On the Matsumoto and Imai's Human Identification Scheme

Chih-Hung Wang, Tzonelih Hwang, and Jiun-Jang Tsai

Institute of Information Engineering
National Cheng-Kung University
Tainan, Taiwan, R.O.C.

Abstract. At Eurocrypt'91, Matsumoto and Imai presented a human identification scheme for insecure channels, which is suitable for human ability of memorizing and processing a short secret. It prevents an intruder from peeping user in typing password on terminal connected to the central computer. However, in this paper, we are going to propose a new attack, called the *replay challenge attack*, where a malicious terminal pretends to be the host and replays the host's challenges to reveal the secret password. A modified scheme will be proposed to avoid this attack.

1 Introduction

It is very often in the computer systems that an end user has to identify himself to a host. Though human identifications using personal characters, such as fingerprint, voice, or the retinal blood-vessel pattern of a human eye, have been developed and applied actually [5], physical devices for special purpose have to be designed and the cost of these devices are very high.

The design of human identifications without the help of any auxilary device has become an important issue. The password authentication scheme, where a log-in user simply memorizes a short secret and presents it to the host for user authentication, however, suffers both *the peeping attacks* where an intruder stands behind the log-in user to peep the typed password and *the replay attacks* where the intruder intercepts the password from the network and then impersonates the same user by replaying the intercepted password.

In 1991, Matsumoto and Imai proposed a human identification scheme for insecure channel to avoid both replay and peeping attacks by use of a simple challenge-response protocol [3]. Each user and the host are assumed to share a common key. Knowing the common key shared with the user, the verifier (the host) can decide whether an answer replied from the prover (the user) is correct or not. In their scheme, what the user has to do are simply to memorize a short secret and perform very simple operations based on the secret.

In this paper, we are going to study the security of Matsumoto and Imai's human identification scheme by proposing a new attack on it. By this attack, a malicious process first pretends to be the host by replaying a challenge to the login user, and then performs the intercepting or peeping attack to reveal the

L.C. Guillou and J.-J. Quisquater (Eds.): Advances in Cryptology - EUROCRYPT '95, LNCS 921, pp. 382-392, 1995.

secret by observing the differences in the responses of the login user. Since this kind of attack is particular useful in the environment where the login process is limited to human ability of performing computations, it is valuable to be pointed out here for consideration of constructing even more secure human identification scheme. In addition, we also proposed a modified scheme to avoid this attack.

The structure of this paper is as follows. In Section 2, we review the Matsumoto and Imai's scheme. Then in Section 3. two attacks are proposed to analyze the security of Matsumoto and Imai's scheme. Section 4 proposes a modified scheme to avoid the replay challenge attack and analyzes the security of the newly modified scheme. Finally, some concluding remarks will be given in Section 5.

2 Matsumoto and Imai's Human Identification Scheme

2.1 Notations and Definitions

The following definitions and notations are used in the entire paper.

- $< n >$: the set of all positive integers less than or equal to n.
- $g \circ f$: the composite function of functions f and g.
- Ω : the whole alphabet.
- Q : the question alphabet; a subset of Ω.
- W : the window alphabet; a subset of Q.
- A : the answer alphabet; a subset of Ω.
- $|S|$: the number of elements in the set S. Note: $2 \leq |A| \leq |W| < |Q| \leq |\Omega|$.
- β : the number of blocks.
- α : a threshold value; $1 \leq \alpha \leq \beta$.
- q_j : the jth question block, which is a bijection from $< |Q| >$ to Q.
- a_j : the jth answer block, which is a surjection from $< |Q| >$ onto A.
- SW : a string of secret word, which is a surjection from $< |W| >$ onto A.
- f_j : the window in q_j, which is an injection from $< |W| >$ into $< |Q| >$ such that

$$f_j = sort(\{i \in < |Q| > | \; q_j(i) \in W \; \})$$

, where the sort function is defined below.

Definition 1 *[3] For a totally ordering finite set $(S, \leq)$, the function sort(S) is defined as a bijection, bf, from $< |S| >$ onto S such that*

$$bf(1) \leq bf(2) \leq \cdots \leq bf(|S|)$$

2.2 Matsumoto and Imai's Scheme

The system determines the parameters $\Omega, Q, W, A, \alpha, \beta$ first. A string of secret word SW is known only to P and V. Then, P can identify himself to V by the following steps.

Step1: V generates β question blocks $q_1, q_2, \cdots, q_\beta$, and sends them to the prover P.

Step2: P selects at least α distinct question blocks out from these β blocks to generate the answer blocks by the following substeps.

Step2.1: For each selected question block q_j, P computes the window

$$f_j = sort(\{i \in < |Q| > | \, q_j(i) \in W \}).$$

Step2.2: P generates the answer blocks

$$a_j\left(f_j(t)\right) = SW(t), \text{for } t = 1, 2, \cdots, |W|.$$

Step2.3 For each r, where $r \in < |Q| >$, and $r \neq f_j(t)$, for $t = 1, 2, \cdots |W|$, P randomly and uniformly selects an elements from A and allocates it to $a_j(r)$.

Step3: For each question block q_j, which is not selected by the prover P at Step2, P allocates a_j with a random block R_j, which is a surjection from $< |Q| >$ onto A.

Step4: P sends the answer blocks $a_1, a_2, \cdots, a_\beta$ to V.

Step5: V verifies the answers as follows.

Step5.1: For each question blocks q_j, V computes the window

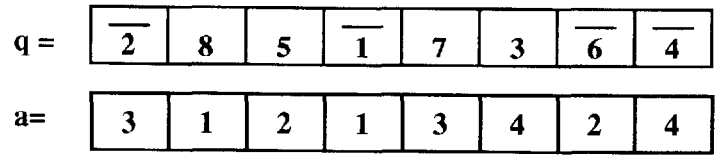
$$f_j = sort(\{i \in < |Q| > | \, q_j(i) \in W \}).$$

Step5.2: If the number of correct answer blocks, i.e., $a_j \circ f_j = SW$, is greater than or equal to α, then V accepts P; otherwise, V rejects P.

2.3 An Example

A simplified example with $\alpha = \beta = 1$ is described here to illustrate Motsumoto and Imai's scheme.

In this example, we use the notations a, q, and f to denote the answer, question and window respectively. Let the question alphabet $Q = \{1, 2, 3, 4, 5, 6, 7, 8\}$, window alphabet $W = \{1, 2, 4, 6\}$ and the answer alphabet $A = \{1, 2, 3, 4\}$. The prover has a string of secret word $SW = 3124$ shared with the verifier. Figure 1 shows the challenge and response between the prover and verifier. The bars in the positions of q show the window positions.



$W = \{1, 2, 4, 6\}$
$SW = 3124$
verify $a \circ f \stackrel{?}{=} SW$

Figure 1: An Example of Answer and Question

If an intruder wants to guess the secret word of the prover, he must guess the window alphabet correctly. Since the window size $|W| = 4$ and the question alphbet size $|Q| = 8$, an intruder has to find the window W in $\binom{|Q|}{|W|} = \binom{8}{4}$ trials according to the analysis in [3].

3 The Attacks on Matsumoto and Imai's Scheme

In this section, we propose two attacks on Matsumoto and Imai's scheme . The first attack is a passive attack where the intruder passively observe the login user's responses to a challenge and then intents to guess his password. The second attack, called the *replay challenge attack*, is an active attack where the intruder actively replay the host's challenges to reduce the number of trials in finding out the window W. Both attacks show lower security levels of the scheme than the original proposed one in [3].

3.1 The passive attack

The idea of this attack is based on the following observation: according to the definitions, SW contains all elements of the answer alphabet A at least once, and the other elements in a_j are also from A. Then, one can reduce the number of trials to reveal the password by this observation. Consider the example shown in Fig. 1 again, since the login user's answer contains all elements of $A = \{1, 2, 3, 4\}$ exactly twice, an intruder simply selects one from two of the same elements to guess the window positions and then reveals the user's password based on the window positions. The trials will be reduced to $\binom{2}{1}^4$ in this case which are less than $\binom{8}{4}$ claimed by Matsumoto and Imai.

For simplicity, we consider the case where $\alpha = \beta = 1$. Let $A = \{e_1, e_2, \cdots e_{|A|}\}$; t_i denote the number of e_i in the answer a, for $i = 1, 2, \cdots, |A|$. Then $|Q| = \sum_{i=1}^{|A|} t_i$. s_i is the number of times the elements e_i appear in the secret word SW . Then, $s_1 + s_2 + \cdots + s_{|A|} = |W|$; $1 \leq s_i \leq Minimum\{|W|, t_i\}$. Thus, the password can be revealed in at most $\sum_{s_1+s_2+\cdots+s_{|A|}=|W|} \binom{t_1}{s_1}\binom{t_2}{s_2}\cdots\binom{t_{|A|}}{s_{|A|}}$ trials.

We will show that the number of trials mentioned above is less than $\binom{|Q|}{|W|}$ in the following theorem.

Theorem 1.

$$\sum_{\substack{s_1+s_2+\cdots+s_{|A|}=|W| \\ 1\leq s_i\leq Minimum\{|W|,t_i\}}} \binom{t_1}{s_1}\binom{t_2}{s_2}\cdots\binom{t_{|A|}}{s_{|A|}} < \binom{|Q|}{|W|}.$$

Proof. Since $|Q| = \sum_{i=1}^{|A|} t_i$ and $|W| = \sum_{i=1}^{|A|} s_i$, we have

$$\binom{|Q|}{|W|} = \sum_{\substack{s_1+s_2+\cdots+s_{|A|}=|W| \\ 0\leq s_i\leq Minimum\{|W|,t_i\}}} \binom{t_1}{s_1}\binom{t_2}{s_2}\cdots\binom{t_{|A|}}{s_{|A|}}.$$

It is obvious that

$$\sum_{\substack{s_1+s_2+\cdots+s_{|A|}=|W| \\ 0\le s_i\le Minimum\{|W|,t_i\}}} \binom{t_1}{s_1}\binom{t_2}{s_2}\cdots\binom{t_{|A|}}{s_{|A|}} > \sum_{\substack{s_1+s_2+\cdots+s_{|A|}=|W| \\ 1\le s_i\le Minimum\{|W|,t_i\}}} \binom{t_1}{s_1}\binom{t_2}{s_2}\cdots\binom{t_{|A|}}{s_{|A|}}.$$

Therefore

$$\binom{|Q|}{|W|} > \sum_{\substack{s_1+s_2+\cdots+s_{|A|}=|W| \\ 1\le s_i\le Minimum\{|W|,t_i\}}} \binom{t_1}{s_1}\binom{t_2}{s_2}\cdots\binom{t_{|A|}}{s_{|A|}}$$

(Q.E.D.)

3.2 The Replay Challenge Attack

In this section, we are going to propose a new attack called the *replay challenge attack* where an attacker impersonates the host to replay an intercepted challenge to the login user and then peeps or intercepts the answers from the user. In an insecure login environment, this attack is considered feasible, since it is not difficult for a malicious node (terminal) to replay a challenge to the end user and then return to the normal state after intercepting the response. For simplicity, we also consider the case where $\alpha = \beta = 1$. By collecting a few answers of the same challenge, the intruder can figure out the secret word of the prover with a high probability . Let $a(i)$, $q(i)$ denote the contents of the ith position in the answer block a and question block q respectively. The positions corresponding to W in the distinct answer blocks for the same question should have the same contents, i.e.,

$$a \circ f = SW = a' \circ f$$

,where a and a' denote these two distinct answer blocks for the same challenge.

Thus, if an intruder discovers that some corresponding positions in the answer blocks a and a' whose contents are distinct, then these positions cannot be the positions of the window W. Thus, we have the following Lemma.

Lemma 1 *Let a and a' be two distinct answer blocks of the same question q . If there exists an i, $i \in < |Q| >$, such that $a(i) \neq a'(i)$, then $q(i) \notin W$.*

Theorem 2. *The window W of Matsumoto and Imai's human identification scheme with $\alpha = \beta = 1$ can be found in $\sum_{k=0}^{|Q|-|W|} \frac{\binom{|Q|-|W|}{k}(|A|-1)^k}{|A|^{|Q|-|W|}} \times \binom{|Q|-k}{|W|}$ expected trials if an intruder replays the same question one time.*

Proof. Let the integer k denote the number of i's such that $a(i) \neq a'(i)$, for $i = 1, 2, \cdots, |Q|$. It is true that $0 \leq k \leq |Q| - |W|$. By Lemma 1, these k positions do not belong to W. In this case, W can be found in $\binom{|Q|-k}{|W|}$ trials. Comparing to the answer a, each of these k positions of a' can only be padded by $(|A|-1)$ possible elements. So, the number of possible permutations for this case is given by $\binom{|Q|-|W|}{k}(|A|-1)^k$. In addition, the other $(|Q| - |W| - k)$ positions of a' must be filled by the same elements of the corresponding positions of a. Thus, the probability for this is $\frac{\binom{|Q|-|W|}{k}(|A|-1)^k}{|A|^{|Q|-|W|}}$ assuming that each element in A is equally likely to be selected by P. Thus, the expected number of trials of finding out W are $\sum_{k=0}^{|Q|-|W|} \frac{\binom{|Q|-|W|}{k}(|A|-1)^k}{|A|^{|Q|-|W|}} \times \binom{|Q|-k}{|W|}$.

(Q.E.D.)

Corollary 1 *Similar to Theorem 2, an intruder can find the window W in*

$$\sum_{k=0}^{|Q|-|W|} \frac{\binom{|Q|-|W|}{k}(|A|^n-1)^k}{|A|^{n(|Q|-|W|)}} \times \binom{|Q|-k}{|W|}$$

expected trials if the same question is replayed n times.

Proof. To locate the window positions in the replayed question, the intruder first figures out the positions which do not belong to the window. He replays the same question n times. Then he constructs a matrix using these answer blocks $\{a^0, a^1, \cdots, a^n\}$ as shown in Figure 2. The element at $(x, y) = (row,\ column)$ is denoted as $a^x(y)$, where $0 \leq x \leq n, 1 \leq y \leq |Q|$. If an intruder finds that the elements in one column, say $\hat{y}$, are not all the same (i.e., $a^0(\hat{y}) = a^1(\hat{y}) = \cdots = a^n(\hat{y})$ is not true), then $q(\hat{y}) \notin W$. The probability of finding k non-window positions is given by

$$\frac{\binom{|Q|-|W|}{k}(|A|^n-1)^k}{|A|^{n(|Q|-|W|)}}.$$

Therefore, the expected number of trials of finding W are

$$\sum_{k=0}^{|Q|-|W|} \frac{\binom{|Q|-|W|}{k}(|A|^n-1)^k}{|A|^{n(|Q|-|W|)}} \times \binom{|Q|-k}{|W|}$$

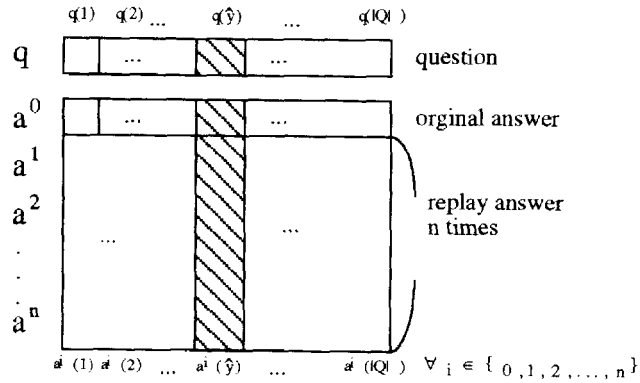


Figure 2. The matrix formed by answer blocks

(Q.E.D.)

Corollary 2 *If the number of times an intruder replays the same question tends to infinity, then the expected number of trials of finding the window W will degrade to* 1

(i.e., $\lim_{n\to\infty} \sum_{k=0}^{|Q|-|W|} \frac{\binom{|Q|-|W|}{k}(|A|^n-1)^k}{|A|^{n(|Q|-|W|)}} \times \binom{|Q|-k}{|W|} = 1$ *).*

We use the following examples to show the validity of our attack.

Test1: $|\Omega| = |Q| = 36$, $|W| = 18$, $|A| = 2$, and $\alpha = \beta = 1$.
Test2: $|\Omega| = |Q| = 50$, $|W| = 10$, $|A| = 3$, and $\alpha = \beta = 1$.

Then, the expected number of trials of finding W are as follows

Value of n	*Expected trials in Test1*	*Expected trials in Test2*
1	3.08×10^7	3.78×10^6
3	8165.32	174.2
5	65.8	3.328
7	5.3	1.19
9	1.73	1.02

Both attacks discussed previously in this section can be combined into one to improve the probability of revealing the secret word. The first phase of the combined attack is to utilize the replay challenge attack to find the positions of q not belonging to the window W. Then use the passive attack as the the second phase to guess the correct positions of window. Let m $(0 \leq m \leq |W|)$ be the number of positions not belonging to the window W found in first step. The number of elements in these positions of answer a can be subtracted from

the corresponding t_i (refer to Section 3.1). Let t_i' be the results of above process and $t_i' \leq t_i$; $\sum_{i=1}^{|A|}(t_i - t_i') = m$. The password can be revealed in at most $\sum_{1 \leq s_i \leq Minimum\{|W|, t_i'\}} \binom{t_1'}{s_1}\binom{t_2'}{s_2}\cdots\binom{t_{|A|}'}{s_{|A|}}$ trials.

4 Modified Identification Scheme

4.1 The Scheme

A modified identification scheme is devised to avoid the replay challenge attack mentioned in the Section 3. However, the computation required to be executed by the human may be too complicate in the modified scheme thus it may not be a practical human identification scheme.

The answer block is divided into two parts. One is called the *window-padding answer portion* denoted as $a_j(f_j(t))$, for $t = 1, 2, \cdots, |W|$; the other is called the *random-padding answer portion* denoted as $a_j(r)$, $1 \leq r \leq |Q|$, $r \neq f_j(t)$, for $t = 1, 2, \cdots, |W|$. In Matsumoto and Imai's scheme, window-padding answer portion is filled with the secret word SW (i.e., $a_j \circ f_j = SW$). However, as discussed previously, once the question is replayed, the window-padding answer portion of the answers remains the same. In order to avoid this weakness, we introduce a new function Υ_j , which can transfer the scheme from deterministic to probabilistic. In our modified scheme, let $|Q|$ be even and $|W| = \frac{1}{2}|Q|$. Υ_j is defined as follows.

Let the function g_j be an injection from $< |W| >$ into $< |Q| >$, such that

$$g_j = sort\left(\{i \in < |Q| > |\ q_j(i) \notin W\}\right) .$$

Definition 2 *The function $Srand_j$ is a bijection from $< |W| >$ onto $< |W| >$, such that*

(i) if $a_j(g_j(i)) < a_j(g_j(i'))$, then $Srand_j(i) < Srand_j(i')$

(ii) if $a_j(g_j(i)) = a_j(g_j(i'))$, and $i < i'$, then $Srand_j(i) < Srand_j(i')$

, where $i, i' \in < |W| >$ and the order of a_j is corresponding to the order of its ASCII code.

Definition 3 *The function Υ_j is a surjection from $< |W| >$ onto A , such that*

$$\Upsilon_j(Srand_j(i)) = SW(i)$$

, where $i \in < |W| >$.

Using the above-mentioned definitions, we modify the Matsumoto and Imai scheme on *Step2.2* and *Step 5.2*. The other steps are the same as those proposed in Matsumoto and Imai's scheme (see also Section 2)

Step2.2: P generates the answer blocks

$$a_j(f_j(t)) = \Upsilon_j(t), \text{ for } t = 1, 2, \cdots, |W| .$$

...

Step5.2: If the number of correct answer blocks, i.e., $a_j(f_j(t)) = \Upsilon_j(t)$, for $t = 1, 2, \cdots, |W|$, is greater than or equal to α, then V accepts P; otherwise, V rejects P.

...

Example 1. Figure 3 illustrates a simplified version (i.e., $\alpha = \beta = 1$) of the newly modified scheme. Let the question alphabet be $Q = \{1, 2, 3, 4, 5, 6, 7, 8\}$, the window $W = \{1, 2, 4, 6\}$, and $A = \{1, 2, 3, 4\}$. The prover shares with the verifier a string of secret word, $SW = 3124$. Figure 3 shows two answers a_1 and a_2 corresponding to the question q.

q =	$\overline{2}$	8	5	$\overline{1}$	7	3	$\overline{6}$	$\overline{4}$
a 1=	4	4	3	2	2	1	1	3

$SW = 3124$

$a \circ g = 4321$

$Srand = 4321$

$\Upsilon = 4213$

a 2=	1	2	1	4	4	1	3	2

$SW = 3124$

$a \circ g = 2141$

$Srand = 3142$

$\Upsilon = 1432$

Figure 3. Example of Answers and Question

4.2 Security Analysis

Here, three attacks will be considered. The first one is the *known-A random attack* proposed in [3] where an attacker knows the function of the protocol and knows the set Ω, Q, $|W|$ and A, but does not know SW and W. The success probability (p_A) of known-A random attack on the modified scheme is given by

$$p_A = \sum_{j=\alpha}^{\beta} \binom{\beta}{j} p^j (1-p)^{\beta-j}$$

, where

$$p = \frac{|A|^{|Q|-|W|}}{|A|^{|Q|} - \sum_{i=1}^{|A|-1} \binom{|A|}{i} (i)^{|Q|}} = \frac{1}{|A|^{|W|}(1 - \sum_{i=1}^{|A|-1} \binom{|A|}{i} (\frac{i}{|A|}))^{|Q|}}$$

This result is equivalent to the security level of Matsumoto and Imai's scheme under the same attack.

The second attack is the passive attack as described in Section 3.1. The intruder has to test at most $\sum_{1 \le s_i \le Minimum\{|W|, t_i\}} \binom{t_1}{s_1}\binom{t_2}{s_2} \cdots \binom{t_{|A|}}{s_{|A|}}$ trials, which are the same as Matsumoto and Imai's scheme to reveal the secret.

The third attack is the *replay challenge attack* as described in Section 3.2. The modified scheme avoids this attack due to the new function Υ_j. For the same question to the end user , the window-padding answer portion of the end user may be changed according to the random-padding answer portion selected by him. It is obvious that the Lemma 1 docs not hold in our modified human identification scheme. Therefore, the replay challenge attack will not be successful on the newly modified scheme.

5 Conclusions

Human identification scheme is an important issue for user identifications in the network environment. The scheme proposed by Matsumoto and Imai is a pioneer work . This paper shows an attack, the replay challenge attack, on Matsumoto and Imai's scheme and proposes a modified scheme to avoid this attack. It requires further research to devise secure and practical human identification schemes.

Acknowledgement. We wish to thank Miss Maujy Peng and the referees of this paper for their useful comments.

References

1. Fiat, A. and Shamir, A., "How to Prove Yourself: Practical Solutions to Identical Solutions to Identification and Signature Problems", *Crypto'86*, 1986.
2. Jennifer G. Steiner B. Clifford Neuman, and Jeffrey I. Schiller, " Kerbero: An Authentication Service for Open Network Systems ", *Usenix Conference Proceedings*, pages 183-190, February 1988.
3. Matsumoto, T. and Imai, H., "Human Identification Through Insecure Channel", *Eurocrypt'91*, 1991.
4. Ohta, K. and Okamoto, T., "A Modification of the Fiat-Shamir Scheme", *Crypto'88*, 1988.
5. Davies, D.M. and Price, W.L., *Security for Computer Networks* , Chapter 7, John Wiley & Sons, 1984.
6. Ross, Sheldon M., *Introduction to Probability Models*, Academic Press, Inc., fifth edition.