数理解析研究所講究録 1222 巻 2001 年 64-73

Subgroup Membership Problem and Its Applications to Information Security

Akihiro Yamamura (山村 明弘) *

1 Subgroup Membership Problem

It is well known that the *subgroup membership problem* for a finitely presented groups is not decidable in general, whereas it is solvable for the class of finite groups or finitely generated abelian groups, However, if we consider more practical computation, that is, the probabilistic polynomial time algorithms (or equivalently the class **BPP**), the membership problem is not trivial even for the class of finite abelian groups. As a matter of fact, several algorithmic problems used in cryptography are characterized as the subgroup membership problem under the assumption that the corresponding membership problem is not in **BPP**. For example, there exists no known probabilistic polynomial time algorithm for the integer factorization or the discrete logarithm problem for some class of finite cyclic groups. The quadratic residue (QR for short) problem and the decision Diffie-Hellman (DDH for short) problem have numerous applications in cryptography, and hence, they have been studied in detail. Following [11], we generalize and formalize them as the *subgroup membership problem* and to show many other algorithmic problems, which are used in public key cryptography, are characterized as the subgroup membership problem as well. Such a unification of algorithmic problems used in cryptography has not been appeared up to date as far as the authors are concerned. Widely used assumptions in cryptography are divided into two groups: the algorithmic assumptions related to the integer factoring (and the QR) and the algorithmic assumptions related to the discrete logarithm problem (and the DDH). The first is originated from the RSA cryptosystem and the second from the Diffie-Hellman key exchange protocol. These two look different and are usually discussed separately. The unified approach to the integer factoring problem and the discrete logarithm problem shed light on the fundamental properties of algorithms required to provide the security. Therefore, we can get better understanding of the algorithmic problems by unified treatment of subgroup membership problems.

1.1 Subgroup Membership Assumption

Determining the membership of a given element of a certain group in its subgroup is not always easy. As a matter of fact, the membership problem of a subgroup in a finitely presented group is not recursive in general. To apply the membership problem to cryptographic schemes such as asymmetric cryptosystems, we require the efficiency of computation for legal participants and the existence of a trapdoor. In this section we consider the subgroup membership problem with a trapdoor, and show that several problems widely used in cryptography are characterized as the subgroup membership problem.

*Communications Research Laboratory, 4-2-1, Nukui-Kitamachi, Koganei, Tokyo, 184-8795 Japan e-mail: aki@crl.go.jp

Let G be a group, and let H be its subgroup. The membership problem is to decide whether or not a given element $g \in G$ belongs to H. We suppose that every element in G has a binary representation of size k, where k is the security parameter. The membership can be decided within polynomial time in k if a certain information, called a *trapdoor*, is provided. The membership of an element $g \in G$ in H can be decided provided the trapdoor, however, the membership cannot be decided with a probability substantially larger than $\frac{1}{2}$ without the trapdoor. We now formalize the subgroup membership problem.

Let k be the security parameter. For the input 1^k, a probabilistic polynomial time algorithm $\mathcal{IG}$ outputs the description of a group G, the description of a subgroup $H \subset G$ and the trapdoor that provides a fast algorithm for the subgroup membership problem of H in G. The algorithm $\mathcal{IG}$ is called the *instance generator*. Every element of G is represented as a binary sequence of length k. Computation of the multiplication in G is performed in polynomial time in k.

The predicate for the membership of a subgroup is denoted by Mem, that is, Mem is defined as follows:

$$\mathrm{Mem}(G, H, x) = \begin{cases} 1 & \text{if } x \in H \\ 0 & \text{if } x \in S \end{cases},$$

where $\mathcal{IG}$ outputs the pair (G, H) for 1^k, x is in G, and $S = G \setminus H$. The *subgroup membership problem* is to compute Mem in polynomial time in k when we inputs 1^k and obtain a pair of groups (G, H) and an element g in G, which is uniformly and randomly chosen from H or G according to the coin toss $b \stackrel{R}{\leftarrow} \{0, 1\}$. If there does not exist a probabilistic polynomial time algorithm that computes Mem with a probability substantially larger than $\frac{1}{2}$, then we say that the membership problem is *intractable*. We also assume that one can choose uniformly and randomly an element from both H and G. This is significant to apply to cryptographic schemes.

The following is trivial, however, it is useful for the construction of an PIR system based on the subgroup membership problem.

Proposition 1.1 *Let G be a group, and let H be a subgroup of G. For any $g \in G$ and $h \in H$, we have $gh \in H$ if and only if $g \in H$.* □

Subgroup Membership Assumption I
For every constant c, and every family $\{C_k \mid k \in \mathbb{N}\}$ of circuits of polynomial size in k, there is an integer K such that for all $k > K$ we have

$$\mathbf{Prob}(C_k(G, H, g) = \mathrm{Mem}(G, H, g)) < \frac{1}{2} + \frac{1}{k^c}, \tag{1.1}$$

where the probability is taken over $(G, H) \leftarrow \mathcal{IG}(1^k)$, $b \stackrel{R}{\leftarrow} \{0, 1\}$, $g \stackrel{R}{\leftarrow} H$ if $b = 1$, $g \stackrel{R}{\leftarrow} S$ if $b = 0$.

The assumption claims that there exists no polynomial size circuit family to compute the predicate Mem. The following is equivalent to the assumption above.

Subgroup membership assumption II
For every constant c, and every family $\{C_k \mid k \in \mathbb{N}\}$ of circuits of polynomial size in k, there is an integer K such that for all $k > K$ we have

$$|\mathbf{P}_H - \mathbf{P}_S| < \frac{1}{k^c}, \tag{1.2}$$

where the probabilities $\mathbf{P}_H$ and $\mathbf{P}_S$ are defined as follows;

$$\mathbf{P}_H = \mathbf{Prob}_{(G,H)\leftarrow\mathcal{IG}(1^k)\,;\,g\stackrel{R}{\leftarrow}H}(C_k(G,H,g)=1)\ ,$$

and

$$\mathbf{P}_S = \mathbf{Prob}_{(G,H)\leftarrow\mathcal{IG}(1^k)\,;\,g\stackrel{R}{\leftarrow}S}(C_k(G,H,g)=1)\ .$$

1.2 Examples

We exhibit several subgroup membership problems: the DDH problem, the QR problem, the rth residue (RR for short) problem studied by Kurosawa and Tsujii [6], the p-subgroup (PSUB for short) problem introduced by Okamoto and Uchiyama [8] and the decisional composite residuosity (DCR for short) problem introduced by Paillier [9]. Recall that the assumption that the QR problem is intractable (QR assumption) is employed to prove the semantic security of Goldwasser-Micali cryptosystem [4], and the assumption that the DDH problem is intractable (DDH assumption) is employed to prove the semantic security of ElGamal cryptosystem. These two have many other applications. The assumption that one of problems above is intractable is employed to prove the semantic security of the corresponding cryptosystem [6], [8], [9], respectively. We also note that the security of the cryptosystem introduced by Naccache and Stern [7] depends on the PSUB assumption as well.

Quadratic Residue Problem

Let p, q be primes. Set $N = pq$. The primes p and q are trapdoor information for the quadratic residue problem, on the other hand, the number N is public information. Let G be the subgroup of $(\mathbf{Z}/(N))^*$ consisting of the elements whose Jacobi symbol is 1, and let H be the subgroup of G consisting of quadratic residues of G, that is, $H = \{x \in G \mid x = y^2 \bmod N \text{ for } y \in (\mathbf{Z}/(N))^*\}$. The quadratic residue problem of H in G is to decide whether or not, a given element $g \in G$, g belongs to H. We can effectively determine the membership of g in H provided that the information p and q are available. No polynomial time algorithm is known for the membership of a randomly chosen element of G in H without the information p and q. Hence, if we define an instance generator for the QR problem as a probabilistic algorithm that outputs two primes p and q of size k and a quadratic non-residue h whose Jacobi symbol is 1 for the input 1^k, then the QR problem is considered as a subgroup membership problem. Note that we can obtain a quadratic non-residue h with Jacobi symbol 1 by using p, q, and that it is possible to uniformly and randomly choose elements from H without the trapdoor information provided h is given.

Decision Diffie-Hellman Problem

Let C be a cyclic group of prime order p. The group C may be a multiplication group of a finite field or a group of rational points of an elliptic curve. Let g be a generator of C. The decision Diffie-Hellman problem is to decide whether or not $h_2 = g_2^a$ for the given quadruple (g_1, h_1, g_2, h_2) of elements in C with $h_1 = g_1^a$ for some $1 \leq a \leq p-1$. If so, we say that (g_1, h_1, g_2, h_2) is a Diffie-Hellman quadruple. The integer a is the trapdoor of the decision Diffie-Hellman problem. Knowing the trapdoor a, we can efficiently decide whether or not $h_2 = g_2^a$.

We show that the DDH problem can be characterized as a subgroup membership problem for a certain group. We set G to be the direct product $C \times C$. Then the input to the DDH problem is (x, y) where $x, y \in G$, that is, $x = (g_1, h_1)$ and $y = (g_2, h_2)$. It is obvious that (g_1, h_1, g_2, h_2) is a Diffie-Hellman quadruple if and only if y belongs to the subgroup $< x >$ of G generated by x. It follows that the DDH problem for the cyclic group C is equivalent to the subgroup membership problem of the group $H =< x >$, where $x = (g_1, g_1^a)$, in the group

$G = C \times C =< g_1 > \times < g_1 >$. Note that, when a generator x of H is given, it is possible to choose uniformly and randomly elements from H without the trapdoor information.

Rth Residue Problem

The RR problem is a natural extension of the QR problem defined as follows. Let p, q be primes, and let e_1, e_2 be odd integers dividing $p-1$ and $q-1$, respectively, such that e_1 is prime to $q-1$ and e_2 is prime to $p-1$. Set $N = pq$ and $r = e_1 e_2$. The primes p and q are the trapdoor information for the RR problem, on the other hand, the number N and r are the public information. Let G be the group $(\mathbf{Z}/(N))^*$, and let H be the subgroup consisting of rth residues of G, that is, $H = \{x \in G \mid x = y^r \bmod N \text{ for } y \in G\}$. The RR problem of H in G is to decide whether or not, a given element $g \in G$, g belongs to H. Thus, the RR is a subgroup membership problem of H in G. We can effectively determine the membership of g in H provided that the information p and q are available. No polynomial time algorithm is known for the membership of a randomly chosen element of G in H without the information p and q. Note that we can obtain an element h such that $h^i \notin \{x^r \bmod N : x \in (\mathbf{Z}/(N))^*\}$ for any $1 \leq i \leq r-1$ by using the trapdoor information, and that we can uniformly and randomly choose an element from H provided h is given.

P-Subgroup Problem

Let p, q be primes such that p does not divide $q-1$. Set $N = p^2 q$ and let g be a random element in $(\mathbf{Z}/(N))^*$ such that the order of $g^{p-1} \bmod p^2$ is p. The primes p and q are trapdoor information for the PSUB problem, on the other hand, the number N, g, k are public information. Let G be a group defined by $G = \{x \mid x = g^m y^N \bmod N \text{ for } m \in \mathbf{Z}/(p) \text{ and } y \in (\mathbf{Z}/(N))^*\}$, and let H be the subgroup defined by $H = \{x \mid x = y^N \bmod N \text{ for } y \in G\}$. The PSUB problem of H in G is to decide whether or not, a given element $g \in G$, g belongs to H. Thus, the PSUB is the membership problem of H in G. We can efficiently determine the membership of g in H provided that the information p and q are available. No polynomial time algorithm is known for the membership of a randomly chosen element of G in H without the information p and q. Note that our description of PSUB is slightly diffrent from Okamoto-Uchiyama [8], where the PSUB is introduced as a variant of the *coset indistinguishability problem*, which we will present in Section 2.3. Naccache and Stern [7] implicitly used PSUB problem in their scheme. Paillier introduces the *decisional composite residuosity* (DCR for short). This is a generalization of [8] and also characterized as a subgroup membership problem.

For other plausible applications of the subgroup membership problem, the reader is also referred to [10] in which the DDH assumption is applied to the cryptographic schemes which only known method to construct is to base on the QR assumption.

1.3 Equivalent Problems

We examine several algorithmic problems equivalent to the subgroup membership problem. Suppose that $\mathcal{IG}$ is an instance generator of a family of groups, and that $\mathcal{IG}$ outputs (G, H) for the input 1^k. We set $S = G \setminus H$. Suppose that t is an integer bounded above by a polynomial in k. Let K_i be the direct product of $t-1$ H's and S, where all jth position $(j \neq i)$ is occupied by H except for ith position, that is, $K_i = H \times H \times \cdots \times \overset{i}{S} \times \cdots \times H$ for every $i = 1, 2, \ldots, t$. Let L be the union of $K_1, K_2, \cdots, K_t$, that is, $L = K_1 \bigcup K_2 \bigcup \cdots \bigcup K_t$.

Pattern Indistinguishability Assumption

The *pattern indistinguishability assumption* is to assume the following holds: for every constant c, every family $\{C_k \mid k \in \mathbb{N}\}$ of circuits of polynomial size in k and all i, j such that $1 \leq i, j \leq n$

there is an integer K such that for all $k > K$ we have

$$|\mathbf{P}_i - \mathbf{P}_j| < \frac{1}{k^c} \ . \tag{1.3}$$

Here the probabilities $\mathbf{P}_i$ and $\mathbf{P}_j$ are defined as follows;

$$\mathbf{P}_i = \mathbf{Prob}_{(G,H)\leftarrow\mathcal{IG}(1^k)\ ;\ (g_1,g_2\cdots,g_t)\stackrel{R}{\leftarrow}K_i}(C_k(G,H,i,g_1,g_2\cdots,g_t)=1) \ ,$$

$$\mathbf{P}_j = \mathbf{Prob}_{(G,H)\leftarrow\mathcal{IG}(1^k)\ ;\ (g_1,g_2\cdots,g_t)\stackrel{R}{\leftarrow}K_j}(C_k(G,H,i,g_1,g_2\cdots,g_t)=1) \ .$$

General Pattern Indistinguishability Assumption

The *general pattern indistinguishability assumption* is to assume the following holds: for every constant c, every family $\{C_k \mid k \in \mathbf{N}\}$ of circuits of polynomial size in k and all $(i_1, i_2, \ldots, i_u)$ and $(j_1, j_2, \ldots, j_u)$, there is an integer K such that for all $k > K$ we have

$$|\mathbf{P}_{(i_1,i_2,\cdots,i_u)} - \mathbf{P}_{(j_1,j_2,\cdots,j_u)}| < \frac{1}{k^c} \ . \tag{1.4}$$

Here the probabilities $\mathbf{P}_{(i_1,i_2,\cdots,i_u)}$ and $\mathbf{P}_{(j_1,j_2,\cdots,j_u)}$ are defined by

$$\mathbf{P}_{(i_1,i_2,\cdots,i_u)} = \mathbf{Prob}(C_k(G,H,x_1,x_2\ldots,x_u)=1) \ ,$$

where the probability is taken over $(G,H) \leftarrow \mathcal{IG}(1^k)$ and $(x_1, x_2 \ldots, x_u) \stackrel{R}{\leftarrow} K_{i_1} \times K_{i_2} \times \cdots \times K_{i_u}$ and

$$\mathbf{P}_{(j_1,j_2,\cdots,j_u)} = \mathbf{Prob}(C_k(G,H,x_1,x_2\ldots,x_u)=1) \ ,$$

where the probability is taken over $(G,H) \leftarrow \mathcal{IG}(1^k)$ and $(x_1, x_2 \ldots, x_u) \stackrel{R}{\leftarrow} K_{j_1} \times K_{j_2} \times \cdots \times K_{j_u}$.

Coset Indistinguishability Assumption

The *coset indistinguishability assumption* is to assume the following holds: for every constant c, every family $\{C_k \mid k \in \mathbf{N}\}$ of circuits of polynomial size in k and every algorithm F that on input (G,H) outputs a pair of elements in G, there is an integer K such that for all $k > K$ we have

$$\mathbf{Prob}(C_k(G,H,g_0,g_1,g)=b) < \frac{1}{2} + \frac{1}{k^c} \ , \tag{1.5}$$

where the probability is taken over $(G,H) \leftarrow \mathcal{IG}(1^k)$, $(g_0,g_1) \leftarrow F(G,H)$, $b \stackrel{R}{\leftarrow} \{0,1\}$ and $g \stackrel{R}{\leftarrow} g_b H$.

Theorem 1.2 *The following are equivalent.*
(1) The subgroup membership assumption I.
(2) The subgroup membership assumption II.
(3) The pattern indistinguishability assumption.
(4) The general pattern indistinguishability assumption.
(5) The coset indistinguishability assumption. □

2 Private Information Retrieval

Chor, Goldreich, Kushilevitz and Sudan [2] introduced the *private information retrieval scheme* for remote database access, in which the user can retrieve the data of user's choice without revealing it. Their scheme attains *information theoretic security*, however, the database must be replicated in several locations where the managers are not allowed to communicate each other. The *computational private information retrieval scheme* was introduced by Chor and Gilboa [3]. Their scheme attains more efficient communication than Chor, Goldreich, Kushilevitz and Sudan's model by sacrificing the information theoretic security, nevertheless, their scheme enjoys computational security by assuming the existence of pseudorandom generators. However, their scheme still needs replication of the database. Kushilevitz and Ostrovsky [5] introduced a computational private information retrieval scheme in which only one database is needed. Their scheme depends on the intractability of the quadratic residue problem. More efficiency, polylogarithmic communication complexity, is attained by Cachin, Micali and Stadler [1]. They assume a number theoretic hypothesis, which they call the Φ assumption, and sacrifice one-round communication and then obtain polylogarithmic communication complexity. However, a rigorous proof of the intractability of the Φ assumption or its equivalence to a widely used assumption like the quadratic residue assumption or the integer factorization is not given in [1].

We briefly review the general scheme of a private information retrieval (PIR for short) scheme. A computational PIR scheme with a single database is a protocol for two players, a user $\mathcal{U}$ and a database manager $\mathcal{DB}$. Both are able to perform only probabilistic polynomial time computation. The database manager $\mathcal{DB}$ maintains a database, which is a binary sequence $X = x_0x_1x_2\cdots x_{n-1}$. The goal of the protocol is to allow $\mathcal{U}$ to obtain the ith bit x_{i+1} of X without leaking any information on x_i to $\mathcal{DB}$. The protocol runs as follows:

Step 1 $\mathcal{U}$ computes a query Query(i) using his random tape (coin toss), which $\mathcal{U}$ keeps secret. Then he sends Query(i) to $\mathcal{DB}$.

Step 2 $\mathcal{DB}$ receives Query(i). He performs a polynomial-time computation for the input X, Query(i) and his random tape. The computation yields the answer Answer(Query(i)). He sends Answer(Query(i)) back to $\mathcal{U}$.

Step 3 $\mathcal{U}$ receives Answer(Query(i)). He performs a polynomial-time computation using the answer Answer(Query(i)) and his private information (his random tape). The computation yields the ith bit x_{i+1} of the database.

Correctness
For any database sequence X and for any query Query(i) for ith bit of X, $\mathcal{U}$ obtains x_i at the end.

Privacy
$\mathcal{DB}$ cannot distinguish a query for the ith bit and a query for the jth bit for all i and j by a polynomial-time (probabilistic) computation with non-negligible probability. Formally, for all constants c, for all database of length n, for any two $1 \leq i, j \leq n$, and all polynomial-size family of circuits C_k, there exists an integer K such that for all $k > K$ we have

$$|\mathbf{Prob}(C_k(\text{Query}(i)) = 1) - \mathbf{Prob}(C_k(\text{Query}(j)) = 1)| < \sigma \ , \tag{2.1}$$

where k is the security parameter of the protocol and $\sigma = \frac{1}{(\text{Max}(k,n))^c}$.

Computation
Computations of both $\mathcal{DB}$ and $\mathcal{U}$ are bounded above by a polynomial in the size n of the database

and the security parameter k.

3 PIR Based on the Subgroup Membership Problem

Following [11], we show that the subgroup membership problem can be applied to a PIR scheme by modifying Kushilevitz and Ostrovsky's scheme [5]. The proposed scheme has the same communication complexity as Kushilevitz and Ostrovsky's scheme whose security depends on the QR assumption. On the other hand, the security of the private information retrieval scheme proposed in this paper is based on the subgroup membership assumption. Therefore, we can construct a private information retrieval scheme based on any algorithmic problems in Section 2.2, in particular, we can use groups of rational points on elliptic curves or multiplicative groups of finite fields under the corresponding DDH assumption. We should remark that all the private information retrieval schemes proposed so far depend on either the existence of pseudorandom number generators or intractability assumption related to the integer factorization. No private information retrieval scheme based on the DDH has been proposed, yet as far as the authors are concerned. Modifying [5], we construct a PIR scheme based on the subgroup membership problem.

3.1 Basic Idea

First of all, we explain the basic idea of the scheme by a simple model. Suppose $\mathcal{DB}$ has the database $X = x_0x_1x_2\cdots x_{n-1}$ and that $\mathcal{U}$ wishes to know the ith bit x_{i-1}. $\mathcal{U}$ chooses group elements $g_0, g_1, g_2, \ldots, g_{i-1}, \ldots, g_{n-1}$ so that $g_j \in H$ for $j \neq i-1$ and $g_{i-1} \in S = G \setminus H$. Then $\mathcal{U}$ sends them all to $\mathcal{DB}$. $\mathcal{DB}$ computes the group element $g = g_0^{x_0} g_1^{x_1} g_2^{x_2} \cdots g_{i-1}^{x_{i-1}} \cdots g_{n-1}^{x_{n-1}}$ and sends it back to $\mathcal{U}$. $\mathcal{DB}$ cannot get to know which of $g_0, g_1, g_2, \ldots, g_{i-1}, \ldots, g_{n-1}$ comes from S if the subgroup membership problem of H in G is intractable. Since $\mathcal{U}$ possesses the trapdoor, he can determine whether or not g lies in H. By Proposition 1, g lies in H if and only if $x_{i-1} = 0$. Therefore, $\mathcal{U}$ can obtain the ith bit x_{i-1}. This simple model illustrates the idea of using the subgroup membership problem, but the communication complexity is still large. We need the trick by [5] to reduce the communication complexity.

3.2 Scheme

Step 0 The user $\mathcal{U}$ inputs 1^k to the instance generator $\mathcal{IG}$ and then gets a pair (G, H) of groups and the trapdoor for the subgroup membership problem of H in G, where k is the security parameter and every element of G is represented by a binary sequence of length k. We assume the subgroup membership assumption of H in G. The group G is shared by both $\mathcal{DB}$ and $\mathcal{U}$. On the other hand, $\mathcal{U}$ keeps the trapdoor information for the subgroup membership problem of H secret. Computations of both $\mathcal{DB}$ and $\mathcal{U}$ are performed in the group G. Let X be the database managed by $\mathcal{DB}$. We suppose that $X = x_0x_1x_2\cdots x_{n-1}$, where $x_i \in \{0,1\}$, and that $n = t^l$, where t, l are positive integers.

Step 1 $\mathcal{U}$ computes a query Query(i) for his desired bit x_{i-1}, where $1 \leq i \leq n$, in the following manner. First, $\mathcal{U}$ computes the t-adic expansion of i. Let $i = \alpha_0$. Then the t-adic expansion of

i is $\beta_l\beta_{l-1}\cdots\beta_2\beta_1$, where

$$\begin{aligned}
\alpha_0 &= \alpha_1 t + \beta_1 \quad 0 \le \alpha_0 \le t^{l-1} - 1, \text{ and } 0 \le \beta_1 \le t-1 \\
\alpha_1 &= \alpha_2 t + \beta_2 \quad 0 \le \alpha_1 \le t^{l-2} - 1, \text{ and } 0 \le \beta_2 \le t-1 \\
\alpha_2 &= \alpha_3 t + \beta_3 \quad 0 \le \alpha_2 \le t^{l-3} - 1, \text{ and } 0 \le \beta_3 \le t-1 \\
&\cdots\cdots \\
\alpha_{l-2} &= \alpha_{l-1} t + \beta_{l-1} \quad 0 \le \alpha_{l-2} \le t-1, \text{ and } 0 \le \beta_{l-1} \le t-1 \\
&\qquad 0 \le \alpha_{l-1} = \beta_l \le t-1 \text{ and } \alpha_l = 0 \ .
\end{aligned} \tag{3.1}$$

For each u $(1 \le u \le l)$, $\mathcal{U}$ chooses uniformly and randomly $t-1$ elements $g_{(u,0)}, g_{(u,1)}, \ldots, g_{(u,\beta_u-1)}, g_{(u,\beta_u+1)}, \ldots, g_{(u,t-1)}$ from H. He also chooses uniformly and randomly $g_{(u,\beta_u)}$ from $S = G \setminus H$. $\mathcal{U}$ defines $Q(u)$ by

$$(g_{(u,0)}, g_{(u,1)}, \cdots, g_{(u,\beta_u-1)}, g_{(u,\beta_u)}, g_{(u,\beta_u+1)}, \cdots, g_{(u,t-1)}) \ , \tag{3.2}$$

that is, $Q(u)$ is a sequence of group elements of G such that the β_uth component is uniformly and randomly chosen from $S = G \setminus H$ and the others are uniformly and randomly chosen from H. Then, $Q(1), Q(2), \ldots, Q(l)$ comprise a query (denoted by Query(i)) for the ith bit x_{i-1} of X, and $\mathcal{U}$ sends Query(i) to $\mathcal{DB}$. Since each $Q(u)$ consists of t group elements from G, $Q(u)$ is represented by $k \times t$ bits. Thus, Query(i) consists of $k \times t \times l$ bits.

Step 2 Receiving Query(i), $\mathcal{DB}$ constructs child databases recursively from the original database X. We regard X as the $t^{l-1} \times t$ binary matrix

$$D(0,\lambda) = \begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_{t-1} \\ x_t & x_{t+1} & x_{t+2} & \cdots & x_{2t-1} \\ & & \cdots & & \\ x_{t^l-t} & x_{t^l-t+1} & \cdots & \cdots & x_{t^l-1} \end{pmatrix},$$

where λ denotes the empty sequence in $\{0,1,2,\ldots,k-1\}^*$. We note that the target bit x_{i-1} is the (α_1, β_1) entry of $D(0,\lambda)$ (α_1 and β_1 are obtained in (3.1)). Denote it by Target$(D(0,\lambda))$.

We recursively define child databases $D(u,s)$, where $1 \le u \le l$ and $s \in \{0,1,2,\ldots,k-1\}^u$. Suppose that we have defined the databases $D(u,s)$ and their target bits Target$(D(u,s))$ and $s \in \{0,1,2,\ldots,k-1\}^u$ for $0 \le u < l-1$. Then we define the databases $D(u+1,s0)$, $D(u+1,s1)$, $\ldots$, $D(u+1,s(k-1))$.

The database $D(u,s)$ is a binary sequence of length t^{l-u}. We regard $D(u,s)$ as a $t^{l-u-1} \times t$ binary matrix. Suppose that

$$D(u,s) = \begin{pmatrix} y_0 & y_1 & y_2 & \cdots & y_{t-1} \\ y_t & y_{t+1} & y_{t+2} & \cdots & y_{2t-1} \\ & & \cdots & & \\ y_{t^{l-u}-t} & y_{t^{l-u}-t+1} & \cdots & \cdots & y_{t^{l-u}-1} \end{pmatrix} .$$

We now construct k child databases, $D(u+1,s0)$, $D(u+1,s1)$, $\ldots$, $D(u+1,s(k-1))$.

Recall that $Q(u)$ consists of t group elements $g_{(u,0)}, g_{(u,1)}, \cdots, g_{(u,t-1)}$ in G (defined in (3.2)). We define a group element g_v for each row $v = 0,1,2,\ldots,t^{l-u-1}-1$ as follows. We set

$$f_{(v,w)} = \begin{cases} g_{(u,w)} & \text{if } D(u,s)(v,w) = 1 \\ 1 & \text{if } D(u,s)(v,w) = 0 \ , \end{cases} \tag{3.3}$$

where $D(u,s)(v,w)$ denotes the (v,w) entry of $D(u,s)$. Then we set

$$f_{D(u,s),v} = \prod_{w=0,1,2,\ldots,t-1} f_{(v,w)} \tag{3.4}$$

for each row $v = 0,1,2,\ldots,t^{l-u-1}-1$. We note that the group element $f_{D(u,s),v}$ $(0 \le v \le t^{l-u-1}-1)$ is of size k, and that $f_{D(u,s),v} \in H$ if and only if $D(u,s)(v,\beta_u) = 0$ by Proposition 1.1. The rth child database $D(u+1,sr)$ $(0 \le r \le k-1)$ is defined to be the sequence consisting of $g_0(r), g_1(r), \ldots, g_{t^{l-u-1}-1}(r)$, where $g_v(r)$ denotes the rth bit of the representation of $f_{D(u,s),v}$. Hence, we have the following matrix equation:

$$\begin{pmatrix} f_{D(u,s),0} \\ f_{D(u,s),1} \\ \cdots \\ f_{D(u,s),t^{l-u-1}-1} \end{pmatrix} = \begin{pmatrix} D(u+1,s0) & \cdots & D(u+1,s(k-1)) \end{pmatrix} \tag{3.5}$$

where each $f_{D(u,s),v}$ is a row vector and each $D(u+1,sr)$ is a column vector. Thus, $D(u+1,sr)$ is a binary sequence of length t^{l-u-1}. We regard it as a $t^{l-u-2} \times t$ binary matrix. Then the target bit for it (denoted by $\mathrm{Target}(D(u+1,sr))$) is defined to be the $(\alpha_{u+1},\beta_{u+1})$ entry of $D(u+1,sr)$ for every $r \in \{0,1,\ldots,k-1\}$ (α_{u+1} and β_{u+1} are obtained in (3.1)).

Step 3 In the last stage of constructing child databases, $\mathcal{DB}$ obtains k^{t-1} databases $D(l-1,s)$ $(s \in \{1,2,\ldots,k\}^{t-1})$. Note that each $D(l-1,s)$ contains t bits. We regard $D(l-1,s)$ as a $1 \times t$ matrix. For each $D(l-1,s)$, we define a group element $A(s)$ as follows. First, we define

$$f_{(0,w)} = \begin{cases} g_{(u,w)} & \text{if } D(l-1,s)(0,w) = 1 \\ 1 & \text{if } D(l-1,s)(0,w) = 0 \end{cases}.$$

Then, we set $f_{D(l-1,s),0} = \prod_{w=0,1,2,\ldots,t-1} f_{(0,w)} = A(s)$. The group element $A(s)$ is of size k for every $s \in \{0,1,2,\ldots,k-1\}^{t-1}$. Then the group elements $A(s)$ $(s \in \{0,1,\ldots,k-1\}^{t-1})$ form the answer $\mathrm{Answer}(\mathrm{Query}(i))$ to the query $\mathrm{Query}(i)$, and $\mathcal{DB}$ sends $\mathrm{Answer}(\mathrm{Query}(i))$ to $\mathcal{U}$.

Step 4 $\mathcal{U}$ receives $\mathrm{Answer}(\mathrm{Query}(i))$ consisting of $A(s)$, where $s \in \{o,1,\ldots,k-1\}^{t-1}$. $\mathcal{U}$ can retrieve the target bit $x_i = \mathrm{Target}(D_{(0,\lambda)})$ in polynomial time in k,n. In fact, the following holds in general.

Theorem 3.1 *For every database $D_{(u,s)}$, where $0 \le u \le l-2$ and $s \in \{1,2,\ldots,k\}^u$, $\mathcal{U}$ can compute* $\mathrm{Target}(D_{(u,s)})$ *in polynomial time in n,k if* $\mathrm{Target}(D_{(u+1,s0)})$, $\mathrm{Target}(D_{(u+1,s1)})$, $\ldots$, $\mathrm{Target}(D_{(u+1,s(k-1))})$ *are given.* □

3.3 Privacy

In the proposed scheme, the query $\mathrm{Query}(i)$ consists of $Q(1)$, $Q(2)$, ..., $Q(l)$, and each $Q(u)$ consists of

$$(g_{(u,0)}, g_{(u,1)}, \cdots, g_{(u,\beta_u-1)}, g_{(u,\beta_u)}, g_{(u,\beta_u+1)}, \cdots, g_{(u,t-1)}),$$

where one of the components is chosen uniformly and randomly from $S = G \setminus H$ and the others are chosen uniformly and randomly from H. The privacy is assured by the inequality

$$|\mathbf{Prob}(C_k(\mathrm{Query}(i)) = 1) - \mathbf{Prob}(C_k(\mathrm{Query}(j)) = 1)| < \sigma,$$

where $\sigma = \frac{1}{(\mathrm{Max}(k,n))^c}$, given in (2.1). This is exactly the general pattern indistinguishability assumption in (1.4) if n is bounded by a polynomial in k. Hence, the privacy of the proposed scheme is guaranteed by the subgroup membership assumption by Theorem 1.2.

3.4 Communication Complexity

In the first step, $\mathcal{U}$ sends Query(i) = $(Q(1), Q(2), \ldots, Q(l))$. Each $Q(u)$ consists of t group elements in G. Since every element in G is represented by a binary sequence of length k, the total bits sent in this stage is $l \times t \times k$. In the second step, $\mathcal{DB}$ sends Answer(Query(i)) consisting of k^{l-1} group elements in G. Therefore, the total bits sent in this stage is $k^{l-1} \times k = k^l$. Consequently, the communication complexity is $ltk + k^l = ln^{\frac{1}{l}}k + k^l$. Suppose that $k = n^c$ and $l = O(\frac{\log n}{\log k})$. Then we have $l = \sqrt{\frac{\log n}{\log k}}$, and $k^l = (2^{\log k})^l = 2^{l \log k} = 2^{\sqrt{\log n \log k}} = 2^{\sqrt{\log nc \log n}} = n^{\sqrt{c}}$. On the other hand, we have $ltk + k^l = k^l(lk + 1) < k^l k^l = (k^l)^2$. Hence, we have $ltk + k^l = (n^{\sqrt{c}})^2$. It follows that the communication complexity is $O(n^c)$.

References

[1] Cachin,C., Micali,S., Stadler,M.: Computationally Private Information Retrieval with Polylogarithmic Communication, Advances in Cryptology. Lecture Notes in Computer Science, Vol. 1592. Springer-Verlag, (1999) 402–414

[2] Chor,B., Goldreich,O., Kushilevitz,E., Sudan,M.: Private Information Retrieval, IEEE Symposium on Foundations of Computer Science. (1995) 41–50

[3] Chor,B., Gilboa,N.: Computationally Private Information Retrieval ACM Symposium on Theory of Computing. (1997) 304–313

[4] Goldwasser,S., Micali,S.: Probabilistic Encryption, J. Computer and System Science **28** (1984) 270–299

[5] Kushilevitz,E., Ostrovsky,R.: Replication Is not Needed: Single Database, Computationally-private Information Retrieval, IEEE Symposium on Foundations of Computer Science. (1997) 364–373

[6] Kurosawa,K, Tsujii,S.: A General Method to Construct Public Key Residue Cryptosystems, Transactions of the IEICE **E-73**, (1990) 1068–1072

[7] Naccache,D., Stern,J.: A New Public-key Cryptosystem, Advances in Cryptology. Lecture Notes in Computer Science, Vol. 1233. Springer-Verlag, (1997) 27–36

[8] Okamoto,T., Uchiyama,S.: A New Public-key Cryptosystem as Secure as Factoring, Advances in Cryptology. Lecture Notes in Computer Science, Vol. 1403. Springer-Verlag, (1998) 308–318

[9] Paillier,P.: Public-key Cryptosystems Based on Composite Degree Residuosity Classes, Advances in Cryptology. Lecture Notes in Computer Science, Vol. 1592. Springer-Verlag, (1999) 223–238

[10] Saito,T., Koshiba,T., Yamamura,A.: The Decision Diffie-Hellman assumption and the Quadratic Residuosity Assumption, IEICE Transactions on Fundamentals of Electronics (1) **E84-A**, (2001) 165–171

[11] Yamamura,A., Saito,T.: Private Information Retrieval Based on the Subgroup Membership Problem, Information Security and Privacy, Lecture Notes in Computer Science, Springer-Verlag, (2001) (to appear)