Where Have You Been? Using Location-Based Security Questions for Fallback Authentication

Alina Hang
Media Informatics Group
University of Munich (LMU)
Germany
alina.hang@ifi.lmu.de

Alexander De Luca
Media Informatics Group
University of Munich (LMU)
DFKI GmbH
Germany
alexander.de.luca@ifi.lmu.de

Matthew Smith
Usable Security & Privacy Lab
University of Bonn
Germany
smith@cs.uni-bonn.de

Michael Richter
Media Informatics Group
University of Munich (LMU)
Germany
michael.richter@campus.lmu.de

Heinrich Hussmann
Media Informatics Group
University of Munich (LMU)
Germany
hussmann@ifi.lmu.de

ABSTRACT

In this paper, we propose and evaluate the combination of location-based authentication with security questions as a more usable and secure fallback authentication scheme. A four weeks user study with an additional evaluation after six months was conducted to test the feasibility of the concept in the context of long-term fallback authentication. The results show that most users are able to recall the locations to their security questions within a distance of 30 meters, while potential adversaries are bad in guessing the answers even after performing Internet research. After four weeks, our approach yields an accuracy of 95% and reaches, after six months, a value of 92%. In both cases, none of the adversaries were able to attack users successfully.

1. INTRODUCTION

Passwords still have a prevalent role in today's world, where they are mostly used in combination with usernames to protect the users' accounts and data. However, the number of these accounts is steadily increasing, confronting users with the challenge to define distinct and secure passwords [1]. When users forget passwords, fallback authentication schemes are required to enable users to regain access to their account and data. While authentication schemes such as passwords have received a lot of attention in the usable security and privacy community, fallback authentication schemes have not seen the same amount of attention.

Most common approaches for fallback authentication rely on email-based password resets or security questions (e.g. [18]). In general, email-based password resets work well, but are not appropriate in all circumstances (i.e. when users forget the password to their email account). Therefore, security questions are often used as an alternative. They take advantage of personal information, assuming that such information are easily remembered by users and at the same time hard to guess by others. However, previous research has shown that the use of security questions comes with a variety of shortcomings with respect to usability and security (e.g. [10]).

To overcome these shortcomings, we propose location-based security questions as an alternative design. Our questions are similar to traditional security questions in the sense that they are based on personal information, but different as they focus on questions about episodic memories with a spatio-temporal context [17] (e.g. *"Where did your first kiss take place?"*) and thus, also differ in the way the answers are provided. Instead of entering them as text, which often comes with issues like repeatability [8], users submit their answers by selecting a location on a map.

Our hypothesis is that location-based questions are easier to recall as they are remembered more vivid than personal facts [17]. Furthermore, using maps for answer input can serve as helpful memory hooks for users to recall their answers to questions (e.g. street crossings, buildings, etc.). In order to test the usability and security of the proposed approach, we conducted a user study over a period of four weeks and evaluated three types of location-based questions: predefined, guided and open questions. All questions were tested with different types of adversaries: close adversaries (i.e. persons that know the user well) and strangers. We also performed an additional evaluation after six months to test the memorability of the presented approach.

The results of our study show that it is hard for persons close to the user as well as strangers to guess or even research the answer to a question. Social networks and search engines do not provide sufficient hints and even if they do, it is difficult to be close enough to the actual location (i.e. to be within a distance of 30 meters). In turn, users are very good in answering their questions. The accuracy values (95% after four weeks; and 91% after six months) of location-based questions are promising, but leave room for improvements with respect to the usability of the approach.

The main contributions of this paper are twofold: (1) we

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

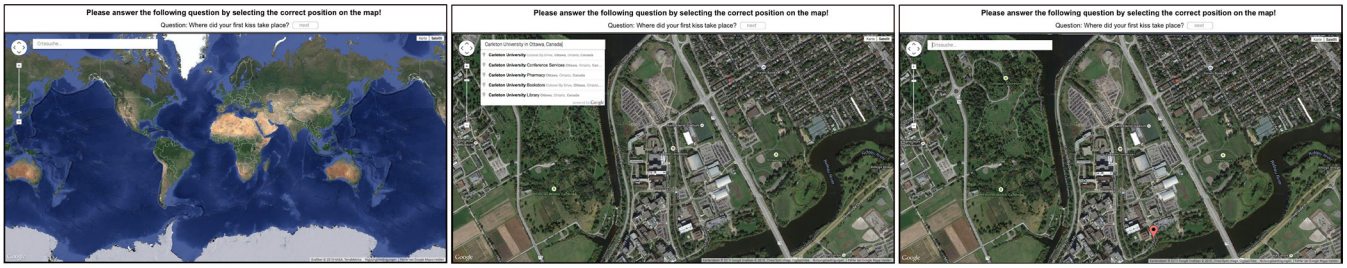

Figure 1: Screenshot of the prototype during authentication for the exemplary question *"Where did your first kiss take place?"*. The authentication always starts with the world map (left). Users then can zoom in on a map section by using the mouse, map controls or the provided text field (center), but the answer must be provided by selecting it on the map using the mouse (right). Translated from German.

present a detailed usability (i.e. memorability) and security analysis of location-based security questions over a period of six months to simulate fallback scenarios and (2) we provide reasons why location-based questions work better than traditional security questions and discuss the potentials and risks of different suggestions for further optimization.

2. RELATED WORK

Fallback authentication usually consists of two phases. In the first phase, the enrollment, users have to provide various information, such as email-addresses, phone numbers or answers to security questions. This information is needed for the second phase when password reset/retrieval is required (e.g. forgotten passwords). The time that elapses between those two phases can be very long.

A commonly used method for fallback authentication is the email-based password reset. In case of password loss, a new password or a reset link is sent to the user's email address. According to Simson Garfinkel [3], this approach works well, but comes with certain shortcomings. For example, it makes the email account a single point of failure. Furthermore, the email address that the user has provided during enrollment might be out of date and thus, not accessible anymore.

Some service providers offer users the possibility to associate their mobile phone number with their accounts. Upon request for password reset, a one-time password is sent to this number, with which the users can temporarily log into their account to define a new password. However, mobile phone numbers are sensitive information that not everyone wants to share with every service provider [5].

Another popular approach is the use of security questions. Such questions and their answers can either be fixed, controlled or open [7]. While fixed questions (i.e. predefined questions) leave little room for the user to make changes, users often lack creativity to handle open questions and thus, come up with questions that are similar to fixed ones. Controlled questions are a combination of both. For example, users can define hints for a question that will be shown when they have to answer the respective question. However, these hints will also be visible for potential adversaries.

Most service providers rely on fixed questions about personal facts (e.g. *"What is your mother's maiden name?"*). In the past, such questions were assumed to be easy to remember by the user and hard to answer by others. The reality is that questions that are easy to remember are often easy to guess, while questions that are hard to guess are also hard to remember [10]. Thus, security questions come with numerous insufficiencies in terms of usability and security. Inapplicability, memorability and ambiguity are one of the key issues worth mentioning with respect to usability [10]. In terms of security, many predefined security questions are researchable (e.g. [4]), can easily be answered by close persons like family and friends (e.g. [6]) or can even be guessed by choosing the most popular answers [12].

In order to overcome these insufficiencies, various alternative solutions have been proposed. For example, Schechter et al. [13] propose a system called social authentication. In case of password loss, users have to contact two or three contacts to retrieve tokens that are part of the authentication process. However, their studies also showed that after a certain time, users could not recall the names of the social contacts they had provided during enrollment.

Since memorability becomes an issue when the time between enrollment and fallback authentication increases, Babic et al. [2] propose a dynamic approach that uses security questions based on recent browser activities. Using implicit data seems promising, but may evoke privacy issues, as users have no power over which information is used.

In summary, it can be said that the design of security questions is a challenging task that in particular tackles issues like memorability and security. Most research so far has focused on the design on question level, neglecting the way the answer is provided.

3. CONCEPT

We suggest an alternative concept to traditional security questions to address their well-known shortcomings (e.g. memorability or repeatability [7, 10]). Our concept focuses on episodic memories with a spatio-temporal context [17] to generate location-based security questions. Psychological research has shown that these kinds of memories are easier to remember than, for example, personal facts due to their more vivid recall [17].

Although traditional security questions also include questions about locations, our concept is different as the answers are not provided as text, but instead, are entered by selecting a location on a map. The way of entering a location into the system is inspired by GeoPass [15]. However, our approach is not an extension of this existing approach, where an arbitrary location is used as a primary password, but instead, we present a novel alternative that combines security

Predefined Questions			
Whereto was your first travel by plane?	(5)	Where have you been camping for the first time?	(1)
Whereto was your longest travel so far?	(5)	Where was your first car accident?	(1)
Where is your favorite beach?	(3)	Where did you park for your driving test?	(1)
Where did your best friend from elementary school live?	(2)	Where did you injure yourself badly for the first time (e.g. broken leg)	(1)
Where was your first time at the sea?	(2)	Where did your best kindergarten friend live?	(0)
Where did you meet your best friend?	(2)	Where did you spend your first vacation?	(0)
Where did your first kiss take place??	(2)	Whereto did you drive in your first driving lesson?	(0)
Where have you been in a dangerous situation?	(2)	Where was your first party?	(0)
Where does a distant relative of yours live?	(1)	Where was your first breakup?	(0)
Whereto did you travel for your first school trip?	(1)	Where was your most embarrassing moment?	(0)
Where was your first job interview?	(1)	Where was your saddest moment?	(0)

Table 1: Overview of the 22 fixed questions used in the study. The values in brackets depict the number of times a question has been selected during the study. Translated from German.

Guided Questions	
Please define a location-based question that refers to a travel destination/vacation destination.	(7)
Please define a location-based question that refers to a personally experienced sport event.	(5)
Please define a location-based question that refers to an event in your childhood.	(4)
Please define a location-based question that refers to an event during your time at university/apprenticeship.	(4)
Please define a location-based question that refers to one of your party experiences.	(3)
Please define a location-based question that refers to something that you did for the first time.	(3)
Please define a location-based question that refers to an event that during your time in school.	(2)
Please define a location-based question that involves another person.	(1)
Please define a location-based question that refers to one of your favorite places.	(1)
Please define a location-based question that refers to an experience that had a strong impact on your life.	(0)

Table 2: Overview of the 10 guidelines for the guided questions used in the study. The values in brackets depict the number of times a category has been selected during the study. Translated from German.

questions with map-based input. This is an important difference, since we argue that map-based input is not a good option to replace passwords (e.g. due to long authentication times), but it is a good option to replace text-based answers.

The context in which location-based questions are supposed to be used (i.e. fallback authentication) represents another difference and thus, imposes harder requirements on the design and evaluation of location-based questions:

Fallback authentication happens less frequently than primary authentication (about once a month or less [14]) so that users should be able to recall the needed information even after longer periods of time. Therefore, it seems advisable to favor cued-based recall over free recall as previous research has shown the superiority of the former [16]. In this concept, we use questions as cues to trigger episodic memories that are associated with a particular location.

Furthermore, in order to authenticate, users have to answer a sequence of location-based questions on a map (instead of remembering an arbitrary location). This is required to reach a certain level of security.

To find the best trade-off between usability and security, we evaluate the concept in four sessions to simulate fallback authentication. We test the memorability of the concept shortly after enrollment as well as one week, three weeks and six months after the last authentication attempt. We evaluate the security of the approach and test it with different types of human adversaries. We further analyze the number of questions that users should answer in order to reach a certain level of security and discuss the implications when users exhaust the number of authentication attempts.

4. THREAT MODEL

We consider three different types of threats to evaluate the presented concept: a) threats by close adversaries, b) threats by close adversaries that use the Internet for researching the answers and, c) threats by strangers that also use the Internet for research to perform educated guesses.

Close adversaries (e.g. partners) have the advantage to know the user well and thus, do not have to rely on plain luck to guess the correct answers. The threat can be increased, when they use additional tools like social networks or search engines for research. This kind of threat can be considered as one of the worst case scenarios for location-based security questions.Threats by close adversaries were shown to be very likely and thus, interesting to consider [9].

The chances for a stranger to guess the correct answer (without any assisting tools for research) is $(\frac{1}{x})^n$, with x being the number of all possible locations on a world map and n depicting the number of questions asked. The answer space is narrowed down when more targeted attacks are considered (e.g. by limiting the answers to the country where the victim lives). Therefore, we also test the performance of adversaries that do not have any prior knowledge about the user (i.e. strangers), but use the Internet to take advantage of information on social networks, telephone directories or results from search engines to make educated guesses.

In the scope of location-based questions, brute force attacks have to be mentioned, where more sophisticated adversaries have the skills to use automated processes to attack the questions by successively guessing one location after another. In order to undermine these attacks, our concept

limits the number of attempts per question to three, which is a common threshold used for fallback authentication systems. A more detailed analysis of the number of attempts will be provided in the result section, while the implications of such a limit will be addressed in the discussion.

5. SECURITY QUESTIONS DESIGN

For the first design of the security questions, we performed a focus group with five participants (all male). They were recruited over bulletin boards, mailing lists and personal communication. Participants were aged between 18-26 years (average: 22 years) and were all students with a background in natural sciences (i.e. computer science, physics and medical engineering).

The participants were invited to our lab and were given a short introduction to fallback authentication and security questions. This was followed by a brief explanation of our concept. We asked participants to discuss advantages and disadvantages of the concept and encouraged them to discuss ideas for location-based security questions.

During the discussion, participants identified promising categories, including *childhood memories* (as these memories lie far in the past so that only few people know about it), *travel / vacation* (as these kinds of questions have a large answer space) and *first time memories* (as they are memorable). Participants also mentioned questions about *big events* (like concerts) or *third parties* (e.g childhood friends).

Since the identified categories are highly individual (not everyone has made similar experiences in the past), participants raised concerns about the applicability of predefined questions. Open questions were also considered as difficult, since users might define questions that are too easy to guess. Therefore, participants suggested to use something in-between those two extremes: guided questions which provide users with a basis to work on, but allow them to personalize the questions (e.g. *define a location-based question that refers to an event in your childhood)*. The concerns comply with the problems discussed in [7].

In our study, we used all three question types (predefined, guided and open) and compared them to each other. For each type (except for open questions), we used the insights from the focus group to design the location-based questions. Altogether, we ended up with 22 predefined questions (see table 1) and 10 guided questions (see table 2).

6. PROTOTYPE

The study application used the Google Maps API (in combination with HTML5 and JavaScript) to obtain location-based information and logged all relevant user interactions (e.g *timestamps*, *selected/defined questions*, *latitude*, *longitude*, etc.). It consisted of three main modes: enrollment, authentication and attack.

6.1 Enrollment

In the enrollment phase, users selected their questions and provided the corresponding answers. The way of enrollment varied for the different question types. For predefined questions users had to select three questions from a list of 22 questions. For guided questions users had to select three out of 10 guidelines from a list. In addition to this, three text fields were provided that allowed users to define a question based on each selected guideline. For open questions users were given three text fields and a brief instruction to define three location-based questions.

Once the questions had been selected/defined, they were consecutively shown to the users. Users were asked to provide the answers to the given questions by selecting a location on the map. Since it may be difficult for some users to find the right region on the world map, they had the possibility to enter an address into a given text field to zoom in on the corresponding map section. However, no position marker was set to ensure that users make their own selection by clicking on the map (see figure 1). This was done to make the selection more individual and thus, more difficult to be guessed. Users were allowed to reposition their marker. The answer was submitted by pressing the *save*-button.

6.2 Authentication

In authentication mode, users were presented with the questions they had selected/defined during enrollment. In order to authenticate, users had to provide the answers by selecting the locations on the map. Again, users had the possibility to enter the location into a text field to zoom in on a particular part of the world map, but a position marker had to be set by clicking on a location on the map. Users had three attempts to submit the correct answer. An answer was considered as correct, when the distance between the selected location and the location provided during enrollment was smaller than 30 meters. This threshold was shown to be useful by Thorpe et al. [15].

6.3 Attack

The only difference to the authentication mode was that the answers were provided by potential adversaries (close ones and strangers) instead of the legit user.

6.4 Map and Zoom Level

For each question, the map was initialized at zoom level 2 and was centered at the position 0.0 / 0.0 (latitude / longitude). Participants always saw the whole world map as a starting point. This was done to avoid influencing users during answer selection and helped to prevent hinting possible location areas for answers to potential adversaries.

In order to submit a location as an answer, users were required to obtain a zoom level that was higher than 16. This value was shown to be useful by Thorpe et al. [15]. In case the zoom level was too small, a pop-up notification informed users to zoom in.

7. USER STUDY

The user study consisted of a short-term evaluation of four weeks (with three sessions) and a long-term evaluation after six months.

7.1 Study Design

For the study design, we used a between-groups design with the independent variable *question type* (three levels: predefined, guided and open). A between-groups design was necessary to prevent biasing users during enrollment (e.g. preventing users to define similar questions to the ones that they encounter for predefined questions).

The prerequisite to participate in the study was to come in pairs and to have a close relationship with each other. We gave participants examples of close relationships during recruitment (e.g. partners, best friends). For each pair, the

participants took over different roles. One acted as legit user, while the other acted as close adversary who tried to attack the questions. In the remainder of this paper, we will refer to legit users shortly as *users* and to participants who attacked the questions as *close adversaries*. It was also possible for participants to take part in both roles, meaning that they acted as users as well as each other's close adversary.

As incentives, participants received gift vouchers of 20€ for users or 5€ for close adversaries. In case they acted in both roles, they received 25€. No incentive was provided when not all required sessions of the short-term evaluation were completed. Participants received additional 5€ gift vouchers when they took part in the long-term evaluation.

7.2 Study Procedure

The study was divided into three sessions and a long-term evaluation. For all sessions, participants were invited to our lab. While users had to attend all sessions for memorability testing, close adversaries only had to come for the first session. The long-term evaluation was conducted online.

7.2.1 First Session

The first session started with a brief introduction to fallback authentication (and security questions), the proposed concept and the study procedure. Then, users were assigned to one of the three groups (predefined, guided or open).

Close adversaries were asked to leave the room and wait, while users did the enrollment for their assigned question type (i.e. selecting/defining the corresponding questions and providing the corresponding answers on a map). Users were asked to select/define and answer three location-based questions. Once the enrollment was completed, we gave users a distraction video (duration about six minutes) after which a short-term memorability test was performed. Users were given three attempts to answer the questions they had just selected/defined. Users were informed whether a question was answered correctly/incorrectly.

For the attack, we asked users to leave the room and invited the close adversaries back in. Adversaries also had three attempts to guess the answers to the selected/defined security questions. For all questions that close adversaries did not answer correctly after three attempts, we gave them a second chance for attack, but this time, they were allowed to use the Internet for research.

In case close adversaries also wanted to participate as users, we paid particular attention that both users completed the enrollment and short-term memorability test first (one after another) before performing the actual attack to avoid influencing users during enrollment. Furthermore, both users were assigned to different groups.

At the end of the study, participants were asked to fill out a questionnaire to collect demographic information, qualitative ratings and also to ask participants to state the closeness of their relationship on a 5-point Likert scale to check their level of agreement.

On a separate form, we asked users to fill out the following information: first name, last name, date of birth and place of birth. This kind of information can usually be spied on (e.g. personal ID) or retrieved from public records. We used this information for educated guessing attacks in which adversaries, that did not know the users, tried to research the answers. We informed users about the purpose of collecting this information and told them that providing the information was optional. However, none of them refused. Furthermore, we paid particular attention that our research complies with the federal (privacy) laws in our country.

7.2.2 Second Session and Third Session

One week after the first meeting, we invited users back to perform another memorability test. Again, users had to answer their three questions from the first session within three attempts. Another memorability test was conducted in a third session that took place three weeks after the second one (i.e. four weeks after the first meeting). Users had to complete the same tasks as for the second session.

7.2.3 Long-Term Evaluation

Six months after the third session, we invited user to a long-term evaluation to simulate a realistic fallback scenario in which a long time between enrollment and required fallback authentication had passed. The procedure was similar to the second and third session, but was done online over Skype to spare users from long travel times to our lab (and thus, to encourage them to participate).

7.2.4 Educated Guessing Attacks

Two persons that were strangers to the user were asked to research the answers to the security questions. We provided them with the users' personal information (i.e. first name, last name, date of birth and place of birth) and a list of the security questions grouped by user. The adversaries had two weeks time to use this information for Internet research. For each question, they had to submit three possible locations and state briefly why they had selected a certain answer.

7.3 Participants

Thirty-two participants (15 female) took part in the user study. Twenty-eight of them acted as both, user and close adversary. Two participants acted as user only and another two acted as close adversary only. Participants were aged between 17-55 years (average: 26 years).

Four of them were high school students, 21 of them were students with different backgrounds (e.g. computer science, business or medicine), 5 were employed (e.g. administration or finance).

The relationships between users and close adversaries were manifold. The majority of pairs were good friends, best friends or partners/spouses. In four cases, the stated relationships did not match. For example, while one person described the other person as good/best friend, the assumed good/best friend stated to be only acquainted/good friends. However, the lines between good friend and good acquaintances or best friends and good friends are hard to draw. Altogether, there was a good agreement among pairs about their relationship. We also asked participants to rate how well they knew each other on a Likert scale ranging from not at all (1) to very well (5). Almost all pairs stated to know each other very well (8 pairs) or at least well (4 pairs). One pair stated to know each other a little. For three pairs, the ratings did not match. Two pairs had a mismatch between very well and well, while one pair had a mismatch between well and a little.

Two additional participants (one female) were recruited to perform educated guessing attacks. They were not related to the users or close adversaries from the user study and thus, strangers to them. They were 29/33 years old. Both of them

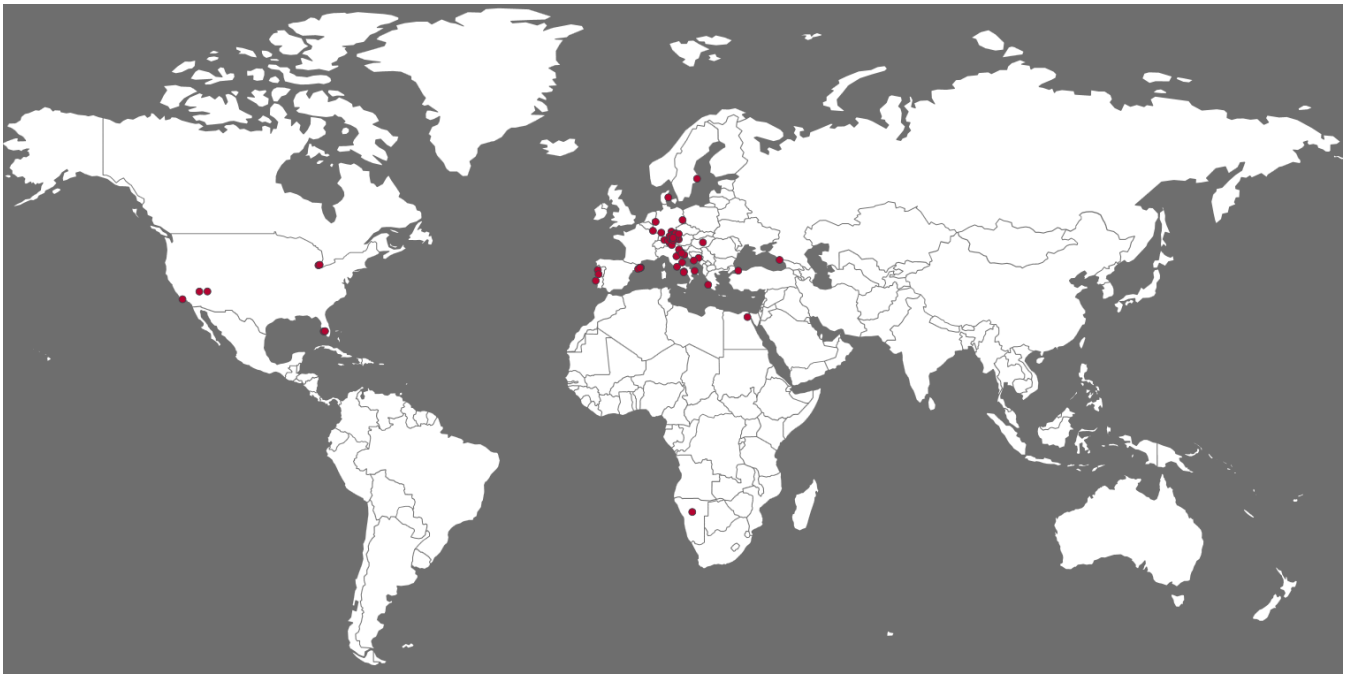

Figure 2: Overview of the distribution of answers provided by users during enrollment on the world map.

were security experts with experience in attacking security systems. They did not receive any incentives, but due to their background, they had a high intrinsic motivation to break the system.

With respect to the long-term evaluation, 24 out of 30 users took part in the experiment. Six users did not reply to our invitation to participate again.

8. RESULTS: SESSIONS 1,2 AND 3

Altogether, users defined or picked 90 location-based security questions (30 predefined questions, 30 guided questions and 30 open questions).

8.1 Question Types

8.1.1 Predefined Questions

Most predefined questions were about travel (e.g. first flight, or longest travel). This is followed by questions that involve a third person (e.g. best friend or first kiss). Table 1 gives an overview of how often each question was selected.

8.1.2 Guided Questions

With respect to guided questions, most users picked out guidelines for questions about travel, followed by questions about sport activities. Table 2 gives an overview of how often a guideline was picked. In terms of content, most sport activities related to first athletic experiences or special achievements (e.g.*"Where did I receive my first sports award?"*, *"Where did I run my first marathon?"*). Examples of the questions about travel are *"Where was your graduation trip?"* or *"Where did I spend my most beautiful summer when I was a child?"*. There was one question that referred to the future (i.e. *"A place where I want to be at least once."*), while all other questions were about the past.

8.1.3 Open Questions

The open questions that our users defined often involved a third person or animal (e.g. *"Where was my tomcat born?"*). They also included special events (e.g. *"Where did I celebrate the victory against Argentina in 2010?"*), first times (e.g. *"Where did your first kiss take place?"*, travel (e.g. *"In which country did I get homesick?"*, education (e.g. *"Where was my final exam?"*) or preferences (e.g. *"Where can I eat my favorite food?"*).

8.2 Amount of Information in a Question

The amount of information that one needs to know to answer a question varied from question to question. For example, the question *"Where is the center of the route to my best childhood friend?"* assumes the knowledge of five pieces of information: Who is the childhood friend? Where does the childhood friend live? Where does the user live? Which route did the user take to his friend (there are probably multiple routes possible)? Where is the center of this route?

All questions require at least the knowledge of one piece of information (i.e. the location of the question). This was the case for 77% of the open and guided questions. Ten questions (17%) required two pieces of information (e.g. the involvement of a third person). Two questions (3%) required three pieces of information, while the remaining two questions (3%) required four or more pieces of information.

8.3 Number of Correct Answers

Users submitted their answers in three sessions (and during the long-term evaluation, but the corresponding results will be reported in another section). Adversaries (close ones and strangers) submitted their answers only in the first session as memorability testing was not relevant for them. Figure 2 gives an overview of the locations that users selected during enrollment. Interestingly, most answers were clus-

	S1	S2	S3
3 Correct Answers	21	20	19
2 Correct Answers	8	9	8
1 Correct Answers	1	1	3
0 Correct Answers	0	0	0
Total	30	30	30

	Questions 1			Questions 2			Questions 3		
	S1	S2	S3	S1	S2	S3	S1	S2	S3
3 Attempts	1	0	1	1	0	0	0	0	1
2 Attempts	0	3	1	1	3	1	1	2	1
1 Attempt	26	25	23	25	24	24	25	25	21
Fail	3	2	5	3	3	5	4	3	7
Total	30	30	30	30	30	30	30	30	30

Table 3: The table overviews for all three sessions (S1, S2 and S3) the number of users that had three, two or one of the three questions correct (left). It also shows for each question and session the number of users that needed one, two or three attempts as well as the number of users that failed for the corresponding question (right).

tered in a geographical area. Table 3 and 4 give an overview of the number of correct answers as well as the number of needed attempts for each question. Note that an answer was considered as correct, when its distance to the actual solution did not exceed a threshold of 30 meters.

8.3.1 Users

In the first session, users answered 80 questions (89%) correctly and failed for 10 questions (11%). Most users provided their answers within one attempt and the majority of users had all of their three questions correct. There were users who failed for some questions (one user failed in two questions; eight users failed in one question), but none of them failed completely.

In most cases, the reason for providing the incorrect answer was precision, meaning that users were close to the correct location, but failed to be within the required threshold (i.e. 30 meters). Only in two cases, the distances to the original locations were over 1000 meters. Those users stated to have forgotten the answer or to have used a location that they did not associate strong memories with.

After one week users recalled 79 of the 90 questions (88%) and failed only for 11 questions (12%). Again, most users needed only one attempt to provide the correct answers. Most of them were able to answer all of their three questions correctly. There is no user who failed in all three questions, but nine users had one incorrect answer and another user had two incorrect answers. Incorrect answers were mostly made by users (nine out of ten) who already failed to provide the correct answers to the same questions in the first session.

Four weeks after the first session, users were able to answer altogether 76 of the 90 questions (84%) correctly. Fourteen questions (16%) were answered incorrectly. The majority of users needed only one attempt to answer a question. No user failed in all three questions and the majority had all questions correct. Again, most users who had difficulties to provide the correct answers in the first and second session, could not provide the correct answers for the same questions in the third session. However, two users who submitted incorrect answers in the first and second session, managed to answer correctly in the third session.

8.3.2 Close Adversaries

Close adversaries answered 6 of the 90 questions (7%) correctly, meaning that they failed to provide the correct answers at most times (93%). No close adversary had more than one correct answer within the set of three questions. The number of attempts needed differed from adversary to adversary. One needed three attempts, one needed two attempts and four needed one attempt.

		Question Type			
	S	**P**	**G**	**O**	**Total**
User	**1**	26	27	27	80
	2	26	26	27	79
	3	27	22	27	76
Close Adversary	**1**	1	3	2	6
Close Adversary (R)	**1**	2	3	3	8
Stranger	**1**	1	2	0	3

Table 4: Overview of the number of correct answers by users, close adversaries, close adversaries with research (R) and strangers for each session (S) and for the three question types: predefined (P), guided (G) and open (O). Adversaries only provided answers in the first session.

Exemplary questions that close adversaries answered correctly are: *"In which street did my grandma live?"* (guessed by spouse) or *"In which building was my first lecture?"* (guessed by university friend). Close adversaries were allowed to research the answers to questions that they had previously answered incorrectly. Only two adversaries succeeded. Each of them found the answer to one question. They needed one and two attempts, respectively.

8.3.3 Strangers

Two strangers tried to research the answers to the questions. Both of them failed most of the time. One of them was able to research the answers to two questions, while the other one succeeded only for one question. They needed one to two attempts. None of them had more than one correct answer within the set of three questions of a user. The questions that they attacked successfully were *"Where was the first time I partied when I was a student?"*; *"Where did I meet my best friend?"* and *"In which building was my first lecture?"* .

The reasons why adversaries selected a particular location are all based on different assumptions. For example, for the last question, the corresponding adversary was assuming that the user was a student at the department he was working at. Thus, he selected common university buildings where students usually have classes. For the question about the best friend, the adversary assumed that the user had met the victim in high school and thus, selected various school buildings in the home town of the user.

Despite the availability of social networks and search engines, both adversaries stated that it was very difficult to research the answers.

8.3.4 Comparison between Users, Adversaries and Question Types

For each session, we conducted a two-way ANOVA to examine the effects of *question type* and *user type* (i.e. user, close adversary and stranger) on the number of correctly answered questions. Simple main effects analysis showed that users were significantly better in answering questions than adversaries for all three session (each $p < 0.01$). No significant effects between the different types of adversaries were found. We also did not find any interaction effects.

8.4 Answer Distances

For the following calculations, we took, for each question, the shortest distance (of all attempts per participant) to the actual solution. This was done to analyze how often users and adversaries were how far from the correct answers. While most answers by users (for all three session) were close to the original location, most answers by adversaries were far from the original location, meaning that it was difficult to guess the correct answer.

8.4.1 Users

The answers of eighty questions from the first session were within a range of 30 meters to the original location (and thus, were answered correctly). For the remaining answers, we found the following distances: Seven had a distance of 40-100 meters; two had a distance of 300-600 meters and one answer had a distance of several kilometers.

Similar observations were made for the second session. Seventy-nine questions were answered correctly and thus in the range of 30 meters. The distances of the incorrect answers from the original location were 40-100 meters for four questions, 100-400 meters for five questions and over one kilometer for two questions.

For the third session, 76 answers were within a distance of 30 meters. In turn, nine answers had a distance between 40-100 meters, two had a distance between 200-600 meters and three had a distance of several kilometers.

8.4.2 Close Adversaries and Strangers

Most close adversaries provided answers that were very far from the actual location. Seventy-two answers had a distance of multiple kilometers (average: 440.4 kilometers). Twelve answers had a distance of 200-900 meters, while six answers were within a distance of 30 meters.

The distance distributions for strangers were similar. The distances were multiple kilometers (average: 1176.7 kilometers) for 167 of the 180 questions (since two adversaries attacked 90 questions each). Nine answers had a distance of 300-800 meters, one answer had a distance of 50 meters and three answers were within 30 meters to the original location.

8.5 Authentication Time

The time measurement for enrollment/authentication started when users pressed the start-button to open the corresponding HTML-page and ended with the submission of the last answer to the last question. On average, users needed four minutes for enrollment. The fastest user needed 1 min 15 s, while the longest enrollment lasted 7 min. In the first session, users needed on average 36 s for authentication (min=12 s; max=214 s). For the second and third session, they needed on average 45 s (min=13 s; max=225 s) and 47 s (min=13 s; max=232 s), respectively.

8.6 Accuracy

Accuracy is a good indicator on how well a system works in terms of usability and security. It takes into account the number of true positives (TP), true negatives (TN), false negatives (FN) and false positives (FP). TP refers to the number of successful authentications by legit users, while TN refers to the number of failed attacks by adversaries. In turn, FN represents the number of unsuccessful authentication attempts by legit users, while FP depicts the number of successful attacks by adversaries. The formula can be described as follows:

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

The formula returns a value between 0 and 1 (100 in percent). A value of 0 means that all authentication attempts by users fail, while all attacks by adversaries succeed. A value of 1 means the opposite and is a desirable result. It should be noted that accuracy values should always be interpreted in combination with the number of FP and FN.

For each session we calculated the accuracy values using a distance threshold of 30 meters. We also took into account two different parameters: a) the number of correct answers that are required in order to authenticate successfully [1..3] and b) the number of maximum attempts that one has to answer a question [1..3]. An overview of the calculations can be found in the appendix A.1 - A.3.

8.6.1 Close Adversaries

For the first session, the best accuracy values, when considering attacks by close adversaries only, are yielded when two correct answers were required and when there were two or three attempts allowed per question. These combinations result in an accuracy value of 98.3% (0FP, 1FN). The accuracy value remains the same after one week (i.e. second session), when allowing up to three attempts to submit the answer to a question. Restricting the number of attempts to two, decreases the accuracy to 96.7% (0FP, 2FN).

In the third session, the accuracy values decrease to 95% (0FP, 3FN) for the combinations of two required answers and two/three attempts for each question. However, still none of the close adversaries succeed in their attack.

The best combinations, the corresponding accuracy values as well as the number of FP and FN remain the same for all three sessions, when allowing close adversaries to research the answers to questions.

8.6.2 Strangers

The best accuracy values, when taking into account attacks by strangers only, are found for two required answers and two or three attempts. These combinations yield an accuracy of 98.3 % (0FP, 1FN). The accuracy value remains stable after one week, when three attempts are allowed. Allowing only two attempts, increases the number of FN and decreases the accuracy to 96.7% (2FN). In the third session, the accuracy value is 95% (0FP, 3FN) for a combination of two required questions and two/three attempts.

Based on the increasing number of FN after the third session, for both types of considered adversaries, we had a closer look at those three users and the distances to the actual solution for the questions that they answered incorrectly. The distances of the first users were between 120 m and 250 m. They were <40 m for the second user and <65 m for the third user.

Increasing the distance threshold to over 250 m would result in more FP and thus, is not reasonable. Also, choosing a distance threshold of 65 m would increase the number of FP to one when attacks by strangers are considered and thus, is not appropriate as well. In turn, using a distance threshold of 40 m does not increase the number of FP and reduces the number of FN by one. This results in an accuracy value of 96.7% (0FP, 2FN) for the third session and a combination of two required answers and two/three attempts.

8.7 Perceived Memorability

During the first session we asked users if they think that they could recall the answers to their question after a longer period of time. Users affirmed this for 78 of 90 questions (87%). They did not agree in three cases (3%) and were neutral for nine of the questions (10%).

During the second and third session, we then asked them to state how well they could recall the answers. For the second session, they stated to have no problems at all for 73 of 90 questions (81%). For 6 questions (7%) they had to think for some time before recalling the answer, and for 11 questions (12%) they had forgotten the answer.

Similar results were found for the third session. For 71 of 90 questions (79%) they had no problems at all. For 8 questions (9%) they had to think about the questions for some time. They had no idea for 11 of the questions (12%).

8.8 Perceived Security

In the first session, users were asked to rate the security of their questions with respect to different types of adversaries. Users provided their ratings on a 5-point Likert scale from strongly disagree (1) to strongly agree (5).

When asked, if they think that their questions are guessable or researchable by close adversaries, the opinions were not clear. For 34 of 90 questions (38%), users thought that their answers were not guessable, while others thought for 35 questions (39%) that they were. The remaining users were neutral.

In terms of researchability, users thought for 48 questions (53%) that their questions were not researchable by close adversaries. Other users did not share this opinion and believed for 27 questions (30%) that the answers could be researched. The remaining users had a neutral position.

For almost all questions (99%) users did not believe that they could be guessed by strangers. They also thought for 85 questions (94%) that they were not researchable by strangers.

8.9 Perceived Ease of Guessing/Researching

During the first session we asked close adversaries to state whether they knew or guessed the answers to the questions. For 46 of 90 questions (51%), adversaries had to guess the answer. Some adversaries had speculations for 25 questions (28%), but only in one case the correct answer was provided. There were some adversaries who thought to know the approximate location for 13 questions (14%), but only in one case the answer was correct. For the remaining six questions (7%), the adversaries were sure about the question's answer and thus, all but two submitted the correct answers.

Interestingly, some close adversaries were sure about their answers as they were part of the actual memory. For example, when they had the same favorite vacation destination as their spouse they tried to attack.

Close adversaries who did not manage to guess the correct answer, were asked if they felt like knowing the answer after they were allowed to research the question. Even after research, the adversaries did not know the answer for 45 of 84 questions (54%). For all these questions the incorrect answer was provided. Some adversaries stated to have had some kind of feeling where the answer might be for 26 questions (31%). Despite their feeling, all but one of these questions were answered incorrectly. Other adversaries thought to know the approximate location after research for 11 questions (13%), but had no correct answers. Only few adversaries were sure about the answers of two questions (2%). All these questions were answered correctly. The adversaries who succeeded in researching the answers stated that they had found the location on social networks.

8.10 Rating of System

In general, users liked the presented concept in terms of time consumption, memorability and security. The majority (21 users) found that location-based security questions are not too time consuming. They also felt that it is more secure than traditional security questions (27 users) as well as more memorable (24 users). All users stated that they would use location-based security questions for their accounts. Fourteen users for all of their accounts, 14 users for their important accounts and 2 users at least for their unimportant accounts.

9. RESULTS: LONG-TERM EVALUATION

9.1 Number of Correct Answers

Six months after the last session, users answered 55 out of 72 questions (76%) correctly. Seventeen questions (24%) were answered incorrectly. The number of attempts needed varied among users. Most of them needed one attempt (11 users for the first and third question, 14 users for the second question), while others needed two attempts (5 for the first question, 7 for the second questions and 2 for the third question) or one attempt (one user for the third question). None of the users failed in all three questions. Eleven users had all questions correct, nine had two correct answers and four had one correct answer. Most incorrect answers were caused by imprecise selections where users were close to the original location, but not within the required threshold. Another reason was that users had forgotten their answers.

Similar to the previous sections, we calculated the answer distances for each question by using the shortest distance of all attempts. Most answers were within a distance of 30 meters to the actual location. However, 17 questions were answered incorrectly. The answer distances were as follows: Five answers had a distance between 40-100 meters, five questions had a distance between 100-700 meters and two answers had a distance of multiple kilometers.

9.2 Accuracy

Accuracy calculation was done as explained previously. An overview of the accuracy calculations can be found in the appendix B. When considering attacks of close adversaries only, the best combination yields an accuracy value of 91.7% (0FP, 4FN) and requires users to answer at least two answers correctly and gives them three/two attempts per question to provide the answer. The same values and parameters work best when only attacks by strangers are considered.

9.3 Perceived Memorability

Users were asked to state how well they were in recalling the answers to their questions. For most of the questions (49 out of 72), users had no problems at all. For ten questions they had to think some time before the answer was recalled, while they had forgotten the answers to 13 questions. The self-assessment complies with the actual performance of the users. Only in four cases, users claimed to have recalled the answers, but gave an incorrect answer instead. Analyzing the distances of the corresponding answers showed that those users were close to the original answers (between 80-213 meters), but not within the required threshold.

9.4 User Feedback

In general, users felt positive about the presented system and found it more usable than traditional security questions. Thus, they would consider using location-based questions in a real-world deployment. However, one of the main concerns that users raised was the precision with which an answer had to be provided. This criticism was not related to the threshold of 30 meters, but instead, the knowledge that such a narrow threshold is given. Several users noted that they would have paid more attention during enrollment if they had known that such a threshold was given.

With respect to the ease with which an answer could be recalled during the different sessions, most users stated to have no major difficulties. They also told us that the ease of recall did not change over time, meaning that answers that they found easy were easy to recall over all three sessions and the other way around. These statements comply with the observations we have made.

Further interesting remarks were made by two participants who were caught by surprise as the map section that they needed had been updated since their last authentication attempt. As a consequence, some orientation points were lost (e.g. buildings) so that they had to think some time before the answer could be provided.

10. DISCUSSION

10.1 Question Type

In our user study we tested three question types: predefined, guided and open. The analysis did not reveal any significant differences, which may be due to the small number of participants per group. Nonetheless, guided questions appear to be the most promising ones of the three.

The lack of guidance for open questions may lead users to define weaker (but not meaning weak) questions than for the other types. For example, two users defined the question *"Where is my mother born?"* which reminds of the common security question *"What is your mother's maiden name"*. This kind of information could be researched through public records, providing hints to potential adversaries (though it is still difficult to select the location within a given distance threshold). In turn, predefined questions do not leave room for users to adapt the questions to their personal needs and thus, they miss the opportunity to phrase a more memorable question. Hence, the use of guided questions seems to be a good trade-off between the two extremes.

10.2 Topics of Question

The topics covered for the different question types (i.e. predefined, guided and open) were similar and ranged from travel, third persons to special activities. The topics were close to the ones that users like to choose for traditional security questions (i.e. preferences and questions about family members) [10]. However, in terms of guessability by close persons, our approach yields much better results (9%) than traditional security questions (38%; e.g. [6]).

More interestingly, allowing users to phrase their own location-based questions (i.e. in case of guided or open questions) gives the questions a more personal notion which is mirrored in the amount of information that is required to answer a question (e.g. *"Where is the center of the route to my best childhood friend?"*). Requiring more information makes it probably even more difficult for adversaries to guess or research the answers.

Thus, when designing location-based security questions, one could think of extending the set of guidelines by encouraging users to create more complex questions. For example, asking them to define a question that involves the center of two locations. However, it is also essential not to limit users too much in phrasing their questions to ensure applicability of the selected guidelines. This is important to avoid users phrasing questions that meet the restrictions of the guidelines, but may not be memorable.

10.3 User Performance

Users in our study were very confident that they will be able to recall their answers after a longer period of time and had a good estimation about their future performance. This is encouraging, since a positive and realistic attitude toward a system will motivate users to pay some effort when defining location-based security questions. A contrary example are commonly used security question that most users are not willing to spend time answering, since they think that they will not remember the answers anyway.

With respect to recall, the majority of our users were good in answering their security questions and only few forgot the answer to a question (even after six months). This shows that our approach works very well in terms of memorability.

10.4 Adversary Performance

Our approach showed promising results in terms of security as the adversaries in our study performed badly and could only attack few of the questions. This was mainly because the answers to the security questions were difficult to research and thus, forced adversaries to guess the answers at most times. In particular, strangers had problems during research. Even close adversaries who thought to have found clues during research failed to provide the correct answers at most times. They either drew the wrong conclusions from their research or were close to the location, but not within the required distance threshold. In comparison to the analysis by Ariel Rabkin [10] where 12% of the security questions sample could be attacked through research, our close adversaries and strangers could only succeed in 2% of the cases.

The biggest threats come from adversaries that share the same or similar experiences. For example, when close adversaries and users have traveled together to the location the question is referring to or when the user and adversary (close one or stranger) have attended the same course of studies in the same city. Most of the questions that the strangers guessed successfully, would not have been possible if the corresponding adversary had not been in the same situation in the past and thus, had some advanced knowledge.

10.5 Answer Precision

Most errors were made in terms of precision, meaning that users were close the the actual location, but not within the required range. However, the problem was, in the majority of cases, not caused by memorability reasons or the strict threshold, but by the assumption of the users that the system was more tolerant of imprecise selections. During the interviews, users were confident that if they had known about this requirement, they would have had less difficulties during authentication.

This means that when designing location-based systems for fallback authentication, it is important to inform users about the required precision to reduce the number of false negatives. Since most precision errors were already done shortly after enrollment (and then repeated in other authentication sessions), one could think of improving the enrollment procedure. For example, the system could ask users to re-enter the location to a question when a marker has been set to verify their answer. This approach is similar to the verification of passwords during registration.

10.6 Answer Distances

An answer was considered as correct when it was within a distance of 30 meters to the actual location. This threshold worked well to distinguish between users and adversaries. Despite the fact that most answers were clustered within a geographical region centered around the user's hometown, most adversaries were not able to guess the answer. They were hundreds of kilometers away from the actual location, supporting the assumption that most of them probably just selected random locations. Even in cases where adversaries stated to know the approximate location, they failed most of the time. This shows that while some users know the region in which an answer has to be in, it is still very hard to know which location within this region the user has selected. Thus, our approach has a very good answer space entropy. In turn, traditional security questions often have a very limited answer space (smaller than 25) [13].

10.7 Perceived Security

The perceived security of users strongly depend on the type of adversary. While they think that most strangers will not be able to guess or research their questions, their opinion is not as clear for close adversaries. In general, users seem to consider close adversaries as more likely to know an answer to a question than strangers. If a close adversary is considered as harmful, probably depends on how often they interact with the user and how much information this user is willing to share with friends in general. In comparison to the actual performance of adversaries, there is no difference in the number of answers they are able to answer, thus our approach works equally well against both types of adversaries.

10.8 Accuracy

To analyze the interplay between usability and security, we calculated the accuracy values for our approach. The best combination requires users to answer at least two out of three questions correctly, allowing them three/two attempts per question. This combination yielded an accuracy value of 91.7% with 4 FN and 0 FP after six months (with an increase of only 1FN in comparison to the third session).

This means that in terms of usability, our approach yielded good values, but leaves room for improvement, since still a few users were not able to authenticate under these conditions. In a real-world deployment, one would have to provide these users an alternative for the fallback authentication. This approach is commonly used for web services where users can select from a set of different fallback authentication schemes.

In terms of security, our evaluation obtained a very desirable result, since no adversaries (close ones as well as strangers) were able to attack successfully. However, we must also take into account that the number of attacks we were able to consider in this paper was limited.

Based on the usability and security insights, it would be interesting to study, if increasing the number of attempts decreases the number of FN, while maintaining the number of FP and if these improvements are resistant to a larger number of attacks, since increasing the number of attempts also means to give adversaries more opportunities for guessing the correct answers. As the answer space of location-based security question is huge, we assume that slightly increasing the number of attempts does not have a big impact on the actual security. However, these questions need to be addressed in the future.

10.9 Limitations

The participants of our focus group were all male which could have had an influence on the identified topics for the design of the questions. Literature on gender differences for autobiographical memories are ambiguous. While some assume no differences, others find woman to have more vivid and precise memories. If the latter is the case, we only have a lower bound for location-based questions, which, however, is good for general applicability.

Although a larger study sample would have been desirable, we opted for long-term participants (opposed to many participants for a one time lab-session), since we believe that this allowed us to get better insights into the potentials and shortcomings of location-based questions. In addition to this, the majority of participants were quite young with diverse backgrounds, but mostly students. Thus, it would be interesting to evaluate the concept with a larger and older sample of participants, since younger and older people remember different types of episodic memories [11]. Therefore, we encourage further studies with a more diverse sample.

Though we were able to re-invite the majority of participants, the dropout rate after six months needs to be mentioned as another limitation.

11. CONCLUSION

In this paper, we proposed the use of location-based security questions as a new approach for fallback authentication, and as an alternative to open text-based security questions that are known for their usability and security issues. We presented the design, implementation and evaluation of this approach and tested the location-based security questions under the worst circumstances. The results reported in this paper highlight the potential of the presented approach.

While users are good in recalling the location-answers to their questions, adversaries (close ones as well as strangers) failed most of the time when attacking these questions. Furthermore, the problems reported by our users are helpful guidelines to be considered when designing location-based

questions for a real-world deployment. Since the accuracy values as well as the number of false positives and false negatives are promising, we believe that the presented approach has the potential to replace commonly used security questions in the future and thus, encourage further research in this area to optimize the questions and the overall parameters for deployment.

12. REFERENCES

[1] A. Adams and M. A. Sasse. Users are not the enemy. *Commun. ACM*, 42(12):40–46, Dec. 1999.

[2] A. Babic, H. Xiong, D. Yao, and L. Iftode. Building robust authentication systems with activity-based personal questions. In *Proc. SafeConfig 2009*, pages 19–24. ACM Press, 2009.

[3] S. L. Garfinkel. Email-based identification and authentication: An alternative to pki? *IEEE Security & Privacy*, 1(6):20–26, 2003.

[4] V. Griffith and M. Jakobsson. Messin' with texas deriving mother's maiden names using public records. In *Proc. ACNS 2014*, pages 91–103. Springer, 2005.

[5] R. Gross and A. Acquisti. Information revelation and privacy in online social networks. In *Proc. WPES '05*, pages 71–80. ACM Press, 2005.

[6] W. J. Haga and M. Zviran. Question-and-answer passwords: An empirical evaluation. *Information Systems*, 16(3):335 – 343, 1991.

[7] M. Just. Designing and evaluating challenge-question systems. *IEEE Security & Privacy*, 2(5):32–39, 2004.

[8] M. Just and D. Aspinall. Personal choice and challenge questions: A security and usability assessment. In *Proc. SOUPS 2009*, pages 8:1–8:11. ACM Press, 2009.

[9] I. Muslukhov, Y. Boshmaf, C. Kuo, J. Lester, and K. Beznosov. Know your enemy: The risk of unauthorized access in smartphones by insiders. In *Proc. MobileHCI 2013*, pages 271–280. ACM Press, 2013.

[10] A. Rabkin. Personal knowledge questions for fallback authentication: Security questions in the era of facebook. In *Proc. SOUPS 2008*, pages 13–23, New York, NY, USA, 2008. ACM Press.

[11] H. L. Roediger and E. J. Marsh. Episodic and autobiographical memory. In A. F. Healy and R. W. Proctor, editors, *Handbook of psychology (Volume 4: Experimental Psychology)*, pages 475–497. John Wiley and Sons, 2003.

[12] S. Schechter, A. J. B. Brush, and S. Egelman. It's no secret: Measuring the security and reliability of authentication via 'secret' questions. In *Proc. SOUPS 2009*, pages 40:1–40:1. ACM Press, 2009.

[13] S. Schechter, S. Egelman, and R. W. Reeder. It's not what you know, but who you know: A social approach to last-resort authentication. In *Proc.CHI 2009*, pages 1983–1992. ACM Press, 2009.

[14] E. Stobert and R. Biddle. The password life cycle: User behaviour in managing passwords. In *Proc. SOUPS 2014*, pages 243–255. USENIX, 2014.

[15] J. Thorpe, B. MacRae, and A. Salehi-Abari. Usability and security evaluation of geopass: A geographic location-password scheme. In *Proc. SOUPS 2013*, pages 14:1–14:14. ACM Press, 2013.

[16] E. Tulving. Availablity versus accessibility of information in memory for words. *Verbal Learing and Verbal Behavior*, 5:381–391, 1966.

[17] E. Tulving. Episodic and semantic memory. In *Organization of Memory*, pages 381–402. Academic Press, 1972.

[18] Yahoo! Help. Recovering a lost or forgotten password. `https://help.yahoo.com/kb/recovering-lost-forgotten-password-sln2047.html` (Accessed: 02/09/2014).

APPENDIX

A. ACCURACY VALUES: SESSION 1, 2 AND 3

A.1 Session 1

A.1.1 Close Adversaries

Table 5: Overview of the accuracy values (A) for the first session. The calculation uses a distance threshold of 30 and takes into account the attacks by close adversaries as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	21	19	16	29	29	28	30	30	30
TN	30	30	30	30	30	30	24	25	26
FP	0	0	0	0	0	0	6	5	4
FN	9	11	14	1	1	2	0	0	0
Accuracy	85,0%	81,7%	76,7%	98,3%	98,3%	96,7%	90,0%	91,7%	93,3%

A.1.2 Stranger

Table 6: Overview of the accuracy values (A) for the first session. The calculation uses a distance threshold of 30 and takes into account the attacks by strangers as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	21	19	16	29	29	28	30	30	30
TN	30	30	30	30	30	30	27	27	29
FP	0	0	0	0	0	0	3	3	1
FN	9	11	14	1	1	2	0	0	0
Accuracy	85,0%	81,7%	76,7%	98,3%	98,3%	96,7%	95,0%	95,0%	98.3%

A.2 Session 2

A.2.1 Close Adversaries

Table 7: Overview of the accuracy values (A) for the second session. The calculation uses a distance threshold of 30 and takes into account the attacks by close adversaries as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	20	20	16	29	28	27	30	30	30
TN	30	30	30	30	30	30	24	25	26
FP	0	0	0	0	0	0	6	5	4
FN	10	10	14	1	2	3	0	0	0
Accuracy	83,3%	83,3%	76,7%	98,3%	96,7%	95,0%	90,0%	91,7%	93,3%

A.2.2 Stranger

Table 8: Overview of the accuracy values (A) for the second session. The calculation uses a distance threshold of 30 and takes into account the attacks by strangers as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	20	20	16	29	28	27	30	30	30
TN	30	30	30	30	30	30	27	27	29
FP	0	0	0	0	0	0	3	3	1
FN	10	10	14	1	2	3	0	0	0
Accuracy	83,3%	83,3%	76,7%	98,3%	96,7%	95,0%	95,0%	95,0%	98,3%

A.3 Session 3

A.3.1 Close Adversaries

Table 9: Overview of the accuracy values (A) for the third session. The calculation uses a distance threshold of 30 and takes into account the attacks by close adversaries as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	19	18	16	27	27	25	30	30	30
TN	30	30	30	30	30	30	24	25	26
FP	0	0	0	0	0	0	6	5	4
FN	11	12	14	3	3	5	0	0	0
Accuracy	81,67%	80,00%	76,67%	95,00%	95,00%	91,67%	90,00%	91,67%	93,33%

A.3.2 Stranger

Table 10: Overview of the accuracy values (A) for the third session. The calculation uses a distance threshold of 30 and takes into account the attacks by strangers as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	19	18	16	27	27	25	30	30	30
TN	30	30	30	30	30	30	27	27	29
FP	0	0	0	0	0	0	3	3	1
FN	11	12	14	3	3	5	0	0	0
Accuracy	81,7%	80,0%	76,7%	95,0%	95,0%	91,7%	95,0%	95,0%	98,3%

B. ACCURACY VALUES: AFTER SIX MONTHS

B.1 Close Adversaries

Table 11: Overview of the accuracy values (A) after six months. The calculation uses a distance threshold of 30 and takes into account the attacks by close adversaries as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	11	10	5	20	20	14	24	24	21
TN	24	24	24	24	24	24	19	19	20
FP	0	0	0	0	0	0	5	5	4
FN	13	14	19	4	4	10	0	0	3
A	72,9%	70,8%	60,4%	91,7%	91,7%	79,2%	89,6%	89,6%	85,4%

B.2 Strangers

Table 12: Overview of the accuracy values (A) after six months. The calculation uses a distance threshold of 30 and takes into account the attacks by strangers as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	11	10	5	20	20	14	24	24	21
TN	24	24	24	24	24	24	22	22	24
FP	0	0	0	0	0	0	2	2	0
FN	13	14	19	4	4	10	0	0	3
A	72,9%	70,8%	60,4%	91,7%	91,7%	79,2%	95,8%	95,8%	93,8%