An Anonymous Authentication Protocol with Single-database PIR

Nakamura, Toru
Graduate School/Faculty of Information Science and Electrical Engineering, Kyushu University

Inenaga, Shunsuke
Graduate School/Faculty of Information Science and Electrical Engineering, Kyushu University

Baba, Kensuke
Research and Development Division, Kyushu University Library

Ikeda, Daisuke
Graduate School/Faculty of Information Science and Electrical Engineering, Kyushu University

他

出版情報：Conferences in Research and Practice in Information Technology. 116, pp.3-8, 2011-01. Australian Computer Society
バージョン：
権利関係：

An Anonymous Authentication Protocol with Single-database PIR

Toru Nakamura[1] **Shunsuke Inenaga**[1] **Kensuke Baba**[2] **Daisuke Ikeda**[1]
Hiroto Yasuura[1]

[1] Graduate School/Faculty of Information Science and Electrical Engineering,
Kyushu University
Moto'oka 744, Nishi-ku, Fukuoka, 819-0395, Japan
Email: {toru, inenaga, yasuura}@soc.ait.kyushu-u.ac.jp
daisuke@inf.kyushu-u.ac.jp
[2] Research and Development Division, Kyushu University Library
10-1, Hakozaki 6, Higashi-ku, Fukuoka, 812-8581, Japan
Email: baba@lib.kyushu-u.ac.jp

Abstract

This paper focuses on anonymous authentication systems in multi-service environment, in which service providers communicate with the central manager in every authentication. Such systems have a merit that the central manager can easily update the database of user information by comparison to the existing anonymous authentication systems without communications between service providers and the central manager. The purpose of this paper is to realize a practical authentication protocol for such systems which satisfies four requirements for security and privacy protection, that is, *correctness*, *impersonation resistance against passive insider*, *anonymity against central manager*, and *anonymity against service providers*. The existing protocol consists of a multi-database PIR scheme, in which there are copies of the same database and none of these copies are allowed to communicate with each other. This paper proposes an authentication protocol which consists of the single-database PIR scheme proposed by Kushilevitz and Ostrovsky. This protocol also realizes all these requirements in the random oracle model. This protocol is more practical since using a single database implies the above-mentioned assumptions for multi-database PIR schemes are not required any more.

1 Introduction

With the increase of the number of services, users are forced to manage more pairs of a user ID (pseudonym) and a password. Hence much attention is recently paid to *authentication systems in multi-service environment*, which enable each user to have only a pair in order to use multiple services with a central manager. For example, single-sign-on systems such as Microsoft's .NET Passport, Shibboleth, and OpenID, have been popular. In this paper, we focus on issues about user privacy such that activity or preference of a user can be revealed by (1) service providers or (2) a central manager. If a user submits his/her ID to multiple service providers and the central manager, information about what, when, and how often a user accesses can be collected. In order to solve such issues, an authentication protocol with anonymity against (1) service providers and (2) a central manager is essential.

Authentication systems in multi-service environment can be classified according to which service providers must communicate with the central manager in every authentication. With respect to authentication systems without such communications, some protocols to realize the both kinds of anonymity are known, such as group signature schemes (Chaum & van Heyst 1991), anonymous credential schemes (Camenisch & Lysyanskaya 2002), and dynamic ID based anonymous authenticated key exchange schemes (Liao & Wang 2009). However, such protocols have a drawback that it is difficult for the central manager to deal with frequent queries to update the database of user information. Hence we focus on authentication systems with communications between service providers and the central manager. The requirements for an authentication system considered in this paper are the following.

- *Correctness*: if a user sends an authentication request with the valid password, every service provider accepts the request.
- *Impersonation resistance against passive insider* [1]: even if an adversary is a service provider, the adversary cannot impersonate a legitimate user.
- *Anonymity against service provider*: it is difficult for any service providers to obtain any information about a user ID.
- *Anonymity against central manager*: it is difficult for any central manager to obtain any information about a user ID.

There are few schemes which satisfy the previous requirements, as far as we know. Nakamura *et al.* (Nakamura et al. 2009) proposed an anonymous authentication protocol which satisfies all the requirements previously described. This protocol is based on *private information retrieval* (*PIR*) schemes (Chor et al. 1998)(Kushilevitz & Ostrovsky 1997). PIR schemes contribute for protecting privacy of a client who makes a query to a database server. Using a PIR scheme, the client can reconstruct an element from the answer which the database server has generated with the query, without the index of the element being revealed to the database server. The authentication protocol consists of a multi-database PIR scheme (Chor et al. 1998). This scheme requires the assumption that there are copies of the same database and none of these copies are allowed to communicate with each other. However, the assumption is not practical.

In this paper, we propose an authentication protocol with a single-database PIR scheme, which does not require copies of the same database. The protocol is called *Single-database PIR based Anonymous Authentication Protocol* (SPAAP for short). The first single-database

[1] In this paper, a "passive and insider adversary" means that an adversary who is restricted to eavesdropping on messages that the service provider obtains.

PIR scheme, which is based on the quadratic residuosity assumption, is proposed by Kushilevitz and Ostrovsky (Kushilevitz & Ostrovsky 1997). The basic idea of realizing the authentication protocol is that (1) a user makes the query related his/her ID and encrypts the query with the public-key of the central manager, (2) the central manager decrypts the query and makes the answer related to the information to verify the user, and (3) the service provider reconstructs the information from the answer, where IDs correspond to indices of the database. If the service provider can obtain the ID, it is impossible to realize anonymity against service providers. However, original Kushilevitz and Ostrovsky's single-database PIR scheme requires an index to reconstruct the element from the answer. Hence the single-database PIR scheme cannot be applied to our protocol. In this paper, we use the special version of Kushilevitz and Ostrovsky's single-database PIR, in which an element of the database can be reconstructed without the index. Furthermore, we prove that SPAAP satisfies all the requirements under the quadratic residuosity assumption and the random oracle assumption (Bellare & Rogaway 1993).

SPAAP is more practical than the existing protocol (Nakamura et al. 2009) since using a single database implies the assumptions for multi-database PIR schemes are not required any more. Therefore, this paper contributes development of anonymous authentication systems in which service providers need to communicate with the central manager from the view point of reducing the impractical assumption.

The organization of this paper is shown as follows. In section 2, we provide some necessary definitions. In section 3, we introduce the definitions of the four requirements of anonymous authentication protocols. In section 4, we show the definition of the special version of single-database PIR and the detail of SPAAP. In section 5, we prove that SPAAP satisfies all the requirements.

2 Preliminaries

2.1 Notations

Let $\mathbb{Z}$ denote the set of integers and $\mathbb{N}$ denote the set of natural numbers. For a finite set X, let $|X|$ denote the number of elements which X contains. For $x \in \mathbb{Z}$, let $\|x\|$ denote the binary length of x. For $k \in \mathbb{N}$, let $[k] = \{1, 2, ..., k\}$. For $a, b \in \mathbb{Z}$, let $a|b$ mean that b is divisible by a. Let $x \circ y$ be the concatenation of bit strings x and y. We denote any polynomial of $n \in \mathbb{N}$ by $p(n)$, and some polynomial by $\mathrm{poly}(n)$.

An *interactive Turing machine* (*ITM*) (Goldreich 2001) is a Turing machine which has a pair of *communication tapes* in addition to a common input tape, a local input tape, an output tape, and a work tape. A *joint computation* of two ITMs is a sequence of pairs of the local configurations. The output of a joint computation is the output of one of the ITMs. The output of a Turing machine $\mathcal{A}$ on an input x is denoted by $\mathcal{A}(x)$. We denote by $\langle \mathcal{A}, \mathcal{B} \rangle$ a joint computation of Turing machines $\mathcal{A}$ and $\mathcal{B}$, and by $\langle \mathcal{A}(y), \mathcal{B}(z) \rangle(x)$ its output on a common input x, a local input y for $\mathcal{A}$, and a local input z for $\mathcal{B}$. We sometimes omit the brackets if the input is empty. In the rest of this paper, we sometimes call a Turing machine $\mathcal{A}$ an "algorithm" $\mathcal{A}$ and a joint computation $\langle \mathcal{A}, \mathcal{B} \rangle$ a "protocol" $\langle \mathcal{A}, \mathcal{B} \rangle$. The idea of a joint computation of two ITMs can be extended straightforwardly to that of three ITMs by two pairs of communication tapes.

For random variables X, Y distributed over a set Z, let

$$\Pr[X = Y] = \sum_{x,y \in Z} \Pr[X = x] \cdot \Pr[Y = y] \cdot \chi(x, y),$$

where χ is a predicate such that $\chi(a, b) = 1$ if $a = b$, and $\chi(a, b) = 0$ otherwise. The output of a probabilistic algorithm $\mathcal{A}$ is determined by given inputs and random sources (called coin tosses). Assuming that coin tosses are given as local inputs, we can regard a probabilistic algorithm as a deterministic algorithm. Let $\mathcal{A}_D$ be a deterministic algorithm corresponding to a probabilistic algorithm $\mathcal{A}$. We assume that coin tosses r is a t-bit string. For random variables X, Y distributed over a set Z and $x, y \in Z$, let

$$\Pr[\mathcal{A}(x) = y] = \frac{|\{r | \mathcal{A}_D(x, r) = y\}|}{2^t},$$

$$\Pr[\mathcal{A}(X) = y] = \sum_{x \in Z} \Pr[X = x] \cdot \Pr[\mathcal{A}(x) = y], \text{ and}$$

$$\Pr[\mathcal{A}(X) = Y] = \sum_{x,y \in Z} \Pr[X = x] \cdot \Pr[Y = y] \cdot \Pr[\mathcal{A}(x) = y].$$

2.2 Indistinguishability

Definition 1 *For any $m \in \mathbb{N}$, two sequences of random variables $X = (X^{(1)}, X^{(2)}, \ldots, X^{(m)})$ and $Y = (Y^{(1)}, Y^{(2)}, \ldots, Y^{(m)})$ whose elements are distributed over $\{0,1\}^{\mathrm{poly}(k)}$ are (computationally)* indistinguishable *if for any $k \in \mathbb{N}$, any probabilistic polynomial-time algorithm $\mathcal{B}$,*

$$|\Pr[\mathcal{B}(1^k, X^{(1)}, X^{(2)}, \ldots, X^{(m)}) = 1] - \Pr[\mathcal{B}(1^k, Y^{(1)}, Y^{(2)}, \ldots, Y^{(m)}) = 1]| < \frac{1}{p(k)}.$$

Definition 2 *A sequence of random variables X which are distributed over $\{0,1\}^{\mathrm{poly}(k)}$ is* constructible *if there exists a probabilistic polynomial-time algorithm $\mathcal{S}$ such that for any $k \in \mathbb{N}$, the sequence of random variables $\mathcal{S}(1^k)$ and X are identically distributed.*

Lemma 1 *For any $k \in \mathbb{N}$, any $m \in \mathrm{poly}(k)$, any constructible sequences of random variables $X = (X^{(1)}, X^{(2)}, \ldots, X^{(m)})$ and $Y = (Y^{(1)}, Y^{(2)}, \ldots, Y^{(m)})$ distributed over $\{0,1\}^{\mathrm{poly}(k)}$, if for any $i \in [m]$, $X^{(i)}$ and $Y^{(i)}$ are indistinguishable, then X and Y are indistinguishable.*

proof: This can be proven easily by the standard hybrid argument (Goldreich 2001).

□

2.3 Quadratic Residuosity Assumption

For $a \in \mathbb{Z}$, let $[[a]] = \{x \in \mathbb{Z} | x \equiv a \bmod n\}$ ($[[a]]$ is called the *residue class* modulo n containing a). For $n \in \mathbb{N}$, let

$$\mathbb{Z}_n^* = \{x | 1 \leq x \leq n, gcd(n, x) = 1\}.$$

The *quadratic residuosity predicate* $\mathcal{W}_n$ is defined as follows:

$$\mathcal{W}_n(y) = \begin{cases} 0 & \text{if } \exists w \in \mathbb{Z}_n^* \text{ such that } w^2 = y \bmod n \\ 1 & \text{otherwise} \end{cases}.$$

For a positive odd n, let $\left(\frac{x}{n}\right)$ denote the Jacobi symbol of $x \bmod n$. Let

$$\mathbb{Z}_n^{+1} = \{x \in \mathbb{Z}_n^* | \left(\frac{x}{n}\right) = +1\}.$$

Let $QR_n^{+1} = \{x \in \mathbb{Z}_n^{+1} | \mathcal{W}_n(x) = 0\}$, $QNR_n^{+1} = \{x \in \mathbb{Z}_n^{+1} | \mathcal{W}_n(x) = 1\}$.

Informally, the Quadratic Reduosity Assumption is the assumption that there is no probabilistic polynomial-time algorithm for computing the predicate $\mathcal{W}_n(x)$. We show the definition of the assumption as follows.

Definition 3 (Quadratic Reduosity Assumption) *For $k \in \mathbb{N}$, let $I_k = \{n | n = \alpha \cdot \beta, \alpha$ and β are distinct primes, $\|\alpha\| = \|\beta\| = k\}$. For any $k \in \mathbb{N}$, any probabilistic polynomial-time algorithm $\mathcal{B}$,*

$$\Pr[\mathcal{B}(N, X) = \mathcal{W}_N(X)] < \frac{1}{2} + \frac{1}{p(k)},$$

where N is a random variable uniformly distributed over I_k and X is a random variable uniformly distributed over $\mathbb{Z}_N^{+1}$.

3 Requirements of Anonymous Authentication Protocol

In this section, we introduce the authentication model which we assume in this paper and the definitions of the four requirements of anonymous authentication protocols.

3.1 Authentication Model

In this paper, we assume an authentication model which consists of the following three types of entities.

- **User**: Let m be the number of the users. Each user is assigned the unique *identifier* $i \in [m]$ and has a *password* $x_i \in \{0,1\}^\ell$ for a natural number ℓ. (Note that ℓ is a polynomial of a security parameter k.)
- **Service provider**: A *service provider* verifies whether the entity who has sent an authentication request is truly the legitimate user.
- **Central manager**: A *central manager* stores the sequence $x = (x_1, x_2, ..., x_m)$ of the passwords of the users. We assume that each password is a random string.

Throughout this paper, we assume that

- each user can communicate only with service providers,
- each service provider can communicate with users and the central manager, and
- the central manager can communicate only with service providers.

Fig.1 is the authentication model that describes which pairs of entities can communicate each other.

We define an authentication protocol as a joint computation $\langle \mathcal{P}, \mathcal{V}, \mathcal{M} \rangle$. $\mathcal{P}, \mathcal{V}$, and $\mathcal{M}$ mean the behaviors of a user, a service provider, and a central manager, respectively. $\mathcal{P}$ takes a pair of an identifier i and a *candidate password* $z \in \{0,1\}^\ell$ as inputs, and $\mathcal{M}$ takes x as an input. After running the authentication protocol, $\mathcal{V}$ outputs $1/0$.

3.2 Requirements

We show the four requirements which an anonymous authentication protocol $\langle \mathcal{P}, \mathcal{V}, \mathcal{M} \rangle$ should satisfy as follows.

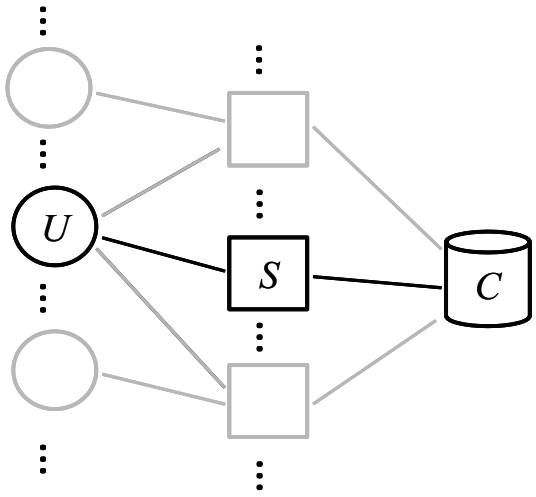

Figure 1: The authentication model describes which pairs of entities can communicate each other. (U: a user, S: a service provider, C: a central manager)

- *Correctness*: for any $k, \ell, m \in \mathbb{N}$, any $i \in [m]$, any $x = \{x_i \mid i \in [m], x_i \in \{0,1\}^\ell\}$,

$$\Pr[\langle \mathcal{P}(1^k, i, x_i), \mathcal{V}(1^k), \mathcal{M}(1^k, x) \rangle = 1] > 1 - \frac{1}{p(k)}.$$

- *Impersonation resistance against passive insider*: for any $k, \ell, m \in \mathbb{N}$ any $i \in [m]$, and any probabilistic polynomial-time algorithm $\mathcal{B}$,

$$\Pr[\langle \mathcal{B}(1^k, T_1), \mathcal{V}(1^k), \mathcal{M}(1^k, X) \rangle = 1] < \frac{1}{p(k)},$$

  where X is a random variable uniformly distributed over $(\{0,1\}^\ell)^m$ and T_1 is a random variable which means a transcript of $\mathcal{V}$'s local tape and read tapes after running $\langle \mathcal{P}(i, x), \mathcal{V}, \mathcal{M}(x) \rangle$ where x is a sample from X.

- *Anonymity against central manager*: for any $k, \ell, m \in \mathbb{N}$, any $i, j \in [m]$, any $z, z' \in \{0,1\}^\ell$, and any probabilistic polynomial-time algorithm $\mathcal{B}$,

$$|\Pr[\mathcal{B}(1^k, T_2) = 1] - \Pr[\mathcal{B}(1^k, T_3) = 1]| < \frac{1}{p(k)},$$

  where X is a random variable uniformly distributed over $(\{0,1\}^\ell)^m$ and T_2 is a random variable which means a transcript of $\mathcal{M}$'s local tape and read tapes after running $\langle \mathcal{P}(i, z), \mathcal{V}, \mathcal{M}(x) \rangle$ where x is a sample from X. Similarly, T_3 means a transcript after running $\langle \mathcal{P}(j, z'), \mathcal{V}, \mathcal{M}(x) \rangle$.

- *Anonymity against service provider*: for any $k, \ell, m \in \mathbb{N}$, any $i, j \in [m]$, any $z, z' \in \{0,1\}^\ell$, and any probabilistic polynomial-time algorithm $\mathcal{B}$,

$$|\Pr[\mathcal{B}(1^k, T_4) = 1] - \Pr[\mathcal{B}(1^k, T_5) = 1]| < \frac{1}{p(k)},$$

  where X is a random variable uniformly distributed over $(\{0,1\}^\ell)^m$ and T_4 is a random variable which means a transcript of $\mathcal{V}$'s local tape and read tapes after running $\langle \mathcal{P}(i, z), \mathcal{V}, \mathcal{M}(x) \rangle$ where x is a sample from X. Similarly, T_5 means a transcript after running $\langle \mathcal{P}(j, z'), \mathcal{V}, \mathcal{M}(x) \rangle$.

4 Our Approach: SPAAP

In this section, we show the anonymous authentication protocol which satisfies all the requirements, called SPAAP. We construct SPAAP with a special version of Kushilevitz and Ostrovsky's single-database PIR schemes (Kushilevitz & Ostrovsky 1997), in which an element of the database can be reconstructed without the index.

4.1 Kushilevitz and Ostrovsky's PIR scheme

For the ease of explanation, we assume that an element of a database is a bit, that is, a database is denoted by $x = x_1 \circ x_2 \circ \cdots \circ x_m \in \{0,1\}^m$. We note that it is easy to modify this simpler scheme to treat a database of ℓ-bit strings (for example, repeating this simpler scheme for ℓ times).

- *Query algorithm* $\mathcal{Q}(\cdot,\cdot)$: $\mathcal{Q}$ is a probabilistic algorithm which receives 1^k and an index $i \in [m]$ (k is a security parameter) as inputs. First, $\mathcal{Q}$ randomly chooses distinct primes α and β whose length is $k/2$. Next, $\mathcal{Q}$ uniformly and randomly chooses m numbers $y_1, ..., y_m \in \mathbb{Z}_n^{+1}$ such that y_j is an element of QNR_n^{+1} if $j = i$, y_j is an element of QR_n^{+1} otherwise, where $n = \alpha \cdot \beta$. Finally, $\mathcal{Q}$ outputs $y_1, ..., y_m$ as a query and (α, β) as a secret.
- *Answer algorithm* $\mathcal{A}(\cdot,\cdot,\cdot)$: $\mathcal{A}$ is a deterministic algorithm which receives 1^k, a database $x \in \{0,1\}^m$, and a query $y_1, ..., y_m \in \mathbb{Z}_n^{+1}$ as inputs. $\mathcal{A}$ computes

$$w_i = \begin{cases} y_i^2 & \text{if } x_i = 0 \\ y_i & \text{if } x_i = 1. \end{cases}$$

  Then, $\mathcal{A}$ outputs as an answer

$$z = \prod_{i=1}^{m} w_i.$$

- *Reconstruct algorithm* $\mathcal{R}(\cdot,\cdot,\cdot)$: $\mathcal{R}$ is a deterministic algorithm which receives 1^k, a secret (α, β), and answer $z \in \mathbb{Z}_n^{+1}$ as inputs. $\mathcal{R}$ outputs 1 if $\mathcal{W}_n(z) = 1$, and outputs 0 otherwise.

The PIR scheme satisfies the following properties under the quadratic residuosity assumption.

- *correctness*: for any $k, m \in \mathbb{N}$, any $x = \{x_i \mid i \in [m], x_i \in \{0,1\}\}$, and any $i \in [m]$,

$$\Pr[\mathcal{R}(1^k, \mathcal{Q}^2(1^k, i), \mathcal{A}(x, \mathcal{Q}^1(1^k, i))) = x_i] > 1 - \frac{1}{p(k)}. \quad (1)$$

- *privacy*: for any $k, m \in \mathbb{N}$, any $i, j \in [m]$, and any probabilistic polynomial-time algorithm $\mathcal{B}$,

$$|\Pr[\mathcal{B}(1^k, \mathcal{Q}^1(1^k, i)) = 1] - \Pr[\mathcal{B}(1^k, \mathcal{Q}^1(1^k, j)) = 1]| < \frac{1}{p(k)}. \quad (2)$$

We prove the following lemma with respect to the PIR scheme. This lemma also holds in the modified scheme for a database $x = \{x_i | i \in [m], x_i \in \{0,1\}^\ell\}$ of ℓ-bit strings. In the rest of paper, a PIR scheme means the modified scheme.

Lemma 2 *If $(\mathcal{Q}, \mathcal{A}, \mathcal{R})$ is the previous described PIR scheme, the following proposition holds: for any $k, m \in \mathbb{N}$, any $i, j \in [m]$, and any probabilistic polynomial-time algorithm $\mathcal{B}$,*

$$\Pr[\mathcal{B}(1^k, \mathcal{Q}^2(1^k, i), \mathcal{A}(1^k, X, \mathcal{Q}^1(1^k, i))) = 1] - \Pr[\mathcal{B}(1^k, \mathcal{Q}^2(1^k, j), \mathcal{A}(1^k, X', \mathcal{Q}^1(1^k, j))) = 1] = 0,$$

where X, X' are random variables uniformly and independently distributed over $\{0,1\}^m$.

proof: Let $I'_k = \{(\alpha, \beta) | \alpha, \beta$ are distinct primes, $\|\alpha\| = \|\beta\| = k\}$. $\mathcal{Q}^2(1^k, i)$ and $\mathcal{Q}^2(1^k, j)$ are (information theoretical) indistinguishable because both of them are random variables uniformly distributed over I'_k.

Let $n = \alpha \cdot \beta$, each $U = U_1 \circ U_2 \circ \cdots \circ U_m$ and $U' = U'_1 \circ U'_2 \circ \cdots \circ U'_m$ be a random variable uniformly distributed over $\{1,2\}^m$. For $1 \leq i \leq m-1$, let each Y_i and Y'_i be a random variable uniformly distributed over QR_n^{+1}. Let each V and V' be a random variable uniformly distributed over QNR_n^{+1}. In the PIR scheme, $\mathcal{A}(1^k, X, \mathcal{Q}^1(1^k, i)))$ corresponds to $Y_1^{U_1} \cdots V^{U_i} \cdots Y_{m-1}^{U_m}$. Similarly, $\mathcal{A}(1^k, X', \mathcal{Q}^1(1^k, j)))$ corresponds to $Y_1'^{U'_1} \cdots V'^{U'_j} \cdots Y_{m-1}'^{U'_m}$.

Since multiplication is commutative,

$$\begin{aligned}
& \Pr[\mathcal{B}(1^k, Y_1^{U_1} \cdots V^{U_i} \cdots Y_{m-1}^{U_m}) = 1] \\
= & \sum_{u \in \{1,2\}^m} \sum_{v \in QNR_n^{+1}} \sum_{b=1}^{m-1} \sum_{y_b \in QR_n^{+1}} \Pr[U = u] \cdot \\
& \Pr[V = v] \cdot \prod_{c=1}^{m-1} \Pr[Y_c = y_c] \cdot \\
& \Pr[\mathcal{B}(1^k, y_1^{u_1} \cdots v^{u_i} \cdots y_{m-1}^{u_m}) = 1] \\
= & \sum_{u' \in \{1,2\}^m} \sum_{v' \in QNR_n^{+1}} \sum_{b=1}^{m-1} \sum_{y'_b \in QR_n^{+1}} \Pr[U' = u'] \cdot \\
& \Pr[V' = v'] \cdot \prod_{c=1}^{m-1} \Pr[Y'_c = y'_c] \cdot \\
& \Pr[\mathcal{B}(1^k, y_1'^{u'_1} \cdots v'^{u'_j} \cdots y_{m-1}'^{u'_m}) = 1] \\
= & \Pr[\mathcal{B}(1^k, Y_1'^{U'_1} \cdots V'^{U'_j} \cdots Y_{m-1}'^{U'_m}) = 1].
\end{aligned}$$

Hence $\mathcal{A}(1^k, X, \mathcal{Q}^1(1^k, i))$ and $\mathcal{A}(1^k, X', \mathcal{Q}^1(1^k, j))$ are (information theoretical) indistinguishable in the PIR scheme. By Lemma 1,

$$\Pr[\mathcal{B}(1^k, \mathcal{Q}^2(1^k, i), \mathcal{A}(1^k, X, \mathcal{Q}^1(1^k, i))) = 1] - \Pr[\mathcal{B}(1^k, \mathcal{Q}^2(1^k, j), \mathcal{A}(1^k, X', \mathcal{Q}^1(1^k, j))) = 1] = 0.$$

□

4.2 SPAAP

We use a public-key encryption scheme and a random oracle as a hash function in order to construct SPAAP.

We show the definition of a public-key encryption scheme (Goldreich 2001) as follows.

Definition 4 *A semantically secure public-key encryption scheme is a triple $(\mathcal{G}, \mathcal{E}, \mathcal{D})$ of probabilistic polynomial-time algorithms satisfying the following conditions.*

- *On input 1^k, algorithm $\mathcal{G}$ outputs a pair of bit strings.*
- *For any pair of (e, d) in the range of $\mathcal{G}(1^k)$, and any $\gamma \in \{0,1\}^*$,*

$$\Pr[\mathcal{D}(d, \mathcal{E}(e, \gamma)) = \gamma] = 1. \quad (3)$$

- *For any $k \in \mathbb{N}$ any $x, y \in \{0,1\}^{\mathrm{poly}(k)}$, and any probabilistic polynomial-time algorithm $\mathcal{B}$,*

$$|\Pr[\mathcal{B}(\mathcal{G}^1(1^k), \mathcal{E}(\mathcal{G}^1(1^k), x)) = 1] - \Pr[\mathcal{B}(\mathcal{G}^1(1^k), \mathcal{E}(\mathcal{G}^1(1^k), y)) = 1]| < \frac{1}{p(k)}. \quad (4)$$

In this paper, we assume that we can regard any hash function as a random oracle (that is, the random oracle model) (Bellare & Rogaway 1993). This assumption is called the *random oracle assumption*. In the random oracle model, all entities can interact with a random oracle $\mathcal{H}$, that is a single function which is uniformly chosen from all possible functions. We note that if the random oracle $\mathcal{H}$ receives the same input, $\mathcal{H}$ answers the same output. We assume that the random oracle outputs m bit strings on inputs ℓ bit strings, where ℓ and m are polynomials of a security parameter k. The following lemma holds.

Lemma 3 *For any $k \in \mathbb{N}$, any $x, y \in \{0,1\}^{\mathrm{poly}(k)} (x \neq y)$, and probabilistic polynomial-time algorithm $\mathcal{B}$,*

$$|\Pr[\mathcal{B}(1^k, \mathcal{H}(x)) = 1] - \Pr[\mathcal{B}(1^k, \mathcal{H}(y)) = 1] = 0.$$

SPAAP $\langle \mathcal{P}, \mathcal{V}, \mathcal{M} \rangle$, which satisfies the all requirements; correctness, impersonation resistance against passive insider, anonymity against central manager, and anonymity against service provider, is shown as follows, where $(\mathcal{Q}, \mathcal{A}, \mathcal{R})$ is the Kushilevitz and Ostrovsky's PIR scheme which described in the previous section.

1. $\mathcal{M}$ computes $(e, d) \leftarrow \mathcal{G}(1^k)$ and publishes e.
2. $\mathcal{P}$ computes $(q, s) \leftarrow \mathcal{Q}(1^k, i)$ and sends $(\mathcal{E}(e, q), s)$ to $\mathcal{V}$.
3. $\mathcal{V}$ sends $\mathcal{E}(e, q)$ to $\mathcal{M}$.
4. $\mathcal{M}$ obtains q by decrypting $\mathcal{E}(e, q)$. $\mathcal{M}$ randomly chooses $c \in \{0,1\}^\ell$ and for any $j \in [m]$ computes $x'_j \leftarrow \mathcal{H}(x_j, c)$. Let $x' = (x'_1, x'_2, \ldots, x'_m)$. $\mathcal{M}$ computes $a \leftarrow \mathcal{A}(1^k, x', q)$ and sends (c, a) to $\mathcal{V}$.
5. $\mathcal{V}$ computes $x'_i \leftarrow \mathcal{R}(1^k, s, a) = \mathcal{H}(x_i, c)$ and sends c to $\mathcal{P}$.
6. $\mathcal{P}$ computes $z' \leftarrow \mathcal{H}(z, c)$ where z is a candidate password, and sends z' to $\mathcal{V}$.
7. $\mathcal{V}$ outputs 1 if $z' = x'_i$, and outputs 0 otherwise.

5 Security Analysis

Theorem 1 *SPAAP has correctness under the quadratic residuosity assumption and the random oracle assumption.*

proof: In Step 2, q is always decrypted by Equality (3). In Step 5, the probability that $x'_i = \mathcal{H}(x_i, c)$ is higher than $1 - 1/p(k)$ by Inequality (1). Hence if $z = x_i$, the probability that $z' = x'_i$ is higher than $1 - 1/p(k)$. □

Theorem 2 *SPAAP has impersonation resistance against passive insider under the quadratic residuosity assumption and the random oracle assumption.*

proof: The main idea of this proof is that an adversary who has no pre-knowledge can simulate the transaction which is given to the service provider.

We prove that by contradiction. It is clearly (information theoretic) hard for any adversary to impersonate a legitimate user, if the adversary can obtain no pre-knowledge about x. That is, for any probabilistic polynomial-time algorithm $\mathcal{B}$,

$$\Pr[\langle \mathcal{B}(1^k), \mathcal{V}(1^k), \mathcal{M}(1^k, X) \rangle = 1] = \frac{1}{2^\ell} < \frac{1}{p(k)}, \quad (5)$$

where X is a random variable uniformly distributed over $(\{0,1\}^\ell)^m$.

The random variable T_1 is $\{\mathcal{E}(\mathcal{G}^1(1^k), \mathcal{Q}^1(i)), \mathcal{Q}^2(i), c, \mathcal{A}((\mathcal{H}(x_1, c), ..., \mathcal{H}(x_m, c)), \mathcal{Q}^1(i)), \mathcal{H}(z, c)\}$, where $x_1, \ldots x_m$ are samples from $\{0,1\}^\ell$, and c is a sample from $\{0,1\}^\ell$. Let T'_1 be $\{\mathcal{E}(\mathcal{G}^1(1^k), 1^{|\mathcal{Q}^1(i)|}), \mathcal{Q}^2(i), c, \mathcal{A}(y_1, \ldots y_m, \mathcal{Q}^1(i)), u\}$, where $y_1, \ldots y_m$ are samples from $\{0,1\}^\ell$, and c and u are samples from $\{0,1\}^\ell$. By Inequality (4), $\mathcal{E}(\mathcal{G}^1(1^k), \mathcal{Q}^1(i))$ and $\mathcal{E}(\mathcal{G}^1(1^k), 1^{|\mathcal{Q}^1(i)|})$ are indistinguishable. By the basic property of a random oracle, $\mathcal{A}(\mathcal{H}(x_1, c), ..., \mathcal{H}(x_m, c), \mathcal{Q}^1(i))$ and $\mathcal{A}(y_1, \ldots, y_m, \mathcal{Q}^1(i))$ are indistinguishable. By Lemma 1, T_1 and T'_1 are indistinguishable.

We assume that SPAAP does not have impersonation resistance against passive insider, that is, there exists some polynomial q and some probabilistic polynomial-time algorithm $\mathcal{D}$ such that

$$\Pr[\langle \mathcal{D}(1^k, T_1), \mathcal{V}(1^k), \mathcal{M}(1^k, X) \rangle = 1] \geq \frac{1}{q(k)}. \quad (6)$$

We derive contradiction by constructing a probabilistic polynomial-time algorithm $\mathcal{D}'$ which takes 1^k as an input and uses the algorithm $\mathcal{D}$ as a subroutine. $\mathcal{D}'$ proceeds as follows.

1. $\mathcal{D}'$ computes$(e, d) \leftarrow \mathcal{G}(1^k)$ and randomly chooses c, y, u.
2. $\mathcal{D}'$ computes $t_2 = \{\mathcal{E}(\mathcal{G}^1(1^k), 1^{|\mathcal{Q}_1(i)|}), \mathcal{Q}^2(i), c, \mathcal{A}(y, \mathcal{Q}^1(i)), u\}$.
3. $\mathcal{D}'$ outputs $\mathcal{D}(1^k, t_2)$.

By Inequality (6), it holds that

$$\Pr[\langle \mathcal{D}'(1^k), \mathcal{V}(1^k), \mathcal{M}(1^k, X) \rangle = 1] \geq \frac{1}{q(k)},$$

because T_1 and T'_1 are indistinguishable. This contradicts to Inequality (5). □

Theorem 3 *SPAAP has anonymity against central manager under the quadratic residuosity assumption and the random oracle assumption.*

proof: We prove that by contradiction. The random variable T_2 is $\{\mathcal{G}^2(1^k), \mathcal{E}(\mathcal{G}^1(1^k), \mathcal{Q}^1(i))\}$ and random variable T_3 is $\{\mathcal{G}^2(1^k), \mathcal{E}(\mathcal{G}^1(1^k), \mathcal{Q}^1(j))\}$. We assume that SPAAP does not have anonymity against central manager, that is, there exists some polynomial q and some probabilistic polynomial-time algorithm $\mathcal{D}$ such that

$$|\Pr[\mathcal{D}(1^k, T_2) = 1] - \Pr[\mathcal{D}(1^k, T_3) = 1]| \geq \frac{1}{q(k)}. \quad (7)$$

We derive contradiction by constructing a probabilistic polynomial-time algorithm $\mathcal{D}'$ which takes 1^k and y as inputs and uses the algorithm $\mathcal{D}$ as a subroutine. $\mathcal{D}'$ proceeds as follows.

1. $\mathcal{D}'$ computes $(e, d) \leftarrow \mathcal{G}(1^k)$.
2. $\mathcal{D}'$ outputs $\mathcal{D}(d, \mathcal{E}(e, y))$.

By Inequality (7), it holds that

$$|\Pr[\mathcal{D}(1^k, \mathcal{Q}_1(1^k, i)) = 1] - \Pr[\mathcal{D}(1^k, \mathcal{Q}_1(1^k, j)) = 1]| \geq \frac{1}{p(k)}.$$

This contradicts to Inequality (2). □

Theorem 4 *SPAAP has anonymity against service provider under the quadratic residuosity assumption and the random oracle assumption.*

proof: The random variable T_4 is $\{\mathcal{E}(\mathcal{G}^1(1^k), \mathcal{Q}^1(i)), \mathcal{Q}^2(i), c, \mathcal{A}((\mathcal{H}(x_1, c), ..., \mathcal{H}(x_m, c)), \mathcal{Q}^1(i)), \mathcal{H}(z, c)\}$, and the random variable T_5 is $\{\mathcal{E}(\mathcal{G}^1(1^k), \mathcal{Q}^1(j)), \mathcal{Q}^2(j), c, \mathcal{A}((\mathcal{H}(x_1, c), ..., \mathcal{H}(x_m, c)), \mathcal{Q}^1(j)), \mathcal{H}(z', c)\}$ where $x_1, \ldots x_m$ are samples from $\{0, 1\}^\ell$, and c is a sample from $\{0, 1\}^\ell$. By Inequality (4), $\mathcal{E}(\mathcal{G}^1(1^k), \mathcal{Q}^1(i))$ and $\mathcal{E}(\mathcal{G}^1(1^k), \mathcal{Q}^1(j))$ are indistinguishable. By Lemma 2 and the basic property of a random oracle, $\mathcal{Q}^2(i), \mathcal{A}((\mathcal{H}(x_1, c), ..., \mathcal{H}(x_m, c)), \mathcal{Q}^1(i))$ and $\mathcal{Q}^2(j), \mathcal{A}((\mathcal{H}(x_1, c), ..., \mathcal{H}(x_m, c)), \mathcal{Q}^1(j))$ are indistinguishable. By Lemma 3, $\mathcal{H}(z, c)$ and $\mathcal{H}(z', c)$ are (information theoretical) indistinguishable. Therefore, T_4 and T_5 are indistinguishable by Lemma 1.

□

6 Conclusions

In this paper, we proposed SPAAP, which consists of the special version of the single-database PIR scheme proposed by Kushilevitz and Ostrovsky, in which an element of the database can be reconstructed without the index. We proved that SPAAP satisfies all the requirements; correctness, impersonation resistance against passive insider, anonymity against central manager, and anonymity against service providers under the quadratic residuosity assumption and the random oracle assumption. SPAAP is more practical than the existing protocol (Nakamura et al. 2009) since using a single database implies the assumptions for multi-database PIR schemes are not required any more.

Acknowledgements

This work was in part supported by CREST-DVLSI of JST. We are grateful for their support.

References

Bellare, M. & Rogaway, P. (1993), Random oracles are practical: A paradigm for designing efficient protocols, *in* 'In Proc. 1st ACM Conference on Computer and Communications Security', ACM Press, pp. 62–73.

Camenisch, J. & Lysyanskaya, A. (2002), Dynamic accumulators and application to efficient revocation of anonymous credentials, *in* 'Advances in Cryptology — CRYPTO 2002', LNCS, Springer-Verlag, pp. 101–120.

Chaum, D. & van Heyst, E. (1991), Group signatures, *in* 'Advances in Cryptology - EUROCRYPT 1991', Vol. 547 of *LNCS*, Springer-Verlag, pp. 257–270.

Chor, B., Goldreich, O., Kushilevitz, E. & Sudan, M. (1998), 'Private information retrieval', *Jornal of the ACM* **45**, 965–982.

Goldreich, O. (2001), *Foundations of Cryptography*, Cambridge University.

Kushilevitz, E. & Ostrovsky, R. (1997), Replication is not needed: Single database, computationally-private information retrieval, *in* 'the 38th Annual Symposium on Foundations of Computer Science', pp. 364–373.

Liao, Y.-P. & Wang, S.-S. (2009), 'A secure dynamic ID based remote user authentication scheme for multi-server environment', *Computer standards and interfaces* **31**(1), 24–29.

Nakamura, T., Inenaga, S., Ikeda, D., Baba, K. & Yasuura, H. (2009), Anonymous authentication systems based on private information retrieval, *in* 'The First Conference on 'Networked Digital Technologies'(NDT2009)', pp. 53–58.