Undercover: Authentication Usable in Front of Prying Eyes

Hirokazu Sasamoto
CMU/CyLab Japan and
Sharp Corporation
sasamoto@itl.tnr.sharp.co.jp

Nicolas Christin
CMU/CyLab Japan
nicolasc@cmu.edu

Eiji Hayashi
CMU/CyLab Japan and
Mitsubishi Research
eij@pam-ya.com

ABSTRACT

A number of recent scams and security attacks (phishing, spyware, fake terminals, ...) hinge on a crook's ability to *observe* user behavior. In this paper, we describe the design, implementation, and evaluation of a novel class of user authentication systems that are resilient to observation attacks.

Our proposal is the first to rely on the human ability to simultaneously process multiple sensory inputs to authenticate, and is resilient to most observation attacks. We build a prototype based on user feedback gained through low fidelity tests. We conduct a within-subjects usability study of the prototype with 38 participants, which we complement with a security analysis.

Our results show that users can authenticate within times comparable to that of graphical password schemes, with relatively low error rates, while being considerably better protected against observation attacks. Our design and evaluation process allows us to outline design principles for observation-resilient authentication systems.

Author Keywords

Usability, Security, Multisensory processes

ACM Classification Keywords

K.6.5 Security and Protection: Unauthorized access; H.5.2 User Interfaces: Haptic I/O

INTRODUCTION

Alice is at an automated teller machine, about to withdraw money. Security cameras monitor the surroundings, and a couple of people are standing in line behind her. Unpleasant thoughts start to cross her mind: Are these real security cameras, not something set up by a thief to record her typing her card's access code? Speaking of which, why is that gentleman behind her staring at the keypad while she is entering her code? And, is that a real banking terminal, or just a very good imitation placed there by crooks with the only goal to steal her credit card information?

CHI 2008, April 5–10, 2008, Florence, Italy.

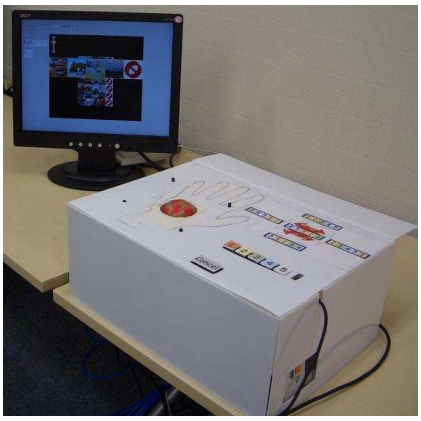


Figure 1. Undercover prototype. A graphical display presents a set of images to the user, who is asked to identify the pictures she selected in advance. A haptic device (here, a trackball), to be covered by the user's hand, randomly determines which button corresponds to each possible answer.

Alice is worried about *observation attacks*, where an unauthorized party can record secret information, such as Alice's access code, and later use it to impersonate its legitimate owner. Alice has a good reason to worry: Banking terminal fraud in the United States only is estimated at $60 million/year [10], and most countries outside the United States do not provide banking customers with any form of legal protection [2].

Contrary to more conventional security threats, such as cryptanalysis or attacks on communication protocols, observation attacks have the particularity of compromising the security of a system by monitoring user behavior rather than subverting the system itself. Observation attacks are not restricted to banking terminals. Under this definition, social engineering attacks such as phishing [7], can be considered as an instance of observation attacks.

We focus here on how to foil observation attacks during *authentication*, the process through which an individual proves her identity to someone else. In addition to withdrawing money, a number of daily activities, such as turning on a cellular phone or accessing a computer system, require authentication. Authentication poses both usability and security challenges. A legitimate user must *always* be able to successfully authenticate (usable authentication), while unauthorized parties should *never* be able to authenticate (secure authentication). Simultaneously achieving both usable and

secure authentication is notoriously difficult (see [5] for a good overview of issues at the interface between usability and security), particularly for systems with a large, varied user base.

Biometric authentication, e.g., by fingerprint, offers a promising alternative from a usability standpoint, but remains extremely vulnerable to certain observation attacks. Fake banking terminals can indeed capture biometric information, just like they capture access codes and credit card numbers [2]. Biometric information could be subsequently replicated to impersonate a legitimate user [17].

In this paper, we investigate the challenges faced in designing authentication schemes resilient to observation attacks. The key idea is that, to thwart observation attacks, at least part of the authentication process must be difficult (or impossible) to observe; a property that can be achieved by designing the system so that the user *has to* hide part of the process to authenticate.

To accomplish this, we consider a novel class of authentication systems, that challenge the user with a puzzle combined with a signal hidden from observers, by sending two simultaneous sensory inputs.

Figure 1 shows a prototype of our proposed *Undercover* authentication system. A graphical display presents several images, and asks which, if any, belongs to an image portfolio previously chosen by the user. A trackball simultaneously sends the user a signal that conditions the mapping of each button to a given answer. To function, the trackball has to be covered by one of the user's hands, making its movement very hard to observe by an outsider.

We make several contributions through this work. First, we propose a novel method to make authentication systems resilient to a general class of observation attacks, including shoulder surfing, fake terminals, or even spyware (programs that surreptitiously record all user input, [22]). We identify the design parameters that play a role in authentication through hand-eye coordination, and build a working prototype based on extensive user feedback gained through low-fidelity tests. Through a within-subjects usability study with 38 participants, we demonstrate that authentication systems can rely on the human ability to mentally combine tactile and visual signals. Finally, we perform a security analysis of Undercover, by trying to capture user portfolios from external observations, and outline some design principles for future observation-resilient authentication systems.

RELATED WORKS

Most authentication schemes are vulnerable to observation attacks because the user has to explicitly input a secret, for instance a password, that positively identifies her. By observing user input, an impostor can capture the secret and later use it to pass for the user.

Research has primarily focused on methods to verify the user knows the secret, without requiring her to explicitly state it. For instance, zero-knowledge proof systems [11] interactively ask a series of questions. The user can answer correctly each question only if she knows the secret, but each answer does not reveal any information about the secret.

The "cognitive trapdoor" game proposed by Roth et al. [19] attempts to implement a zero-knowledge proof system. In this game, the user's password is a number. The system randomly assigns colors (black and white) to digits, graphically displays the color assignment, and asks the user if the first digit of her password is white or black. The process is then repeated by re-assigning colors at random for each digit. The probability an impostor who does not know the password gains access to the system exponentially decreases in the number of digits in the password.

Other recent work on authentication schemes resilient to observation attacks [16,24] uses images, called "graphical passwords," as authentication tokens. The rationale for using graphical passwords is that they are arguably easier to memorize than long strings of text or numbers [6, 18]. During an initial setup phase, the user selects a few images to constitute a private portfolio. She later authenticates by identifying her portfolio images from a set of decoy images.

To thwart observation attacks, instead of directly selecting portfolio images, the user can input information derived from to the portfolio images and their location, either by keyboard [16], or graphically [24]. Alternatively, users can mentally compute a path formed by their portfolio images, and give an answer based on that mentally computed path [23].

To be effective, graphical password schemes resilient to observation attacks [16, 23, 24] require a significant number of objects (decoys and portfolio images) to be displayed simultaneously, which challenges their usability. In addition, while all techniques discussed so far [16, 19, 23, 24] are resilient to single observation attacks, repeated observations can yield authentication tokens (number or graphical passwords), by comparing user inputs from different sessions [6, 12].

Another technique conceals user input by solely relying on the user's gaze to enter a password [14]. While promising, the approach requires expensive eye-tracking equipment. Additionally, a rogue eye-tracker hidden close to the authentication device could capture user passwords. Varying finger pressure is another possible input resilient to observation attacks [15], but the authors' own user study shows that participants find it challenging to use.

Compared with existing work, our method advocates a radically different approach to the problem. Instead of trying to conceal user input, we instead hide the *authentication challenges* users are presented with. If an authentication session can be thought of a question-answer exchange ("What is the password? – 'Buddy'."), proposals so far have focused on trying to hide the answer ('Buddy'). We conjecture hiding (part of) the question is more usable and secure.

OBJECTIVES

While our main stated goal is resilience to observation attacks, a practical authentication system must fulfill a number of security and usability objectives.

Security objectives

At a high level, the authentication system should be resilient against observation attacks, while being as secure as comparable systems against other attacks.

No unauthorized access. The authentication system should not grant access to an impostor with a probability higher than that of obtaining the correct selection of authentication tokens by pure chance. We aim for a level of security comparable to that provided by authentication systems using a four-digit personal identification number (PIN). That is, an impostor should not gain access to the system with a probability higher than 1/10,000.

Resilience to remote observation. A third-party able to record from a distance, e.g., using video equipment, one or more authentication sessions should not be able to gain knowledge sufficient to impersonate a legitimate user.

Resilience to internal observation. A third-party able to record keystrokes, mouse clicks, or the contents of authentication display by commandeering the authentication terminal itself, e.g., using spyware or a fake terminal, should not be able to gain knowledge sufficient to impersonate a legitimate user.

Resilience to repeated observations. The authentication scheme should be resilient to long-term, repeated observations of multiple authentication sessions. Observing a user log in N times should yield as little information as observing the same user log in just once.

Resilience to social engineering. Users should not be able to easily reveal their authentication tokens to a third party, be it by mistake or by choice.

None of the security properties above listed restricts the size of the authentication system to be devised. These requirements could equally apply to banking terminals, PCs, or even mobile devices.

Usability objectives

To be viable for systems with a large, varied, user base, an authentication system has to satisfy a set of usability constraints.

Low error rate. Legitimate users' mistakes in the authentication process should be a rare event. This requirement demands that the authentication tokens be easily memorable, and that the system be intuitive to use.

Fast authentication. Any user should be able to complete an authentication session in a limited amount of time. To ensure the scheme has an authentication time comparable to that of graphical password authentication schemes [6], we set the upper bound on the authentication session to one minute.

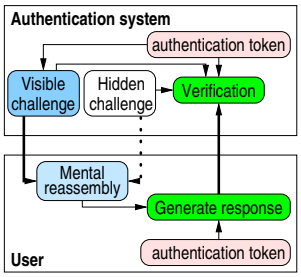


Figure 2. Overview of the design. The authentication system sends both a visible and a hidden challenge to the user, who reassembles them mentally, and answers the reassembled challenge. Only the visible challenge and the response can be eavesdropped on, and cannot be correlated.

Rapid training. The authentication system should be intuitive enough so that a user with no prior training can understand how to use it in five minutes or less.

Physically safe. The authentication system should never cause any physical harm to the user even if operated incorrectly. This requirement prohibits the use of sharp parts. Further, it is highly desirable that users do not feel the slightest discomfort while using the system.

DESIGN

We design a challenge-response authentication system. That is, the system asks the user a question (challenge), to which the user has to properly respond to prove her identity. To thwart observation attacks, rather than hiding the response, we choose to hide the challenge.

Completely hiding the challenge may prove a difficult task, especially when considering usability factors. As illustrated in Figure 2, we instead choose to hide *part* of the challenge, by breaking it into two halves. The first half of the challenge is conveyed through a visible (and hence, observable) channel, while the second half of the challenge is conveyed through a hidden channel. The user mentally reassembles both parts of the challenge, and, using her authentication token(s) as an input, answers the reassembled challenge. The authentication system can verify the answer by combining its own knowledge of the authentication token(s) with its knowledge of what was sent on the hidden channel.

Going back to our earlier example, instead of asking "What is the password?" we use the following kind of challenge-response exchange. We ask aloud, for instance, "Does your password contains a 'd'?" (visible challenge) while whispering to the user's ear "Tell a lie," or "Tell the truth" (hidden challenge). If an outsider is unable to decipher the whisper in the ear, the user's answer does not give her any clue whether there is a 'd' in the password or not. We have reduced the problem from hiding the complete challenge to hiding one (or a few) bit(s) of information.

Visible and hidden challenges

If the hidden challenge is kept perfectly secure, that is, if 1) it cannot be observed by outsiders, and 2) its contents is perfectly random (e.g., the decision whether to ask the user to

lie is made by flipping a coin), then the system is equivalent to a one-time pad, which is mathematically proven impossible to break [21].

In practice, guaranteeing perfect security of the hidden challenge is a difficult task. We do not address in this paper the question of random input generation, and instead refer the reader to the literature available on the subject (see for instance [21] for an overview). The main question we tackle here is how to make the channel conveying the hidden challenge impossible to observe by an outsider.

The above example of a hidden auditive channel (whisper in the ear) is potentially vulnerable to remote eavesdropping using high quality recording equipment such as parabolic microphones. Implementing a hidden visual channel may be challenging for similar reasons.

On the other hand, a tactile challenge can be made hard to eavesdrop on, and limited amounts of information can be easily encoded as tactile signals. A particularly interesting feature of a tactile challenge is that the user needs to make physical contact with the communication channel to determine the signal being transmitted. That physical contact, e.g., pressing the palm against a surface, makes the user hide the channel from outsiders without having to think about it. Hence, such a device naturally requires the user to adopt a secure practice.

Our usability requirements mandate that authentication tokens must be easily memorable. Self-chosen graphical passwords, where the user selects a collection of personal images (portfolio) as an authentication token, are an appealing candidate. They have indeed been shown to be easy to memorize [6, 18], and can also be made resilient to social engineering through the use of distortion [13].

In addition, using self-chosen graphical passwords makes it very difficult for a crook to build a fake authentication terminal. Because the terminal needs to display the user portfolios, the crook would need to know *in advance* all images with which all users may possibly wish to authenticate. Failing that, people would quickly realize something is wrong with the terminal.

Last, combining the tactile and visual challenges must be a straightforward operation for the user. Graphical passwords can rely on a trivial authentication function, asking if any of the displayed pictures is the user's, which should facilitate the mental operation to perform.

Design parameters

The next question to tackle is how to combine the visual and tactile challenges. Studies, e.g., [3, 8], have showed that tactile stimuli could reinforce visual or audio-visual stimuli. We refer the reader to the literature, e.g. [4], for a more thorough description of our current understanding of the interactions between different sensory inputs.

The complexity of reassembling information from the tactile and visual channels depends on the amount of information conveyed through the tactile channel. We face a trade-off between the amount of information users must memorize and the amount of information users must process.

Indeed, the less information the tactile channel contains (e.g., two values as in the "Tell a lie/Tell the truth" example from before), the easier it is to implement, and the easier it is for the user to perform the mental reassembly task. However, the less information comes from the tactile channel, the more authentication tokens the user has to memorize. If the tactile channel is limited to two values, the visual challenge can only be a yes/no question, e.g., "Is this picture part of your portfolio?" To get security comparable to a four-digit-PIN-based system, users would have to successfully complete 14 authentication challenges in a row, which means memorizing at least 14 different pictures.[1]

Conversely, if the tactile device can convey five values, the visual challenge can become "Which of these five pictures belongs to your portfolio?" Because $5^6 > 10,000$, such a challenge reduces the number of portfolio pictures to memorize to 6. However, the haptic device implementing the tactile challenge becomes more complex, and the mental operation the user performs is probably more challenging.

IMPLEMENTATION

We explore the trade-offs between the different design parameters by building a number of mock prototypes, and conducting a series of informal, low fidelity tests with a limited number of users. Results of the low fidelity tests lead us to iterate our design before implementing a full prototype, which we later describe.

Low fidelity tests

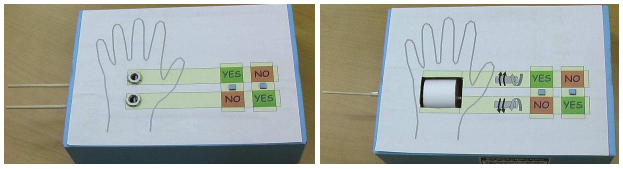
For brevity, we only detail here the two most important low fidelity tests we conduct. The objective of the first test is to determine which haptic device is most usable. The second low fidelity test helps us decide the size of the information channel the haptic device implements.

We construct our low fidelity prototypes using reusable cardboard boxes for the haptic device, and a PowerPoint display for the visual challenges. These reusable prototypes provide a realistic approximation of the user interface of the Undercover prototype.

Determining the type of haptic device. We identify (punctual) contact and friction as the two main design alternatives for the haptic device.[2] In the first low fidelity test, we compare how fast and accurately users manage to authenticate, using two different prototypes relying on contact and friction, respectively.

[1] An impostor selecting at random has a 50% chance of being right at each authentication stage. Thus, the number of stages n must satisfy $1/2^n \leq 1/10000$ which implies $n \geq 14$. Moreover, each stage has to be statistically independent from the others, which imposes at least 14 different portfolio images.

[2] Other possibilities, such as temperature changes, are likely to be more observable, e.g., with an infrared reader.



(a) Contact-based prototype (b) Friction-based prototype

Figure 3. Prototypes used in the first low-fidelity test to determine which type of haptic device is more usable.

	Participant	1	2	3	4
	Age	30s	40s	50s	50s
	IT experience	yes	yes	no	no
Pins	**Auth. time (s)**	73	60	110	100
(Contact-based)	**Error rate (%)**	0	7	21	21
Cylinder	**Auth. time (s)**	70	45	113	95
(Friction-based)	**Error rate (%)**	0	0	14	7

Table 1. Results of the first low-fidelity test. Participants authenticate in general faster, and with fewer errors, using the friction-based prototype (cylinder).

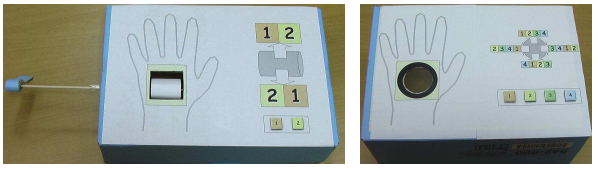
The first prototype, represented in Figure 3(a), implements a punctual contact with the user, thanks to two pins manually actioned by the experimenter. Exactly one of the two pins is up at any given time (i.e., both pins cannot be simultaneously up or down). Participants place their left hand over the pins, and press one of two buttons to answer a challenge given on a PC screen (not shown in the figure). As the figure shows, the meaning of the two answer buttons depends on which pin is lifted. If the top pin is lifted, the left button means "NO." If the bottom pin is raised, the left button means "YES."

The second haptic prototype, shown in Figure 3(b), relies on friction. Here, the meaning of the answer buttons is given by the sense of rotation of a cylinder, also manually actioned by the experimenter. For instance, if the cylinder rotates upwards, the left button means "YES."

The simulated authentication process consists of seven challenges. Each challenge, presented on the PC screen, simply asks "Press YES," or "Press NO," while the experimenter actions the haptic device to generate random tactile challenges.

We conduct this first low-fidelity test with four participants, ages ranging from the 30s to the mid-50s, and with different backgrounds (only two of the participants have background in information technology). Each participant goes through a complete authentication session with each prototype. We measure the time to complete all seven challenges, check the error rate, and receive feedback from the participants.

Our findings, summarized in Table 1, are that 1) younger participants authenticate faster, and more accurately than older participants, and 2) participants authenticate faster and more accurately using the friction-based (cylinder) prototype. Comments from the participants indicate their hand gets numb after a number of trials with the pin-based prototype, and that the cylinder-based prototype is easier to use. One participant acknowledges some difficulty in concentrating on both the screen and the tactile display simultaneously.



(a) Two-value tactile channel (b) Four-value tactile channel

Figure 4. Prototypes used in the second low-fidelity test to determine how complex the information in the hidden channel can be made without affecting usability. (The trackball is not shown in this picture.)

We repeat a similar test with two other prototypes, using four pins, and a trackball that can rotate in four directions. Results confirm that friction-based prototypes outperform contact-based prototypes.

Dimensioning the tactile channel. The more authentication tokens users have to remember, the more likely they are to make mistakes. By increasing the amount of information the tactile channel conveys, we can reduce the number of required authentication tokens, which could help us achieve lower error rates. The second low-fidelity test helps us dimension the amount of tactile information users can comfortably process.

We use two friction-based prototypes made out of cardboard boxes. The first prototype, shown in Figure 4(a), uses a cylinder, which provides two possible values (up, down) for the tactile channel. The second prototype, shown in Figure 4(b), uses a trackball, which provides four possible values (up, down, left, right) for the tactile channel.

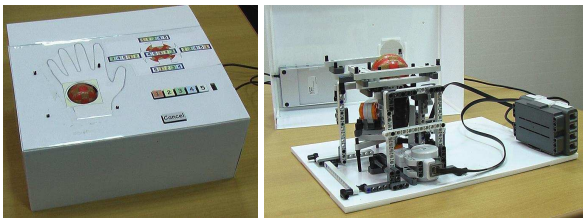
In each authentication stage, the display shows a row of two or four pictures depending on the prototype used. Users are asked to press a button, given the location of their portfolio image on the screen, and the movement of the haptic device. Buttons are color-coded to assist users in making the correct selection, and a mapping table is printed on the haptic device. For instance, if the cylinder of Figure 4(a) is moving down, and the portfolio image is on the left, the user should press the button marked "2."

Authentication with the cylinder-based prototype requires fourteen challenges, each consisting of two pictures, one of them being in the user's portfolio. Given the increased size of the tactile channel, the trackball-based prototype uses only seven challenges. Because this test only aims to quantify how users interpret the information coming from the hidden channel, the PC screen directly shows the solution using check marks and crosses in lieu of actual portfolio or distractor images, and the experimenter speaks out the movement of the haptic device in addition to manually actioning it.

We hold this second test with three participants, who were not involved in the first low-fidelity test. Ages range from early 30s to mid-40s; one of the participants does not have any IT background. Each participant tries a complete authentication session with both prototypes. As in the first low-fidelity test, we record the time to complete the full authentication phase, the error rates, and gather user feedback.

	Participant	1	2	3
	Age	30s	40s	30s
	IT experience	no	yes	yes
2 values	Authentication time (s)	45	45	22
	Error rate	1/14	0	0
4 values	Authentication time (s)	20	20	13
	Error rate	1/7	0	0

Table 2. Results of the second low-fidelity test. Errors marginally increase with a 4-value tactile channel, while considerably lowering the authentication time.



(a) Outside view (b) Inside view

Figure 5. Haptic device and associated machinery. The authentication system software, hosted on an external PC (not shown), pilots two servo motors that in turn govern the trackball movement.

We report our results in Table 2.

The main lesson from this test is that a four-value tactile channel drastically reduces the total authentication time. Even considering that the larger size of the tactile channel allows to reduce the number of authentication stages, the result is surprising, in that the added complexity does not slow down users. Likewise, the number of errors made are comparable, which further evidences that users can use a four-value tactile channel almost as accurately as a two-value tactile channel. This is confirmed by feedback gained after the session – users tell us they prefer the more complex version because it shortens the whole experiment, without being more difficult.

Further low fidelity tests inform us that adding a "vibrate" mode to the trackball, which brings a fifth possible value to the tactile channel (up, down, left, right, vibrate), has actually no impact on the ability of people to select the right button, while reducing the number of challenges with which the users need to be presented.

Undercover prototype

Based on the outcome of our low fidelity tests, we undertake the realization of a full-scale prototype of our Undercover authentication system, which we show in Figure 1. The authentication system is hosted on a PC, which interfaces with an external haptic device through USB connections. The haptic device consists of a trackball, whose movement is governed by a Lego Mindstorm NXT [1] robot. As shown in Figure 5, the haptic device is embedded in a plastic case, which shields the underlying machinery from users. The plastic case also contains five numeric buttons and a Cancel button for users to enter their answers to the challenges given on the PC display.

We try to make the Undercover prototype look as much as possible as a stand-alone device rather than a PC-controlled system. Our main use case here is banking terminal authentication, so that we do not have stringent spatial constraints. Our prototype is 37cm × 26cm × 16cm, not including the screen. The design can be easily miniaturized to adapt Undercover to mobile devices. For instance, piezoelectric components can be used to create miniaturized haptic devices with a functionality similar to our trackball.

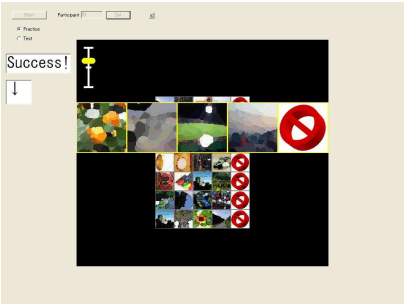


Figure 6. Visual channel. Users proceed through a series of seven challenges. The current challenge is zoomed in. Users are asked to identify which, if any, of the four pictures belong to their portfolio. The fifth ("NO") symbol is used when none of the pictures are the user's.

In our prototype, users rely on a portfolio of five images as authentication tokens. These images are self-selected, for instance, they are pictures the user has taken with a digital camera. The authentication process consists of a series of seven visual challenges. Each challenge consists of four images, and a fifth image showing "None." Users are asked to identify the location of their portfolio image among the five possibilities; if none of the displayed images is in the user's portfolio, the user is expected to select "None." The addition of the "None" input, as opposed to a fifth image, allows us to obtain a number of possible inputs larger than 10,000, while reducing the number of portfolio to memorize down to five. So, there are a total of 28 pictures shown throughout an authentication session, 5 of them being portfolio images, and the 23 others being distractor images. We never repeat any picture to make the system resilient to multiple observations, so that there are 20,480 possible inputs to the system.[3] Hence, a brute force attack is harder to convey here than on a traditional four-digit-PIN-based system.

Figure 6 shows how the visual channel presents challenges to the user. The current challenge is zoomed in, but the user can see the other challenges in the background. On the left side, indicators show if the previous challenge was successfully answered, and what the movement of the trackball was. This feedback is only given during a practice phase, and is disabled during an actual authentication session.

The user proceeds from one challenge to the next by pressing one of the five buttons, or the "Cancel" button to go back to the previous challenge, situated next to the trackball.

[3] 20,480 is obtained by noting that exactly two of the seven stages have to be answered "None". We can then only consider the four other possibilities (i.e., excluding "None") for the five remaining stages, yielding $\binom{7}{2} \cdot 4^5 = 20,480$ possibilities.

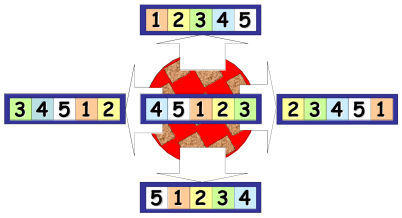


Figure 7. Map shown to the users to explain how each position on the screen maps to a different button, depending on the trackball movement. The center position corresponds to the "vibrate" mode.

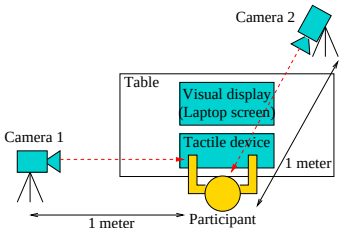


Figure 8. Experimental setup. Participants are in an isolated room, only accompanied by the experimenter. Two tripod-mounted cameras record each participant's hands and eye movements, as well as any noise in the room.

The trackball is governed by two servo motors to move in five directions (up, down, left, right, vibrate). The first motor rotates the ball by friction along an horizontal axis. The second motor rotates the first motor by 90 degrees when needed. Vibration is realized by quickly alternating the sense of rotation of the first motor. We paste small pieces of cork onto the surface of the trackball to make it more rugged and improve usability [9].

To assist users in combining the movement of the trackball with the visual display, a map, shown in Figure 7, is pasted on the plastic case hosting the trackball. The map describes how each position on the screen corresponds to each button, according to the trackball movement. Buttons are both numbered and color-coded to reinforce user perception. As an example, if the portfolio image is the third image from the left, and the trackball moves left, the map tells the user she should press the white button "5."

EVALUATION

We conduct a formal usability test to evaluate the Undercover prototype. A total of 38 people (4 students and 34 government employees) participate in our study. Participants' ages range from early twenties to late fifties with a median age in the mid-thirties. All participants have some information technology background but not necessarily computer expertise. None of the 38 participants has taken part in any of our low-fidelity tests. The usability test relies on a within-subject design, that is, each participant goes through the same set of experiments.

Procedure

Prior to the experiment, participants are asked to bring five personal digital images, and to select a four-digit PIN. Our usability test consists of a single session divided between a training phase, a control phase, and two authentication phases. All sessions are conducted in an isolated room, with a set-up as described in Figure 8. Each participant is alone in the room with the experimenter during the whole test. The complete usability test takes about 30 minutes per participant, including 5 minutes of training.

Figure 9. Visual challenges. During the training phase, the user is given the answer, e.g. here, the third image from the left. In the first type of authentication phase, the original portfolio images are placed among a set of distractor images. In the second type of authentication phase, all images are distorted. If none of the pictures is a correct answer, participants are to select the rightmost "NO" sign.

To help us with a security analysis of the system, which we detail later, two cameras record each participant's hand and eye movement and any noise in the room, allowing us to carry out a powerful observation attack. We do not tell users why these cameras are here, other than mentioning everything is recorded for experimental purposes. We tell each participant that they should use the authentication system as if it were a banking terminal. Also, we play music in the room to cover the small motor noise, to avoid any bias in the participants' perception of the trackball movement.

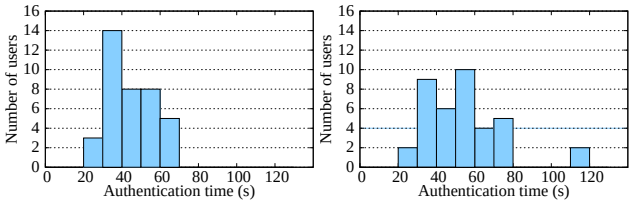
Training phase. The training phase consists of seven authentication challenges with five pictures displayed per challenge. As in our low-fidelity tests, instead of showing portfolio and distractor images, the visual display directly gives the solution, as shown in Figure 9(a). Users are to select the picture denoted by a circle. If only slashed circles are displayed, participants should select the rightmost "NO" sign.

Authentication phases. Each participant goes through two different authentication phases. Both phases consist of seven challenges containing five pictures each and are similar save for the type of pictures used, as shown in Figure 9(b) and (c).

In the first type of authentication phase, the five pictures displayed consist of at most one portfolio image and at least four distractor images.[4] In the second type of authentication phase, the mix of portfolio-distractor images is the same, but all images are distorted prior to the experiment, using an oil painting filter.

The distortion operation helps make the system resilient to

[4]To be more precise, exactly one portfolio image is displayed in five of the seven challenges, and there are no portfolio images displayed in the other two challenges.



(a) Undercover (b) Undercover w/ distortion

Figure 10. Distribution of authentication times (lower is better). The median authentication times are 35 seconds (variance = 403.57) for the training phase, 32 seconds (variance = 144.35) for Undercover, and 45 seconds (variance = 415.39) for Undercover with distortion.

social engineering attacks [13]. Knowing that the user owns a white dog, for instance, could lead an attacker to correctly identify a picture of a white dog as part of the user's portfolio. With distortion, however, the attacker cannot figure out whether the image represents a white dog, a snowman, or a chicken. Conversely, the legitimate user can quickly identify her distorted portfolio image [13]. Thus, using distortion arguably enhances security. The main question we try to answer through this experiment is whether the cognitive load imposed on users – recognizing a distorted picture, and combining this information with the input from the haptic device – becomes too high to ensure relatively low error rates and authentication times.

Control phase. The control phase is used to compare the results obtained with the Undercover prototype with those obtained for a four-digit PIN authentication system. We build a PIN-based authentication device of similar shape and size as the Undercover prototype. The device consists of a single numerical keypad, and mimics a classical banking terminal. Participants are asked to enter their PIN to authenticate. User input is shown on a screen using stars in lieu of the actual numbers keyed in. Because PINs are self-selected, we expect low (or null) error rates, and very short authentication times. We run this control experiment to have a more precise idea of the baseline values for both metrics.

To reduce the impact of fatigue or training effects in our measurements, the order in which both authentication phases and the control phase take place is randomized for each user.

Usability evaluation

We rely on measurements of authentication times and error rates to assess the usability of our system.

Figure 10 provides the distribution of authentication times for Undercover, along with the median authentication times. Median authentication times remain below our self-imposed one-minute criterion, but a few users take longer to authenticate, especially when picture distortion is added to the authentication scheme. In comparison, PIN-based authentication has a median authentication time of 3.2 seconds (variance = 1.1), which gives an idea of the added complexity of Undercover. Authentication with original, non-distorted pictures is significantly easier than authentication with distorted pictures, as evidenced by the lower median authentication times and variance, and shorter distribution tail.

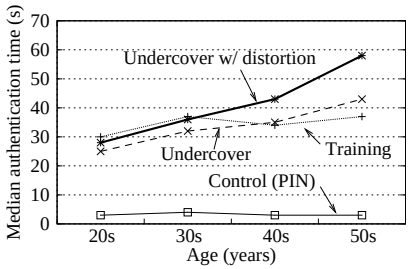


Figure 11. Median authentication times vs. age (lower is better).

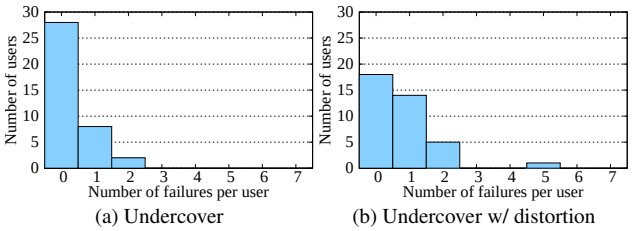
Authentication type	Control (PIN)	Original pictures	Distorted pictures
Failed challenges	1/152	12/266	29/266
Failed sessions	1/38	10/38	20/38

Table 3. Error rates. Failure to answer a single challenge implies failure of the whole session.

In post-experiment interviews, we ask users how they feel about the length of the authentication procedure, using a five-point scale, where 1 is "short" and 5 is "long." 17 participants consider the process to be a bit lengthy (answers 4 or 5), 18 participants consider it acceptable (answers 1, 2 or 3), and 3 do not answer. 15 of the users answering 4 or 5 further indicate that authentication times should be below 15 seconds.

Figure 11 plots the median time to authenticate against the participants' age. Authentication times are largely independent of age for PIN-based authentication. On the other hand, older participants take slightly more time authenticating with Undercover. This may be due to the mental reassembly process of the visual and tactile information becoming slightly slower with age, similar to effects previously observed [20], or to a certain loss in tactile perception as the experiment progresses. Authentication times are largely independent of the amount of information technology experience of each participant.

Table 3 reports the authentication error rates. Each authentication session consists of seven challenges for Undercover, and four challenges (corresponding to each digit) for the PIN-based authentication. Because failing to answer correctly a single challenge leads to failing the whole authentication session, even a low error rate over all challenges can translate into fairly high authentication failure rates. In other words, seven challenges per session may still be too many.



(a) Undercover (b) Undercover w/ distortion

Figure 12. Distribution of number of failures per user (lower is better).

Figure 12 elaborates on the error rates, by showing the distribution of failures per user, and invites optimism. While the addition of distortion seems to put too high a cognitive load on some users (e.g., the participant who makes five mistakes), most participants make zero or only one mistake over an entire session. Considering how unfamiliar participants are with the scheme, and the limited amount of training they receive, compared to PIN-based systems they have used their entire life, this result seems encouraging. With additional training, we expect people to further lower their error rates.

Finally, a graph we omit for brevity shows that error rates increase with age but remain independent of information technology experience, which confirms the insight Figure 11 provides: Undercover is slightly more usable for younger users.

Security evaluation

By videotaping each participants' hand and eye movement, and given our own knowledge of the authentication system, we can conduct a powerful observation attack on Undercover. The questions this security evaluation attempts to answer are 1) what could cause an observation attack (on the evaluated system) to be successful, and 2) what is the extent of the damage to which a successful observation attack can lead. In particular, we want to find out if we can recover some of our participants' authentication tokens (image portfolios or PINs).

First, we are able to recover *all* PINs the users selected, thereby confirming that PIN-based authentication is insecure. Even the more security conscious users who slightly try to cover their right hand with their left hand fall victim to at least one of the two cameras. A cover on top of the keypad would reduce usability, as users could hardly see what they type, and would not prevent observation attacks, as it would still be possible to place a small camera inside the cover.

Next, a vast majority of participants fail to completely cover the trackball, so that we can frequently infer its movement. This could be easily fixed, by fitting the user's hand into a given position, e.g., using a glove-shaped cover. Such a tight cover should not affect usability, while preventing small cameras to be placed inside.

Further, a number of users involuntarily leak information in three different manners. Some participants point the map pasted on the plastic case, revealing either trackball movement or portfolio image location. Some others move their hands to get a better sense of the movement of the trackball thereby revealing information about its sense of rotation. Last, a couple of users say something out loud about the movement of the trackball or their portfolio images.

Table 4 shows how frequently information leaks occur, and the consequences. 9 participants out of 38 leak information at one point or another (some participants leak information in multiple ways). The most frequent problem is the user pointing at the map pasted on the plastic case, which very often leads to revealing the challenge answer to an observer.

Pointing at the map	Other hand movement	Saying something	**Challenges affected**	**Correct observer guesses**
•			78/532 (14.7%)	71/78 (91%)
	•		34/532 (6.4%)	26/34 (76.5%)
		•	5/532 (0.9%)	1/5 (20%)
•		•	8/532 (1.5%)	8/8 (100%)
•	•		16/532 (3.0%)	12/16 (75%)
	•	•	0/532 (0.0%)	N/A
•	•	•	1/532 (0.2%)	1/1 (100%)

Table 4. Types of information leaks and their consequences. Information leaked during a challenge usually allows an observer to correctly guess the solution to the challenge, but information leaks occur rarely with Undercover. In comparison, we could recover the PINs of all 38 participants.

However, over the total number of challenges ($7 \times 2 \times 38 = 532$) the frequency of information leaks remains low. Moreover, we cannot find any correlation between information leaks and user error rates. That is, participants who leak information do not have better (or worse) authentication success rates as participants who do not. This result is encouraging, as it shows that getting rid of information leaks, e.g., through education (for instance by pasting a "Don't point at the map" warning next to the map), would not reduce usability of the system.

Finally, outside of the lab, we expect users to disclose far less information. Indeed, some of the participants who leak information are acting as if the experiment is an exam, and are apparently seeking confirmation or encouragement from the experimenter.

DISCUSSION AND CONCLUSIONS

This paper tackles the issue of securely and easily proving one's identity, even while being spied on. We present what we believe to be the first authentication system that achieves resilience to observation attacks by hiding the challenges posed by the system, rather than the responses coming from the users.

We demonstrate feasibility of our proposed technique with a proof-of-concept implementation. Our Undercover authentication system exploits the human ability to combine tactile and visual information to answer relatively complex challenges. We build a prototype of Undercover for banking terminal authentication, by going through an iterative design extensively relying on user feedback. Undercover could also be implemented on small devices.

A within-subjects usability study with 38 participants, complemented by a security assessment shows that such an authentication system is viable. Authentication times remain comparable with those of graphical password authentication schemes, which is acceptable for half of the users (even though future work must strive to reduce these authentication times).

Error rates, while not negligible, are encouraging, and should further decrease as users get more familiar with the system. Resilience to observation attacks is considerably enhanced, and can be further improved.

Our study also leads to a number of observations and questions that we can use to outline design guidelines for future systems.

A key finding from the security standpoint is the importance of minimizing the amount of information users can involuntarily disclose. For instance, in our prototype, hand movement may reveal otherwise hidden information. More generally, authentication systems may have a large number of "covert channels" leaking information. For instance, different input keys make different clicking sounds [25] revealing what the user is typing. Likewise, the motor noise in our prototype needs to be covered to ensure the trackball movement cannot be inferred from it. The advantage of systems like Undercover is that possible information leaks (either from the user, or from the system) are much more limited in scope, and seem much easier to address than in traditional authentication systems.

Finally, the degree of complexity that two independent sensory signals can present while being successfully reassembled by a majority of people comes a bit as a surprise. While most cognitive studies thus far focused on reinforcement of a visual signal with a tactile or auditive aid, our study shows that people are also very skilled at dissociating and recombining independent sensory inputs. We do believe this finding can have an impact far beyond the security application discussed in this paper.

REFERENCES

1. LEGO.com Mindstorm NXT home. `http://mindstorms.lego.com`.
2. R. Anderson. Why cryptosystems fail. In *Proc. ACM CCS'93*, 215–227, 1993.
3. P. Blamey, R. Cowan, J. Alcantara, L. Whitford, and G. Clark. Speech perception using combinations of auditory, visual, and tactile information. *J. Rehab. Res. and Dev.*, 26(1):15–24, 1989.
4. G. Calvert, C. Spence, and B. Stein, editors. *The Handbook of Multisensory Processes*. MIT press, 2004.
5. L. Cranor and S. Garfinkel, editors. *Security and Usability: Designing Secure Systems That People Can Use*. O'Reilly Media, 2005.
6. R. Dhamija and A. Perrig. Déjà vu: A user study, using images for authentication. In *Proc. 9th USENIX Sec. Symp.*, 2000.
7. R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. In *Proc. ACM CHI'06*, 581–590, 2006.
8. A. Diederich, H. Colonius, D. Bockhorst, and S. Tabeling. Visual-tactile spatial interaction in saccade generation. *Exp. Brain Res.*, 148(3):328 – 337, 2003.
9. E. Gamzu and E. Ahissar. Importance of temporal cues for tactile spatial-frequency discrimination. *J. Neuroscience*, 21(18):7416–7427, 2001.
10. L. Giesen. ATM fraud: Does it warrant the expense to fight it? *Banking Strategies*, 82(6), 2006.
11. S. Goldwasser, S. Micali, and C. Rackoff. The knowledge complexity of interactive proof-systems. In *Proc. ACM STOC'85*, 291–304, 1985.
12. P. Golle and D. Wagner. Cryptanalysis of a cognitive authentication scheme. In *Proc. 2007 IEEE Symp. Sec. Privacy*, 66–70, 2007.
13. E. Hayashi, N. Christin, R. Dhamija, and A. Perrig. Mental trapdoors for user authentication on small mobile devices. Tech. Rep. CMU-CyLab-07-011, Carnegie Mellon Univ., 2007.
14. M. Kumar, T. Garfinkel, D. Boneh, and T. Winograd. Reducing shoulder-surfing by using gaze-based password entry. In *Proc. SOUPS'07*, 2007.
15. B. Malek, M. Orozco, and A. El Saddik. Novel shoulder-surfing resistant haptic-based graphical password. In *Proc. EuroHaptics'06*, 2006.
16. S. Man, D. Hong, and M. Mathews. A shoulder-surfing resistant graphical password scheme. In *Proc. Int. Conf. Sec. Mgmt.*, 105–111, 2003.
17. T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. Impact of artificial gummy fingers on fingerprint systems. In *Proc. SPIE*, vol. 4677, 275–289, 2002.
18. W. Moncur and G. Leplâtre. Pictures at the ATM: exploring the usability of multiple graphical passwords. In *Proc. ACM CHI'07*, 887 – 894, 2007.
19. V. Roth, K. Fischer, and R. Freidinger. A PIN entry method resilient against shoulder surfing. In *Proc. ACM CCS'04*, 236–245, 2004.
20. T. Salthouse. The processing speed theory of adult age differences in cognition. *Psych. Rev.*, 103(3):403–428.
21. B. Schneier. *Applied Cryptography: Protocols, Algorithms, and Source Code in C*. Wiley Computer Publishing, 2nd edition, 1995.
22. S. Shukla and F. Nah. Web browsing and spyware intrusion. *Comm. ACM*, 48(8):85–90, 2005.
23. D. Weinshall. Cognitive authentication schemes safe against spyware. In *Proc. 2006 IEEE Symp. Sec. Privacy*, 295–300, 2006.
24. S. Wiedenbeck, J. Waters, L. Sobrado, and J.-C. Birget. Design and evaluation of a shoulder-surfing resistant graphical password scheme. In *Proc. AVI'06*, 177–184, 2006.
25. L. Zhuang, F. Zhou, and J. D. Tygar. Keyboard acoustic emanations revisited. In *Proc. ACM CCS'05*, 373–382, 2005.