Usability of Image/Concept Authentication For Communication Over An Insecure Channel

Nathaniel Granor
Advisor: Michael Fischer
May 4, 2009

Abstract

Passwords are an inherently weak form of security. As demand for authentication systems has grown, passwords have become the ubiquitous preferred method of verifying identity over insecure networks because of their simplicity. However, the need for secret information - the password itself - to be sent through a potentially insecure system is an inherent flaw in the model, which allows security to be easily breached through eavesdropping and over-the-shoulder attacks. Expanding on existing research, we develop and implement an image-based authentication system that avoids these problems and study its usability. In our system, which resembles a zero-knowledge protocol, users are given secret pass-concepts which are categories that could be present in or absent from images. In each authentication session, users are challenged with a different set of images and asked to determine how many of their concepts are present in each image. We find that the image-based system requires careful refinement to serve as useful replacement for traditional passwords, as there is often ambiguity in the interpretation of images.

Background

The standard password model is so common as to make the following description superfluous for most readers. A human interacting with a computing device needs to prove his identity. He inputs a pre-arranged password (a string of characters) via a keyboard and the computing system checks his input against that stored in memory. There are many variations: the user might be communicating over the internet to log into an application on another server, or simply logging into his workspace. The password may be encrypted after it is typed and before it is sent anywhere. There may be rules governing the permutation of characters allowed, the length of the string or the inclusion or exclusion of particular characters. The password may or may not be displayed on the screen as it is typed. Regardless of these differences, no such system is immune from certain low or no-tech attacks.

At the most basic level, a passerby might watch over the shoulder and observe the keys pressed on the keyboard. A piece of malicious software running on the computing device or a piece of malicious hardware attached to the computer might surreptitiously capture the keyboard input and send it off to another user, or, in the case of passwords that are displayed on the screen, capture images of the screen. If the password is sent over a network or the internet, eavesdroppers may be able to steal the password, or if it is encrypted, attempt to decrypt it.

An entirely new kind of authentication system is needed; one that is more robust and does not have a single link that an eavesdropper can discover to infiltrate the system. One suggestion is one-time passwords [6]. Such a system might be similar to the one-time-pads of pre-digital era: two copies of a pad with pseudo-randomly-generated numbers were cre-
ated and distributed. These numbers were used as keys in cryptographic systems. Without a pad’s twin, decryption was impossible. However, it seems impractical to expect users to carry around pads of future passwords (or separate devices with stored passwords) and such devices would be subject to loss and theft [4].

We can envision a digital equivalent of the one-time pad: in this system, the human has device that implements a secret algorithm. When prompted with a challenge from the authenticator, the user enters the challenge into his device, which produces a suitable response. Once again, the material required is too expensive and impractical for most usage scenarios. However, these systems contain some original suggestions.

First, there is the notion of a Challenge-Response protocol. A password system is a special case of Challenge-Response, where the challenge and response are always the same (Challenge: “What is the password?”, Response: “42”). A more robust variant, like the system described above, has different challenges that require different responses. The secret itself (the algorithm) is never transmitted. In such a situation, an eavesdropper who was following along would gain only very limited information about the secret.

The system described above resembles a zero knowledge protocol. In a zero-knowledge protocol, a challenge is selected at random from the challenge domain, and the response is something very basic: a number, or even a single bit of information. In theory, an eavesdropper can learn absolutely nothing about the secret from observing the challenges and responses. If the response is a single bit of information, then an eavesdropper trying to cheat his way into the system has a 50% chance of producing a correct response to a given challenge about the secret. As successive challenges are issued, the eavesdropper’s chance of randomly guessing the correct response vanishes. [1]

But the zero-knowledge system described above relies on complex mathematical calculations. We cannot expect typical computer users to regularly perform such calculations on their own. We seek a variant of Challenge-Response that takes advantage of the sorts of computation at which humans excel.

Jameel, Lee and Lee [5] propose a system using images, and later, the internal semantic structure of images, as the domain of an authentication system. In such a system, a user might have a concept as their secret. The challenge would be a series of images each of which either do or do not instantiate the concept. The user then responds with “yes” or “no” for each image. For instance, a user’s secret concept might be blue rhinoceros or hunger. The challenges could be photographs or even compound images consisting of randomly selected and placed clip-art images. The user would decide whether each challenge image displayed a blue rhinoceros, or hunger.

The security of such a system is difficult to quantify, as it would require discretizing abstract notions. One large security issue is that the system would not, in practice, be zero-knowledge, as an eavesdropper might be able to see the challenge images as well as the responses. With sufficient knowledge of the system (like a database of potential concepts that are available as secrets), an eavesdropper might be able to perform statistical calculations or a learning algorithm to guess the secret.

However, some additional strategies might be employed to increase security and make the problem of recovering a users secret concept from observed challenge-response interactions intractable. First, the user could be encouraged to obscure their response [2], or, more appropriately in this case, to include one intentionally incorrect response out of the set. Additionally, a user could be given an arbitrary (and variable) number of different secret concepts, and be asked to evaluate challenges based on the union or the intersection of the concepts. Once again, there are a multitude of possible variations to such a system. The key notion, however, is that we can conceive of authentication systems which are easily usable by people but don’t involve sending the secret information through an insecure channel.

In this project, we attempt to implement such a system and perform user testing to study the tradeoff between security and usability.
Implementation

Our authentication system is titled “Image/Concept Authentication”. The key idea of the system is that the “secret” shared by the user and authentication service is never transmitted. In this case, the secret amounts to a set of one or two pass-concepts. The concepts are well-known categories that could be present in or absent from an image, like airplane and flower. In an authentication session, the server provides a challenge consisting of a set of images randomly selected from a database of pre-labeled images. The user’s response indicates how many of his or her pass-concepts are exhibited in each image. In our study, an interaction with our system consisted of a training phase, a setup phase, an authentication phase, and a survey phase.

Pre-process

We implemented Image/Concept Authentication as an internet-based PHP/MySQL application. Our MySQL database contains tables for users, concepts, images, and the relationships between the three.

The selection of images and concepts for the database is an essential part of the system. The pass-concepts must be obvious, clear ideas that can be represented in images. In our implementation, the set of concepts was derived from the set of images we were able to obtain, and refined by a first round of user trials. Concepts that users had trouble understanding were eliminated from the system.

Our images were photographs taken from royalty-free, open-copyright stock photo websites. During a first round of user testing, we attempted to eliminate images from our set which were blurry or vague. Subjectivity in the concept and image sets destroys the usability of the authentication system.

After the selection of the concept and image set, the arduous task of labeling began. Initially, we attempted to give a boolean label to each image/concept pair. We discovered that this led to some ambiguous labelings. A picture of a hand holding a flower, for instance, is a clear positive example for the concepts hand and flower, and a clear negative example for airplane or eye, but it is an ambiguous example for a concept like animal. The hand obviously belongs to a human, and humans are animals. But it remains unclear whether animal requires a whole animal, or perhaps refers only to non-human animals. Thus, we adjusted our model to allow an “unknown” labeling for a given image/concept pair. So the image described above might be shown to users whose pass-concepts are like hand or airplane, but a user with animal as a pass-concept would never be shown that image because of its “unknown” labeling on the concept animal.

Phase 1: Training

Upon visiting our test website, users were first given a brief introduction to the Image/Concept Authentication system. They were then shown an example of its use. This portion of the process was both a practical necessity and a theoretical one. Since systems like this one are not well-known, it was important to provide study participants with instruction. Indeed, the success of any similar implementation will probably hinge upon the ability to show users how it works in a clear, simple way.

Phase 2: Account Setup

After the introduction and example, participants were asked to freely choose a username. Like in a traditional username/password scenario, the username must be unique in the database, since it is used as a key to retrieve the user’s pass-concepts.

By design, there are many fewer concepts than users in an Image/Concept system. Therefore, multiple users will have the same pass-concepts or even the same set of pass-concepts. To maintain security, the concepts should remain fairly evenly distributed between users. Therefore, we could not allow the users free choice between all concepts in the database. The simplest solution is to assign pass-concepts to each user based on the current distribution when a user signs up. However, we theorized that users would find it easier to remember a concept they had selected than one randomly assigned. As a compromise, users could be given a choice between several pass-concepts. By showing the user the least-commonly
chosen pass-concepts to choose from, we can maintain a relatively even distribution of concepts to users, and users can still select concepts that they find easier to remember. Assignment versus choice of pass concept was one of the parameters tested in the usability study. The number of concepts assigned to each user also varied between 1 and 2. To aid users in the selection and memorization of their concepts, they were shown images from the database that positively exhibited each of their concepts.

Phase 3: Logging In

Participants in our study were asked to perform 5 log-in attempts. Some elected to attempt more, and some stopped after fewer than 5. We randomly varied the number of images shown in each attempt between 4 and 7. In designing the login page, our primary considerations were speed, clarity and simplicity. As with traditional login systems, we tried to keep the entire activity to a single webpage. The images were confined to a single width (though some varied in height) and arranged in a 2-row grid. A text box positioned below each image received the users’ responses. When focused (i.e., when the cursor was inside a given textbox), each text box was highlighted by a colored border. To minimize the number of clicks required, focus was automatically shifted to the next text box after the user entered a character.

In our first round test, we determined that the one-stage login, where all login activity happens on a single page view, was not feasible for our system. In a one-stage setup, the challenge images must be selected and displayed before the username is known to the authentication service. This means the images must be selected entirely at random from the image set. But concepts are not represented uniformly in the image set. For instance, there may be few enough images that exhibit the concept crane in the total image set that the expected number of crane images in a challenge set is zero. Indeed, if the concept set is large, as it would need to be in a full-scale implementation, the majority of concepts would be prone to this situation. A hacker’s best attack would be to simply guess all zeros. It is imperative, therefore, that we know the pass-concepts, and by extension the username, before selecting the challenge image set. With a known username, the service can produce an image set with an arbitrary probability of demonstrating the user’s pass-concepts.

Phase 4: Survey

After each login attempt, participants were shown the result of their login attempt (Success or Failure) and then completed a brief survey about their experience. They were asked to compare the difficulty of the task and time spent on it to a traditional login system. They were also asked how they would feel if Image/Concept Authentication were deployed in an online banking service, an email service, a social networking site, an e-commerce site, or at their home or work computer terminal. Finally, users were allowed to leave general comments.

Additionally, the system tracked the time elapsed from when the user submitted their username until they submitted their responses. Survey responses, as well as the experimental parameters and results of the login attempt were stored in a log table. We also tracked statistics regarding the effectiveness of each image and concept. Issuing a survey between each login attempt allowed us to track users’ attitudes over time. It also provided a way to break up their login attempts. Finally, when users failed authentication, we “punished” them with an 8-second delay to simulate the real frustration of a failed login attempt.

Security of Image/Concept Authentication

On each login attempt, the user is shown $n$ images, where $n$ is arbitrarily selected by the service. Each image can contain $0$ to $m$ of the user’s secret pass concepts. The number of possible responses is therefore $(m+1)^n$. The probability of randomly guessing the correct response is $\frac{1}{(m+1)^n}$ which can be made arbitrarily small by adjusting $m$ and $n$. However, these values must remain relatively small to keep the authentication process manageable for users. The highest values of $m$ and $n$ in our study were 2 and 7,
yielding a probability of $\frac{1}{2187}$ of randomly guessing the correct response. In contrast, a traditional username/password scenario has a much lower probability of randomly guessing the correct response. If $a$ and $b$ represent the minimum and maximum password lengths respectively, and a password may contain only alpha-numeric characters, the chance of randomly guessing a password is

$$\sum_{i=a}^{b} 36^i.$$

In short, the traditional username/password method seems extremely secure against brute-force attacks, where image/concept is rather weak. This is mitigated by several factors. First, in practice, most users of username/password authentication do not follow best practices to keep their account secure. Specifically, users often choose relatively predictable passwords that incorporate words or personal data. Hackers often use tables of likely username/password combinations to greatly increase their odds [8]. Additionally, brute force attacks can be defended by blocking offending IP-addresses or temporarily suspending a user’s account. Unlike traditional methods, Image/Concept Authentication can also defend against over-the-shoulder attacks, man-in-the-middle attacks, and phishing schemes.

**Over-The-Shoulder Attacks**

In an over-the-shoulder attack, a hacker watches the actions of the user during authentication either literally over-the-shoulder or by means of malicious hardware or software installed on the user’s terminal or on a network. In the worst case, a hacker can log the keystrokes of the user and gain complete access to their account by reproducing the series of keystrokes observed. Image/Concept authentication sidesteps this attack by presenting a randomly selected challenge that changes with each attempt. Viewing the details of a single login attempt provides very little useful information because the particular set of challenge images is extremely unlikely to be shown again.

**Man-In-The-Middle Attacks**

In a man-in-the-middle attack, a hacker intercepts an authentication attempt before it reaches the secure service. He then relays the authentication data to the service, acting as a conduit between the user and the service. Once again, he has gained complete access to the user account, without the knowledge of either legitimate party [7]. In Image/Concept Authentication, a man-in-the-middle attack would succeed granting a hacker access to a user account, but only for the session he was intercepting. He could not use the information intercepted to login again later, as with the Over-The-Shoulder Attacker.

**Phishing Schemes**

In a phishing scheme, a hacker sets up a fraudulent website that mimics the appearance of a credible website. He then attempts to direct users of the target website to the phishing site, often through emails with obscured hyperlinks. If a user enters their login information at the fake site, their credentials are successfully stolen. The deceived user may not realize they have fallen victim to a phishing attack for some time [3]. Image/Concept Authentication is relatively immune to phishing schemes. A fraudulent site would not be able to provide a challenge that incorporated images with the user’s pass-concepts. Even if they did, the login response provided by the user would yield only the correct labels for the images in the challenge and not the pass-concepts themselves.

**Obfuscation**

The protection against over-the-shoulder and man-in-the-middle attacks springs from the system’s resemblance to a zero-knowledge protocol. But Image/Concept Authentication is not itself a zero-knowledge protocol, because the challenges issued by the authenticator do contain potentially useful information. We believe it to be a safe assumption that a computer could not extract the semantic contents of an image. A human can do so easily. This means that, with enough over-the-shoulder data, a human adversary might be able to deduce a user’s
pass-concepts (or concepts similar enough to provide access to the account). Features can be added to the system to protect against this case. We suggest two obfuscation methods: intentional error and plausible distractors.

To confuse any potential observers, the system could be altered to allow users a single mistake in their authentication response. Users could then be encouraged to include such a mistake each time they log in. An observer would have no way of determining which piece of the response, if any, was garbage [2]. Additionally, when selecting the challenge images, the system could include an additional pass-concept (one not belonging to the user) and, with an arbitrary probability, choose images that match this distracting concept.

Results and Discussion

During our study, 195 distinct users performed 724 login attempts. Users were only successful at logging in 68% of the time. Overall, users found the system somewhat more difficult to use and more time-consuming than traditional username/password logins. Users were not happy with the prospect of seeing Image/Concept Authentication put into common use, as shown in figure 1.

User comments supported these findings. The following comments are fairly representative of the overall dataset:

“I don’t quite know why I failed.”

“It’s important to use non-ambiguous images. On my last log-in, I failed, but I’m not sure why, unless I should have considered a helicopter to be an airplane…”

“It is hard to identify what is or isn’t an image of the assigned word. My word was ‘stem.’ Is a unit of grass a ‘blade’ or a ‘stem?’ Does a leaf have a stem, or is the thing that I perceive as a stem just part of the leaf?”

Users also seemed worried about the overall security of the system in their comments:

“I still don’t feel as if I trust this for authentication. With a password, the hacker has to guess (say) 8 things from a list of (say) 62+ different options (a-zA-Z0-9, and punctuation), so brute forcing it is difficult even assuming I don’t have a dictionary word as password. With this, the most I saw was 6 things, so it’s is only 64 different attempts before they can get in. Maybe it responds with different images if I fail to connect, but even with that, there’s a 1 in 64 chance of getting in randomly so they can just keep on trying.”

“$3^3 * 3^3 = 81$. Insecure.”

Some users did seem to appreciate the potential of a system like this one:

“Most of the time, this would be in the way, because I relied on cached passwords. I can see it being extremely valuable when traveling, though, or otherwise using a computer I don’t own.”

A more in depth analysis of the test data does reveal a few promising facts. Removing the 8 worst-performing concepts from the concept-set yields an
11% improvement in the overall success rate. Perhaps through a process of pruning and refinement, we could achieve a desirable success rate. Figure 2 shows the success rates of the complete data set and the set pruning the worst-performing concepts. Figure 3 shows the success rates of each concept.

Furthermore, there did not seem to be much difference between the opinions of those users with 1 pass-concept and those with 2, as indicated in figure 4. This might indicate that even higher numbers of pass-concepts could be assigned, which would allow an increase in the number of bits of information sent with each response and thus an increase in overall security.

Also, the login success rate rose from 60% to 72% between users’ 1st and 5th attempts. The users that made a 6th attempt (there were only 22 of them) succeeded 86% of the time. Figure 5 displays this result. These numbers may be a little misleading, as frustrated users were probably more likely to quit early. However, they do seem to indicate that people get much better at this task with practice.

Conclusions

Image/Concept Authentication is not ready for prime time. The main problem seems to be the difficulty
in providing unambiguous labelings of image/concept pairs. Even if users agreed with the labelings 95% of the time, they would still be expected to fail at least 1 in every 3 login attempts (assuming 7 images were used on each attempt). Users don’t seem to mind the larger numbers of images per login or assigned concepts necessary to provide an adequate level of security - or at least, they mind the ambiguous labelings a lot more.

To a certain extent, it may not be possible to remove all subjectivity from the image/concept labelings. Results would probably improve after a continued process of pruning and refinement - removing the images and concepts most prone to ambiguous interpretations. One novel idea would be to use a large scale voting mechanism to provide new labelings to the system. This might be achieved through surveys or web-based games. The reCAPTCHA service, which acts as a test to keep bots from filling out forms on the internet, implements a similar idea. In that system, users prove they are human by correctly identifying words scanned from books that fail automated character recognition algorithms. Each user is shown 2 words at a time: one word with a known label, and one with an as yet undetermined label. The user does not know which word is which. The previously identified word is used to prove the user’s identity as a human, and the response to the unlabeled word is expected to be accurate if the user got the labeled one correct. In this way, the reCAPTCHA system continuously adds new challenges to its database. One could imagine a similar solution for Image/Concept, where each set of challenge images includes one unlabeled image. If the user logs in correctly, their response to the unknown image/concept pair is assumed to be correct (or becomes a vote).

If the problem of ambiguous labelings can be solved, Image/Concept Authentication would probably become an acceptable security tool - perhaps serving as a secondary login system for users who suspect their computers are compromised or are using an insecure connections.

References


