Quantitative Evaluation of Unlinkable ID Matching Schemes

Nohara, Yasunobu
Kyushu University

Inoue, Sozo
Kyushu University

Baba, Kensuke
Kyushu University

Yasuura, Hiroto
Kyushu University

出版情報：Proc. of the 2005 ACM Workshop on Privacy in the Electronic Society, pp.55-60, 2005-11. ACM SIGSAC
バージョン：
権利関係：

Quantitative Evaluation of Unlinkable ID Matching Schemes

Yasunobu Nohara Sozo Inoue Kensuke Baba Hiroto Yasuura
Kyushu University
6-1 kasuga-koen, Kasuga-shi
Fukuoka, 816-8580 Japan
{nohara,sozo,baba,yasuura}@c.csce.kyushu-u.ac.jp

ABSTRACT

As pervasive computing environments become popular, RFID devices, such as contactless smart cards and RFID tags, are introduced into our daily life. However, there exists a privacy problem that a third party can trace user's behavior by linking device's ID.

The concept of unlinkability, that a third party cannot recognize whether some outputs are from the same user, is important to solve the privacy problem. A scheme using hash function satisfies unlinkability against a third party by changing the outputs of RFID devices every time. However, the schemes are not scalable since the server needs $O(N)$ hash calculations for every ID matching, where N is the number of RFID devices.

In this paper, we propose the *K-steps ID matching scheme*, which can reduce the number of the hash calculations on the server to $O(\log N)$. Secondly, we propose a quantification of unlinkability using conditional entropy and mutual information. Finally, we analyze the K-steps ID matching scheme using the proposed quantification, and show the relation between the time complexity and unlinkability.

Categories and Subject Descriptors

F.2.3 [**Analysis of Algorithms and Problem Complexity**]: Tradeoffs between Complexity Measures; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Security, Measurement

Keywords

RFID Security, Privacy, Degree of Unlinkability, K-steps ID Matching

WPES'05, November 7, 2005, Alexandria, Virginia, USA.

1. INTRODUCTION

RFID (Radio Frequency IDentification) is a technology that identifies humans and objects using *RFID devices*, which are silicon chips with IDs and radio frequency functions. In RFID, the server identifies an RFID device by *ID matching*, compareing the outputs of RFID devices with IDs stored in the server. As pervasive computing environments are becoming popular, RFID devices, such as contactless smart card and RFID tag are introduced into our daily life.

Privacy is one of the serious problems related to RFID. *Location Privacy Problem* is a privacy issue that a third party can trace user's behavior by reading and linking device's ID.

The concept of *unlinkability*, that a third party cannot recognize whether some outputs are from the same user, is important to solve location privacy problem. Randomized Hash Lock Scheme[1] and Extended Hash-chain Scheme[2, 3] satisfy unlinkability against a third party, by using hash function. These schemes are suitable for an RFID system, since the implementation cost of a hash function must be low. However, these schemes are not scalable since the server needs $O(N)$ hash calculations for every ID matching, where N is the number of RFID devices.

In this paper, we propose *the K-steps ID matching scheme*, which can reduce the number of the hash calculations on the server to $O(\log N)$. Secondly, we propose the quantification of unlinkability using conditional entropy and mutual information. Finally, we analyze the K-steps ID matching scheme using the proposed quantification, and show the relation between the time complexity and unlinkability.

The remainder of this paper is organized as follows. Section 2 describes related works. Section 3 presents the quantification of unlinkability. Section 4 presents the K-steps ID matching scheme. Section 5 evaluates the proposed scheme. Section 6 concludes this paper with summary.

2. RELATED WORK FOR UNLINKABLE ID MATCHING

Randomized Hash Lock Scheme[1] and Extended Hash-chain Scheme[2, 3] satisfy unlinkability against a third party, by using a hash function. These schemes are suitable for RFID systems since the implementation cost of a hash function must be low.

Let N be the number of the RFID devices in an RFID system, and M be the set of the RFID devices. And the ID id_i of an RFID device M_i is a string of length L over a finite alphabet Σ for $1 \leq i \leq N$. We assume that if $i \neq j$ then

$id_i \neq id_j$ for $1 \leq i, j \leq N$, and $2^L \gg N$. For $s, t \in \Sigma^*$, we denote by $s\|t$ the concatenation of s and t.

2.1 Randomized Hash Lock Scheme

In this scheme, a hash function H, a ROM, and a pseudo-random number generator are embedded onto the RFID devices.

An RFID device M_i stores id_i in the ROM. The server stores the IDs id_i $(1 \leq i \leq N)$ of all devices. The ID matching protocol of this scheme is as follows.

STEP1: The RFID device M_i generates a random number R, and sends $X = H(id_i\|R)$ and R to the server.

STEP2: The server finds id_i which corresponds to X by checking $X = H(id_i\|R)$ for $1 \leq i \leq N$.

$X = H(id_i\|R)$ is not fixed since R changes every time. Hence 2^L hash calculations are necessary when a third party tries to get id_i from X and R. It is computationally hard to calculate the hash function 2^L times. Therefore, this scheme has unlinkability against a third party.

2.2 Extended Hash-chain Scheme

In this scheme, two different hash functions H, G, a ROM, and a nonvolatile memory are embedded onto the RFID devices.

An RFID device M_i stores id_i in the ROM, and a secret information $cs_i^1 \in \Sigma^{L'}$ and a count number k in the nonvolatile memory. The server stores the pair (id_i, cs_i^1) $(1 \leq i \leq N)$ of all devices. The ID matching protocol of this scheme is as follows.

STEP1: The RFID device M_i sends $X = H(id_i\|cs_i^k)$ and k to the server. The RFID device M_i updates $cs_i^{k+1} \leftarrow G(cs_k)$ and $k \leftarrow k+1$.

STEP2: The server finds the corresponding id_i to X by checking $X = H(id_i\|cs_i^k)$ for all $1 \leq i \leq N$.

$X = H(id_i\|cs_i^k)$ is not fixed since cs_i^k changes every time. Hence $2^{L+L'}$ hash calculations are necessary when a third party tries to get id_i from X and R. It is computationally hard to calculate the hash function $2^{L+L'}$ times. Therefore, this scheme has unlinkability against a third party.

Moreover, it is computationally hard to get $cs_i^{k'}(k' < k)$ even if id_i and cs_i^k are tampering. Therefore, the scheme satisfies the forward security, in which any RFID device cannot be traced from past ID information even if the secret information is tampering.

2.3 Problems of Existing Hash-based Schemes

In Randomized Hash Lock scheme and Extended Hash-chain scheme, the server needs to calculate $H(id_i\|R)$ for $1 \leq i \leq N$ in every ID Matching. It means these schemes are not scalable since the server needs $O(N)$ hash calculations.

Avoine *et al.*[4] developed a specific time-memory trade-off that reduces the amount of computation in the Enhance Hash-chain scheme[2, 3]. However, Avoine *et al.*'s scheme[4] needs a heavy pre-calculation.

3. DEGREE OF UNLINKABILITY

In this section, we formalize *the degree of unlinkability* to evaluate unlinkability quantitatively.

3.1 Decoder

We introduce decoders to formalize attacks against unlinkability.

Let M_{index} be the set $\{1, 2, \cdots, N\}$ of numbers. We assume that each RFID device outputs a set of strings. Let O_i be the set of the outputs of $M_i \in M$ and O the set of the outputs from the RFID devices, that is,

$$O = \bigcup_{M_i \in M} O_i.$$

We assume $O_i \cap O_j = \phi$ if $i \neq j$ for any $i, j \in M_{index}$.

A *decoder* D is a function from O to M_{index} s.t.

$$i = j \implies D(o) = D(o')$$

for any $i, j \in M_{index}$, any $o \in O_i$, and any $o' \in O_j$.

A *complete decoder* D_{comp} is a special decoder s.t.

$$i = j \iff D_{comp}(o) = D_{comp}(o')$$

for any $i, j \in M_{index}$, any $o \in O_i$, and any $o' \in O_j$.

A *naive decoder* D_{naive} is one of the decoders s.t.

$$D_{naive}(o) = const$$

for any $o \in O$ and a $const \in M_{index}$.

For a decoder D, $c(x)$ is the number of $i \in M_{index}$ s.t. $D(o) = x$ for any $o \in O_i$. Intuitively, $c(x)$ can be seen as the number of RFID devices decoded to x by the D. Note that $c(x)$ is uniform in case of D_{comp}, i.e., $c(x) = 1$ for any x. For D_{naive}, $c(x) = N$ for $x \in M_{index}$ and $c(y) = 0$ for the other $y \in M_{index}$.

3.2 Modeling Information Sources of RFID Systems

Here, we formalize information sources of RFID systems and attacks against unlinkability.

When the system is implemented, the outputs from RFID devices are observed by single/multiple readers located in any place. Some of the readers might be used by attackers who try to break unlinkability. In the following, we describe an operation model which represents such a situation.

The environment of an RFID system is modeled as the following: the environment

1. lets *operation outputs* $V_o = \phi$,
2. lets *random number set* $V_b = \phi$,
3. chooses M_i randomly from M, adds i to V_b, and
4. chooses o randomly from O_i, adds o to V_o, and
5. repeats 3 to 5 finite times.

Only V_o is assumed to be observable from anyone.

An attack against unlinkability from an attacker is modeled as the following: the attacker, who has D,

1. lets *attack results* $V_y = \phi$,
2. for each $o \in V_o$ introduced above, obtains $D(o)$, and adds it to V_y.

We introduce information sources for the model. Let B be the information source which has V_b as an informataion source variables and a uniform probability, and Y be the information source which has V_y as an information source variables.

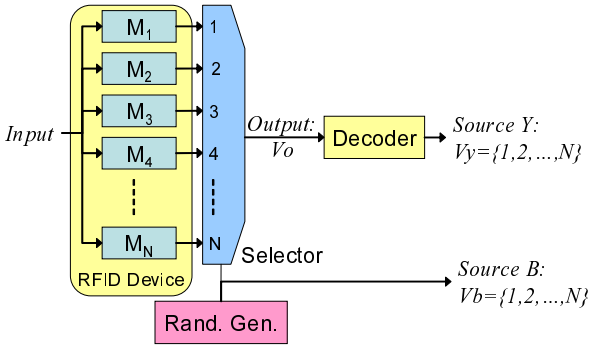

Figure 1: The environmental model of an RFID system

Figure 1 illustrates the model described above. From N RFID devices, one RFID device $M_i \in M$ is selected randomly according to the output $b \in V_b$ of a random number generator. Next, M_i outputs $o \in O_i$ randomly.

An attacker reads the outputs from RFID devices, and inputs each of them to the decoder the attacker has. From the outputs of the decoder, the attacker guesses which $b \in V_b$ each output corresponds to.

3.3 Degree of Unlinkability

We define the degree of unlinkability against an attacker using entropy. Let $p(y)$ be the probability of y at Y, and $p(b|y)$ the conditional probability of b at B on condition that $y \in V_y$ is obtained at Y. The *degree of unlinkability against an attacker* $U_{attacker}$ is defined to be the conditional entropy

$$H(B|Y) = \sum_{y \in V_y} p(y) H(B|y),$$

where

$$H(B|y) = -\sum_{b \in V_b} p(b|y) \log_2 p(b|y).$$

From the definition of $c(y)$, we can see $p(b|y) = \frac{1}{c(y)}$ for $c(y)$ b's in V_B, and $p(b|y) = 0$ for $(N - c(y))$ b's.

Thus,

$$H(B|y) = -\sum_{i=1}^{c(y)} \frac{1}{c(y)} \log_2 \frac{1}{c(y)} = \log_2 c(y).$$

Moreover, $p(y) = \frac{c(y)}{N}$ since $c(y)$ out of N RFID devices outputs y. Thus,

$$\begin{aligned} U_{attacker} &= \sum_{y \in V_y} \frac{c(y)}{N} H(B|y) \\ &= \frac{1}{N} \sum_{y=1}^{N} c(y) \log_2 c(y) \end{aligned} \quad (1)$$

The *mutual information* $I(B;Y)$ is defined as follows.

$$\begin{aligned} I(B;Y) &= H(B) - H(B|Y) \\ &= \log_2 N - \frac{1}{N} \sum_{y=1}^{N} c(y) \log_2 c(y) \end{aligned} \quad (2)$$

$U_{attacker} = H(B|Y)$ denotes the uncertainty of B when the elements of Y are given, and the attacker can determin B from the elements of Y if $H(B|Y)$ bits of information is given. In other words, $U_{attacker}$ represents the average information to link one output from an RFID device with the output history by the attacker. Therefore, we can use $U_{attacker}$ to measure unlinkability.

$I(B;Y)$ denotes the information of B given the elements of Y. In other words, $I(B;Y)$ represents the information that is revealed to an attacker who reads the outputs Y of RFID devices.

$U_{attacker}$ is depended on the attaker's ability of making decoders. When the attacker uses D_{comp}, $U_{attacker} = 0$, and $I(B|Y) = \log_2 N$, since $c(D_{comp}(o)) = 1$ for $\forall o \in O$. When the attacker uses D_{naive}, $U_{attacker} = \log_2 N$, and $I(B;Y) = 0$, since $c(D_{naive}(o)) = N$ for $\forall o \in O$. We can see that *the system has unlinkability against the attacker* iff $U_{attacker} = \log_2 N$.

3.4 Related Work for Modeling Unlinkability

Díaz *et al.*[6][8] and Serjantov *et al.*[7] use entropy to quantify anonymity[1] in communication systems. Probability distributions to anonymity set is calculated by analyzing the connection structures of network[6][7] or data contents on network[8]. And they calculate entropy from the probability distribution to measure anonymity.

Steinbrecher *et al.*[9] quantify unlinkability using entropy. They calculate entropy from the given probability distribution between links. However, the steps calculating the probability distribution is not included.

In our scheme, the outputs of the RFID devices are analyzed by a decoder. Moreover, we measure the relation between analytical result and true value using conditional entropy and mutual information.

4. K-STEPS ID MATCHING SCHEME

In this section, we describe the K-steps ID matching scheme, which was first presented in [10]. First, we explain the basic ideas for reducing the time complexity. And then, we show a method for generating IDs and a protocol for ID matching.

A similar approach reducing the number of hash caluculations by constructing a tree of IDs is proposed by Molnar *et al.*[11]. Compared to the literature, our work (1)considers the structure of IDs more generally, and (2)addresses the optimal form of the tree.

4.1 Reducing Time Complexity

For reducing the time complexity, we adopt the idea of grouping IDs. First, classify all RFID devices into groups of size α, and assign an ID called a group ID to each RFID device in a group. When an ID matching are executed, the number of hash calculation is reduced to $\frac{N}{\alpha}$ by each RFID device sending its group ID.

However, this approach aggravate unlinkability for a third party, since group IDs will be exposed to attackers. For the problem, we adopt the idea to hash group IDs not only IDs, and to prevent group IDs from being exposed to a third party. This approach preserves unlinkability. The time complexity of the servers becomes $\frac{N}{\alpha} + \alpha$ which is larger than the previous approach by α.

Furthermore, we adopt the above approaches recursively. Based on this approaches, we make an improvement for

[1]Anonymity is the state of being not identifiable within a set of subjects, the anonymity set[5]

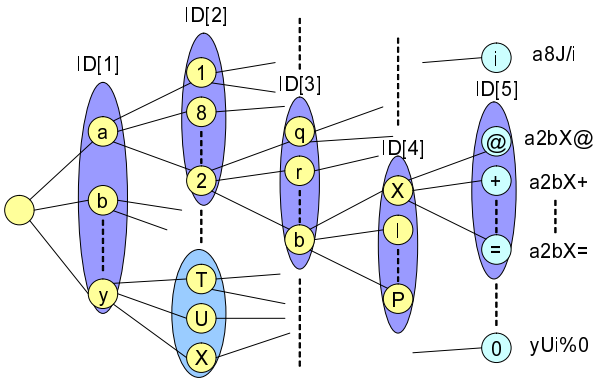

Figure 2: An ID structure for the K-steps ID matching scheme

generating IDs, and reduce the time complexity on the server by utilizing the property of the tree.

4.2 Generation of IDs

We use a labeled tree of depth K such as the tree illustrated in Figure 2. The tree has N leaves, and each leaf corresponds to an RFID device. For each node has a unique label. An ID id_i of an RFID device corresponding to a leaf node is defined to be the sequence of labels from the root node to the leaf node, e.g. a2bX@ in Figure 2.

In the following, the k-th ($1 \leq k \leq S_i$) label of M_i is denoted by $id_i[k]$, where S_i is the depth of the leaf of i, and $1 \leq S_i \leq K$.

4.3 Protocol

In the K-steps ID matching scheme, the server recognizes an ID from an output of RFID devices as the following protocol.

STEP1: The RFID device M_i generates a random number R. Then, M_i sends $(R, H_1, H_2, \ldots, H_K)$ to the server, where H_k is $H(id_i[k]\|R)$ if $1 \leq k \leq S_i$, and a random number R_k if $S_i + 1 \leq k \leq K$.

STEP2: The server performs as follows:

- **STEP2-1:** let Z be the root of the labeled tree;
- **STEP2-2:** compute $H(L_i\|R)$ for each child L_i of Z and find H_k s.t. $H(L_i\|R) = H_k$, and let Z be the child L_i;
- **STEP2-3:** output the label corresponds to Z as the ID of the RFID device if Z is a leaf, and let $k = k+1$ and return to STEP2-2 otherwise.

In Step1, the RFID device M_i sends a random number as H_k for $S_i < K$, which prevents the unlinkability against a third party from being worse by the difference of the number of H_k's.

In case where $K = 1$, the protocol and the structure of IDs of the proposed protocol correspond to the Randomized Hash Lock scheme[1].

5. EVALUATION

In this section, we analyze the relation between the time complexity and unlinkability for the K-steps ID matching scheme. In the concrete, we consider

- the expected number of hash calculations on the server, and
- the degree of unlinkability against a third party and a user.

Although the former depends on other factor (for example, the time for string matching), in practice, the hash calculation is the most essential. As to the latter, we gave the definition in the previous section. Both of them depend on (at least)

- the number N of the leaves,
- the depth K, and
- the number α_n of the edges for each node n

of the labeled tree for IDs.

We first find the number α_n of the edges with respect to each node which minimizes the number of hash calculations on the server (which depends on the number N of the leaves), and then, make clear the relation between the degree of unlinkability and the depth K of the tree.

5.1 Number of Hash Calculations

We consider the expected number of hash calculations on the server for a single ID matching.

We assume that the number of hash calculations necessary for an ID matching with m candidates is m. The ID matching with the labeled tree is solved by K ID matchings with α_n candidates. We also assume that any leaf is of the same depth K and $N^{\frac{1}{K}}$ is integer.

LEMMA 1. *The expected number of hash calculations for an ID matching with the labeled tree of depth K with N leaves is at least $KN^{\frac{1}{K}}$ for any $N > 0$ and any $1 \leq K \leq N$.*

PROOF. By the induction for K. If $K = 1$, the number of hash calculations is N. Let ν be the root of the tree and $\nu_1, \nu_2, \ldots, \nu_{\alpha_\nu}$ the children of ν. We denote by $g_K(N)$ the expected number of hash calculations for an ID matching with a tree of depth K with N leaves. Then,

$$g_K(N) = \alpha_\nu + \sum_{i=1}^{\alpha_\nu} \frac{n_i}{N} g_{K-1}(n_i),$$

where n_i is the number of leaves in the sub-tree whose root is ν_i for $1 \leq i \leq \alpha_\nu$ and $\sum_{i=1}^{\alpha_\nu} n_i = N$. By induction hypothesis, $g_{K-1}(n_i)$ is at least $(K-1)n_i^{\frac{1}{K-1}}$ for each $1 \leq i \leq \alpha_\nu$. Hence,$g_K(N)$ is at least $\alpha_\nu + \frac{K-1}{N}\sum_{i=1}^{\alpha_\nu} n_i^{\frac{K}{K-1}}$. Since $\frac{K}{K-1}$ is larger than unity and $\sum_{i=1}^{\alpha_\nu} n_i = N$, the expected number is minimal if $n_i = \frac{N}{N/\alpha_\nu}$ for any $1 \leq i \leq \alpha_\nu$. Therefore, we have only to consider the case where $g_K(N)$ is of the form

$$\begin{aligned} &\alpha_\nu + \frac{K-1}{N} \cdot \alpha_\nu \cdot \left(\frac{N}{\alpha_\nu}\right)^{\frac{K}{K-1}} \\ &= \alpha_\nu + (K-1) \cdot N^{\frac{1}{K-1}} \cdot {\alpha_\nu}^{-\frac{1}{K-1}}. \end{aligned}$$

This is minimal only if $1 - N^{\frac{1}{K-1}} {\alpha_\nu}^{-\frac{K}{K-1}} = 0$ and therefore $\alpha_\nu = N^{\frac{1}{K}}$. Thus, the minimal number of $g_K(N)$ is

$$N^{\frac{1}{K}} + (K-1)N^{\frac{1}{K-1}} N^{\frac{1}{K}-\frac{1}{K-1}} = KN^{\frac{1}{K}}.$$

By the previous proof, the number of hash calculations for a single matching is the minimal number $KN^{\frac{1}{K}}$ if the number of each edge is $\alpha_n = N^{\frac{1}{K}}$. Therefore, in the rest of this paper, we consider the labeled tree in which any node has the same number α of children. Then, $KN^{\frac{1}{K}} = \alpha \log_\alpha N$. Therefore, the previous lemma implies the following theorem.

THEOREM 1. *The K-steps ID matching protocol can find an ID in N candidates by $O(\log N)$ time.*

5.2 Unlinkability

We calculate the degree of unlinkability against a third party and against a user, respectively, by the result of Section 3. We assume that an attacker can get an output $o = (R, H_1, H_2, \cdots, H_K)$.

5.2.1 Unlinkability against a Third Party

A third party has no ID in the set of the N IDs. By the assumption of hash function, an attacker can get no information from o, that is, $U_{TP} = \log_2 N$, $I(B; Y_{TP}) = 0$.

THEOREM 2. *An RFID system with the K-steps ID matching protocol has unlinkability against any third party.*

5.2.2 Unlinkability against a User

A user has its own ID id_i and can make a decoder $D_{user} : O \to M_{index}$ performs as follows:

STEP1: compute $H(id_i[k] \| R)$ by id_i and compare with H_k in o for each $1 \le k \le K$;

STEP2: let β be the number of matches H_k and $H(id_i[k] \| R)$ for $0 \le \beta \le K$;

STEP3: output $y = \beta + 1$.

We consider the number $c(\beta + 1)$ of RFID devices which is decoded as y. If $\beta = K$, a single ID (which is the ID of the user) is decided as the result of ID-matching for the K sub-ID's. Therefore, $c(K + 1) = 1$. If $0 \le \beta \le K - 1$, there exist $\alpha^{K-\beta} - \alpha^{K-\beta-1} = \alpha^{K-\beta-1}(\alpha - 1)$ candidates as the result of β ID-matchings. Therefore, $c(y) = \alpha^{K-y}(\alpha - 1)$ for $1 \le y \le K$. Since $\beta \le K$, $c(y) = 0$ for $K + 2 \le y \le N$. Thus,

$$c(y) = \begin{cases} \alpha^{K-y}(\alpha - 1) & (1 \le y \le K) \\ 1 & (y = K + 1) \\ 0 & (K + 2 \le y \le N). \end{cases} \quad (3)$$

By Eq. 1 and Eq. 3, the degree of unlinkability against a user U_{user} is

$$\begin{aligned} & \frac{1}{N} \sum_{y=1}^{N} c(y) \log_2 c(y) \\ = \; & \frac{1}{N} \sum_{i=0}^{K-1} \alpha^i (\alpha - 1) \log_2 (\alpha^i (\alpha - 1)) \\ = \; & \frac{\alpha - 1}{N} \left(\log_2 \alpha \sum_{i=0}^{K-1} i\alpha^i + \log_2 (\alpha - 1) \sum_{i=0}^{K-1} \alpha^i \right). \end{aligned}$$

Since $\alpha^K = N$, by the fact that

$$\sum_{i=0}^{K-1} i\alpha^i = \frac{K\alpha^K}{\alpha - 1} - \frac{\alpha(\alpha^K - 1)}{(\alpha - 1)^2}$$

Table 1: The number of hash calculations and the degree of unlinkability in the ID-matching protocols with a hash function

	Hash Lock, Hash-Chain	K-steps
H1	$O(N)$	$O(\log N)$
H2	$O(1)$	$O(\log N)$
U1	$\log_2 N$	$\log_2 N$
U2	$\log_2 N$	less than $\log_2 N$

and

$$\sum_{i=0}^{K-1} \alpha^i = \frac{\alpha^K - 1}{\alpha - 1},$$

we have

$$\begin{aligned} & U_{user} \\ = \; & \left(\frac{K\alpha^K}{N} - \frac{\alpha(\alpha^K - 1)}{N(\alpha - 1)} \right) \log_2 \alpha \\ & + \frac{\alpha^K - 1}{N} \log_2 (\alpha - 1) \\ = \; & \left(K - \frac{N - 1}{N} \cdot \frac{\alpha}{\alpha - 1} \right) \log_2 \alpha \\ & + \frac{N - 1}{N} \log_2 (\alpha - 1) \\ = \; & \log_2 N + \\ & \frac{N - 1}{N} \left(\log_2 (\alpha - 1) - \frac{\alpha}{\alpha - 1} \log_2 \alpha \right) \end{aligned} \quad (4)$$

and moreover

$$\begin{aligned} U_{user} = \; & \log_2 N \\ & + \frac{N - 1}{N} \cdot \log_2 (N^{\frac{1}{K}} - 1) \\ & - \frac{N - 1}{N} \cdot \frac{N^{\frac{1}{K}}}{K(N^{\frac{1}{K}} - 1)} \log_2 N. \end{aligned} \quad (5)$$

If we use the tree in which α is large enough,

$$U_{user} \simeq \log_2 N$$

and

$$I(B; Y_{user}) \simeq 0.$$

THEOREM 3. *In an RFID system with the K-steps ID matching protocol, if the number of the each node's children in the labeled tree are large, the degree of unlinkability against a user can be regarded to be $\log_2 N$, where N is the number of the leaves in the tree.*

5.3 Comparison with Other Protocols

Table 1 shows (H1) the number of hash calculations on the service server and (H2) on an RFID device, and (U1) the degree of unlinkability against a third party and (U2) against a user, with respect to the K-steps ID matching protocol (K-steps) and two existing protocols: the randomized hash lock scheme[1] (Hash Lock) and the extended hash-chain scheme[2, 3] (Hash-Chain).

The existing protocols[1][2, 3] for ID matching with a hash function take $O(N)$ time and therefore can not be applied for

a large N. Our protocol is suitable for a practical system in the sense of the time complexity. However, as to the time complexity on an RFID device, our protocol requires $O(\log N)$ hash calculations in an ID matching despite that the some existing protocols[1][2, 3] require $O(1)$ times [2]. And U2 of our protocol is less than that of the existing protocols.

From the above-mentioned, there is a relationship that H2 and U2 are going worse when H1 is improved. Therefore, we should select the adequate parameter K in a practical system, where H1, H2 and U2 are balanced.

6. CONCLUSION

In this paper, we proposed the K-steps ID matching scheme, which can reduce the number of calculations of hash function on the server to $O(\log N)$. Secondly, we proposed the quantification of unlinkability using conditional entropy and mutual information. Finally, we analyzed the K-steps ID matching scheme using the proposed quantification, and showed the relation between the time complexity and unlinkability.

7. ACKNOWLEDGMENT

This work has been partly supported by the Grant-in-Aid for Creative Scientific Research No.14GS0218 and the 21st Century COE Program. We are grateful for their support.

8. REFERENCES

[1] Stephan A. Weis, Sanjay E. Sarma, Ronald L. Rivest, and Daniel W. Engels, "Security and Privacy Aspects of Low-Cost Radio Frequency Identification Systems", International Conference on Security in Pervasive Computing 2003, LNCS 2802, pp.201-212, 2004.

[2] Miyako Ohkubo, Koutarou Suzuki and Shingo Kinoshita, "Cryptographic Approach to a Privacy Friendly Tag", RFID Privacy Workshop@MIT, 2003.

[3] Miyako Ohkubo, Koutarou Suzuki, Shingo Kinoshita, "Hash-Chain Based Forward-Secure Privacy Protection Scheme for Low-Cost RFID", Proc. of the 2004 Symposium on Cryptography and Information Security(SCIS2004), Vol.1, pp.719-724, Jan. 2004.

[4] Gildas Avoine and Philippe Oechslin, "A Scalable and Provably Secure Hash-Based RFID Protocol", 2nd International Workshop on Pervasive Computing and Communications Security(PerSec2005), pp.110-114, Mar. 2005.

[5] Andreas Pfitzmann and Marit Köhntopp, "Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology", International Workshop on Design Issues in Anonymity and Unobservability, LNCS2009, pp.1-9, 2000.

[6] Claudia Díaz, Stefaan Seys, Joris Claessens, and Bart Preneel, "Towards measuring anonymity", Workshop on Privacy Enhancing Technologies 2002, LNCS 2482, pp. 54-68, 2002.

[7] Andrei Serjantov and George Danezis, "Towards an Information Theoretic Metric for Anonymity", Workshop on Privacy Enhancing Technologies 2002, LNCS 2482, pp. 41-53, 2002.

[8] Claudia Díaz, Joris Claessens, Stefaan Seys, and Bart Preneel, "Information Theory and Anonymity", Proceedings of the 23rd Symposium on Information Theory, pp. 179-186, May. 2002.

[9] Sandra Steinbrecher and Stefan Köpsell, "Modelling unlinkability", Workshop on Privacy Enhancing Technologies 2003, LNCS2760, pp. 32-47, 2003.

[10] Yasunobu Nohara, Sozo Inoue, Kensuke Baba, and Hiroto Yasuura, "Unlinkable ID Matching Protocol for Large-scale RFID Systems", Proc. of the 2005 Symposium on Cryptography and Information Security (SCIS2005), Vol.3, pp.1567-1572, Jan. 2005. (in Japanese)

[11] David Molnar and David Wagner, "Privacy and Security in Library : RFID Issues, Practices, and Architectures", Proc. of Computer and Communication Security 2004, pp.210-219, Oct. 2004.

[2]This means the calculation time in tags is incresed. However, this doesn't mean the implementation cost of the tags is increased since this calculations don't have to be done at once.