calculated the accuracy values when $y = \{0, 1, 2, ...x\}$ errors were allowed. In theory, this should be done for all possible combinations with x apps. However, due to computation limitations, we had to follow another approach. Thus, we repeated the analyses a thousand times (each time with a different subset of x apps that was randomly selected) to compare the changes in accuracy values yielded. Please note that for $x = 40$ the attacks by one adversary could not be considered, as the corresponding adversary did not have enough attacks (i.e. was only quizzed 37 apps). For $x = 45$ only the data of 14 participants and 14 adversaries was considered.

The best accuracy values were reached when users are quizzed about 40 and 45 apps and when they were allowed to make at most 5/6 errors. Table 5 gives an overview of how often the different accuracy values occurred for 40 (left) and 45 (right) apps, when the analyses were repeated 1000 times.

With 40 quizzed apps and 5 allowed errors, we reached accuracy values between 79.3%-100%, with 96.6% accuracy appearing most often (369 out of 1000 times). In addition to this, we also calculated the overall accuracy, meaning that all 1000 calculations were considered as authentication attempts by users and adversaries (=$1000 * 15 + 1000 * 14 = 29000$ authentication attempts). The overall accuracy is 95%, with 747 FP and 722 FN. This means that 2.6% of all authentication attempts by legit users failed, while 2.5% of attacks succeeded.

Similar observations can be made for 45 quizzed apps and 6 allowed errors. We reached accuracy values between 78.6 % - 100%, with 96.4% accuracy appearing most often (371 out of 1000 times). The overall accuracy is 95%, with 824 FP and 529 FN, resulting in 3% of failed authentication attempts and 1.9% of succeeded attacks.

DISCUSSION

Limited App Usage

Some participants had only few apps installed on their devices. Three of them owned less than 20 apps. In a real world deployment, it is difficult to generate enough questions for them. To circumvent the lack of data, it is possible to fill the missing number of apps with library apps. However, adversaries who are familiar with their victim's smartphone habits (and thus know that the victim is not a heavy app user) may be tempted to mark all apps as not installed and, depending on the tolerance level for errors, may succeed with their attack.

Therefore, it is advisable to offer users an alternative for fallback authentication in case not enough data is available. This is already common practice in the Internet and should be adapted in the mobile world. For example, users that do not have access to their email account to reset their password, have sometimes the possibility to answer security questions instead. However, when designing alternatives, we must keep in mind that with each additional option, we add a potential point of failure that could be exploited by adversaries.

Authentication Time

On average, our users needed 2.4 seconds to answer one question. When quizzing 45 apps, this results in an overall authentication time of about two minutes. For a primary authentication scheme, this would be unacceptable. However, in the context of fallback authentication (which happens less frequently), this seems to be a reasonable time and is comparable to the time needed for current methods (e.g. waiting times for lock screen lockouts that increase exponentially).

Study participants were significantly faster in identifying their apps than adversaries so that one might think of adding a time limit for each question in future implementations. This would improve the security of the proposed approach, as adversaries are left with less time to think about or to research the answer to a question (for example, to research if an app is paid or free when they assume that their victim is not willing to pay for apps or whether the user has liked the app in social networks as an indicator that the app is actually installed).

However, it is also important to keep the user's situation in mind. Lockouts are most likely to be considered as annoying and stressful so that the time limits should not be too narrow to avoid causing more pressure on the user. Nonetheless, the security aspects, in the context of fallback authentication, should have a higher priority.

Popularity of Apps

Adversaries often assumed that popular apps (e.g. Facebook) were installed on the victim's device. This is a reasonable assumption and may have different implications.

A possibility is to filter these apps or to weigh the answers according to their popularity (e.g. less popular apps contribute more to the overall authentication score). Another option is to exploit this strategy to encourage adversaries to select the wrong answers, for example, by adding popular apps to the quiz that users do not have installed so that adversaries are likely to make the wrong decisions.

Similar Apps

Study participants had difficulties in distinguishing different versions of the same app (e.g. premium vs. free). Thus, we suggest to merge these apps so that it will be easier for users to identify them. This means that in case users own two versions of an app, they are considered as one. For example, instead of quizzing users about the apps *Weather Premium* or *Weather Free*, they should only be asked about the *Weather* app to avoid confusion.

The occurrence of similar apps was also used by adversaries for their guessing strategies. For example, in case they marked an app as installed, they were less likely to mark a similar app as installed as well. This is a reasonable strategy and should be exploited to improve the design of our approach. For example, apps from the library that are similar to the ones installed on the user's device should be quizzed first to increase the likelihood that adversaries mark actually installed apps as not installed. However, particular attention has to be paid that the randomness of the quiz is not hurt, to avoid adversaries learning how the system works.

Both observations mean that the selection of library apps needs to be done carefully. Apps should be similar enough to influence the selection of adversaries, but at the same time, need to be distinct enough to be differentiated by the user.

Library Apps

Our prototype used a static app library and was based on the latest data from the Google Play Store during the time of the study. However, in a real world deployment, such a library needs to be dynamic to keep the library up-to-date. For example, top apps and app icons change frequently and thus, need to be adapted accordingly to prevent adversaries from recognizing library apps during authentication. The library should also consist of more apps to offer a larger variety.

Security

The results showed that quizzing users about 40/45 apps with an error tolerance of 5/6 yielded the best accuracy values (about 95%). With these parameters, we also observed that there are app combinations that make users fail their authentication or allow adversaries to succeed in their attack.

With 40 questions, we reach a higher theoretical security level than PIN, which is often used for primary authentication ($(\frac{1}{2})^{40} = 0.9 * 10^{-12} < 0.0001$). This is important to prevent adversaries to circumvent the primary authentication by attacking the fallback solution to gain access to the device.

However, similar to PIN, the actual security level of our approach does not correspond to its theoretical security level. This means that the results are promising, but do not yet meet the requirements for an actual authentication system. Since the presented approach is an initial exploration, we think that implementing the suggested improvements can help to prevent adversaries from guessing the correct answers and thus, influences the accuracy values positively.

Privacy

Although the use of installed apps seems promising, it also has privacy implications that needs to be discussed. Hang et al. [5] found that users do not mind when other people see the apps they have. However, our approach does not only use installed apps, but also includes library apps.

As a consequence, adversaries may get the wrong impression of a user, when incriminating apps are shown that the user does not possess. This needs to be taken into account when designing the library. Furthermore, users should have the option to deactivate the fallback scheme when they do not feel comfortable in potentially sharing with others which apps they use.

LIMITATIONS

Even though the design and evaluation of the survey as well as the concept were carefully done, there are limitations in this exploratory approach which need to be addressed.

Since we focused on the general population of mobile device users, we were able to give a good overview on the number of lockout experiences. At the same time, we were not able to gather deeper insights into the diversity of such experiences. Even if we feel confident that the results of our survey already are a good motivation for the research on alternative fallback mechanisms, more detailed evaluations of lockout experiences are still needed. In addition, it has to be noted that our survey was based on self-reported MTurk data. Therefore, our findings are not necessarily representative to all smartphone users.

The presented concept is a good example for one possible design alternative for usable and universal fallback mechanisms on smartphones. However, this work is at an early stage. Therefore, we were not yet able to achieve sufficient security and potential privacy problems need to be further investigated. In addition, further analysis should compare the proposed mechanism to already deployed concepts.

Due to the exploratory character of the study, the analysis was based on a small sample of users. Most of our participants were young and tech-savvy. Nevertheless, we argue that this work is well suited to give relevant insights in this topic. It represents a first important step for the development of novel fallback authentication concepts for mobile devices.

CONCLUSION

In this paper, we presented the results of an online survey and 12 semi-structured interviews. It is one of the first to address the topic of fallback authentication in the wild and provides a general overview of fallback experiences on smartphones.

We found that fallback authentication happens infrequently and that most users are satisfied with current solutions. Nonetheless, there was a minority of users who could not cope with current schemes due to the lack of understanding or due to special circumstances. These users are an interesting target group that should be taken into account when improving existing schemes or designing new fallback schemes for fallback authentication. Thus, we proposed several design recommendations, including aspects like independency, immediacy or guidance.

Based on these insights, we explored the ability of users to distinguish apps that are and are not installed on their devices and tested the suitability of this idea for the design of a fallback authentication scheme for smartphones. The results were promising, but, with an overall accuracy of 95%, leaves room for improvements. This includes, for instance, that specific attacker strategies, like choosing popular apps, should be considered in the design.

ACKNOWLEDGMENTS

We want to thank Victoria Müller for her help in conducting the online survey and semi-structured interviews.

REFERENCES

1. Anne Adams and Martina Angela Sasse. 1999. Users Are Not the Enemy. *Commun. ACM* 42, 12 (1999), 40–46.
2. Murray W Enkin and Alejandro R Jadad. 1998. Using anecdotal information in evidence-based health care: Heresy or necessity? *Annals of Oncology* 9, 9 (1998), 963–966.
3. Steven Furnell. 2007. An assessment of website password practices. *Computers & Security* 26, 7 (2007), 445 – 451.
4. Alina Hang, Alexander De Luca, and Heinrich Hussmann. 2014. Using Icon Arrangement for Fallback Authentication on Smartphones. In *CHI EA '14*. ACM, 2467–2472.
5. A. Hang, A. De Luca, and H. Hussmann. 2015. I Know What You Did Last Week! Do You? Dynamic Security Questions for Fallback Authentication on Smartphones. In *Proc. CHI'2015*. ACM Press, 1383–1392.
6. Marian Harbach, Emanuel von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. 2014. It's a hard lock life: A field study of smartphone (un) locking behavior and risk perception. In *SOUPS'14*. USENIX Association, 213–230.
7. Markus Jakobsson, Erik Stolterman, Susanne Wetzel, and Liu Yang. 2008a. Love and Authentication. In *Proc. CHI'2008*. ACM Press, 197–200.
8. Markus Jakobsson, Liu Yang, and Susanne Wetzel. 2008b. Quantifying the Security of Preference-based Authentication. In *Proc. Workshop DIM'08*. ACM, 61–70.
9. Mike Just. 2004. Designing and Evaluating Challenge-Question Systems. *Security & Privacy* 2, 5 (2004), 32–39.
10. Patrick Gage Kelley. 2010. Conducting Usable Privacy & Security Studies with Amazon's Mechanical Turk. In *USER Workshop at SOUPS*. 11.
11. Ildar Muslukhov, Yazan Boshmaf, Cynthia Kuo, Jonathan Lester, and Konstantin Beznosov. 2013. Know Your Enemy: The Risk of Unauthorized Access in Smartphones by Insiders. In *Proc. MobileHCI'13*. ACM, 271–280.
12. Ahmad Rahmati, Clayton Shepard, Chad Tossell, Mian Dong, Zhen Wang, Lin Zhong, and Philip Kortum. 2011. Tales of 34 iPhone Users: How they change and why they are different. *arXiv preprint* (2011).
13. Stuart Schechter, A. J. Bernheim Brush, and Serge Egelman. 2009a. It's No Secret: Measuring the Security and Reliability of Authentication via 'Secret' Questions. In *Proc. SOUPS 2009*. ACM Press, Article 40.
14. Stuart Schechter, Serge Egelman, and Robert W. Reeder. 2009b. It's Not What You Know, but Who You Know: A Social Approach to Last-resort Authentication. In *Proc.CHI 2009*. ACM Press, 1983–1992.

15. Dirk Van Bruggen, Shu Liu, Mitch Kajzer, Aaron Striegel, Charles R. Crowell, and John D'Arcy. 2013. Modifying Smartphone User Locking Behavior. In *Proc. SOUPS'13*. ACM, Article 10, 14 pages.

16. Emanuel von Zezschwitz, Paul Dunphy, and Alexander De Luca. 2013. Patterns in the Wild: A Field Study of the Usability of Pattern and Pin-based Authentication on Mobile Devices. In *Proc. MobileHCI'13*. ACM, 261–270.