Image-Feature Based Human Identification Protocols on Limited Display Devices*

Hassan Jameel[1], Riaz Ahmed Shaikh[1], Le Xuan Hung[1], Yuan Wei Wei[1], Syed Muhammad Khaliq-ur-rehman Raazi[1], Ngo Trong Canh[1], Sungyoung Lee[1], Heejo Lee[2], Yuseung Son[3], and Miguel Fernandes[3]

[1] Department of Computer Engineering, Kyung Hee University, 449-701 Suwon, South Korea
{hassan,riaz,lxhung,weiwei,raazi,ntcanh,sylee}@oslab.khu.ac.kr
[2] Department of Computer Science and Engineering, Korea University Anam-dong, Seongbuk-gu, Seoul 136-701, South Korea
heejo@korea.ac.kr
[3] Institute for Graphic Interfaces, Ehwa Womans University, Seoul 120-750, South Korea
{yssohn,mfernandes}@igi.re.kr

Abstract. We present variations and modifications of the image-feature based human identification protocol proposed by Jameel et al with application to user authentication on mobile devices with limited display capabilities. The protocols introduced are essentially reduced versions of the original protocol with a minor tradeoff between security and usability. However, the proposed protocols are not aimed for computation and memory restrained devices. A brief user survey highlights the usability. By employing realistic assumptions pertaining to mobile devices, we show that the protocols are secure under the conjectured difficulty of extracting the secret feature from the observation of images and their binary answers. The adversary considered is strictly passive.

1 Introduction

Secure user authentication (identification) protocols, in which a human user securely authenticates himself/herself to a remote server, are of utmost importance in today's increasingly connected world. Presently, the most prevailing form of user authentication is *password based* authentication. Alternative forms of authentication have been proposed but are not commonly deployed, generally due to their relative difficulty of use. Matsumoto [2], proposed a threat model in the user authentication scenario in which the adversary has more powers than conventional threat models for authentication. The adversary not only has passive and active access to the communication channel between the user and the server, but has also access to the computer terminal being used by the user and

* This work was supported by the IT R&D program of MIC (Ministry of Information and Communication)/IITA (Institute of Information Technology Assessment). [2005-S-604-02, Realistic Virtual Engineering Technology Development].

K.-I. Chung, K. Sohn, and M. Yung (Eds.): WISA 2008, LNCS 5379, pp. 211–224, 2009.

the alphanumerics being entered by the user. Furthermore, the user does not have access to any additional trusted hardware. Obviously, under such circumstances, traditional authentication protocols fail miserably. In accordance with the terminology used in [2], we will call user authentication protocols in which an *unassisted* human user authenticates to a remote server as *Human Identification Protocols*.

Human identification protocols secure under Matsumoto's threat model have been proposed in literature but are too impractical for humans due to high memory requirements and difficulty of use. One such protocol was proposed by Jameel et al in [1]. Their protocol is built on the diversity of things (features) an image describes. One of these features as a secret shared between the user and the server and the user has to answer '1' if a given picture contains that feature and answer '0' otherwise. On an authentication session, a series of pictures are shown to the user who constructs a binary answer string according to a predefined secret order. The protocol, as it is, cannot be used on mobile devices with small display units. In this paper, we present significant variations of the protocol in [1], so as to make it practical on mobile devices. The protocols presented are secure under the conjectured difficulty of the problem of finding the secret feature among a given set of images, which is different from the one presented in [1]. We present arguments in support of this difference and also illustrate general weaknesses and shortcomings in the original protocol as well as argue its specific inadequacies if applied to mobile devices in its original form. Our additional contribution is a comprehensive description of the threat model in [2] and a small-scale user survey based on an implementation of one of the proposed protocols which aims at demonstrating its usability.

Related Work. The first attempt at constructing a secure human identification protocol dates back to Matsumoto [2]. After this scheme was broken in [4], Matsumoto proposed another scheme in [5]. A variation of this scheme whose security is based on a mathematically hard problem was proposed by Hopper and Blum in [7], which is commonly known as the HB protocol. But HB protocol is impractical for human users as is acknowledged by the authors themselves. Other attempts, which loosely resemble this category of protocols are proposed in [4], [6] and [8]. However, the tradeoff between security and usability remains intact. Weinshall [12] proposed a slightly different protocol in which the user has to remember images instead of alphanumeric secrets. The protocol was cryptanalyzed and broken in [13]. Usage of images as memory aids is also employed in other graphical authentication schemes such as DeJa Vu [9], Passface [10], Point & Click [11] and [3]. They require little or no numerical computation whatsoever. As an example, the basic theme of [10] is to present the user a series of images, a subset of which is the set of secret images. The user is authenticated if his/her selection of secret images among the given set of images is correct. On the other hand, in [11], the user is authenticated if he/she clicks on the correct secret location in the given picture. [3] works similarly by letting the user draw the secret symbol or figure on a display device. Evidently, these purely graphical

schemes are not secure against shoulder-surfing also known as "peeping" attacks [8]. An observer noting the actions of the user can know the secret readily. For a comprehensive survey of all these protocols, see [8]. Jameel et al [1] proposed a slightly different concept in which the internal properties of images are used as secrets. After a secret has been chosen, pictures which satisfy the properties are presented randomly with pictures which do not. The user has to answer the pictures according to the secret property. It is conjectured that finding out the secret property is a hard problem for adversaries. This concept is employed in this paper with new protocols different from the one proposed in [1].

2 Matsumoto's Threat Model

Matsumoto proposed a threat model in [2] and attempted to devise a human identification protocol secure in this model. Figure 1 shows this threat model pictorially. We have a human user $\mathcal{H}$ who wants to authenticate to a remote server $\mathcal{C}$. $\mathcal{H}$ is equipped with a malicious computing device. $\mathcal{H}$ and $\mathcal{C}$ are connected via an insecure communication channel. The adversary $\mathcal{A}$ enjoys the following capabilities:

P-Channel : Passive access to the communication channel.
A-Channel : Active access to the communication channel.
P-Device : Passive access to the software and hardware of the computing device.
A-Device : Active access to the software and hardware of the computing device.
P-Display : Passive observation of the visual display unit of the computing device.
P-User : Passive observation of users' inputs and actions.

Notice that the adversary has all these capabilities *during* an authentication session. We shall call this model, Full-MTM. A real world example of this model can be visualized as a user attempting to authenticate to a remote server using a computer in a public internet cafe. The computer can potentially be infected by trojan horses and key-loggers. There can be an added risk of shoulder-surfers [8] peeping at the user's input. Hidden cameras, network snoops and spoof websites can complete a real world realization of the Full-MTM.

An interesting line of research is to find human identification protocols secure in relaxed versions of Full-MTM. As an example, consider a variant of the Full-MTM in which $\mathcal{A}$ has only two capabilities: P-Display and P-User. A user entering his/her PIN number on an Automated Teller Machine is a perfect candidate for this threat model. Needless to say, the traditional *PIN number protocol* succumbs completely in the said threat model. Indeed, this protocol is only secure when the adversary has just the P-Display capability. The variation of Full-MTM considered in this paper assumes the adversary to possess all but two capabilities: A-Channel and A-Device. We call this variant P-MTM, to acknowledge that the adversary is strictly passive.

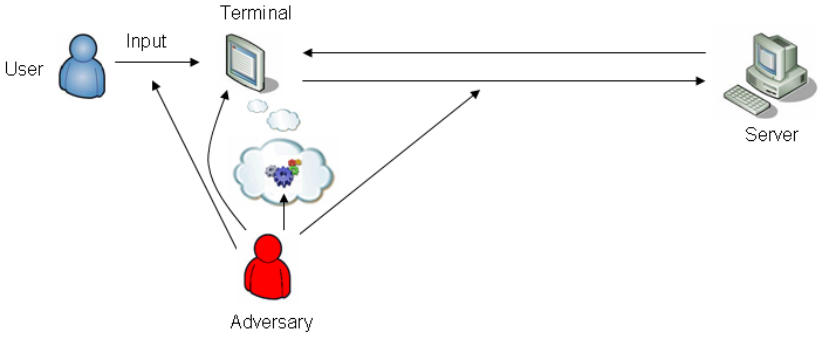

Fig. 1. Matsumoto's Threat Model

3 Human Identification Protocol Based on Images and their Features

Jameel et al [1] proposed a human identification protocol in which $\mathcal{H}$ and $\mathcal{C}$ share a common secret feature (question) which has a binary answer when applied to any image. An example secret feature can be 'an arrow'. So, by looking at the picture in Figure 1, $\mathcal{H}$ has to answer the following question: 'Does the image contain an arrow?'. During an authentication session, $\mathcal{C}$ sends a group of images such that with probability 1/2 each image satisfies the secret feature. $\mathcal{H}$'s response consists of a binary string which is constructed according to a secret permutation, also shared between $\mathcal{H}$ and $\mathcal{C}$. As an example, let the secret permutation be $**2**1345*$. The $*$'s represent the *don't care positions*. $\mathcal{C}$ sends a group of 10 images labeled from 0 to 9. $\mathcal{H}$ replies by constructing a binary string of length 10 by placing random bits in the positions of $*$'s and answering the images in the order specified by the secret permutation and in accordance with the secret question. Here, '1' means that the image satisfies the secret feature and '0' means that it does not.

The job of the adversary is to extract the secret feature. It is conjectured in [1], that since the answer string contains shuffled answers and random bits, it is "very hard" to extract the secret feature, although no concrete value for the hardness of this problem is given. The security of the protocol is reduced from the conjectured security of the said hard problem. Although the authors discuss the resiliency of the scheme against a couple of active attacks, they do not give a reductionist argument [14] against active adversaries. Therefore, we can only safely assume that the protocol in [1] is secure under P-MTM.

We notice the following weak points and deficiencies in the above mentioned protocol:

– The permutation string, if long, can be hard to remember for most of the people. Even with the length '10', as used in [2], remembering the location of the don't care positions is difficult.

- It is not known how a human user can generate random bits. Most users would just put all 0's or all 1's in the don't care positions of the answer string. This is also acknowledged by the authors in [1]. It is also not evident whether the inclusion of random bits adds to the security of the protocol or not.
- Two users with the same secret feature might differ when answering the same image. Thus, a feature/image pair is subject to $\mathcal{H}$'s interpretation which might differ from $\mathcal{C}$'s interpretation. As a result, we should allow for user error in the protocol.
- The suggested parameters in [2] are not suitable for mobile devices. Presenting 10 pictures at a time is very bothersome on the user even if he or she can *scroll* around all the images using a mobile device.
- In [2], whenever $\mathcal{C}$ presents an image to $\mathcal{H}$, it discards it from the repository and never uses it again. We can argue that if the repository of images for a given feature is large, the probability of the same picture being presented in succession is very small. We can thus get rid of this impractical requirement.
- The conjectured difficulty of extracting the secret features from images is justified due to the presence of the secret permutation. However, the underlying mechanism of shuffling the answer bits and employing random bits is very basic and can be cracked once the secret feature is found [15]. The adversary at least has the knowledge that on the average half of the images will contain the secret feature. We can similarly conjecture that the problem is hard even with out the presence of random bits and shuffled answers, although intuitively the hardness assumption will be stronger than in [1]. On the brighter side, this can allow us to construct more practical protocols on mobile devices.

In what follows, we will assume $\mathcal{H}$ to possess a mobile device and the goal is to authenticate $\mathcal{H}$ to $\mathcal{C}$ under P-MTM. We begin by rigorously defining the modified version of the conjecture presented in [1].

4 Preliminaries: Definitions and Conjectures

Denote the three parties by $\mathcal{H}$, $\mathcal{C}$ and $\mathcal{A}$ as in the previous section. $\mathcal{C}$ has a set of Images I and a set of features Q, whose contents are kept secret. Define the function: $f : Q \times I \rightarrow \{0, 1\}$ that takes as argument a $q \in Q$ and an $i \in I$ and maps the pair to 0 or 1. This function is implemented by $\mathcal{C}$. For all $q \in Q$ and $i \in I$, $b \leftarrow \mathcal{H}(q, i)$ represents $\mathcal{H}$ returning an answer bit b after taking as input q and i.

Conjecture 1. Let $q \overset{R}{\leftarrow} Q$ and $i \overset{R}{\leftarrow} I$. Then,

$$\Pr\left[b \leftarrow \mathcal{H}(q, i) ; b = f(q, i)\right] \geq 1 - e \geq \frac{1}{2}$$

Here $0 \leq e \leq \frac{1}{2}$ is the average error probability. For the adversary $\mathcal{A}$ we have the following conjecture:

Conjecture 2. Let $q \xleftarrow{R} Q$ and $b_1 b_2 \cdots b_r \xleftarrow{R} \{0,1\}^r$. Let $i_1, \ldots, i_r$ be sampled from I, such that: $f(q, i_1) || \cdots || f(q, i_r) = b_1 \cdots b_r$. Let $b \xleftarrow{R} \{0,1\}$ and $i \leftarrow I$ such that $f(q,i) = b$ and $i \neq i_t$ for $1 \leq t \leq r$. Then,

$$\frac{1}{2} \leq \Pr\left[b' \leftarrow \mathcal{A}\left(i, (b_1, i_1), \ldots, (b_r, i_r)\right); b' = b\right] \leq \frac{1}{2} + \delta(r)$$

Here $0 \leq \delta(r) \leq \frac{1}{2}$ is a non-decreasing function of r. Notice that the adversary is shown corresponding pairs of images and their binary answers, except for the last image.

Notes. We have not defined $\mathcal{H}$ and $\mathcal{A}$ as turing machines as they could be either humans or a combination of human or machine in the case of the adversary. Conjecture 2 differs from the one presented in [2], in that we do not attribute the hardness of the problem to the extraction of the secret feature. Instead, we have related it to guessing the correct answer of an image after observing a few pairs of images and their binary answers. Furthermore, we do not involve the time consumed by $\mathcal{A}$ in conjecturing the hardness of the problem. The reason being that if $\mathcal{A}$ is a computer program, it is very hard to construct a program that will be able to solve the problem at hand. Instead, $\mathcal{A}$ needs a great deal of human assistance. We cannot say for certain that the longer the human adversary spends time in contemplating the images the better the result will be, since the time is not being used to do numerical computations.

We define the concept of human identification protocol on the basis of the definition of an identification protocol given in [7]. Define the result of the interaction between $\mathcal{H}$ and $\mathcal{C}$ with inputs x and y respectively, by $\langle \mathcal{H}(x), \mathcal{C}(y) \rangle$. For the sake of identification protocols, $\langle \mathcal{H}(x), \mathcal{C}(y) \rangle$ would be either `accept` or `reject`.

Definition 1. *A* human identification protocol *is an interactive protocol between a human $\mathcal{H}$ and a probabilistic program $\mathcal{C}$, both having auxiliary inputs, such that:*

$$\Pr\left[\langle \mathcal{H}(z), \mathcal{C}(z) \rangle = \texttt{accept}\right] \geq p$$
$$\Pr\left[\langle \mathcal{H}(z'), \mathcal{C}(z) \rangle = \texttt{reject}\right] < 1 - p$$

where $z' \neq z$.

For a human identification protocol to be useful, the probability p in Definition 1 should be high. We will say that a candidate protocol is *logically complete* if it satisfies the above definition. Of course, we need another definition for the human usability of such protocols. We take this definition from [7] again, as a reference:

Definition 2. *A human identification protocol is* (α, β, τ)-human executable*, if at least α proportion of the human population can perform the protocol computations unaided and correctly with probability greater than β and with an average time τ.*

5 Proposed Protocols

We begin with the first protocol[1] which is a direct consequence of the main idea. From here onwards, we assume the sets I and Q to be large enough so that the same image is not presented again in quick succession.

5.1 Protocol P1

SETUP. $\mathcal{C}$ samples $q \xleftarrow{R} Q$. $\mathcal{C}$ and $\mathcal{H}$ share q as a secret.
PROTOCOL.

- $\mathcal{C}$ initializes $j \leftarrow 0$.
- Repeat k times
  - $\mathcal{C}$ samples a bit $b \xleftarrow{R} \{0,1\}$. Randomly picks an image i from I such that $f(q,i) = b$. $\mathcal{C}$ sends i to $\mathcal{H}$.
  - $\mathcal{H}$ sends the response bit a to $\mathcal{C}$, where $a \leftarrow \mathcal{H}(q,i)$.
  - If $a = b$, $\mathcal{C}$ updates $j \leftarrow j + 1$.
- If $j \geq s$, $\mathcal{C}$ outputs `accept`. Else $\mathcal{C}$ outputs `reject`.

Logical Completeness. We see that the protocol is logically complete if Conjectures 1 and 2 hold.

Claim 1. *If Conjectures 1 and 2 hold, Protocol* P1 *is logically complete with probability p, subject to the following conditions:*

$$p \leq \sum_{j=s}^{k} \binom{k}{j} (1-e)^j e^{k-j}$$

$$1 - p > \sum_{j=s}^{k} \binom{k}{j} \left(\frac{1}{2} + \delta(k)\right)^j \left(\frac{1}{2} + \delta(k)\right)^{k-j}$$

Proof. We see that if $\langle \mathcal{H}(q), \mathcal{C}(q) \rangle =$ `accept`, then $\mathcal{H}(q,i)$ should be equal to $f(q,i)$ at least s times. From Conjecture 1, the probability that $\mathcal{H}(q,i)$ equals $f(q,i)$ is $\geq 1 - e$. The sum of the probabilities of s or more successes, can be found through the binomial probability distribution. Thus for a given probability p of the event $\langle \mathcal{H}(q), \mathcal{C}(q) \rangle =$ `accept`, the sum of the probabilities of s or more successes should be greater than or equal to p.
The event $\langle \mathcal{H}(q'), \mathcal{C}(q) \rangle =$ `accept` can happen with probability:

$$\sum_{j=s}^{k} \binom{k}{j} \left(\frac{1}{2} + \delta(0)\right)^j \left(\frac{1}{2} - \delta(0)\right)^{k-j}$$
$$\leq \sum_{j=s}^{k} \binom{k}{j} \left(\frac{1}{2} + \delta(k)\right)^j \left(\frac{1}{2} - \delta(k)\right)^{k-j}$$

Thus this probability has to be less than $1 - p$, for Protocol P1 to be logically complete. □

Table 2 in Appendix A shows the values of s against different values of p, e and k.

[1] A similar protocol is described in an unpublished technical report [15].

5.2 Protocol P2

Denote by $\sigma[m]$, the set of all permutations from $\{0, 1, \ldots, m\}$ to itself. Let $\sigma \in \sigma[m]$ be a generic permutation. $\sigma(j)$ denotes the jth element of σ. Let Grid be a data structure that holds an ordered sequence of m images. The operation '+' means the append operation when applied to Grid.
SETUP. $\mathcal{C}$ samples $q \overset{R}{\leftarrow} Q$ and $\sigma \overset{R}{\leftarrow} \sigma[m]$. $\mathcal{C}$ and $\mathcal{H}$ share q and σ as a secret.
PROTOCOL.

- $\mathcal{C}$ initializes $j \leftarrow 0$.
- Repeat k' times
  - $\mathcal{C}$ initializes Grid $\leftarrow \phi$. $\mathcal{C}$ samples $b_1 b_2 \cdots b_m \overset{R}{\leftarrow} \{0,1\}^m$.
  - For $1 \leq t \leq m$:
    - $\mathcal{C}$ randomly picks an image $i_{\sigma(t)}$ from I such that $f\left(q, i_{\sigma(t)}\right) = b_t$. $\mathcal{C}$ updates Grid $\leftarrow$ Grid $+ i_{\sigma(t)}$.
  - $\mathcal{C}$ sends Grid to $\mathcal{H}$.
  - $\mathcal{H}$ initializes the answer string $a \leftarrow null$.
  - For $1 \leq t \leq m$:
    - $\mathcal{H}$ updates $a \leftarrow a || \mathcal{H}\left(q, i_{\sigma(t)}\right)$.
  - $\mathcal{H}$ sends a to $\mathcal{C}$.
  - For $1 \leq t \leq m$, if $a(t) = b(t)$, $\mathcal{C}$ updates $j \leftarrow j + 1$.
- If $j \geq s$, $\mathcal{C}$ outputs `accept`. Else $\mathcal{C}$ outputs `reject`.

Logical Completeness

Claim 2. *If Conjectures 1 and 2 hold, Protocol* P2 *is logically complete with probability p, subject to the following conditions:*

$$p \leq \sum_{j=s}^{mk'} \binom{mk'}{j} (1-e)^j \, e^{mk'-j}$$

$$1 - p > \sum_{j=s}^{mk'} \binom{mk'}{j} \left(\frac{1}{2} + \delta(mk')\right)^j \left(\frac{1}{2} + \delta(mk')\right)^{mk'-j}$$

Proof. The first part of the proof is similar to the first part of the last claim with k replaced by mk'. For the second part, we see that the probability of the event $\langle \mathcal{H}(q'), \mathcal{C}(q) \rangle =$ `accept` would be less than or equal to the corresponding probability of the same event in Protocol P1, since now the introduction of the permutation σ adds extra error probability for someone without the knowledge of q and σ. The result follows immediately by once again replacing k by mk'. □

For the allowable values of s, see Table 2 in Appendix A.

6 Security of the Protocols

Let $\mathcal{A}$ be an adversary in P-MTM. Let $\mathcal{A}'$ be another adversary with only the P-Channel capability. Informally speaking, in order to demonstrate the security of the protocols, we will first show that adversary $\mathcal{A}$ has no real advantage over adversary $\mathcal{A}'$. After that, we will attempt to reduce the security of the protocols to Conjecture 2. Once that is accomplished, we will say that the protocol is secure under P-MTM with an associated probability. Let $T(\mathcal{H}(.), \mathcal{C}(.))$ denote the transcript of messages sent between $\mathcal{H}$ and $\mathcal{C}$ in a single run. In other words, it represents *one* challenge-response pair.

Definition 3. *We say that the adversary $\mathcal{A}$ is $\lambda(r)$-almost equivalent to the adversary $\mathcal{A}'$ for the human identification protocol $(\mathcal{H}, \mathcal{C})$, if:*

$$\begin{aligned} &\Pr\left[\langle \mathcal{A}\left(T^r\left(\mathcal{H}(q), \mathcal{C}(q)\right)\right), \mathcal{C}(q)\rangle = \texttt{accept}\right] \\ \leq &\Pr\left[\langle \mathcal{A}'\left(T^r\left(\mathcal{H}(q), \mathcal{C}(q)\right)\right), \mathcal{C}(q)\rangle = \texttt{accept}\right] + \lambda(r) \end{aligned}$$

Definition 4. *A human identification protocol $(\mathcal{H}, \mathcal{C})$ is (p', r) secure against the adversary $\mathcal{A}'$, if:*

$$\Pr\left[\langle \mathcal{A}'\left(T^r\left(\mathcal{H}(q), \mathcal{C}(q)\right)\right), \mathcal{C}(q)\rangle = \texttt{accept}\right] \leq p'$$

Definition 4 is taken from [7]. We can now relate the security of the protocol to Conjecture 2 in a straightforward manner. Notice that these proofs are not reductionist arguments in the formal sense, as we have not defined the adversary as being a turing machine.

Claim 3. *Protocol* P1 *is (p', r) secure under* P-MTM*, where:*

$$p' = \sum_{j=s}^{r} \binom{r}{j} \left(\frac{1}{2} + \delta(r)\right)^j \left(\frac{1}{2} - \delta(r)\right)^{r-j}$$

Proof. See Appendix B.1.

Claim 4. *Protocol* P2 *is (p'', r) secure under* P-MTM*, where $p'' \leq p' + \delta(r)$ and:*

$$p' = \sum_{j=s}^{r} \binom{r}{j} \left(\frac{1}{2} + \delta(r)\right)^j \left(\frac{1}{2} - \delta(r)\right)^{r-j}$$

Proof. See Appendix B.2.

7 Implementation

We implemented the protocols by employing a small set of test images and features. The server side maintained a database of features which contained a record of features, the corresponding question and the number of images pertaining to the features. The server also maintained an archive of images related to the features. The images used were scaled down to be more feasible for use on a mobile device using wireless communication. An example feature used was: "cartoon". The corresponding question was: "Does the picture contain a cartoon?". The image archive contained two disjoint sets of images for the feature "cartoon". The images which satisfied the feature were labeled *yes-images* and the ones that did not were labeled as *no-images*. The administrator had the privilege to add, remove or update features and/or images in the feature database and the image archive.

The client side was implemented on a PDA. A small prototype for Protocol P2 was made, with $m = 4$, but was not used for subsequent study. Instead Protocol P1 was fully implemented and a user survey was made. Each pass of the protocol implemented, consisted of an image displayed on the PDA showing two buttons at the bottom, labeled "yes" or "no", as shown in Figure 2. The values $k = 10$ and $s = 8$ were chosen, which meant that the user would be accepted if he or she answers correctly at least 8 times. A user was registered by first assigning a unique *ID* to the user after which he or she was given the secret feature. Once registration was done, the users could freely test the identification protocol.

The users used in the survey were all graduate school students. They were asked to attempt Protocol P1 3 times. The average time taken by the user as well as the number of errors was noted down. The number of times the users failed to authenticate was recorded. After the experiment, the users were asked to describe their experience of using the protocol in terms of the difficulty of usage

Fig. 2. One pass of Protocol P1 as implemented on the PDA

Table 1. User Statistics and Experience

Success Percentage and Average Time

Users	Attempts	Successes	Success %	e	Avg. Time
5	15	12	80	0.113	25.6 sec

Difficulty of Usage

Total	Easy	Normal	Difficult
5	4	1	0

Lengthiness

Total	Normal	Little Long	Too Long
5	1	4	0

and lengthiness. Both these characteristics were divided into three qualitative categories as shown in Table 1.

The average time taken by the users came out to be 25.6 seconds, with the best time being 21 seconds and the worst being 38 seconds. The maximum number of errors in an authentication session was 4. It should be noted that since the number of images were small, some images were repeated a small number of times. A user erring in one of the images would err again with the same image. This will certainly not be the case if the set of images is large. The average time of 25.6 seconds meant that the user on average spent 2.56 seconds per image. Since 80 percent of the users described executing the protocol as easy, we can roughly associate the probability 0.80 with the variable α in Definition 2. Also, the total successful attempts were 12 out of 15 which allows us to loosely associate a probability 0.80 with the parameter β. Thus we can state that Protocol P1 is $(0.80, 0.80, 2.56k)$-human executable, where k is the protocol's security parameter. We do acknowledge the extremely low number of test subjects due to resource limitations. Such a small portion might not be a complete representative of the human population. Nevertheless we can get a rough idea about the usability of the protocol. Finally it should be noted that most users described the protocol as a *little lengthy*. This opinion is sure to change into *very lengthy* if the value of k is chosen to be somewhere around 20. We acknowledge this as a drawback of the protocol.

8 Conclusion

While secure human identification protocols have been proposed in the literature for normal desktop computers, it is an interesting line of research to construct protocols that can be used on mobile devices. Mobile devices among other constraints have a smaller display unit. We have proposed variations of the protocol proposed in [1] that are practical for devices with smaller display units. However, these protocols are inherently heavy on devices with limited memory and computational power. Overcoming this limitation remains an open problem as most

of the human identification protocols are graphics-based which inherently makes them unsuitable for resource constraint devices. The proposed protocols are secure under a slightly different conjecture from the original one. The protocols are usable if used sparingly, as the number of rounds in one authentication session is required to be high for strong security. Furthermore, it is not evident how the protocols will be secure under an active adversary without compromising the usability. It is therefore desirable to propose human identification protocols secure against active and passive adversaries alike as well as being practical on resource limited devices. The proposed protocols in this paper are a small step in that direction.

References

1. Jameel, H., Shaikh, R.A., Lee, H., Lee, S.: Human identification through image evaluation using secret predicates. In: Abe, M. (ed.) CT-RSA 2007. LNCS, vol. 4377, pp. 67–84. Springer, Heidelberg (2006)
2. Matsumoto, T., Imai, H.: Human Identification through Insecure Channel. In: Davies, D.W. (ed.) EUROCRYPT 1991. LNCS, vol. 547, pp. 409–421. Springer, Heidelberg (1991)
3. Jermyn, I., Mayer, A., Monrose, F., Reiter, M., Rubin, A.: The design and analysis of graphical passwords. In: 8th USENIX Security Symposium (1999)
4. Wang, C.H., Hwang, T., Tsai, J.J.: On the Matsumoto and Imai's Human Identification Scheme. In: Guillou, L.C., Quisquater, J.-J. (eds.) EUROCRYPT 1995. LNCS, vol. 921, pp. 382–392. Springer, Heidelberg (1995)
5. Matsumoto, T.: Human-computer cryptography: An attempt. In: 3rd ACM Conference on Computer and Communications Security, pp. 68–75. ACM Press, New York (1996)
6. Li, X.-Y., Teng, S.-H.: Practical Human-Machine Identification over Insecure Channels. Journal of Combinatorial Optimization 3, 347–361 (1999)
7. Hopper, N.J., Blum, M.: Secure Human Identification Protocols. In: Boyd, C. (ed.) ASIACRYPT 2001. LNCS, vol. 2248, pp. 52–66. Springer, Heidelberg (2001)
8. Li, S., Shum, H.-Y.: Secure Human-computer Identification against Peeping Attacks (SecHCI): A Survey. Unpublished report, available at Elsevier's Computer Science Preprint Server (2002)
9. Dhamija, R., Perrig, A.: Deja Vu: A User Study using Images for Authentication. In: Proc. of the 9th USENIX Security Symposium, pp. 45–58 (2000)
10. Passfaces Corporation: White Paper. The Science behind Passfaces (2005), `http://www.passfaces.com`
11. Sorensen, V.: PassPic - Visual Password Management (2002), `http://www.authord.com`
12. Weinshall, D.: Cognitive Authentication Schemes Safe Against Spyware (Short Paper). In: IEEE Symposium on Security and Privacy, pp. 295–300 (2006)
13. Golle, P., Wagner, D.: Cryptanalysis of a Cognitive Authentication Scheme. Cryptology ePrint Archive, Report 2006/258, `http://eprint.iacr.org/`
14. Bellare, M.: Practice-Oriented Provable-Security. In: Okamoto, E. (ed.) ISW 1997. LNCS, vol. 1396, pp. 221–231. Springer, Heidelberg (1998)
15. Jameel, H., Lee, H., Lee, S.: Using Image Attributes for Human Identification Protocols. Technical Report, CoRR abs/0704.2295 (2007), `http://arxiv.org/abs/0704.2295`

A Numerical Values of Parameters in Protocol P1 and P2

Table 2 illustrates different allowable values for the parameters in Protocol P1 and P2[2]. So, for instance, if $p \geq 0.90$ and $k = 10$ is desired, s should be equal to 8, with the error e being less than or equal to 0.1 and the adversary's advantage $\delta(10) \leq 0.05$. The '∞' symbol indicates that $\delta(k)$ is undefined for the corresponding choice of parameters. In other words, adversary's probability of success will always be higher than $1 - p$ for these choices. Hence Protocols P1 and P2 are not logically complete under these parameter values.

Table 2. Numerical Values of the Parameters

p	e	k	s	$\delta(k)$	p	e	k	s	$\delta(k)$	p	e	k	s	$\delta(k)$
0.90	0.10	10	8	≤ 0.05	0.80	0.10	10	8	≤ 0.11	0.70	0.10	10	9	≤ 0.27
		15	12	≤ 0.10			15	13	≤ 0.23			15	13	≤ 0.27
		20	16	≤ 0.13			20	17	≤ 0.24			20	17	≤ 0.27
		25	21	≤ 0.20			25	21	≤ 0.24			25	22	≤ 0.31
	0.20	10	6	∞		0.20	10	7	∞		0.20	10	7	≤ 0.0655
		15	10	∞			15	11	≤ 0.0923			15	11	≤ 0.1322
		20	14	≤ 0.0327			20	15	≤ 0.1335			20	15	≤ 0.1675
		25	17	≤ 0.0327			25	18	≤ 0.1176			25	19	≤ 0.1895
	0.30	10	5	∞		0.30	10	6	∞		0.30	10	6	∞
		15	8	∞			15	9	∞			15	10	≤ 0.0648
		20	11	∞			20	12	∞			20	13	≤ 0.0658
		25	15	∞			25	16	≤ 0.0358			25	16	≤ 0.0672

B Security of the Protocols

B.1 Proof of Claim 3

There is no computation done by $\mathcal{H}$ or the computing device except for displaying the image. Now, if the display unit is *large enough* to display the whole image, then $\mathcal{A}$ has no advantage over $\mathcal{A}'$ whatsoever. Therefore, $\mathcal{A}$ is 0-almost equivalent to $\mathcal{A}'$, where $\lambda(r) = 0$ is the constant function. Let us assume that $\mathcal{A}'$ defeats the protocol with probability $> p'$. We construct an adversary $\mathcal{B}$ that uses $\mathcal{A}'$ to violate Conjecture 2. In the *training* phase, $\mathcal{B}$ simply provides $\mathcal{A}'$ the images being given to it and with probability $1 - e$ it provides the correct answers to these images to $\mathcal{A}'$. When $\mathcal{A}'$ completes the training phase, $\mathcal{B}$ provides the image i to $\mathcal{A}'$ whose answer $\mathcal{B}$ has to guess. Whenever, $\mathcal{B}$ gets an answer it outputs the answer and halts. Since:

$$p' = \sum_{j=s}^{r} \binom{r}{j} \left(\frac{1}{2} + \delta(r)\right)^{j} \left(\frac{1}{2} - \delta(r)\right)^{r-j}$$

[2] Notice that for Protocol P2, $k = mk'$.

This means that the probability that $\mathcal{B}$'s guess is correct is $> \frac{1}{2} + \delta(r)$, where r is the number of runs the adversary $\mathcal{A}'$ needs in the training phase. This contradicts Conjecture 2 and the result follows. □

B.2 Proof of Claim 4

The computing device has to display images in the correct order which is also visible to $\mathcal{A}'$. $\mathcal{H}$ just needs to recall σ and hence does not have to do any computation. If the display unit is large enough to display all m images at the same time then adversary $\mathcal{A}$ has no advantage over $\mathcal{A}'$ in this regard as well. However, if the computing device has a smaller display unit (a mobile device), then $\mathcal{H}$ has to scroll left and right to answer the images in the order specified by σ. Thus $\mathcal{A}$ has an advantage over $\mathcal{A}'$, but which cannot be more than $\delta(r)$, or else it will violate Conjecture 2. Therefore, $\mathcal{A}$ is $\delta(r)$-almost equivalent to $\mathcal{A}'$, where $\lambda(r) = \delta(r)$. Suppose now that $\mathcal{A}'$ defeats the protocol with probability $> p'$. We can construct an adversary $\mathcal{B}$ which uses $\mathcal{A}'$ in the same way as in the proof of Claim 3 except now it samples a random $\sigma \xleftarrow{R} \sigma[m]$ and feeds $\mathcal{A}'$ with the images and their answers in accordance with σ. By an argument similar to the previous section we can show that the probability that $\mathcal{B}$'s guess is correct is $> \frac{1}{2} + \delta(r)$, where $r \equiv 0 \bmod m$ is the number of images or answers shown to $\mathcal{A}'$ in the training phase. Since this contradicts Conjecture 2, we can say that Protocol P2 is (p', r) secure against the adversary $\mathcal{A}'$ and (p'', r) secure under P-MTM, where $p'' \leq p' + \delta(r)$. □