Photographic Authentication through Untrusted Terminals

As a technique for authentication through untrusted public Internet terminals, photographic authentication relies on a person's ability to recognize personal photographs. The results from a study demonstrate that this technique can reasonably withstand replay attacks, a common vulnerability of many existing techniques.

Trevor Pering, Murali Sundar, John Light, and Roy Want
Intel Research

Public Internet access points provide a convenient means to access the Internet, but they pose considerable security risks. For example, an unscrupulous attacker could modify a public system to capture all of a user's keystrokes using devices such as the Key Katcher.[1] One way to enhance security on such terminals is through photographic authentication, a technique that relies on personal photographs for authenticating user access. The photographic-authentication prototype described in this article authenticates users by requiring them to identify a few personal photographs among a set of random ones.

Although not as mathematically secure as multicharacter passwords, the photographic-authentication technique resists replay attacks because it uses a different subset of pictures for each login attempt. You cannot simply replay the results from one login sequence to repeat authentication. Most current authentication techniques only protect against unscrupulous network communication, a strategy that relies on a faulty assumption.[2] By presenting different images each time from a large corpus of personal photographs, a photographic-authentication system can mitigate the effectiveness of replay attacks.

This authentication technique works in conjunction with a trusted "home server" that stores the user's photographs and account information. First, the users identify themselves to the system, initiating the authentication process with their home server. After successful authentication, the home server passes the necessary credentials to the desired Web-service host. Because the home server manages the authentication process, the access terminal does not gain access to any unnecessary information, such as the users' photographic databases.

We set up a prototype authentication system and tested eight participants who attempted to recognize their pictures, provided in advance. At each step, the system presented photographs such as the ones shown in Figure 1. The users selected the images that belonged to them. We repeated this process several times to achieve a statistical sampling. For comparison, a separate set of participants, given a cheat sheet representing information captured from a compromised public terminal, tried to break the authentication scheme. The base results demonstrate that photographic authentication is a viable technique that can reasonably withstand most replay attacks.

Motivation and premise

The primary motivation for photographic authentication is the need for more secure login mechanisms that grant or deny access through untrusted terminals. The emerging mobile Internet will rely heavily on public infrastructure, increasing the need for alternative authentication techniques. Also, the increased prevalence of digital photography and the ease with which people can recognize photographic images[3] also provides motivation to develop a photographic-authentication system for such purposes.

Published by the CS and Communications Society ■ 1536-1268/03/$17.00 © 2003 IEEE

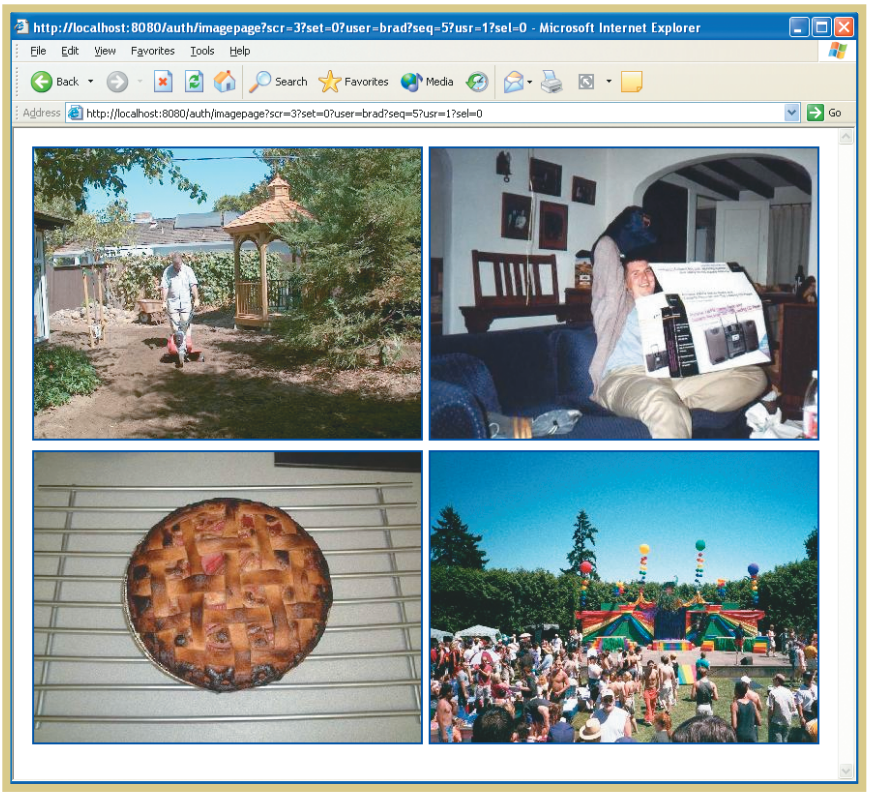


Figure 1. Prototype photographic authentication Web browser interface. Users must select the images that belong to them.

One additional risk associated with using public infrastructure is that an assailant can potentially capture all information being entered and displayed, not just the data from the authentication process. For example, when users check the status of their bank accounts, they are potentially compromising both their account balance and account number. However, it is generally only necessary to display the account balance, not both. Such casual data, such as the bank account balance without the account number, could be suitably protected with a photographic-authentication scheme because it is private but not high security.

A highly secure authentication technique would be overkill for such a terminal because secure authentication in itself does not guarantee the security of the data accessed. Photographic authentication aims to be "secure enough" for casual data by providing the necessary level of security without compromising ease of use. Ideally, the complete system would not even allow a user to access high-security data through an untrusted terminal. In other words, just because you already showed your badge to enter work doesn't mean you should leave your wallet on your desk.

The popularity of digital photography has recently exploded because of the widespread availability of affordable consumer-grade cameras and computers capable of manipulating photographs. As a result, many people have substantial personal digital photograph collections. Furthermore, as cameras become more affordable and easier to use, more people will possess large personal image collections, and digital storage capacities are rapidly increasing, providing ample space to save images. For the users who have them, these images can form a convenient authentication system that doesn't require much configuration.

The Personal Server,[4] the mobile device that inspired this article, provides a mobile user experience by wirelessly connecting with PCs and displays found nearby in the environment, rather than using a small-screen display on the device itself. Developing secure authentication techniques that do not require the user to juggle small devices, such as an authentication key, is an important step in making such systems usable and widely accepted.

Security overview

Theoretically, the photographic-authentication implementation presented here is about as secure as a six-digit password. This means that there is about a 1 in 10^6 chance that random guessing will be successful, a smaller chance than that of the personal identification numbers (PINs) of present-day ATM machines, which have a 1 in 10^4 chance of being randomly guessed, assuming you have the ATM card. In contrast, strong-text passwords, or even the weak passwords typically used on the Web, are many thousands of times more secure when subjected to a randomized guessing attack—approximately 10^{12} combinations for a six-character alphanumeric-punctuation password. The real vulnerability of photograph-based authentication is not numeric, but cognitive. In a cognitive attack, the attacker uses knowledge about the user.

Another technique would be to require users to carry a portable electronic device, such as a PDA[5] or SecurID card (www.rsasecurity.com/products/securid) as a trusted authentication mechanism that would let them safely log in to an untrusted terminal using a one-time key generated by the device. Although attractive from a security standpoint, this technique is quite complicated from a user's perspective: Users must retrieve the device, activate it, and manually type in the appropriate code. Additionally, they must not forget, lose, or break the device. In contrast, photographic authentication is streamlined: Users simply walk up to a terminal and select from a few sequences of images presented to them on the screen.

Photographic authentication is well suited to providing access through semitrusted or untrusted terminals where a user might want to access information only a few times while not implicitly trusting the access point. Photographic authentication is also well

TABLE 1
Primary users and image-set descriptions.

User	Gender & Age	Photos	Collection description
A	M, 22	48	Group team-building activities, primarily collections of people
B	M, 62	140	House maintenance projects, family gatherings, and test photos
C	F, 34	150	Vacation, family and friends, landscapes, and still-life compositions
D	M, 29	204	Extensive world travel and landscape photos, occasionally group shots
E	M, 30	397	Travel, landscapes, social events, and photographic experimentation shots
F	M, 26	457	Mainly images of people, including friends, family, and social events
G	F, 30	1,171	Everything, including friends, events, trips, landscapes, and so forth
H	F, 28	1,293	Everything, including friends, events, trips, landscapes, and so forth

suited to trusted environments because it provides an easier means to access information than text-based authentication. In any case, photographic authentication is not a strict replacement for other techniques but instead provides front-line authentication for access to some kinds of data. We designed our prototype system only to prevent *acute* attacks that target any user logging into the compromised terminal, and not *conspiracy* attacks that target a specific user regardless of the access terminals used.

There is a fundamental difference between knowledge-based authentication systems, such as text passwords, and recognition-based ones, such as photographic authentication. Password systems use a unique piece of knowledge—the password—to perform the authentication process, while recognition systems use a challenge-response sequence, such as selecting an image from a challenge set. The challenge-response technique is similar to the underlying security mechanisms used for logging into remote computers over a potentially insecure network: The remote system challenges the local system, which forms a response using local knowledge.

Although still based on passwords, the knowledge itself in a traditional challenge-response mechanism never leaves the secure access terminal and can't be deducted from the response. Changing this model so that the endpoint is untrusted—a public-access terminal—highlights the need for a true end-to-end challenge-response system, where the final endpoint is the user instead of the machine.

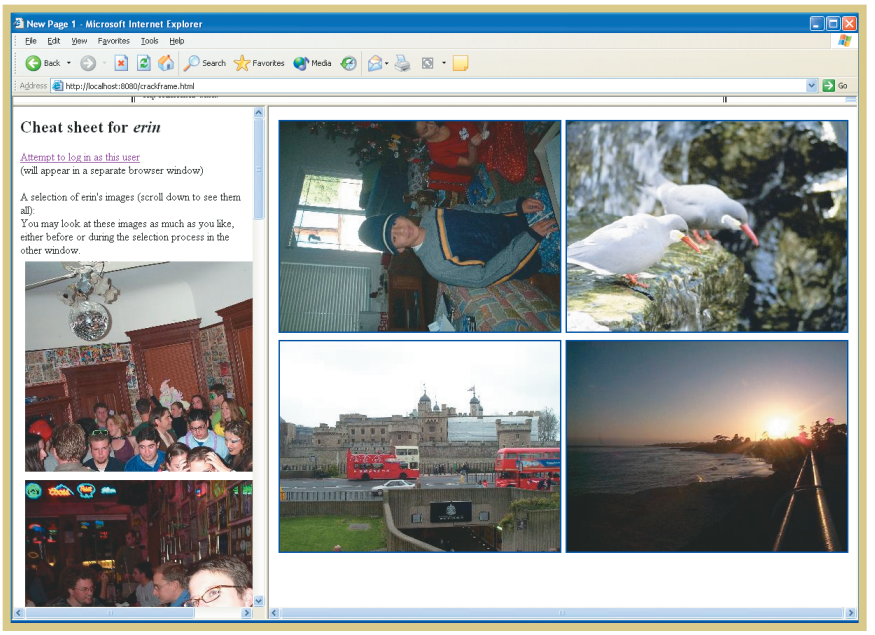


Figure 2. Experimental attack setup. Attackers view their cheat sheet in the scrollable left frame and the selection choices in the right frame.

Experimental evaluation

Two sets of experiments helped us evaluate photographic authentication. First, we simulated a standard login process to see whether photographic authentication is feasible, and then we simulated an attack against the system to see if it would hold up under a reasonable replay attack. In preparation for the tests, we solicited image collections from the eight users listed in Table 1. These eight people formed the basis for the primary authentication test, while a separate set of 12 people conducted the attack experiment against the eight primary image sets.

Because we did not place any restrictions on the number of images in a set, our experiment had a wide range of image sets, from a minimum of 48 to a maximum of 1,293. We took the incorrect images—the ones not belonging to the authenticating user—from the other users' image sets. We conducted both experiments through a Web interface and logged all transactions. The server recorded the identification success rate and the trial time, the data that best characterizes the interaction.

Our prototype presents images four at a time using a Web browser interface. For each set, the users click on the image they identify as their own (or belonging to the victim, in the case of an attacker), after which the system presents them with another set of four images. The system

Related Work

There are two bodies of work related to photographic authentication: graphical passwords and authentication over untrusted channels. The basic idea of using graphical systems for authentication is not new. But no prior research has explored using personal photograph collections. Similarly, some previous work has investigated displaying information on untrusted terminals, but this work has not dealt with the authentication process.

Several projects have explored image-based passwords, most notably in the form of recognizing preselected images[1] based on abstract art or people's faces. Although image-based, these systems are still fundamentally password-based because they require the user to remember some predetermined set of images through an up-front configuration process. Photographic authentication, on the other hand, uses a larger set of preknown images to define the acceptable image set, which makes it both less venerable to replay attacks and easer for the user.

There is a large body of work along the lines of automatic image recognition and classification,[2] which could potentially be used for programmatically breaking this authentication scheme. However, all of these systems match images on the basis of color or feature recognition, not semantics or association. For example, a pictorial search on the Ponte Vecchio bridge in Italy will return images of other bridges, not other images of Italy. Therefore, these techniques would not be effective at identifying a user's personal image collection, which will be defined more by semantics and less by visual similarity.

There are several techniques that rely on a trusted wearable component for authentication interaction. In addition to the smart-card-style systems mentioned in the main article, another system[3] relies on a small wearable camera to authenticate on-screen interactions. Basically, the system watches over your shoulder to determine the acceptability of an authentication request. Another system uses a small wristwatch-sized interface for authentication.[4]

REFERENCES

1. R. Dhamija and A. Perrig, "DejaVu: A User Study Using Images for Authentication," *Proc. 9th Usenix Security Symp.*, Usenix, Aug. 2000, pp. 45–58.
2. Y. Chen and J. Wang, "A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, Sept. 2002, pp. 1252–1267.
3. D. Clarke et al., "The Untrusted Computer Problem and Camera-Based Authentication," *Proc. Pervasive 2002*, Springer-Verlag, 2002, pp. 114–124.
4. J. Al-Muhtadi, D. Mickunas, and R. Champbell, "Wearable Security Services," *Proc. Int'l Conf. Distributed Computing Systems*, IEEE CS Press, Los Alamitos, Calif., 2001, pp. 266–271.

shows the users 10 image groups in sequence, which represents one login attempt. Figure 2 illustrates the corresponding attacker display layout. For the primary experiment, users completed 10 groups of images (for a total of 100 selections), while the attackers completed eight groups (one for each target). The system chooses individual images completely at random, which means they could repeat within a trial or set.

We did not manually manipulate images, nor did we use any metadata, such as the date the photo was taken. We converted all images to 400 × 300 resolution, clipped as necessary. We automatically rejected images that were too small, too tall, or too wide. We also automatically rejected images that compressed to less than 5 Kbytes because these images tend to be blank pictures or obvious photographic mistakes. We did not manually evaluate images, which meant that the image sets included multiple versions of the same general image and that some of the pictures appeared rotated because they were portrait images taken in landscape mode.

Figures 3a and 3b shows the recognition rates and trial times, respectively. Trial times do not include the first trial. This graph shows that attackers fared significantly worse than the primary users at recognizing images. We quantized results by trial, each of which consists of 10 individual selections. Error bars, when shown, represent a 95 percent confidence interval.

Authentication experiment

We designed the primary authentication test to see whether users could correctly distinguish their own images from those of others. We asked the users who contributed photographs to complete 10 trials to identify their images. After the test, we asked the users four qualitative questions to capture their overall impression of the authentication system.

From the base results, we determined that users can quickly and accurately identify their own pictures. The two users with a very large number of images (G and H) exhibited slightly slower recognition times. User D (who is known for being meticulous) also showed significantly slower times. Users could quickly and correctly identify the very first set of their images and did not require any learning.

At the end of each user trial, we asked users to qualitatively rank their experience. The users' perception of the process was mixed. Some thought that it was more secure than traditional text-based password techniques, while others thought it was less secure. Overall, however, everybody had fun using the technique and found it easy to use. Some people actually

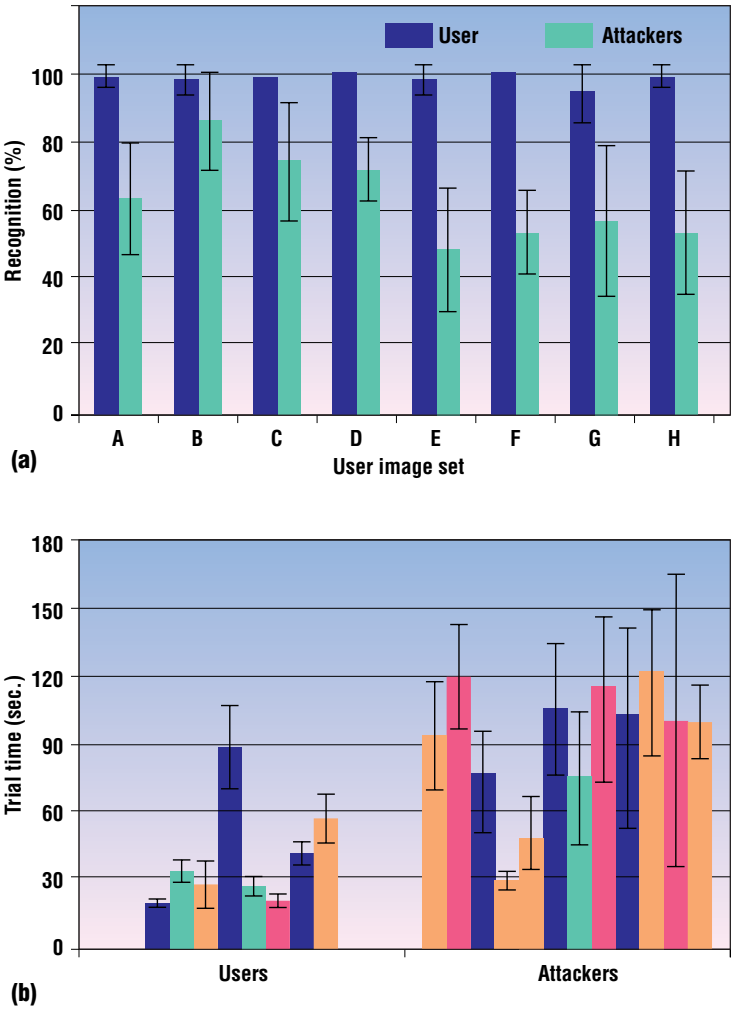


Figure 3. (a) User and attacker recognition rates for individual picture selection. Error bars indicate a 95 percent confidence interval. Each data point summarizes 10 user or 8 attacker trials. (b) User and attacker trial times. Attacker trial times do not include the first trial. Error bars indicate a 95 percent confidence interval. Each data point summarizes 10 user or 7 attacker trials.

enjoyed the process and had a good laugh because they were shown pictures they hadn't seen for quite some time.

Attack experiment

We designed the login attack to simulate an attack on a user account by someone who had snooped on a previous authentication session by that user. We allowed attackers to look at all the sample images before and during their login attempt. Each attacker had one trial against each primary user, for a total of eight sessions conducted in a random order.

Figure 4 shows the basic statistics for attackers, which highlights the great variability of success rate and speed. This variability highlights the notion that a trained perpetrator could perform significantly better than an average user at breaking into a photographically protected account. The first attacker trial (the first user they attempted to compromise) was significantly slower than subsequent trials, which is not surprising given that there were no practice rounds. The reported times in Figure 4 do not include this first trial.

Base attack recognition rates, stratified by user (Figure 3a), indicate that most users' image sets are relatively immune to attack. Although the recognition rates are seemingly high (above 80 percent in one case), the high rates represent recognition of only one picture. A complete login would require multiple successful recognitions. The exact number required for access would depend on the level of security desired. User B's data seems particularly susceptible to the simulated replay attack. In fact, two attackers achieved 100 percent recognition against User B. Post-experiment analysis indicates that the associated image set is extremely thematic, showing many pictures of the same house construction project.

Discussion

Security is the prime concern of any authentication mechanism. Because our technique is based on recognition, rather than memorization, there are no security leaks generated by people writing down passwords, which is a common weak link with many strong-encryption techniques. However, there still are many ways the system can be compromised. Because this technique is fundamentally different from most existing authentication mechanisms, it is difficult to extrapolate from previous experience with security systems to determine exactly how significant these attacks would be. It is also difficult to determine what other less-obvious attacks exist.

One deficiency of the attack experiment is that it relied on unskilled participants who are not necessarily representative of people who would be trying to break into the system. The varied success levels of the attackers, shown in Figure 4, further support this observation. Some people seem better at it than others. A professional photographer trained in psychoanalysis might be extremely adept at identifying user pictures. The people we selected for the attack role in this test would therefore represent

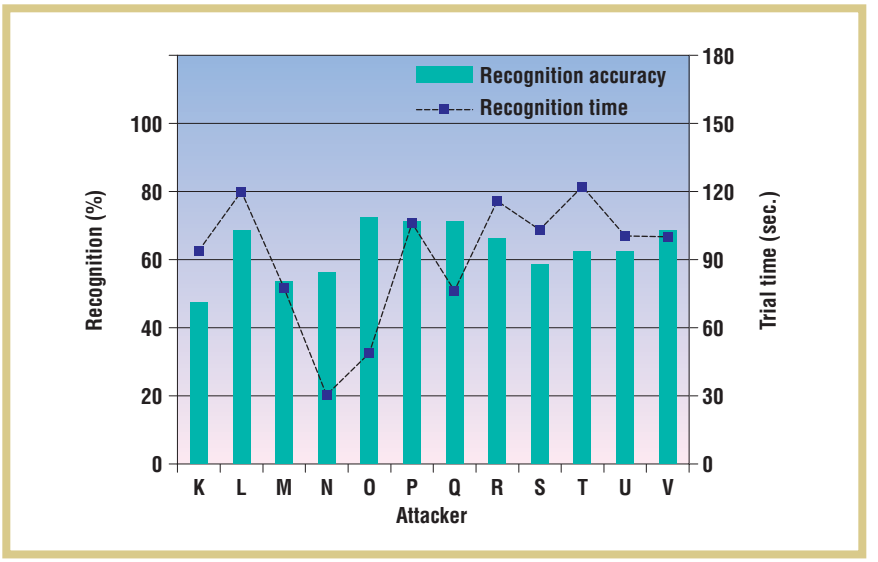


Figure 4. Recognition rate and trial time by attacker. The time results do not include the first trial.

the casual attacker, one who is unscrupulous but not necessarily skilled at the task.

Replay attacks

As mentioned earlier, replay attacks, also known as *observer attacks*, consist of capturing part of a communication between two entities and playing back that information at a later time to compromise the system. In context, this translates to the untrusted access point capturing all input and output data. Specifically, if the attacker could capture all keystrokes, a Web site password entered in clear text on the keyboard could be used later to access the same site.

Photograph authentication is well suited to resist replay attacks through untrusted terminals by varying the challenge image set each time. Single-use key systems, such as SecurID, also prevent replay attacks but require the user to possess and use a trusted hardware component to generate local keys. Photographic authentication is not completely immune to replay attacks because the images from one attempt might provide enough information to deduce the correct images in following attempts.

Cognitive attacks

There are two kinds of cognitive attacks against photographic authentication: similarity attacks and knowledge attacks. Similarity attacks involve determining whether two images are pictures of the same thing (although not necessarily from the same time or angle). Automatic image recognition algorithms could easily be applied to an image set to strengthen it against similarity attacks. Knowledge attacks, on the other hand, use specific pieces of knowledge, such as knowing about a trip to Paris, to identify related pictures.

For text passwords, password attacks based on the victim's birthday are a classic example of a cognitive attack. Users are commonly instructed not to use their birthday as a password. Photographic authentication is somewhat sensitive to knowledge attacks because of the strong correlation between users' lives and the pictures they keep. Almost any high-level information about where users have been or what they do could give clues as to what pictures they might have.

The trial times shown in Figure 3 indicate that limiting the time allowed for authentication might be a useful technique for detecting cognitive attacks and manual replay attacks. A cognitive attack requires the perpetrators to think about the selections they are making instead of just picking images they recognize. This distinction means that it might be advantageous to disregard login attempts if the images are recognized either inaccurately or too slowly.

Coincident attacks

A coincident attack is one in which an unscrupulous agent or proxy running on the untrusted terminal has access to a user's data in parallel to the user actively operating the system. In other words, a coincident attack is one where it is impossible for the remote system to tell the difference between a legitimate access and a temporally coincident unauthorized one from the same source. This is not an authentication problem, but rather a problem inherent in untrusted terminals.

The window for a coincident attack begins after a successful authentication and ends when the user either explicitly logs out of the system or times out (which might never happen if the unscrupulous agent keeps the connection alive). Although potentially very damaging, coincident attacks are tricky to implement and would compromise only the data made available through an untrusted terminal, which can be deliberately limited.

Compromised attacks

A compromised attack is one in which the system's integrity has already been compromised. For example, the attacker has cracked the password or identified the picture set. The question then becomes how to fix the system so that it is again secure. In the case of text passwords, it is quite easy to fix the problem: The user simply selects a new password. In fact, high-security systems typically require users to change passwords regularly.

Fixing a compromised photographic-authentication system is more difficult because a user can't forget pictures they have seen and suddenly recognize new ones. One way to work around this problem is to use a series of image subsets for the authentication process. When one subset becomes compromised, the user simply rotates to the next set. Or the subsets could simply rotate periodically.

Polling attacks

A polling attack is one in which the authentication server is repeatedly accessed to gather information about the authentication account. In the case of text passwords, a polling attack is similar to random or dictionary attacks, where trial passwords are thrown at the authentication mechanism to guess the correct password. In the case of photographic authentication, this kind of attack could be used to glean the entire set of images used for authentication. The key, however, is to present the images so that the attacker can't deduce which images are the correct ones.

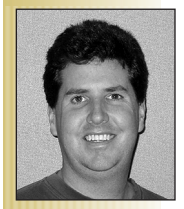
There are several promising avenues for our future work. We plan to explore alternate image presentation and techniques for generating challenge image sets. We could also improve the effectiveness of the challenge set by preprocessing images to remove obvious similarities between pictures. Specifically, we could apply programmatic image-recognition algorithms to filter images on the basis of color histograms or face recognition. We also plan to explore using trial time to filter attacks.

People without personal photographs obviously must acquire some before using this kind of authentication system. Because the basic premise is that people can easily recognize their own images, simply buying an image library will not work (unless a user invests a significant amount of time to become familiar with the images). Someone could probably acquire a suitable collection by borrowing a digital camera and taking the requisite amount of pictures. But because this activity requires considerable effort, it runs counter to the spirit of a low-configuration system. Unfortunately, people not accustomed to taking pictures might be really bad at it and end up with a set of easily identifiable pictures. Admittedly, only two out of the four authors of this article possessed a significant digital image collection.

A key observation about attacking this authentication system is that attackers need only identify which pictures are incorrect or do not belong to the specific user. This raises the question about where the decoy images come from. If the decoy image collection is collected from images found on the Web, an attacker could compare challenge images to find out which ones are not from the specified user. One solution to this problem is to use other users' authentication images to form the appropriate image sets. Although technically viable, this solution relies on users being willing to share their authentication images within a trusted group. However, it is not clear how well this solution would work if the system were to support a larger group.

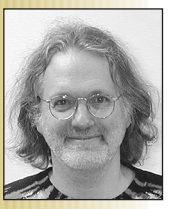
Photographic authentication is a novel technique for dealing with public infrastructure, an emerging concern as mobile and fixed-infrastructure systems continue to evolve and merge. By capitalizing on advances in consumer-grade digital photography, photographic authentication can increase the confidence and spontaneity with which people can use public infrastructure.

the AUTHORS

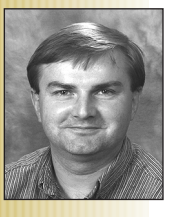


Trevor Pering is a research scientist at Intel Research, where he is a member of the Ubiquity project. His research interests include usage models, power management, novel form factors, and software infrastructure for mobile and ubiquitous computing. He received a PhD in electrical engineering from the University of California, Berkeley. He is a member of the ACM. Contact him at trevor.pering@intel.com.

Murali Sundar is a senior researcher at Intel Research. His research interests include programming language design, type systems, and ubiquitous computing. He received an MS in industrial engineering from Texas Tech University. Contact him at murali.sundar@intel.com.



John Light is a staff researcher at Intel Labs. His research interests include ubiquitous computing and 3D user interfaces. He received a BS in mathematics and a BA in psychology from California State University, Los Angeles. Contact him at john.light@intel.com.



Roy Want is a principal engineer at Intel Research, where he manages the Ubiquity project. His research interests include ubiquitous computing, wireless protocols, hardware design, embedded systems, distributed systems, automatic identification, and microelectromechanical systems. He received a PhD in computer science from Cambridge University. Contact him at roy.want@intel.com.

REFERENCES

1. I.L. Paulson, "Key Snooping Technology Causes Controversy," *Computer*, Mar. 2002, p. 27.
2. R. Anderson, "Why Cryptosystems Fail," *Comm. ACM*, Nov. 1994, pp. 215–227.
3. R. Harbor, "How We Remember What We See," *Scientific American*, vol. 222, no. 5, May 1970, pp. 104–112.
4. R. Want et al., "Personal Servers: Changing the Way We Think about Ubiquitous Computing," *Proc. Ubicomp 2002*, Springer-Verlag, 2002, pp. 194–209.
5. D. Balfanz and E. Felten, "Hand-Held Computers Can Be Better Smart Cards," *Proc. Usenix Security 99*, Usenix, 1999, pp. 15–24.