A PIN-Entry Method Resilient Against Shoulder Surfing

Volker Roth
OGM Laboratory LLC
USA
vroth@ogmlabs.com

Kai Richter
ZGDV
Germany
kai.richter@zgdv.de

Rene Freidinger
Technical University Darmstadt
Germany

ABSTRACT

Magnetic stripe cards are in common use for electronic payments and cash withdrawal. Reported incidents document that criminals easily pickpocket cards or skim them by swiping them through additional card readers. Personal identification numbers (PINs) are obtained by shoulder surfing, through the use of mirrors or concealed miniature cameras. Both elements, the PIN and the card, are generally sufficient to give the criminal full access to the victim's account. In this paper, we present alternative PIN entry methods to which we refer as *cognitive trapdoor games*. These methods make it significantly harder for a criminal to obtain PINs even if he *fully observes the entire input and output of a PIN entry procedure.* We also introduce the idea of *probabilistic cognitive trapdoor games*, which offer resilience to shoulder surfing even if the criminal records a PIN entry procedure with a camera. We studied the security as well as the usability of our methods, the results of which we also present in the paper.

Categories and Subject Descriptors

D.4.6 [**Operating Systems**]: Security and Protection—*access controls, authentication*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*user-centered design, interaction styles, screen design*; H.1.2 [**Models and Principles**]: User/Machine Systems—*human factors*; K.4.4 [**Computers and Society**]: Electronic Commerce—*security*; K.6.5 [**Management of Computing and Information Systems**]: Security and Protection—*authentication*

General Terms

Measurement, Security, Human factors

Keywords

PIN, password, shoulder surfing, ATM, cognitive trapdoor games

1. INTRODUCTION

Each time a user withdraws money from an ATM or unlocks his cell phone, he types the identical four-digit PIN number sequence. Anyone who observes this procedure e.g., by looking over the shoulder of a user, can easily memorize the PIN. In conjunction with stolen or skimmed material such as magnetic stripe cards, account numbers printed on receipts, or mobile devices, criminals easily gain access e.g., to a victimized user's bank account [4, 33] or telecommuncations services.

Requiring users to memorize longer or multiple PIN sequences would have a detrimental effect on recall, and obviousy, no substantial improvement will be achieved for as long as the entered information remains constant. Likewise, requiring users to perform complicated mathematical calculations [35] when entering PINs is unreasonable. All this would raise the rate of erroneous PIN entries, which would in turn annoy users and thereby reduce the acceptance of the technology. Moreover, service and operation costs e.g., in the retail banking sector would increase due to a growing number of requests to reset PINs which are commonly blocked after three false entries.

The question then is, can the PIN entry method be redesigned in way that renders it more secure while still being easily usable? Towards an answer to that problem, we consider what we call an interactive *cognitive trapdoor game.* The key idea behind such a game is that it is easily won if the PIN is known, and is hard to win otherwise. Therefore, knowledge of the PIN constitutes the trapdoor. Additionally, being able to observe the game must not yield sufficient information to substantially improve the observer's ability to win subsequent instances of the game. Here, we assume that the observer's resources are *bounded by the cognitive capabilities of a human* e.g., by the capacity of a human's short term memory. Hence, the term cognitive trapdoor game.

In this paper, we present two variants of such a game (see §3). The principal idea is to present the user the PIN digits as two distinct sets e.g., by randomly coloring half of the keys black and the other half white. The user must enter in which set the digit is by pressing either a separate black or white key. Multiple rounds of this game are played to enter a single digit and it is repeatedly played until all digits are entered. The verifier e.g., the automatic teller machine (ATM), determines the entered PIN digits by intersecting the chosen sets. However, no individual round uniquely identifies the entered PIN digit. Any observer must quickly perceive and note, or memorize and process information from all rounds to derive the entered PIN. Our hypothesis is that this task can be designed so that it exceeds the cognitive capabilities of a human observer who does not know the genuine PIN whereas a human who knows the PIN can perform the task easily. In order to verify our hypothesis, we conducted two user studies: one study investigates the security of our methods whereas the second investigates their usability. In the end, the usability of a mechanism determines to a large degree its practicality. The importance of human factors in the design of security mechanisms is signified by the growing number of publi-

CCS'04, October 25-29, 2004, Washington, DC, USA.

cations and workshops with a focus on that topic [37, 26, 12, 22, 5, 25, 34, 31].

We found that our PIN entry methods provide significant robustness against shoulder surfing (see §4) and that users may accept our methods despite a disadvantage in terms of usability when compared to the contemporary PIN entry method (see §5), likely for reason of their recognition of the security benefits [25].

However, criminals increasingly employ miniature cameras e.g., to observe and record the PIN numbers entered by users [30, 9, 6, 36] at ATMs. A cognitive trapdoor game is ineffective against such an attack because the observer may determine the user's input in the same way the ATM would do it. This brings us to the notion of a *probabilistic cognitive trapdoor game.* Such a game exhibits all the properties of a cognitive trapdoor game but additionally the user's input is ambiguous. The uncertainty about the entered PIN yields a limited success probability of the observer winning future instances of the game despite the complete recording of a previous session (see §3.4). Although, recordings of multiple sessions may allow the observer to derive the genuine PIN.

Notably, our methods also provide protection against attacks directed against the PIN pad itself. For instance, by spraying suitable chemicals onto a PIN pad, an attacker may determine the set of digits that a user entered, and even the sequence [7] of digits. This attack fails against our input method because the user's responses are binary in nature. With a high probability, the user will have to press both response buttons utilized by our methods, which yields no additional information to the attacker. The binary response method also facilitates the implementation of user input in a fashion that is easy to operate while being hard to observe visually (because the user himself does not need to steer his fingers as much and therefore the response buttons can be easier obstructed from view).

As already hinted, our approach bases on the limitation of the capacity of a human's short term memory. A variety of mechanisms can be built around this limitation and probably other cognitive limitations as well. The research question we would like to put forward is what other cognitive trapdoor games may exist to let a human authenticate himself to a machine in the presence of observers.

2. THREAT MODEL

Authentication by a PIN occurs in different settings and with different devices including automatic teller machines (ATM), point of sale (POS) terminals, mobile phones, portable digital assistants (PDA), prepaid phone cards, or door locks in areas with increased security requirements. We model PIN authentication more abstractly as a game between three parties: a machine interrogator, a human oracle, and a human observer.

The objective of the oracle is to authenticate himself to the interrogator by his PIN. The objective of the interrogator is to decide whether the oracle knows the correct PIN by asking the oracle questions. The observer observes all interactions between the oracle and the interrogator; his objective is to impersonate the oracle in subsequent games with the same interrogator. The game also involves a setup phase with a trusted dealer who, using a confidential and authentic channel, distributes a token to the oracle, and a master secret to the interrogator (alternatively, the interrogator may query the dealer online during each game over an authentic and confidential channel). The token, typically a magnetic stripe [10] or chip card, contains information that uniquely identifies the oracle. It also contains information by which the interrogator can verify whether the oracle's input is correct and matches his identity.

We assume that the observer cannot verify the correctness of a given PIN unless he also knows the master secret (and we assume he does not). Additionally, we expect that the interrogator keeps a record of how often an oracle successively inputs a false PIN. If the count reaches three, the interrogator voids the oracle's authorization until the oracle receives a new PIN from the trusted dealer. Let the observer possess (a perfect copy of) the oracle's token (obtained by theft or skimming). Most important, we assume that the resources of the observer are computationally and memory bounded by the cognitive capacity of a human, particularly the short term memory (STM). In §3.4, we consider the case that the observer has the capability to record all interactions between the oracle and the interrogator without error for a single game.

Compared to an actual implementation, we make idealized and simplified assumptions. For instance, we assume that the PINs are generated in a uniformly distributed fashion. Actually, the PIN distribution e.g., of the Eurocheque Card system was shown to be skewed considerably [20, 16], to a large degree because recommendations in the applicable standards [15] were not fully adhered to. Otherwise, our model corresponds closely to what is typically found in the ATM world.

Having explained our assumptions and threat model, we continue by describing our alternative PIN entry methods.

3. PIN ENTRY METHODS

The general principle we apply is to consecutively display the PIN digits to the human oracle as two distinct sets. The oracle chooses the set in which the current PIN digit is. The interrogator then determines the correct PIN digit by intersecting the chosen sets, and the algorithm is repeated until the oracle entered all digits. The input and output methods determine the difficulty of the cognitive task that must be accomplished by the oracle and the observer. Below, we describe two designs of such a task: the *immediate oracle choice* variant and the *delayed oracle choice* variant. Our hypothesis is that in our designs the task is of limited cognitive complexity if the PIN is known, and of significant cognitive complexity otherwise. Hence, the term *cognitive trapdoor game.* We discuss and compare the properties of our designs in §3.3. Both variants achieve significantly better resilience against shoulder surfers without automatic recording devices than contemporary PIN entry methods (see §4 for experimental evidence). In §3.4, we describe a modification which provides limited resilience even if shoulder surfers record all inputs and outputs with a camera.

3.1 Immediate Oracle Choices

Let $\mathcal{A}$ be the the alphabet of PIN digits. The algorithm of the immediate oracle choice variant starts with a set $Q \leftarrow \mathcal{A}$ of probable PIN digits, and executes $n = \lceil \log_2 |\mathcal{A}| \rceil$ rounds. Typically, the alphabet consists of the digits from 0 to 9 and a PIN has $l = 4$ digits; hence $|\mathcal{A}| = 10$ and $n = 4$. In each round, the algorithm randomly partitions the set of $q = |Q|$ remaining probable PIN digits into two sets L and R of sizes $\lceil q/2 \rceil$ and $\lfloor q/2 \rfloor$. For the purpose of obfuscation, L and R are randomly augmented with digits that were eliminated in previous rounds so that both sets are of size $|\mathcal{A}|/2$. The two augmented sets are then displayed in a distinguishable fashion e.g., by coloring the keys of digits in L black and the keys of digits in R white. Here, we assume that the PIN pad allows us to change the color of keys. A round concludes when the oracle chooses one of the sets by pressing a separate black or white button.

We refer to this variant as "immediate choice" because the oracle must select a set without hesitation. After n rounds, only one digit remains and can be deduced unambiguously. Entering a PIN with l digits requires a total of $l \cdot n$ correct interactions, or 16 in a typical scenario.

A formal implementation is given in Fig. 2, which uses two functions π and γ. Let $\pi : \mathcal{A}^* \rightarrow \mathcal{A}^*$ be a random permutation of an

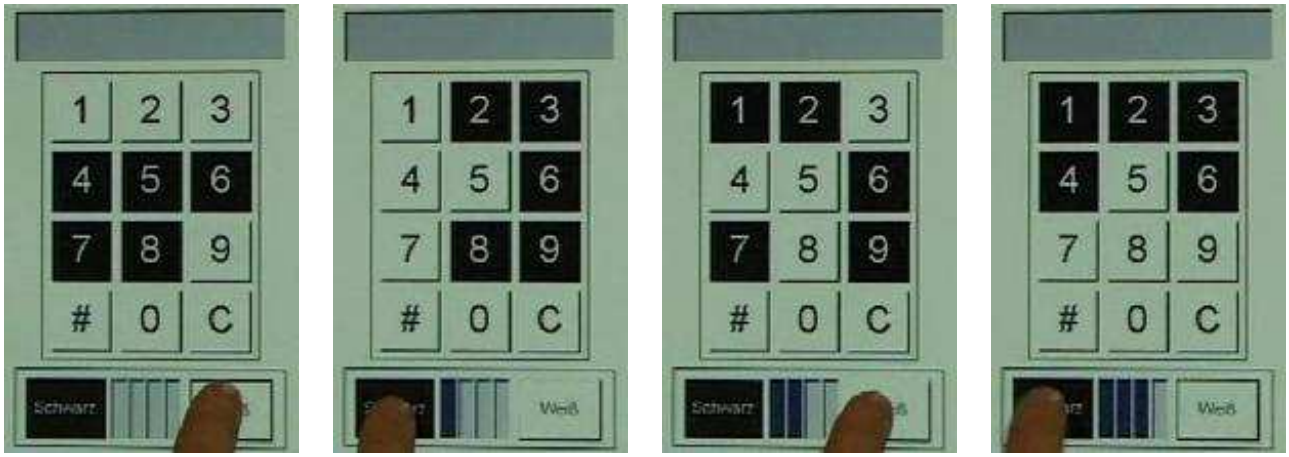


Figure 1: The above sequence illustrates the immediate oracle choice variant. The oracle is presented two partitions colored black and white, and must input in which partition the current PIN digit is. In the example above, the oracle enters the PIN digit '3'.

ordered set of PIN digits, and let γ be a function that partitions a set of q digits into two sets with sizes $\lceil q/2 \rceil$ and $\lfloor q/2 \rfloor$. More formally, let $\gamma : \mathcal{A}^* \rightarrow \mathcal{A}^* \times \mathcal{A}^*$.

A variety of display variants can be devised. We were particularly interested to study the usability of our methods. In order to minimize the influence of external factors on the results of our evaluation, we retained the original PIN pad layout and assumed that the color of digits can be changed from black on white background to the inverse and vice versa. This also has the advantage that subjects easily locate the position of digits on the pad (as opposed to e.g., the ScramblePad [2] which permutes the positions of digits on the PIN pad). Hence, one set is displayed as white on black, and the other as black on white. The result is a scattered pattern of black and white patches distributed over the pad. Two additional buttons (left for black and right for white) are required to input the choice of partition. Figure 1 shows the layout of our prototype implementation and illustrates an example input sequence for the digit 3. The text field at the top indicates the number of PIN digits already entered (using asterisks as in a regular setting) whereas the four vertical fields between the left and right button count the number of responses already given (which is at most four).

3.2 Delayed Oracle Choices

If the oracle responds slowly then the partitions are exposed longer to the observer. The longer the exposure is the easier is it for the observer to memorize or manually record a partition (which also uniquely identifies the second partition). As a remedy, we devised another approach which we call "delayed oracle choices." In that approach, n rounds are displayed consecutively with a predetermined exposure period of 0.5 seconds. The display is cleared subsequently and only then do the left and right input buttons appear. Using these buttons, the oracle must consecutively input the coloring that his PIN digit had in these n rounds. Rather than requiring the oracle to input its choice immediately, we delay the input until after exposure of all rounds. Consequently, the oracle cannot influence the exposure time. On the other hand, the oracle has only a limited period of time to determine the color of the current PIN digit in each round, and he must memorize the color sequence. Again, this procedure is repeated until all PIN digits are entered.

The delayed oracle choice variant requires that a set of n pairs of random partitions are precomputed such that an input pattern occurs at most once. Otherwise, the input of a PIN digit would be ambiguous. Furthermore, the number of black and white patches should be equal in each round to maintain symetric probabilities (this also holds for the first algorithm). A formal implementation of this variant is given in Fig. 3.

3.3 Cognitive Complexity Analysis

For both variants it holds that if the observer cannot observe the oracle's input then he has no information on the entered PIN even if he observes all output. On the other hand, if the observer can observe all input and output then he can deduce the entered PIN easily. Therefore, the security of the variants rests entirely on the ability of the observer to memorize or record the input and output. We begin by assuming that the observer has no automatic recording devices such as a concealed camera (although he may use e.g., manual tools such as paper and pencil).

The cognitive capabilities of a human have interesting limitations. Recently, Vogel and Machizawa [32] discovered a relationship between neural activity and memory capacity and found neurophysiological evidence that the human visual short term memory (STM) is limited to three to four symbols. In their measurements, they used a delay of one second between exposure and recall. Few subjects they tested had the capacity to hold five symbols in their STM. This is even less than the findings of Miller [19] who suggested that the capacity of the STM is limited to 7 ± 2 symbols. However, retention of items in the STM, and transfer to long term memory (LTM), appears to be critically dependent on the ability to rehearse the information in the STM [23]. A constant stream of new information in short succession, as in the case of our mechanism design, is known to impede rehearsal and thus later recall.

An effect that increases the capacity of the STM is referred to as *chunking* [21]. Multiple items are grouped together and represented as a single item that occupies one "slot" in the STM. For instance, an American can probably remember the sequence 149217761941 easily because it can be grouped into three chunks of four digits. Each of the chunks represents a year in which a historic event occurred that is significant to Americans, and which can be represented as a single item. It is not entirely clear to us at this point whether chunking may have an impact on our mechanisms. All 30240 black and white five digit numeric patterns are equally likely, it appears that the opportunities for frequent chunking are marginal. In summary, the limitations of humans' STM are a promising starting point for devising cognitive trapdoor games although there may

1: $(L, R) \leftarrow \gamma(\pi(\mathcal{A}))$
2: $(O, P) \leftarrow (\emptyset, \emptyset)$
3: **for** $i = 1, \cdots, n$ **do**
4: display $L \cup P$ and $R \cup O$
5: input choice $\in$ (left,right)
6: **if** choice $=$ left **then**
7: $(O, P) \leftarrow \gamma(\pi(O \cup P \cup R))$
8: $(L, R) \leftarrow \gamma(\pi(L))$
9: **else**
10: $(O, P) \leftarrow \gamma(\pi(O \cup P \cup L))$
11: $(L, R) \leftarrow \gamma(\pi(R))$
12: **end if**
13: **end for**
14: return L

Figure 2: The generic PIN symbol input algorithm is shown below. The sets L (black) and R (white) of probable PIN symbols are either eliminated or halved in each round depending on the oracle's choice. All other sets are working sets. The number of PIN symbols must be even for the algorithm to be correct.

1: $(P_{0,0}, P_{0,1}) = \gamma(\pi(\mathcal{A}))$
2: **for** $i = 1, \cdots, n$ **do**
3: $(L', L'') \leftarrow \gamma(\pi(P_{i-1,0}))$
4: $(R', R'') \leftarrow \gamma(\pi(P_{i-1,1}))$
5: $P_{i,0} \leftarrow L' \cup R''$
6: $P_{i,1} \leftarrow R' \cup L''$
7: display $P_{i,0}$ and $P_{i,1}$
8: **end for**
9: input $b_1, \cdots, b_n \in \{0, 1\}$
10: $Q \leftarrow \bigcap_{i=1}^{n} (P_{i,b_i})$
11: **if** $|Q| \neq 1$ **then**
12: return error
13: **end if**
14: return $q \in Q$

Figure 3: The principal delayed oracle choice algorithm is shown below. For each round, partitions of black ($P_{i,0}$) and white ($P_{i,1}$) digits are precomputed. A PIN digit is identified by intersecting the partitions that the oracle chooses in each round. All other sets are working sets.

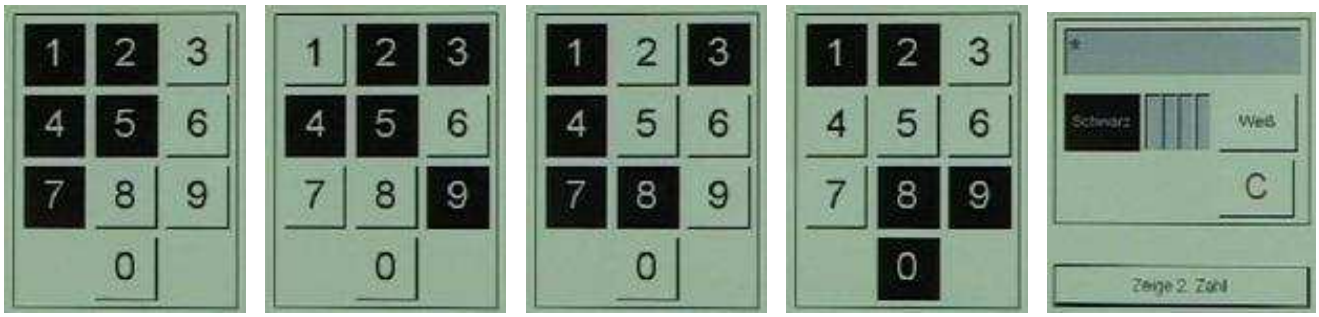


Figure 4: The above sequence illustrates the delayed oracle choice variant. Subsequent to the display of the four patterns on the left, the panel to the right appears. The oracle then has to enter the color sequence of the current PIN digit.

be alternative approaches.

In the immediate oracle choice variant (IOC), the oracle must retrieve a PIN digit from long term memory and must decide on the color that digit has before responding. In the delayed oracle choices variant (DOC), the oracle must remember a sequence of four colors in its short term memory. In both cases can the oracle focus its gaze on the position of the current PIN digit on the PIN pad, which eliminates the need to maintain awareness of the digit itself. The IOC variant is well within the cognitive capacity of a healthy human, the DOC variant is within practicable bounds.

In the IOC variant, the observer must remember at least five symbols (the new information presented in each round) and the response of the oracle, and he must be able to record this information at the same rate at which it is presented. If the observer does not record the information but attempts to emulate the interrogator's processing then he must additionally remember his current hypothesis of what the set of probable PIN symbols is. Furthermore, he must mentally intersect his hypothetical set of solutions with the recently presented set of digits. In the DOC variant, the observer has no means of pruning the set of possible PIN symbols before the oracle inputs his answers. This amounts to memorizing information worth at least 20 symbols.

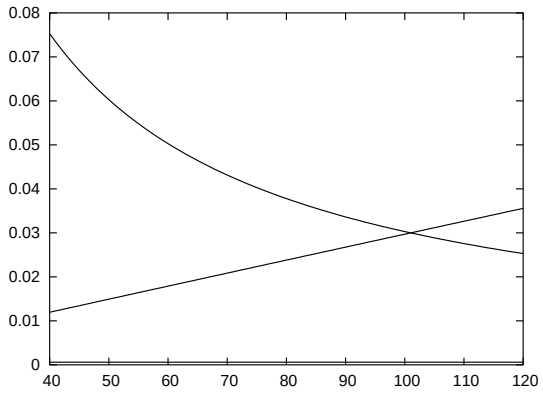
We refrained from using more than two colors because this complicates the decision task of the oracle. It also facilitates the task of the observer because the sizes of the individual sets become smaller. In a degenerate case e.g., if ten colors were used, the PIN pad becomes functionally identical to a scrambled PIN pad. Additionally, multiple colors increases the risk that people with impaired color perception cannot use the system.

In order to verify our hypothesis that the asymmetry of the cognitive overhead of the oracle's and the observer's task fulfills our requirements for a cognitive trapdoor game, we conducted two studies. First, we measured subjects' ability to record and guess PIN digits in recorded PIN entry procedures. Second, we studied subject's ability to enter PINs using our designs. Additionally, we measured how well subjects accepted our designs. The results of our studies are presented in §4 and §5 respectively.

3.4 Recording Resilience

Criminals increasingly employ concealed miniature cameras to observe and record the PINs entered by victims [36, 30, 6]. The question that we asked ourselves was whether our designs could be modified to be robust even if the observer can record all input and output. Ideally, this amounts to devising the analog of a cryptographic zero knowledge protocol between the interrogator and the oracle.

An essential ingredient of recording resilience is that the oracle's input must not uniquely identify the PIN. Otherwise, the observer would be able to impersonate the oracle in future games without any probability of failure. For instance, assume that each color input sequence identifies exactly two PIN digits (e.g., the input sequence white, black, black, white for digit 3 in Fig. 4 also identifies a second digit, only five input sequences are used overall). Whichever digit the oracle enters, the interrogator and the observer are left with two equally probable choices per digit. Given that PIN



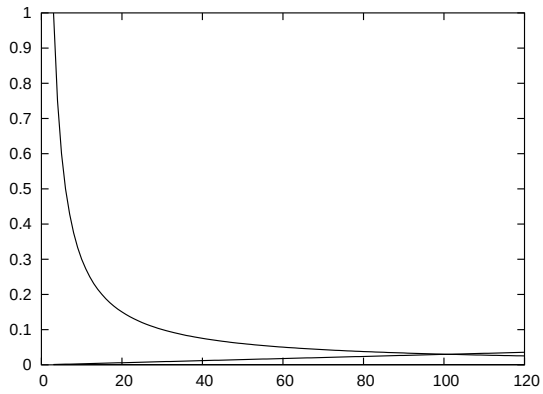


Figure 5: The left graph shows the probability of success for guessing PINs from a known shadow set (Z, approaching intersection from top), guessing PINs randomly from the PIN space (Z', approaching intersection from bottom), and guessing input sequences randomly (Z'', barely visible at the bottom). The size of the shadow set is plotted on the abscissa, the probabilities are plotted on the ordinate. The PIN space is of size 10,000. The right graph is an excerpt of the left graph.

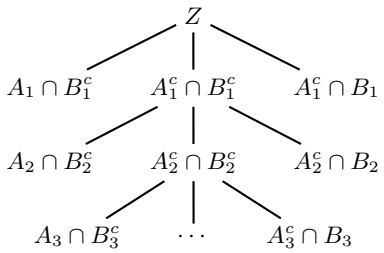


Figure 6: The decision tree to compute the probability of successful impersonation of an oracle by an adversary.

numbers are of length $l = 4$ digits this yields $2^l = 16$ equally probable PIN numbers one of which is the correct number. We refer to the 16 probable PIN numbers as *shadows*. The interrogator can efficiently verify all shadows based on the master secret, and he may authorize the oracle if one of the shadows is valid. However, the observer can not decide which of the shadows is the correct one because this requires the master secret known only to the interrogator (compare to §2). This leaves the observer with three attempts to impersonate the oracle choosing candidate PINs from the intercepted shadows.

Assume that the observer knows the set $\mathcal{S}$ of $s = |\mathcal{S}|$ shadows and follows a simple strategy to impersonate the oracle: he tries one PIN in the shadow set after the other in a random order. The observer succeeds if, in $n = 3$ or fewer attempts:

Event A he guesses the correct PIN from the shadow set, or
Event B he guesses a wrong PIN but one of the shadows of the wrong PIN is the correct PIN.

Let Z be the event that the observer is successful in n or fewer attempts. The probability $\Pr[Z]$ can be calculated based on a decision tree (see Fig. 6) which iterates the successful events.

Note that the observer cannot guess a correct and a wrong PIN simultaneously and therefore in all attempts k, $A_k \cap B_k$ is the empty event ϕ. The events at each node of the tree are mutually independent conditional to their parent node and thus the observer's probability to succeed in n or fewer attempts is the sum of the probabilities of the leaves in the tree, or more precisely:

$$\Pr[Z] = \sum_{k=1}^{n}((\Pr[A_k \cap B_k^c] + \Pr[A_k^c \cap B_k]) \cdot \prod_{i=1}^{k-1} \Pr[A_i^c \cap B_i^c])$$

The individual probabilities for each event can be calculated based on the observation that, conditional to picking a PIN $x_k \in \mathcal{S}$, A_k and B_k are independent experiments. Therefore it holds that:

$$\Pr[A_k \cap B_k^c] = \Pr[A_k] \cdot \Pr[B_k^c]$$
$$\Pr[A_k^c \cap B_k] = \Pr[A_k^c] \cdot \Pr[B_k]$$

$$\Pr[A_k] = \frac{1}{s-k+1} \qquad \Pr[B_k] = \frac{1}{N-s+1}$$

Unfortunately, the introduction of shadows also increases the probability that an adversary succeeds to impersonate the oracle by randomly guessing a PIN without knowing a shadow set. Let C_k be the event that the adversary guesses the correct PIN from the entire set of PINs in the kth attempt. By similar considerations as summarized above, it holds that the success probability Z' can be derived by substituting C_k for A_k in Z where:

$$\Pr[C_k] = \frac{1}{N-k+1}$$

Now consider that, rather than reducing the number of distinct color input sequences, we reduce the number of rounds per game so that s shadows remain overall. This has the added benefit of speeding up the PIN entry procedure. More precisely, for N possible PINs and s shadows, the information that the oracle must enter to generate the shadow set is reduced from $\log_2(N)$ bits to $t = \log_2(N) - \log_2(s)$ bits when compared to the contemporary PIN entry method. This means that the number of rounds of input is reduced to t and thereby the information available to the observer is minimized. Assume that the adversary (without knowing a shadow set) picks input patterns randomly, then his probability Z'' to impersonate the oracle in n or fewer attempts becomes:

$$\Pr[Z''] = \sum_{k=1}^{n}(\frac{1}{2^t} \cdot (1 - \frac{1}{2^t})^{k-1})$$

Figure 5 shows plots of the graphs of the probabilities for Z, Z', and Z'' for different numbers of shadows given a PIN space of size

10,000. The size of the shadow set is plotted on the abscissa, the probabilities are plotted on the ordinate. Around a shadow set size of 100, the probability of Z'' breaks even with that of Z (approaching from above) at a probability of approximately 0.03. This means that a shadow set size of about 100 is the reasonable maximum and yields approximately a 3% chance that an adversary impersonates an oracle with or without knowledge of a shadow set. The probability of Z' remains small until the size of the shadow set approaches the size of the PIN space, at which point the probabilities of Z and Z' merge and approach 1 (not shown in the graphs).

At the same time, the number of rounds an oracle has to play the cognitive trapdoor game is theoretically about halved, which would significantly improve the usability of our PIN entry method beyond the evaluation results we present in §5. Although in practice, the round design limits the sizes of shadow sets to powers of 2. It would be advantageous to select the size of the PIN space as a power of 2 as well to obtain optimal security tradeoffs. Due to the probabilistic nature of this recording resilience extension we refer to such games as *probabilistic cognitive trapdoor games*.

One caveat remains, though. In a typical scenario, the interrogator resets its counter of false attempts once a PIN is entered correctly. Hence, the observer may probe one or two PINs taken from the shadow set. If these attempts fail then he waits until the oracle again entered his genuine PIN. At this point, the interrogator resets his false attempts counter and the observer can probe one or two more PINs from the shadow set. This strategy may be continued until the observer identified the genuine PIN. In order to avoid this attack, the interrogator must display the recorded number of false attempts before the game, so that the oracle is alerted. A consequential denial of service condition due to intentional entry of false PINs can be avoided by amending the identifying information stored on the oracle's token with a salt. Hence, the token cannot be forged from obvious identifying information (such as account numbers printed also on balance sheets) but must be stolen first (in which case invalidation of access is in the best interest of the oracle).

4. SECURITY EVALUATION

We were curious to measure the security of our oracle methods against shoulder surfing attacks, and conducted a user study towards this end. Our expectation was that, compared to the regular method, the two oracle methods would be more difficult to follow by an observer. We decided to test this hypothesis under optimal conditions for the attacker.

We implemented all three alternatives, the regular PIN pad (REG), the immediate oracle choice (IOC), and the delayed oracle choice (DOC), using the Microsoft .Net framework and the C# language. The software logged all user input for subsequent analysis. We deployed it on a kiosk system with touch screen interface, running a Windows PC. The screen resolution was 1024×768 pixels and the size of the PIN entry window had a size of 640×480 pixels. As a preparation, we filmed the entering of ten random sequences of four-digit PIN numbers for each method using a Sony digital camera which recorded directly to a hard disk. The view port of the camera was chosen so that the entire PIN pad was visible as well as the fingers of the person entering the numbers. Care was taken not to unnecessarily obstruct the display during PIN entry. With this approach, we intended to provide optimal conditions for an attack. Additionally, we produced three separate example films for the purpose of explaining all input methods to the subjects with whom we conducted the study.

We recruited 8 students of the local university as subjects. All subjects were on natural science tracks and had experience in mathematics and methodology. We first briefed the subjects with the example films and by a written explanation of the principles of the methods. They were instructed to use paper and pencil or any other readily available tools (except for a camera) suitable to improve guessing performance. Discussion on strategies between participants was encouraged by the experimenter. One of our goals was to study strategies that the subjects came up with. The participants were motivated by promising a reward for the first correct guess on an oracle method.

Correct digits	4	3	2	1
Regular	100	0	0	0
Delayed OC	0	0	5	7.5
Immediate OC	0	0	5	8.75

Figure 7: This table shows the guessing rate in percent of tries.

Subsequent to the briefings, the three films were shown to the participants as a group (the films were projected to a screen in front of the group). The film was stopped between each PIN entry sequence and a short break was offered in order to allow participants to write down their guesses and for reflection and discussion of their strategies. The experimental sequence for the group was first to show the ten sequences with the regular PIN entry method, followed by the immediate, followed by the delayed oracle choice method. The intention behind this fixed procedural sequence was to allow the participants to elaborate on their strategies and to refine them in preparation for what we assumed to be the more difficult methods. In order to prevent fatigue, we offered breaks between the films. The films totalled a length of approximately 10 minutes. After each film, the participants were interviewed about their impression of the method and the strategies they had used and elaborated.

All eight participants were able to complete the study and there was no visual or understanding problem in following the contents of the films. As for the results of the attack, no participant was able to guess even one of the PIN numbers entered with one of the oracle choice methods while all participants were able to correctly guess the PIN numbers entered with the regular method in each of the ten trials. However, some participants succeeded in guessing one or two digits of some of the PIN numbers entered with the oracle choice methods (see Table 7 for a summary of results).

The isolated successes appear to be a result of the strategies employed by the participants. In four cases, the subject has focussed on one or two randomly chosen numbers and compared the input to the pattern of those numbers. Another strategy of subjects was to capture the distribution of black and white buttons as a pattern that they sketched on paper. Some participants used prepared stencils as an aide to mark black and white buttons. However, there was no significant difference in guessing performance between strategies.

5. USABILITY EVALUATION

Having gained insights into the observer's task and the relative security of the oracle choice methods, we directed our attention towards the users of such a system. Backed by the encouraging results of our first study we conducted a second study with the goal to quantify and compare the relative usability of all three PIN entry methods. We did not expect that the two oracle choice methods are more usable than the regular input method which enjoys an overwhelming degree of familiarity in modern societies. Without any doubt both oracle methods are more complex and take longer to accomplish than the regular method. However, Sasse [25] found that users do weight costs versus benefits in their judgement of a technology. Along this line, we wished to learn whether our methods constitute an acceptable tradeoff.

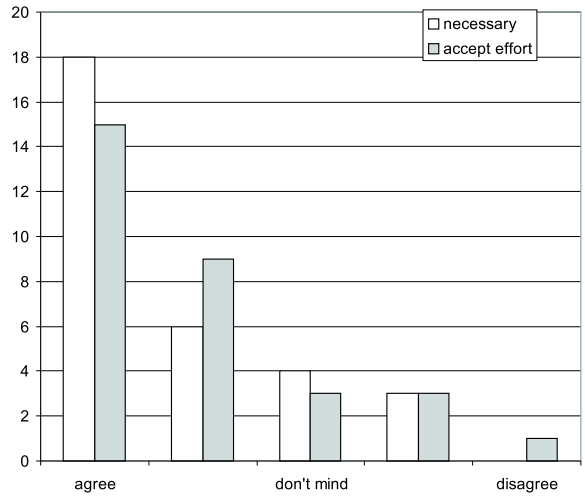


Figure 9: Frequency of the ratings on two statements expressing the necessity to improve security and the acceptance of additional effort in PIN entry.

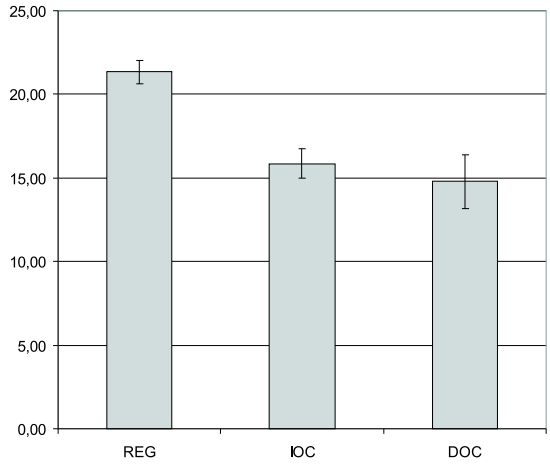


Figure 8: Usability rating of the users of the three systems on seven items of the SUS. The values represent a summed score but not the SUS index. The regular condition is rated significantly more usable than the oracle methods.

5.1 Method

Thirty-seven participants with academic education were recruited for this study. Four participants were females and all participants were about 20 to 30 years old. We chose a demographically homogenous group of participants in order to limit the required number of subjects, and to maximize the impact of the condition factor on variance. Records from three participants had to be eliminated due to missing data, and the data of one participant was excluded because his bad physical pre-test condition was an outlier, which left us with a set of 33. Each participant was randomly assigned one input method of which he had to complete 10 input cycles (number of subjects for REG: $n = 9$; for IOC: $n = 14$; for DOC: $n = 10$). All participants had to perform their input on the same kiosk system that we used in our first study.

We asked the participants to substitute an easily memorable four digit number for their real PIN e.g., based on their birthday or another well-known number, in order to compensate for the mental salience of the real PIN. As dependent variables, user condition (pre- and post test), acceptance, time used for entry and the error rate were collected. User condition was assessed using the short scale of the BMS [24] which indicates the work load in four subscales: physiological fatigue, concentration, motivation, emotional state. Values are summed to scores for each subscale. This test was applied before and after a participant performed the PIN entry in order to obtain a measure of fatigue caused by the method. The usability was measured with a subset of items taken from the *Software Usability Scale* (SUS) [8]. Additionally we asked the participants several questions about how acceptable and secure they found the tested input method was. The time and error rates were extracted from the log files that our software recorded during the sessions.

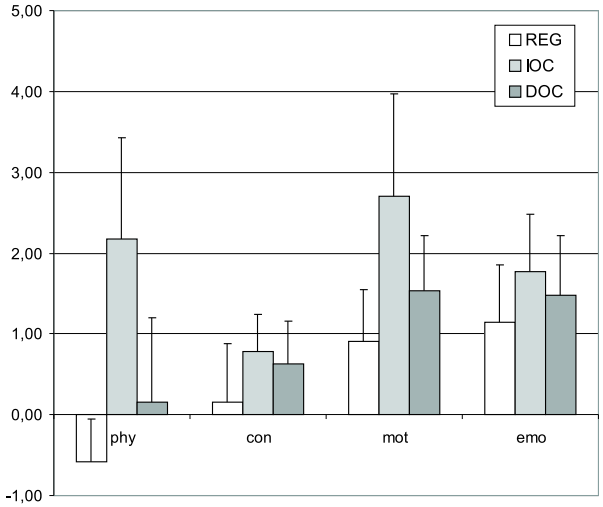


Figure 10: Change in user condition (phy = physiological load, con = concentration load, mot = motivational effort, emo = emotional effort) for the three conditions during the usage. Higher values indicate a higher load and fatigue level.

5.2 Results

User condition. We found that the operation of all the three systems caused some degree of fatigue for the user. Figure 10 shows a consistently higher load of the oracle methods, especially the IOC. However, there was no significant difference between conditions based on the Kruskal-Wallis test for independent samples (phy: $\chi^2(2) = 3.165$, $p = 0.2$; con: $\chi^2(2) = 1.51$, $p = 0.46$; mot: $\chi^2(2) = 2.34$, $p = 0.31$; emo: $\chi^2(2) = 0.13$, $p = 0.93$;). A Wilcoxon test for dependent samples revealed a significant fatigue effect between the pre test and post test condition in motivation and emotional state (for motivation: IOC: $Z = -2.27$, $p = 0.02$; DOC: $Z = -2.55$, $p = 0.01$; for emotion: IOC: $Z = -2.23$, $p = 0.03$;)

Usability rating. The summed score of the seven SUS subscales shows a clear advantage of the regular method when compared to IOC and DOC. A Mann-Whitney-U test revealed that both oracle methods were rated significantly less usable than the regular method while there was no significant difference between IOC and DOC ($REG *$ IOC:, $Z = -3.51$, $p < 0.01$; $REG *$ DOC: $Z = -2.75$, $p < 0.01$; $IOC *$ DOC: $Z = -0.47$, $p = 0.67$).

Acceptance. Most of the participants agreed with the statement "I think that it is necessary to improve security in PIN entry methods" and "I would accept additional effort in order to have more secure methods" (see also Fig 9 for more details). This result is in accordance with the fact that, even though other usability measures and the entry times were to the disadvantage of the oracle methods, the agreement on the statement "I would favour having such a system deployed in real life" was higher for the oracle methods than for the regular one (REG $= 1.86$, IOC $= 2.93$, DOC $= 2.00$) although this result failed to reach significance ($F(2, 28) = 3.785$, $p = 0.087$). The perceived security was significantly higher for both oracle methods than for the regular method (REG $= 1.44$; IOC $= 3.36$; DOC $= 3.5$) as shown by a Mann-Whitney-U test (REG $*$ IOC: $Z = -3.42$, $p < 0.01$; REG $*$ DOC: $Z = -3.58$, $p < 0.01$; IOC $*$ DOC: $Z = 0$, $p = 1$).

Entry times. It takes users about ten times longer to enter a PIN with an oracle method than with the regular method. Figure 11 shows that after a steep learning curve, all systems settle to a normal level of performance after the third trial. The average overall entry time for the last five iterations is in seconds: REG $= 2.797$, IOC $= 23.228$, DOC $= 25.676$. On the level of single keystrokes, the logs show that both the DOC method and the regular method converge to about 700 to 900 milliseconds, while users perform much slower in the IOC condition. The average of the last five iterations is in milliseconds: REG $= 705$, IOC $= 876$, DOC $= 2746$. In the regular condition users exhibit a learning curve on the level of key strokes whereas this is not the case in the oracle conditions. This is also illustrated in Fig 12.

Error probability. We also found that the mean error probability of all three methods during the last three trials does not differ significantly after learning has taken place (Pr[REG] $= 0$; Pr[IOC] $= 0.09$; Pr[DOC] $= 0.2$; $F(2, 28) = 2.117$; $p = 0.139$).

5.3 Interpretation

The significant improvement of security that we found in our first study has also been perceived by the participants of the second study. This seems to have caused a higher although not significantly different ($p = 0.07$) acceptance rating of the two oracle choice methods between both studies.

This result is surprising in consideration of the user condition and usability tests as well as the results of the time and error data. While there were significant differences in the usability ratings and the entry times we found only a slight and insignificant advantage of the regular method in user condition and errors. The advantage of the regular method in all those measures is obvious and has been expected. However, what is surprising is:

1. the obvious advantage of the regular method has not lead to a clear differentiation of the methods in the quantitative data;
2. users appear to accept the alternative oracle choice methods despite the higher mental effort that is required.

We conclude that the participants noticed the superior security of the oracle choice methods and hence their assessment of costs versus benefits [25] is in favor of these methods.

The persistently higher keystroke latency in the IOC condition, when compared to the REG and DOC condition, provides further interesting insight in the cognitive processes during the task. The IOC method forced subjects to maintain a high level of concentration throughout the entry procedure because the changing patterns require immediate responses. In the DOC condition, the automatic playback of patterns forced subjects to memorize the color sequence in STM until the subject could enter it. At the time of entry, subjects merely had to recite the sequence and therefore the entry process was faster. Interestingly though, the total time in the DOC condition was slower than IOC even though the presentation time for each pattern was 500 ms which is about the minimum time required to perceive and store the patterns. This indicates that participants found a more direct way to process patterns into adequate responses in the IOC condition than in the DOC condition. The higher memory load caused by the DOC codition may also explain the slightly higher error rate.

6. RELATED WORK

Human cognition has been investigated intensely with the goal to enhance recall of passwords e.g., cognitive passwords [38], word associations [27], pass phrases [28], images [11], or pass faces [3]. None of this work considered cognitive capabilities and limitations in the context of PIN or password entry. However, the importance of human factors in the design of security mechanisms has been increasingly recognized recently [37, 26, 12, 22, 5, 25, 34, 31].

Wilfong [35] filed a US Patent on a PIN entry method in which, put simply, the interrogator challenges the oracle with a random PIN. The oracle is required to add modulo 10 each digit of his genuine PIN to the digits of the random PIN, and must enter the outcome. The interrogator inverts the calculation by subtracting the random PIN from the entered one. This method is not suited to prevent shoulder surfing attacks since there is no principal asymmetry in the task of the oracle and the observer. The observer observes the challenge and the entered PIN as if it were a regular PIN and easily calculates the genuine PIN from that information.

Matsumoto and Imai [18] presented a human-computer authentication scheme which is based on the following idea: the terminal and the user secretly share a string w and an ordered set v of symbols. The terminal presents a random string $x = x_1, \cdots x_n$ to the user as a challenge. If symbol $x_i = v_j$ for some j then the user replaces x_i with w_j and with a randomly chosen symbol from the answer alphabet otherwise. The correctness of the substitution is verified by the terminal. The mental effort required to memorize both the the password and the set and to perform the appropriate substitutions appears to be considerable when compared to a regular PIN entry procedure. This also holds for the scheme of Li and Teng [17] which requires users to memorize and operate on three keys with different functions each of which has 20-40 bits worth of information.

Hopper and Blum [13, 14] also devised a human-computer authentication scheme. Their principal idea is as folllows: the terminal and the user share a secret vector s of length n. In round $i = 1, \cdots, k$ the terminal displays randomly generated query vectors x_i to the user who computes and enters the parity of $s \cdot x_i$ with probability $1 - p \in (0, \frac{1}{2})$ and enters $1 - s \cdot x_i$ with probability p. For a sufficient number of rounds, the terminal can authenticate the user with overwhelming probability whereas learning the parity becomes $\mathcal{NP}$-Hard in the presence of errors. The same principle can be used to let the terminal decide when the user should make an

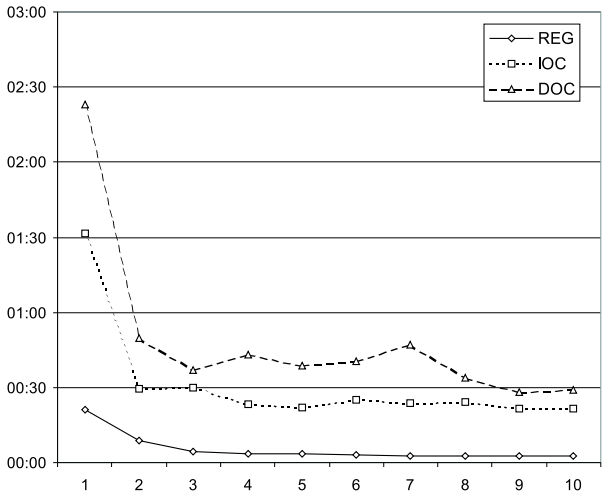


Figure 11: The development of the total entry time needed for the three methods over the three trials. The regular method is the least time consuming method followed by the IOC and the DOC.

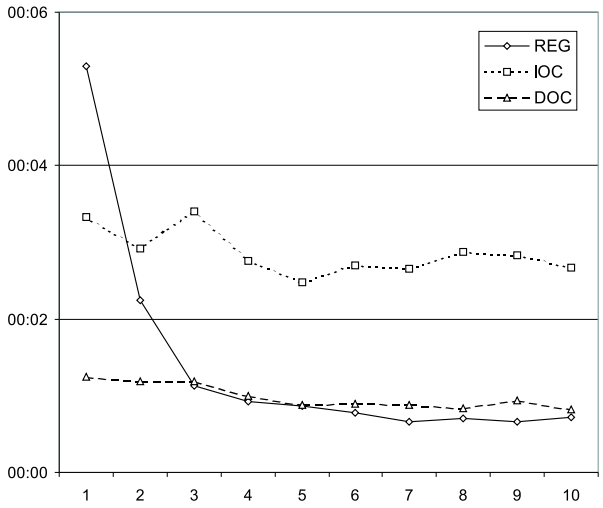


Figure 12: Time users take for single keystrokes in the different conditions. REG and DOC converge to a similar speed while IOC is consistently slower.

error, which simplifies the human's task. However, this requires an additional secret s' (which can be derived from s based on a simple calculation without compromising security). This scheme would be secure against camera recording of $s^{O(n/\log n)}$ challenges and thus multiple recordings for a large enough n whereas our scheme does not provide adequate security against multiple recordings. However, the operations that must be performed by humans in Hopper and Blum's scheme are complicated and the size of the secret that a user must remember is prohibitively large. The authors concluded that the scheme, albeit coming close, is not usable in practice [14].
The schemes that we referenced above [18, 17, 14] concentrate on an information theoretic notion of security in the presence of an observer with recording capabilities that are unlimited for practical purposes. All schemes are without doubt more secure than our approach, and are based on a precise notion of security whereas we base the security of our solution on the imprecise notion of limited cognitive capabilities. However, all those schemes differ substantially from the common PIN entry procedure, required complex mental operations and memorization of significantly more information than the four digits that are typical in practice. We, on the other hand, do not require that users memorize more information than has been previously the case. Despite the fact that the operations that we ask users to perform are simple compared to the calculations of these other schemes, the additional workload is at the limit of what we found users would accept. In this light, we doubt that the other schemes are as usable as our approach. However, we intend to conduct further usability studies with the goal to compare those mechanisms.
Swivel Technologies [1] provides technology whereby the user is presented a permutation and is required to enter his PIN accordingly permuted. The technology is meant to thwart key logging attacks but could be used against observers as well, although it is not resilient against camera recording. An observer must remember the permutation and the PIN entered, all of which is in plain sight. The amount of information that must be remembered is less than in our scheme, and the security of that scheme is likely also less. Hirsch Electronics [2] provides a PIN pad with LED numeric display integrated in the keys. The displayed digits are scrambled randomly before PIN entry so that the positions of PIN digits changes. Furthermore, the vertical and horizontal angle of light emission from the LEDs is restricted so that the digits are visible only within a narrow viewing angle in front of the display. Both mechanisms complicate the task of a human shoulder surfer although a cleverly placed miniature cameras might still be able to record the PIN pad layout. However, PIN entry becomes more complicated as well since the user has to locate the PIN digits on the pad. This is detrimental to recall since users frequently report that they remember PINs by position rather than purely by number. The ScramblePad suffers from the disadvantage that terminals must be retrofitted with it whereas our solution merely requires a software patch and may even coexist with the current procedure. However, we would like to compare the security and usability of the aforementioned technologies and our solution more closely in a future study.

7. CONCLUSION AND OUTLOOK

Towards a PIN entry method that is robust against shoulder surfing, we proposed two variants of an interactive challenge-response protocol to which we refer as *cognitive trapdoor games*. The essential feature of such a game is that it is easily won if the PIN is known, and hard to win otherwise. The cognitive capabilities of a human are generally not sufficient to derive the genuine PIN through observation of the entire game's input and output. As a defense against automatic recording e.g., by miniature cameras, we proposed another variant which maintains a certain level of uncertainty about the genuine PIN even if automatic recording devices are deployed. Due to its probabilistic nature, we refer to this variant as a *probabilistic cognitive trapdoor game.*
In order to assess the security and usability of our PIN entry methods, we conducted two user studies. The results of these studies support our hypothesis that our methods provide superior resilience against shoulder surfing which is of significant value when entering PINs in a public environment. Among the variants, the immediate oracle choice method has shown considerable advantages over the delayed oracle choice method concerning security, usability, accep-

tance, entry times, and error rates, despite the fact that it seems to cause the most effort for the user. However, the additional effort, when compared to the regular PIN entry method, turned out to be offset by users' subjective and objective security advantages gained by this method, which supports Sasse's notion of users' cost versus benefit calculation [25].

Some of the shortcomings of the immediate oracle choice method, such as the longer total execution time and the resulting higher fatigue level can be improved by applying the recording resilience strategy described in §3.4. This decreases the number of rounds required for PIN entry, and hence improves the usability and error rate.

Future work will address the issue of demographical influences on usability of the system. Above all, the usability for elderly people and people with little computer experience will be in the focus when testing the revised versions of our methods. We are also interested to compare the usability of our approach with other approaches such as the ones we discussed in §6. Furthermore, we would like to investigate the effects of additional strategies for displaying the partitions of PIN digits.

Lastly, we would like to encourage researchers to investigate what other cognitive trapdoor games may exist and how these principles can be employed to secure a wider range of systems and applications in a user-oriented fashion.

8. ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful comments which led to some corrections and improvements of our paper, particularly the mathematical modelling of probabilities and our coverage of related work. We also thank Simson Garfinkel for bringing the related technology of Swivel Technologies to our attention. Thanks also go to Stephen Wolthusen and Michael Arnold for reviewing our revised modelling of probabilities.

9. REFERENCES

[1] `http://www.swiveltechnologies.com`, July 2004.

[2] `http://www.hirschelectronics.com/Products_ScramblePads.asp`, July 2004.

[3] Passfaces. `www.realuser.com`, Apr. 2004.

[4] ANDERSON, R. Why cryptosystems fail. In *Proc. 1st ACM Computers and Communications Security Conference* (Fairfax, Virginia, USA, Nov. 1993).

[5] BALFANZ, D. Usable access control for the world wide web. In *Proc. Nineteenth Annual Computer Security Applications Conference* (Dec. 2003), IEEE, pp. 406–415.

[6] BRADER, M. Shoulder-surfi ng automated. Risks Digest 19.70, Apr. 1998.

[7] BRIER, E., NACCACHE, D., AND PAILLIER, P. Chemical combinatorial attacks on keyboards. *International Association for Cryptographic Research ePrint Archive 2003*, 217 (2003).

[8] BROOKE, J. SUS: A quick and dirty usability scale. In *Usability evaluation in industry*, P. Jordan, B. Thomas, B. Weerdmeester, and I. McClelland, Eds. Taylor and Francis, London, 1996, pp. 189–194.

[9] COLVILLE, J. Atm scam netted $620,000 australian. Risks Digest 22.85, Aug. 2003.

[10] COUNT ZERO. Card-o-rama: Magnetic stripe technology and beyond. *Phrack*, 37 (1992).

[11] DHAMIJA, R., AND PERRIG, A. Déjà vu: A user study using images for authentication. In *Proc. 9th USENIX Security Symposium* (Denver, CO, USA, Aug. 2000).

[12] DOURISH, P., AND REDMILES, D. An approach to usable security based on event monitoring and visualization. In *Proc. New Security Paradigms Workshop* (Virginia Beach, VA, USA, Sept. 2002), ACM, pp. 75–81.

[13] HOPPER, N. J., AND BLUM, M. A secure human-computer authentication scheme. Technical Report CMU-CS-00-139, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, May 2000.

[14] HOPPER, N. J., AND BLUM, M. Secure human identifi cation protocols. In *ASIACRYPT* (2001), C. Boyd, Ed., vol. 2249 of *Lecture Notes in Computer Science*, Springer Verlag, pp. 52–66.

[15] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *Banking – Personal Identification Number (PIN) management and security – Part 1: Basic principles and requirements for online PIN handling in ATM and POS systems*, May 2002. TC 68/SC 6.

[16] KUHN, M. Probability theory for pickpockets – ec-PIN guessing. Available at `http://www.cl.cam.ac.uk/~mgk25/`, 1997.

[17] LI, X.-Y., AND TENG, S.-H. Practical human-machine identifi cation over insecure channels. *Journal of Combinatorial Optimization 3*, 4 (1999).

[18] MATSUMOTO, T., AND IMAI, H. Human identifi cation through insecure channel. In *EUROCRYPT* (1991), D. W. Davies, Ed., vol. 547 of *Lecture Notes in Computer Science*, Springer Verlag, pp. 409–421.

[19] MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review 63* (1956), 81–97.

[20] MÖLLER, B. Schwächen des ec-PIN-Verfahrens. Available at `http://www.informatik.tu-darmstadt.de/TI/Mitarbeiter/moeller`, Feb. 1997. Manuscript.

[21] MURDOCK, B. B. The retention of individual items. *Journal of of Experimental Psychology 62* (1961), 618–625.

[22] PATRICK, A., LONG, C., AND (ORGANIZERS), S. F. Workshop on human-computer interaction and security systems at acm chi 2003. Web pages at URL `http://www.andrewpatrick.ca/CHI2003/HCISEC/index.html`, Apr. 2003.

[23] PERTERSON, L. R., AND PETERSON, M. J. Short-term retention of individual verbal items. *Journal of of Experimental Psychology*, 58 (1959), 193–198.

[24] PLATH, H.-E., AND RICHTER, P. Ermüdungs-Monotonie-Sättigung-Stress (BMS). Tech. rep., Psychodiagnostisches Zentrum, Dresden, Germany, 1984.

[25] SASSE, M. A. Computer security: Anatomy of a usability, and a plan for recovery. [22].

[26] SMETTERS, D. K., AND GRINTER, R. E. Moving from the design of usable security technologies to the design of useful secure applications. In *Proceedings of the New Security Paradigms Workshop* (Virginia Beach, VA, USA, Sept. 2002), ACM, pp. 82–89.

[27] SMITH, S. L. Authenticating users by word association. *Computers & Security 6* (1987), 464–470.

[28] SPECTOR, Y., AND GINZBERG, J. Pass-sentence – a new approach to computer code. *Computers & Security 13* (1994), 145–160.

[29] STIRZAKER, D. *Elementary Probability*, 2nd ed. Cambridge University Press, 2003.

[30] SUMMERS, C., AND TOYNE, S. Gangs preying on cash machines. BBC News Online, Oct. 2003.

[31] TOM MARKOTTEN, D. G. User-centered security engineering. In *Proc. 4th NordU Conference* (Helsinki, Finland, Feb. 2002).

[32] VOGEL, E. K., AND MACHIZAWA, M. G. Neural activity predicts individual differences in visual working memory capacity. *Nature 428* (Apr. 2004), 748–751.

[33] WEINSTOCK, C. Atm fraud. Risks Digest 4.86, May 1987.

[34] WHITTEN, A., AND TYGAR, J. D. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *Proc. 9th USENIX Security Symposium* (August 1999).

[35] WILFONG, G. T. Method and apparatus for secure PIN entry. US Patent #5,940,511, United States Patent and Trademark Offi ce, May 1997. Assignee: Lucent Technologies, Inc. (Murray Hill, NJ).

[36] WOOD, D. Spain uncovers hi-tech cashpoint fraud. BBC News Online, Jan. 2003.

[37] YEE, K.-P. User interaction design for secure systems. In *Proc. 4th International Conference on Information and Communications Security* (Singapore, Dec. 2002), R. Deng, S. Qing, F. Bao, and J. Zhou, Eds., vol. 2513 of *Lecture Notes in Computer Science*, Springer Verlag, pp. 278–290. ISBN 3-540-00164-6.

[38] ZVIRAN, M., AND HAGA, W. J. Cognitive passwords: The key to easy access control. *Computers & Security 9* (1990), 723–736.