Symmetric Subgroup Membership Problems

Kristian Gjøsteen

Department of Matematical Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway, kristian.gjosteen@math.ntnu.no

Abstract. We define and discuss symmetric subgroup membership problems and their properties, including a relation to the Decision Diffie-Hellman problem. We modify the Cramer-Shoup framework, so that we can derive a chosen ciphertext secure cryptosystem in the standard model from symmetric subgroup membership problems. We also discuss how chosen ciphertext secure hybrid cryptosystems based on a symmetric subgroup membership can be constructed in the standard model, giving a very efficient cryptosystem whose security relies solely on the symmetric subgroup membership problem.

Key words: public key encryption, hybrid encryption, standard model, subgroup membership problem.

1 Introduction

Public key cryptography was first proposed by Diffie and Hellman [5]. The most general security notion for public key cryptosystems is security against adaptive chosen ciphertext attacks (CCA) [10]. While many efficient schemes achieve this in the random oracle model, Cramer and Shoup [2, 4] designed the first efficient scheme to achieve this security level in the standard model.

The security proofs for many public key cryptosystems essentially rely on subgroup membership problems. The most famous subgroup membership problem is the Decision Diffie-Hellman problem [1], on which the Cramer-Shoup cryptosystem relies. Yamamura and Saito [11] catalogued many subgroup membership problems that have appeared in the literature. Cramer and Shoup [3] gave a framework for turning general subgroup membership problems into secure cryptosystems, generalising their previous work and giving several new instances with interesting properties.

We study *symmetric subgroup membership problems* (Sect. 2), and show how they relate to the Decision Diffie-Hellman problem (Sect. 3). We also extend the framework of Cramer and Shoup to make efficient use of symmetric subgroup membership problems, giving very efficient cryptosystems secure against chosen ciphertext attacks in the standard model (Sect. 4). Finally, we discuss new developments in hybrid encryptions (Sect. 5) and construct a very efficient cryptosystem provably chosen ciphertext secure in the standard model, relying solely on the symmetric subgroup membership problem.

1.1 Notation

If S is a non-empty finite set, we denote by $\mathbb{N}_S$ the set $\{0, \ldots, |S| - 1\}$.

Let X be a distribution on a set S. We denote by $x \leftarrow X$ the act of sampling x from S according to the distribution X. The notation $x \leftarrow S$ is used to denote sampling x from S according to the uniform distribution. We denote by $x \leftarrow s$ the assignment of the value s to x.

We use the following notation to describe new distributions. Let $X_1, \ldots, X_n$ be distributions on sets $S_1, \ldots, S_n$, and let $f : S_1 \times \cdots \times S_n \to S$ be a function. Then by

$$X = \{f(x_1, \ldots, x_n) \mid x_1 \leftarrow X_1, \ldots, x_n \leftarrow X_n\}$$

we denote the distribution on S defined by

$$\Pr[x = s \mid x \leftarrow X] = \Pr[f(x_1, \ldots, x_n) = s \mid x_1 \leftarrow X_1, \ldots, x_n \leftarrow X_n] \; .$$

The distance between two distributions X and Y on a set S is

$$\mathrm{Dist}(X, Y) = \frac{1}{2} \sum_{s \in S} |\Pr[X = s] - \Pr[Y = s]| \; .$$

We say that two distributions X and Y are δ-close if $\mathrm{Dist}(X, Y) \leq \delta$.

2 Symmetric Subgroup Membership Problem

A subgroup membership problem consists of a finite abelian group G along with a proper, non-trivial subgroup K. The problem is to decide if a group element $x \in G$ is in K or in $G \setminus K$. We denote this subgroup membership problem by $\mathcal{SM}_{(G,K)}$, and the advantage of an adversary A is

$$\mathrm{Adv}_A^{\mathcal{SM}_{(G,K)}} = |\Pr[A(G, K, x) = 1 \mid x \leftarrow K] - \Pr[A(G, K, x) = 1 \mid x \leftarrow G \setminus K]| \; .$$

Let G be a finite abelian group, and let K and H be subgroups of G such that $K \cap H = \{1\}$ and $G = KH$. Then $K \times H \simeq G$, and the isomorphism is simply the group operation: $(k, h) \mapsto kh$. If $\gcd(|K|, |H|) = 1$, then if $d \equiv |H|^{-1} \pmod{|K|}$, we get that $c \mapsto (c^{|H|d}, c^{1-|H|d})$ is the inverse map. So anyone who knows $|K|$ and $|H|$ can compute the reverse isomorphism.

The *symmetric subgroup membership problem* $\mathcal{SSM}_{(G,K,H)}$ is the subgroup membership problem $\mathcal{SM}_{(G \times G, K \times H)}$. It is easy to show that distinguishing $K \times H$ is equivalent to distinguishing either K or H or both, and that considering maximum advantages for algorithms using some fixed amount of resources, we get

$$\mathrm{Adv}^{\mathcal{SM}_{(G,K)}} - \frac{|K| - 1}{|G| - 1} \leq \mathrm{Adv}^{\mathcal{SSM}_{(G,K,H)}} \leq \mathrm{Adv}^{\mathcal{SM}_{(G,K)}} + \mathrm{Adv}^{\mathcal{SM}_{(G,H)}} \; .$$

We shall assume that there are efficient algorithms available for sampling the subgroups K and H from a distribution that is δ-close to the uniform distribution, for some negligible $\delta \geq 0$. Typically, these algorithms simply choose a random exponent and exponentiate a generator for the subgroup. If δ cannot be zero, it is always easy to make δ arbitrarily small.

We describe two instances of the symmetric subgroup membership problem.

Let $n = pq$ be an RSA modulus, and let G be a group of order n. Let K be the subgroup of order p and let H be the subgroup of order q. Then we have a symmetric subgroup membership problem $\mathcal{SSM}_{(G,K,H)}$.

If $p' = 2n + 1$ is prime, the set of quadratic residues in $GF(p')^*$ is exactly such a group structure, and it seems plausible that it gives a hard symmetric subgroup membership problem. It was discussed in [8] and [9]. We could also consider $p' = 2sn + 1$ for some small integer s, with little additional complexity.

As an alternative, let a, b, c, d, $p = 2ab + 1$ and $q = 2cd + 1$ be primes, let $n = pq$, and let G be the subgroup of $\mathbb{Z}_n^*$ with Jacobi symbol 1. Let K be the subgroup of order $2ac$ and H be the subgroup of order bd. It is plausible that the resulting symmetric subgroup membership problem is hard. Note also that ac can be made much smaller than bd.

To see how this group structure can be used, we briefly describe a key encapsulation method (KEM) [4], and show that it is secure against passive attacks if and only if the symmetric subgroup membership problem is hard.

The key generation algorithm simply selects a suitable symmetric subgroup membership problem $\mathcal{SSM}_{(G,K,H)}$, and outputs a public key (G, K, H). The private key is $(G, |K|, |H|)$.

To sample a symmetric key and encipher it, $(x, y) \in G \times G$ is sampled (almost) uniformly at random from $K \times H$, using the sampling algorithms for $\mathcal{SSM}_{(G,K,H)}$. The key is (x, y) and the ciphertext is the product xy.

To decipher $c \in G$, the knowledge of $|K|$ and $|H|$ is used to compute $(x, y) \in K \times H$ such that $c = xy$, as described above.

It is clear that distinguishing the decryption (x, y) of a ciphertext c from a random pair $(x', y') \in G \times G$ such that $x'y' = c$ is equivalent to deciding the symmetric subgroup membership $\mathcal{SSM}_{(G,K,H)}$.

To discuss the efficiency of the above KEM, we shall compare it with three other schemes. The first is the cryptosystem proposed in [9] (NBD), the second is Diffie-Hellman in G (DH/G), and the third is Diffie-Hellman in the subgroup K (DH/K).

It was shown in [9] that NBD is secure if the symmetric subgroup membership problem is hard. Sect. 3 will show that Diffie-Hellman in G is not less secure than the above KEM. Sect. 4 will show that Diffie-Hellman in K can be turned into a cryptosystem with messages in G that is secure if the symmetric subgroup membership problem is hard.

DH/G requires two exponentiations in G to encrypt, and one to decrypt. DH/K requires two exponentiations in K to encrypt, and one to decrypt. NBD requires one exponentiation in K and one in H to encrypt, and approximately 1.3 exponentiations in G to decrypt. Our KEM requires essentially one exponentiation in K and one in H, both to encrypt and decrypt.

As we can see, Diffie-Hellman in K is the best option, especially if exponentiations in K can be made cheaper than exponentiations in H.

3 The Decision Diffie-Hellman problem

We keep the notation introduced in Sect. 2. Let x be a generator for G. The *Decision Diffie-Hellman* (DDH) problem is to distinguish the two distributions $\{(x, x^u, x^v, x^{uv}) \mid u, v \leftarrow \mathbb{N}_G\}$ and $\{(x, x^u, x^v, x^w) \mid u, v, w \leftarrow \mathbb{N}_G\}$. Some definitions require $w \not\equiv uv \pmod{|G|}$, but the difference is negligible. The advantage of an algorithm A taking four group elements as input and answering 0 or 1 against DDH is defined to be

$$\mathrm{Adv}_A^{\mathcal{DDH}_G} = \frac{1}{2}|\Pr[A(x, x^u, x^v, x^{uv}) \mid u, v \leftarrow \mathbb{N}_G] - \Pr[A(x, x^u, x^v, x^w) \mid u, v, w \leftarrow \mathbb{N}_G]| \ .$$

We shall need the following result later on, so we state it as a separate lemma.

Lemma 1. *Let G be a finite cyclic group, and let K and H be non-trivial subgroup of G such that $K \cap H = \{1\}$ and $G = KH$. Let g be a generator for K. Consider the two distributions given by $U = \{(g^u, y, y^u) \mid u \leftarrow \mathbb{N}_G, y \leftarrow G \setminus K\}$ and $V = \{(g^u, y, y^u z \mid u \leftarrow \mathbb{N}_G, y \leftarrow G \setminus K, z \leftarrow H\}$. Then*

$$\mathrm{Dist}(U, V) \leq \frac{|H| - \phi(|H|)}{|H|} \ .$$

Proof. Let $u_1 = u \bmod |K|$ and $u_2 = u \bmod |H|$, and let $y = y_1 y_2$ with $y_1 \in K$, $y_2 \in H$. It is easy to see that

$$U = \{(g^{u_1}, y_1^{u_1} y_2^{u_2}) \mid u_1 \leftarrow \mathbb{N}_K, u_2 \leftarrow \mathbb{N}_H, y_1 \leftarrow K, y_2 \leftarrow H \setminus \{1\}\}$$

and

$$V = \{(g^{u_1}, y_1^{u_1} y_2^{u_2} z) \mid u_1 \leftarrow \mathbb{N}_K, u_2 \leftarrow \mathbb{N}_H, y_1 \leftarrow K, y_2 \leftarrow H \setminus \{1\}, z \leftarrow H\} \ .$$

With $U' = \{y_2^{u_2} \mid u_2 \leftarrow \mathbb{N}_H, y_2 \leftarrow H \setminus \{1\}\}$ and $V' = \{y_2^{u_2} z \mid u_2 \leftarrow \mathbb{N}_H, y_2 \leftarrow H \setminus \{1\}, z \leftarrow H\}$, it is clear that

$$\mathrm{Dist}(U, V) = \mathrm{Dist}(U', V')$$

and that V' is exactly the uniform distribution on H. If y_2 is a generator, then U' is also uniformly distributed on H. It follows that

$$\mathrm{Dist}(U', V') \leq \frac{|H| - \phi(|H|)}{|H|},$$

which concludes the proof. □

Theorem 1. *Let $\mathcal{SSM}_{(G,K,H)}$ be a symmetric subgroup membership problem such that G is cyclic, and suppose that the sampling algorithms for K and H are δ-close to uniform. Let A be an algorithm that decides the Decision Diffie-Hellman problem in G. Then for any $\delta' > 0$ there are algorithms A_1, A_2 and A_3*

that use A once as an oracle and otherwise do $O(\log 1/\delta')$ exponentiations in G, such that

$$\begin{aligned}\mathrm{Adv}_A^{\mathcal{DDH}_G} \le{} & \mathrm{Adv}_{A_1}^{\mathcal{SM}_{(G,K)}} + \mathrm{Adv}_{A_2}^{\mathcal{SM}_{(G,K)}} + \mathrm{Adv}_{A_3}^{\mathcal{SM}_{(G,H)}} + \\ & \frac{|G|-\phi(|G|)}{|G|-|K|} + \frac{|K|-\phi(|K|)}{|K|} + \frac{|H|-\phi(|H|)}{|H|} + \frac{|G|-\phi(|G|)}{|G|-|H|} + \\ & 7\delta' + 4\delta \ .\end{aligned}$$

Proof. We shall need the following three experiments.

Experiment 1.
Input: A, G, $x \in G$

1. $u, v, w \leftarrow \mathbb{N}_G$.
2. $y \leftarrow x^v$.
3. $b \leftarrow \{0,1\}$.
4. If $b = 1$, then $z \leftarrow y^u$, otherwise $z \leftarrow x^w$.
5. $b' \leftarrow A(x, x^u, y, z)$.
6. If $b = b'$, output 1, otherwise output 0.

Output: 0 or 1.

Experiment 2.
Input: A, G, $y \in G$

1. $u, v \leftarrow \mathbb{N}_G$.
2. $x \leftarrow K$.
3. $b \leftarrow \{0,1\}$.
4. If $b = 1$, then $z \leftarrow y^u$, otherwise $z \leftarrow y^w$.
5. $b' \leftarrow A(x, x^u, y, z)$.
6. If $b = b'$, output 1, otherwise output 0.

Output: 0 or 1.

Experiment 3.
Input: A, G, $h \in G$

1. $u, v \leftarrow \mathbb{N}_G$.
2. $x \leftarrow K$, $y \leftarrow G \setminus K$.
3. $b \leftarrow \{0,1\}$.
4. If $b = 1$, then $z \leftarrow y^u h$, otherwise $z \leftarrow y^w$.
5. $b' \leftarrow A(x, x^u, y, z)$.
6. If $b = b'$, output 1, otherwise output 0.

Output: 0 or 1.

In each experiment, Step 1 and 2 requires sampling certain elements from certain uniform distributions. It may be impossible to implement these steps, but we can implement approximations.

For Step 1, we note that the numbers sampled are used as exponents. Therefore, we can sample uniformly from a larger set to get an element distribution close to uniform. The cost is exponentiating to the larger exponent, but it is easy to show that for any $\delta' > 0$, $O(1/\log \delta')$ extra work suffices for a δ'-close to uniform distribution.

For Step 2, we simply use the algorithms provided by the subgroup membership problem, which are δ-close to uniform.

Consider first Experiment 1. If the input x is a generator for G, then this experiment measures the advantage of A against DDH. Let T_1 denote the event that the experiment outputs 1 when the input x is sampled from $G \setminus K$. An easy computation shows that

$$\mathrm{Adv}_A^{\mathcal{DDH}_G} \le |\Pr[T_1] - 1/2| + \frac{|G|-\phi(|G|)}{|G|-|K|} \ . \tag{1}$$

Let T_1' denote the event that the Experiment 1 outputs 1 when the input x is sampled from K. By the comments above, we can use Experiment 1 to construct a distinguisher A_1 for K, and

$$|\Pr[T_1] - \Pr[T_1']| \le \mathrm{Adv}_{A_1}^{\mathcal{SM}_{(G,K)}} + 3\delta' \ . \tag{2}$$

Next, we consider Experiment 2. Let T_2' be the event that Experiment 2 outputs 1 when the input y is sampled from K. Suppose the input x to Experiment 1 and y to Experiment 2 are sampled uniformly from K. In either case, if the x sampled generates K, the two experiments proceed identically. In other words,

$$|\Pr[T_1'] - \Pr[T_2']| \leq \frac{|K| - \phi(|K|)}{|K|} . \tag{3}$$

Let T_2 be the event that Experiment 2 outputs 1 when the input y is sampled from $G \setminus K$. As above, we can use Experiment 2 to construct a distinguisher A_2 for K, and

$$|\Pr[T_2] - \Pr[T_2']| \leq \mathrm{Adv}_{A_2}^{\mathcal{SM}(G,K)} + 2\delta' + \delta . \tag{4}$$

Then we consider Experiment 3. Let T_3' be the event that the experiment outputs 1 when the input h is sampled from H. When the input y to Experiment 2 is sampled from $G \setminus K$ and the input h to Experiment 3 is sampled from H, Lemma 1 shows that

$$|\Pr[T_2] - \Pr[T_3']| \leq \frac{|H| - \phi(|H|)}{|H|} . \tag{5}$$

Let T_3 be the event that the experiment outputs 1 when the input h is sampled from $G \setminus H$. As above, we can use Experiment 3 to construct a distinguisher A_3 for H, and

$$|\Pr[T_3] - \Pr[T_3']| \leq \mathrm{Adv}_{A_3}^{\mathcal{SM}(G,H)} + 2\delta' + 3\delta . \tag{6}$$

To conclude, we need only observe that in Experiment 3, when the input h is sampled from $G \setminus H$ and y is a generator, the distribution of z is independent of b, and therefore

$$|\Pr[T_3] - 1/2| \leq \frac{|G| - \phi(|G|)}{|G| - |H|} . \tag{7}$$

Combining equations (1)–(7) proves the theorem. □

4 Chosen ciphertext security

4.1 Hash proof systems

We give a brief presentation of hash proof systems. It is only superficially different from [3], so we refer the reader there for further details.

Let G be a set, and let K be a subset of G. We say that W is a *witness set* for K if there is an easily computable bijection $\rho : W \to K$. This bijection allows one to prove that an element $x \in G$ really is in K by presenting an element $w \in W$ such that $\rho(w) = x$. This obviously assumes that it is easy to recognise elements of W.

For two sets S, S', denote by $\mathrm{Map}(S, S')$ the set of maps from S to S'. Let L be a group. We are interested in looking at maps from G to L. There is a natural map $\mathrm{Map}(G, L) \to \mathrm{Map}(K, L)$ given by restriction. From ρ we get a bijection $\rho^* : \mathrm{Map}(K, L) \to \mathrm{Map}(W, L)$. We also denote the natural map $\mathrm{Map}(G, L) \to \mathrm{Map}(W, L)$ by ρ^*.

A *projective hash family* is a tuple (G, K, L, W, ρ, M), where G is a set, K is a subset of G, L is a group, W is a witness set for K with isomorphism ρ, M is a subset of $\mathrm{Map}(G, L)$, and for any $f \in M$, the image of K under f is a subgroup of L. We also suppose that L has a subgroup L', such that $L' \cap f(K) = \{1\}$ and $L = L'f(K)$. This gives us a subgroup membership problem $\mathcal{SM}_{(L,L')}$. (This corresponds to the definition sketched in Section 8.2.4 of [3].)

Let (G, K, L, W, ρ, M) be a projective hash family. The projective hash family is *ϵ-universal* if for any $f' \in \rho^*(M)$, $x \in G \setminus K$ and $y \in L$, we have that

$$\Pr[f(x) = y \wedge \rho^*(f) = f' | f \leftarrow M] \leq \epsilon \Pr[\rho^*(f) = f' | f \leftarrow M] \ .$$

The projective hash family is *ϵ-universal-2* if for any $f' \in \rho^*(M)$, $x_0 \in G \setminus K$, $x \in G \setminus (K \cup \{x_0\})$ and $y, y_0 \in L$, we have that

$$\begin{aligned}\Pr[f(x) = y \wedge f(x_0) = y_0 \wedge \rho^*(f) = f' | f \leftarrow M] \\ \leq \epsilon \Pr[f(x_0) = y_0 \wedge \rho^*(f) = f' | f \leftarrow M] \ .\end{aligned}$$

It is clear that ϵ-universal follows from ϵ-universal-2.

Let (G, K, L, W, ρ, M) be a projective hash family. Define the two distributions

$$\begin{aligned}U &= \{(x, \rho^*(f), f(x)) \mid x \leftarrow G \setminus K, f \leftarrow M\}, \\ V &= \{(x, \rho^*(f), f(x)y) \mid x \leftarrow G \setminus K, f \leftarrow M, y \leftarrow L'\} \ .\end{aligned}$$

We say that the projective hash family is *ϵ-smooth* if

$$\mathrm{Dist}(U, V) \leq \epsilon \ .$$

A *hash proof system* Π for a subgroup membership problem $\mathcal{SM}_{(G,K)}$ is a projective hash family (G, K, L, W, ρ, M), along with efficient algorithms for sampling W and M δ'-close to uniform, and for evaluating the hash functions on points in G and W.

An *extended hash proof system* $\hat{\Pi}$ for $\mathcal{SM}_{(G,K)}$ is a projective hash family $(G \times S, K \times S, \hat{L}, W, \hat{\rho}, \hat{M})$, where S is some set depending on G, along with efficient algorithms for sampling W and $\hat{M}$ δ'-close to uniform, and for evaluating the hash functions on points in $G \times S$ and $W \times S$.

A (extended) hash proof system Π ($\hat{\Pi}$) is ϵ-smooth (ϵ-universal-2) if the projective hash family is ϵ-smooth (ϵ-universal-2).

Let $\mathcal{SSM}_{(G,K,H)}$ be a symmetric subgroup membership problem such that G is cyclic, and suppose that a generator g is available for K. We shall describe a hash proof system Π and an extended hash proof system $\hat{\Pi}$ for $\mathcal{SSM}_{(G,K,H)}$. The group L will be G, and $L' = H$.

Let $W = \mathbb{Z}_{|K|}$ and $\rho([w]) = g^w$. Let $L = G$ and let $L' = H$. Since G is cyclic, the homomorphism group $\mathrm{Hom}(G, G)$ is isomorphic to $\mathbb{Z}_{|G|}$, and we let

$M = \mathrm{Hom}(G, G)$. For any $f \in M$, a useful description of the function $\rho^*(f)$ is the group element $f(g)$, since for any $[w] \in W$, $f(g^w) = f(g)^w$. The projective hash family is $(G, K, G, \mathbb{Z}_{|K|}, \rho, \mathrm{Hom}(G, G))$, with the obvious sampling and evaluation algorithms.

By Lemma 1, this hash proof system is ϵ-smooth, for $\epsilon = 1 - \phi(|H|)/|H|$.

The extended hash proof system $\hat{\Pi}$ is slightly more complicated. Let ℓ be the smallest prime dividing $|H|$. We shall suppose that for some sufficiently large l, a 1-1 function $h : G \times G \to \{0, \dots, \ell - 1\}^l$ is available. Then $\hat{M}$ is the set of functions of the form

$$\hat{f}(x, e) = f_0(x) \prod_{i=1}^{l} f_i(x)^{\gamma_i},$$

where $h(x, e) = (\gamma_1, \dots, \gamma_l)$, and $f_i \in \mathrm{Hom}(G, G)$.

The witness set for $K \times G$ is $\mathbb{Z}_{|K|} \times G$, and the map $\hat{\rho}$ is given by $\hat{\rho}([w], e) = (g^w, e)$, where g is a generator for K. It is clear that

$$\hat{\rho}^*(\hat{f})([w], e) = f_0(g)^w \prod_{i=1}^{l} f_i(g)^{w\gamma_i},$$

where $h(g^w, e) = (\gamma_1, \dots, \gamma_l)$. So a useful description of the function $\hat{\rho}^*(\hat{f})$ is the tuple $(s_0, s_1, \dots, s_l) = (f_0(g), f_1(g), \dots, f_l(g))$.

By Theorem 3 of [3], the extended hash proof system $\hat{\Pi}$ described above is $1/\ell$-universal-2. Just as in [4], it is possible to replace the 1-1 function h with a collision resistant hash function, to get a computationally secure construction with $l = 1$.

4.2 The cryptosystem

The standard goal for a public key cryptosystem is indistinguishability of ciphertexts against a adaptive chosen ciphertext adversary. We consider adversaries A consisting of a pair of algorithms (A_1, A_2), where A_1 receives the public key and outputs a pair of messages (m_0, m_1). A_2 then receives an encryption of one of the messages and must decide which one. Both A_1 and A_2 are allowed to have arbitrary ciphertexts decrypted (the challenge ciphertext excepted, obviously). If T is the event that A decides correctly, we say that A wins the game, and its advantage is defined to be

$$\mathrm{Adv}_A^{CCA} = |\Pr[T] - 1/2| \ .$$

Suppose we have a subgroup membership problem $\mathcal{SM}_{(G,K)}$, a hash proof system Π for $\mathcal{SM}_{(G,K)}$, and an extended hash proof system $\hat{\Pi}$ for $\mathcal{SM}_{(G,K)}$ such that the projective hash families are (G, K, L, W, ρ, M) and $(G \times L, K \times L, \hat{L}, W, \hat{\rho}, \hat{M})$, respectively.

We derive the cryptosystem CS' described in Fig. 1 from the two hash proof systems. Note that M, $\hat{M}$ and W are sampled using the algorithms from the hash proof systems.

Key generation.
Input: $\mathcal{SM}_{(G,K)}$, Π, $\hat{\Pi}$.

1. $f \leftarrow M$, $\hat{f} \leftarrow \hat{M}$.
2. $sk \leftarrow (G, L, \hat{L}, f, \hat{f})$.
3. $pk \leftarrow (G, W, L, \hat{L}, \rho, \rho^*(f), \hat{\rho}^*(\hat{f}))$.

Output: (pk, sk).

Encryption.
Input: pk, $m \in L$.

1. $w \leftarrow W$.
2. $x \leftarrow \rho(w)$.
3. $y \leftarrow \rho^*(f)(w)$.
4. $e \leftarrow ym$.
5. $\hat{y} \leftarrow \rho^*(\hat{f})(w, e)$.

Output: $(x, e, \hat{y})$.

Decryption.
Input: sk, $(x, e, \hat{y})$.

1. $\hat{y}' \leftarrow \hat{f}(x, e)$.
2. If $\hat{y}' \neq \hat{y}$, output $\perp$.
3. $y \leftarrow f(x)$.
4. $m \leftarrow ey^{-1}$.
5. Output m.

Output: A message m or $\perp$.

Fig. 1. The cryptosystem CS'

Key generation.
Input: $\mathcal{SSM}_{(G,K,H)}$, $g \in K$.

1. $(k, k_0, k_1, \ldots, k_l) \leftarrow \{0, \ldots, |G| - 1\}^{l+2}$.
2. $(s, s_0, s_1, \ldots, s_l) \leftarrow (g^k, g^{k_0}, g^{k_1}, \ldots, g^{k_l})$.
3. $pk \leftarrow (G, g, s, s_0, s_1, \ldots, s_l, h)$.
4. $sk \leftarrow (G, k, k_0, k_1, \ldots, k_l, h)$.

Output: (pk, sk).

Encryption.
Input: pk, $m \in G$.

1. $w \leftarrow \{0, \ldots, |K| - 1\}$.
2. $x \leftarrow g^w$.
3. $y \leftarrow s^w$.
4. $e \leftarrow ym$.
5. $(\gamma_1, \ldots, \gamma_l) \leftarrow h(x, e)$.
6. $\hat{y} \leftarrow s_0^w \prod_{i=1}^{l} s_i^{w\gamma_i}$.

Output: $(x, e, \hat{y}) \in G \times G \times G$.

Decryption.
Input: sk, $(x, e, \hat{y}) \in G \times G \times G$.

1. $(\gamma_1, \ldots, \gamma_l) \leftarrow h(x, e)$.
2. $\hat{y}' \leftarrow x^{k_0} \prod_{i=1}^{l} x^{k_i \gamma_i}$.
3. If $\hat{y} \neq \hat{y}'$, then output $\perp$.
4. $y \leftarrow x^k$.
5. $m \leftarrow ey^{-1}$.

Output: A message $m \in G$ or $\perp$.

Fig. 2. The cryptosystem CS' instantiated with a symmetric subgroup membership problem $\mathcal{SSM}_{(G,K,H)}$

The security analysis closely follows the analysis in [3].

Suppose that Π is ϵ-smooth, that $\hat{\Pi}$ is ϵ'-universal-2, that the sampling algorithms for Π and $\hat{\Pi}$ are δ'-close to uniform, and that the sampling algorithms for the subgroup membership problems are δ-close to uniform.

Suppose $A = (A_1, A_2)$ is a chosen ciphertext adversary against CS'. We shall use the following experiment to construct a distinguisher A' for (G, K). Again, note that M and $\hat{M}$ are sampled using the algorithms from the hash proof systems.

Experiment 4.
Input: $A = (A_1, A_2)$, (G, K), Π, $\hat{\Pi}$, $x_0 \in G$.

1. $f \leftarrow M$, $\hat{f} \leftarrow \hat{M}$.
2. $sk \leftarrow (G, L, \hat{L}, f, \hat{f})$.
3. $pk \leftarrow (G, W, L, \hat{L}, \rho, \rho^*(f), \hat{\rho}^*(\hat{f}))$.
4. Initialise decryption oracle $\mathcal{D}_{sk}$.
5. $(m_0, m_1, s) \leftarrow A_1(pk)$, giving A_1 access to $\mathcal{D}_{sk}$.
6. $b \leftarrow \{0, 1\}$.
7. $y_0 \leftarrow f(x_0)$, $e_0 \leftarrow y_0 m_b$, $\hat{y}_0 \leftarrow \hat{f}(x_0, e_0)$.
8. Initialise restricted decryption oracle $\mathcal{D}'_{sk}$.
9. $b' \leftarrow A_2(pk, m_0, m_1, s, x_0, e_0, \hat{y}_0)$, giving A_2 access to $\mathcal{D}'_{sk}$.
10. If $b = b'$, output 1, otherwise output 0.

Output: 0 or 1.

Note that Steps 1–3 do exactly as the key generation algorithm would do.

Let T' be the event that Experiment 4 outputs 1 when the input x_0 is in K. Since Step 7 produces exactly the same result as the encryption algorithm when the input $x_0 \in K$, it is clear that the only difference between Experiment 4 and a real attack is that x_0 has been sampled uniformly from K, and not via the sampling algorithm for W. Since Experiment 4 outputs 1 when the adversary wins, we have that

$$\mathrm{Adv}_A^{CCA} \leq |\Pr[T'] - 1/2| + \delta', \tag{8}$$

since the sampling algorithm for W is δ'-close to uniform.

Let T be the event that Experiment 4 outputs 1 when the input x_0 is in $G \setminus K$. It is clear that from Experiment 4 we can derive an algorithm A' for distinguishing K from $G \setminus K$ such that

$$|\Pr[T'] - \Pr[T]| \leq \mathrm{Adv}_{A'}^{\mathcal{SM}_{(G,K)}} . \tag{9}$$

To analyse the event T, we shall make a series of modifications to Experiment 4. We number the modified experiments as $4'$, $4''$, etc. Note that these modifications need not be efficiently implementable.

First modification We change Step 1 so that f and $\hat{f}$ are sampled from the uniform distribution, and not using the algorithms provided by the hash proof systems.

Let T_1 be the event that Experiment $4'$ outputs 1 when the input x_0 is in $G \setminus K$. Since the algorithms provided by the hash proof systems were δ'-close to uniform, we obviously have that

$$|\Pr[T] - \Pr[T_1]| \leq 2\delta' \ . \tag{10}$$

Second modification We change the decryption oracles so that they refuse to decrypt a ciphertext $(x, e, \hat{y})$ if $x \notin K$. Let T_2 be the event that Experiment $4''$ outputs 1 when the input x_0 is in $G \setminus K$.

It is clear that this modification only affects the outcome if the adversary produces a valid ciphertext $(x', e', \hat{y}')$ with $x \notin K$, so $|\Pr[T_2] - \Pr[T_1]|$ is upper-bounded by the probability of this happening.

Since $\hat{\Pi}$ is ϵ'-universal-2, we can show, using the same arguments as in [3], that if A_1 and A_2 make Q decryption queries in total, then

$$|\Pr[T_2] - \Pr[T_1]| \leq Q\epsilon' \ . \tag{11}$$

Third modification We change Step 7 to be

7. $y' \leftarrow L'$, $y_0 \leftarrow f(x_0)$, $e_0 \leftarrow y_0 m_b y'$, $\hat{y}_0 \leftarrow \hat{f}(x_0, e_0)$.

Let T_3 be the event that Experiment $4'''$ outputs 1 when the input x_0 is in $G \setminus K$.

Since A_1 and A_2 cannot query the decryption oracle with ciphertexts $(x, e, \hat{y})$ where $x \notin K$, their only information about f is $\rho^*(f)$. Since Π is ϵ-smooth, we get that

$$|\Pr[T_3] - \Pr[T_2]| \leq \epsilon \ . \tag{12}$$

Fourth modification We change Step 7 to be

7. $y' \leftarrow L \setminus L'$, $y_0 \leftarrow f(x_0)$, $e_0 \leftarrow y_0 m_b y'$, $\hat{y}_0 \leftarrow \hat{f}(x_0, e_0)$.

Let T_4 be the event that Experiment $4''''$ outputs 1 when the input x_0 is in $G \setminus K$.

It is quite clear that if y' had been sampled uniformly from L, then there would be no information about m_b present in the ciphertext, and the probability that Experiment $4''''$ output 1 when the input x_0 was in $G \setminus K$ would be $1/2$. Since Experiment $4''''$ samples from $L \setminus L'$, we get that

$$|\Pr[T_4] - 1/2| \leq \frac{2|L'|}{|L|} \ . \tag{13}$$

We need to bound $|\Pr[T_4] - \Pr[T_3]|$. To do this, we introduce another experiment.

Experiment 5.
Input: $A = (A_1, A_2)$, (G, K), Π, $\hat{\Pi}$, $y' \in L$.

Steps 1–6 are as in Experiment 4.

7. $x_0 \leftarrow G \setminus K$, $y_0 \leftarrow f(x_0)$, $e_0 \leftarrow y_0 m_b y'$, $\hat{y}_0 \leftarrow \hat{f}(x_0, e_0)$.

Steps 8–10 are as in Experiment 4.

Output: 0 or 1.

It is quite clear that we can repeat the two first modifications to Experiment 4 on Experiment 5, and the analysis remains the same. Let R' be the event that Experiment $5''$ outputs 1 when the input y' is in L', and let R be the event that Experiment $5''$ outputs 1 when the input y' is in $L \setminus L'$.

If the input y' to Experiment $5''$ is in L', then it behaves exactly as Experiment $4'''$. Hence, $\Pr[R'] = \Pr[T_3]$.

If the input y' to Experiment $5''$ is in $L \setminus L'$, then it behaves exactly as Experiment $4''''$. Hence, $\Pr[R] = \Pr[T_4]$.

It is clear that we from Experiment 5 can derive an algorithm A'' to distinguish L' from $L \setminus L'$, by sampling x_0 not uniformly from $G \setminus K$, but via the subgroup membership problem's algorithms, and that

$$|\Pr[T_4] - \Pr[T_3]| = |\Pr[R] - \Pr[R']| \leq \mathrm{Adv}_{A''}^{\mathcal{SM}_{(L,L')}} + 2\delta' + \delta + Q\epsilon' \ . \qquad (14)$$

Summing up Combining (8)–(14), we have proved the following theorem.

Theorem 2. *Let CS' be the cryptosystem described above, based on a subgroup membership problem $\mathcal{SM}_{(G,K)}$ and hash proof systems Π and $\hat{\Pi}$. Let L be the group associated to G by Π, and let L' be the subgroup of L. Suppose that Π is ϵ-smooth, that $\hat{\Pi}$ is ϵ'-universal-2, that the sampling algorithms for Π and $\hat{\Pi}$ are δ'-close to uniform, and that the sampling algorithms for the subgroup membership problem are δ-close to uniform. Then for any chosen ciphertext adversary A against CS', we have that*

$$\mathrm{Adv}_A^{CCA} \leq \mathrm{Adv}_{A'}^{\mathcal{SM}_{(G,K)}} + \mathrm{Adv}_{A''}^{\mathcal{SM}_{(L,L')}} + 5\delta' + \delta + 2Q\epsilon' + \epsilon + \frac{2|L'|}{|L|},$$

where A' and A'' are algorithms that invoke each stage of A once, and Q is the number of decryption queries made by A

It is clear that when instantiated with the hash proof systems described in Sect. 4.1, then if the extended hash proof system is removed, the cryptosystem CS' reduces to Diffie-Hellman in K, and the above proof is easily modified to show that it is secure, as was claimed in Sect. 2.

Finally, we briefly discuss the performance of the scheme when instantiated with the hash proof systems described in Sect. 4.1 (using a hash function instead of a 1-1 function) and the symmetric subgroup membership problems discussed in Sect. 2.

Two things should be noted. For encryption, three exponentiations in CS' are in K, while the fourth exponent has bit length equal to the length of the hash value used. Second, when $\mathbb{Z}_n^*$ is used, K can be made very small compared to G. It is not unreasonable that for a t bit security level, $\log_2 |K| \approx 4t$ is sufficient.

The length of the hash should be $2t$. This means that the work required for an exponentiation corresponds roughly to one exponentiation with exponent bit length $14t$. For 80 bit security level this is 1120, and 1792 for 128 bit security level. This compares well with the corresponding modulus lengths 1024 and 3096.

For decryption, slightly more than two exponentiations in G are required (exactly two if $GF(2n+1)^*$ is used and $|G| = n$ is known). If the order of K is known to the private key holder, then roughly three exponentiations in K are required, but since they are all to the same base, the actual cost is smaller, say roughly equivalent to two exponentiations. For $\mathbb{Z}_n^*$, this corresponds to one exponentiation in G with exponent bit length $8t$.

Of course, if $\mathbb{Z}_n^*$ is used and the factorisation of n is known to the private key holder, Chinese remainder tricks are also available.

Compared to the instantiations of the Cramer-Shoup construction given in [3], our two instantiations are significantly faster, except for the elliptic curve variants of Cramer-Shoup. Asymptotically, they are faster than our variants, but at 80 bit security level, our variants would seem to have an advantage, at least for encryption.

5 Hybrid encryption

When a key encapsulation method is all that is required, the Cramer-Shoup key encapsulation method [4] using a subgroup of a finite field will be faster than our two constructions in the previous section. However, recent advances in [7] and [6] show that it is possible to construct secure hybrid encryption schemes from key encapsulation methods that are by themselves not secure.

The basic idea is that an ϵ-universal-2 hash proof system by itself will do, when its output is split into two bit strings, where one is used as a key for a symmetric cryptosystem, and the other is used as a key to a message authentication code.

We sketch a variant of this construction based on the symmetric subgroup membership problem in $\mathbb{Z}_n^*$. We do not believe that it will be faster than other instantiations, but we believe it is possible to construct a very fast cryptosystem based only on the hardness of the subgroup membership problem, which is in itself interesting.

The basic scheme requires five parts, a subgroup membership problem, a key derivation function, a MAC, a symmetric encryption scheme, and a hash function. Note that there are information theoretically secure MACs and symmetric encryption schemes.

The subgroup membership problem is based on $\mathbb{Z}_n^*$, where $n = (2ab+1)(2cd+1)$ as described in Sect. 2. To simplify things, G shall be the subgroup of quadratic residues. (It may be possible to use the subgroup with Jacobi symbol 1 instead.) We are given a generator g for K, of order ac.

The key derivation function $\kappa : G \to \{0,1\}^{l_1} \times \{0,1\}^{l_2}$ should return bit strings indistinguishable from random when applied to group elements sampled uniformly at random from certain subsets of G. Universal hashing techniques should provide an information-theoretically secure key derivation function.

The interesting point, however, is the hash function. What we need is a hash function $h : G \to \mathrm{Hom}(G, G)$ that is target collision resistant, where we count as a collision two homomorphisms that happen to be the same on any subgroup of G (this is why we restrict to the quadratic residues, and why $GF(2n+1)^*$ cannot be used).

Note that $\mathrm{Hom}(G, G) \simeq \mathbb{Z}_{\phi(n)/2}$. The hash function is simply $h(x) = x$, since $x \in G$ can be represented by an integer in the set $\{1, \ldots, n-1\}$ (we will consider the group elements to be integers when convenient). We claim that the advantage of any collision finder against this hash function is less than $\mathrm{Adv}^{\mathcal{SSM}_{(G,K,H)}}$.

So suppose we have some algorithm that on input of G and g outputs distinct x_1, x_2 such that $h(x_1)$ and $h(x_2)$ collide on some subgroup of G. We consider all possibilities in turn.

If they collide on G itself, this means that $x_1 \equiv x_2 \pmod{abcd}$, or that $abcd$ divides $x_1 - x_2$. Let z be any element with Jacobi symbol -1. Then $z^{x_1 - x_2}$ must be congruent to 1 modulo p and -1 modulo q, or vice versa. In other words, $z^{x_1 - x_2}$ gives a factorisation of n.

If they collide on K or H, but not both, then ac or bd divides $x_1 - x_2$, but not both. This may not lead to a factorisation of n, but it is clear that any multiple of $ac = |K|$ or $bd = |H|$ can be used to distinguish K or H.

If they collide modulo a, but not modulo c, or vice versa, we use the subgroup membership problems sampling algorithm to get an element $z \in K$. Unless we by chance have already got a factorisation of n, $z^{x_1 - x_2}$ will give us one. Likewise, for b and d.

This proves the claim. (Note that we prove collision resistance, which is stronger than target collision resistance.)

The key generation algorithm takes as input G and g. It samples k_0, k_1 uniformly at random from $\{1, \ldots, \lfloor n/4 \rfloor\}$. The public key is $(G, g, s_0, s_1) = (g^{k_0}, g^{k_1})$, the private key is (G, k_0, k_1).

The encryption algorithm takes the public key as input, as well as a message encoded as a bit string. It samples w uniformly at random from $\{1, \ldots, N\}$ (where N is sufficiently much larger than $|K|$). It computes $x = g^w$, $x' = s_0^{2w} s_1^{2wh(x)}$. Then it applies the key derivation function to x' to get encryption and MAC keys. It uses the encryption key to encrypt the message into ciphertext e and the MAC key to compute a tag t for e. The ciphertext is (x, e, t).

The decryption algorithm computes $x^{2(k_0 + h(x)k_1)}$ and applies the key derivation function to the result. It checks the tag t with the derived MAC key, and if it is correct, decrypts the ciphertext e with the encryption key and outputs the result.

The security analysis for this scheme should be essentially the same as in [6], which is very similar to the proof in Sect. 4. Note that the extra squaring makes

Key generation.	Encryption.	Decryption.
Input: $G \subseteq \mathbb{Z}_n^*, g \in G$.	Input: pk, $m \in G$.	Input: sk, (x, e, t).
1. $(k_0, k_1) \leftarrow \{0, \ldots, \lfloor n/4 \rfloor\}^2$.	1. $w \leftarrow \{0, \ldots, \lvert K \rvert - 1\}$.	1. $x' \leftarrow x^{2(k_0 + h(x)k_1)}$.
2. $(s_0, s_1) \leftarrow (g^{k_0}, g^{k_1})$.	2. $x \leftarrow g^w$.	2. $(\kappa_1, \kappa_2) \leftarrow kdf(x')$.
3. Select kdf.	3. $x' \leftarrow s_0^{2w} s_1^{2wh(x)}$.	3. $t' \leftarrow \mathcal{T}(\kappa_2, e)$.
4. $pk \leftarrow (G, g, s_0, s_1, kdf)$.	4. $(\kappa_1, \kappa_2) \leftarrow kdf(x')$.	4. If $t \neq t'$, output $\perp$.
5. $sk \leftarrow (G, k_0, k_1, kdf)$.	5. $e \leftarrow \mathcal{E}(\kappa_1, m)$.	5. $m \leftarrow \mathcal{D}(\kappa_1, e)$.
	6. $t \leftarrow \mathcal{T}(\kappa_2, e)$.	6. Output m.
Output: (pk, sk).	Output: (x, e, t).	Output: A message m or $\perp$.

Fig. 3. The hybrid cryptosystem using a symmetric cryptosystem $(\mathcal{E}, \mathcal{D})$ and MAC algorithm $\mathcal{T}$

the cryptosystem benignly malleable, in the sense that (x, e, t) and $(-x, e, t)$ both decrypt to the same message. This is not a security problem.

Compared to the scheme described in Sect. 4, the encryption cost measured in total exponent length is $8t + \log_2 n$. For 80 bit security level, this is roughly 1664, and 4120 for 128 bit security level. The decryption cost is roughly 480 and 768, respectively. The advantage is that we only depend on the subgroup membership problem.

6 Concluding remarks

We have defined and discussed symmetric subgroup membership problems. The main result of the theoretic discussion is a relation between the Decision Diffie-Hellman problem and the symmetric subgroup membership problem.

Then we have designed and analysed a chosen ciphertext secure public key cryptosystem based on a symmetric subgroup membership problem, by extending the framework of Cramer and Shoup. The resulting scheme is quite efficient compared to other instances of the Cramer-Shoup framework, although it requires a new hardness assumption.

Finally, we have sketched how to design a hybrid cryptosystem with chosen ciphertext security based only on a symmetric subgroup membership problem. In the immediate aftermath of CRYPTO'04, not relying on a target collision resistant hash function seems to be a conservative move. The full security proof for this cryptosystem will appear at a later time.

References

1. D. Boneh. The Decision Diffie-Hellman problem. In *Proceedings of the Third Algorithmic Number Theory Symposium*, volume 1423 of *LNCS*, pages 48–63. Springer-Verlag, 1998.

2. Ronald Cramer and Victor Shoup. A practical public key cryptosystem secure against adaptive chosen cipher text attacks. In Hugo Krawczyk, editor, *Proceedings of CRYPTO '98*, volume 1462 of *LNCS*, pages 13–25. Springer-Verlag, 1998.
3. Ronald Cramer and Victor Shoup. Universal hash proofs and a paradigm for adaptive chosen ciphertext secure public-key encryption. In Lars R. Knudsen, editor, *Proceedings of EUROCRYPT 2002*, volume 2332 of *LNCS*, pages 45–64. Springer-Verlag, 2002.
4. Ronald Cramer and Victor Shoup. Design and analysis of practical public-key encryption schemes secure against adaptive chosen ciphertext attack. *SIAM Journal on Computing*, 33(1):167–226, 2003.
5. W. Diffie and M. E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 22:644–654, 1976.
6. Rosario Gennaro and Victor Shoup. A note on an encryption scheme of Kurosawa and Desmedt. Cryptology ePrint Archive, Report 2004/194, 2004. `http://eprint.iacr.org/`.
7. K. Kurosawa and Y. Desmedt. A new paradigm of hybrid encryption scheme. In Matt Franklin, editor, *Proceedings of CRYPTO 2004*, volume 3152 of *LNCS*. Springer-Verlag, 2004.
8. W. Mao. Fast Monte-Carlo primality evidence shown in the dark. Technical Report HPL-1999-30R1, HP Laboratories, October 1999.
9. Juan Manuel González Nieto, Colin Boyd, and Ed Dawson. A public key cryptosystem based on the subgroup membership problem. In S. Quing, T. Okamoto, and J. Zhou, editors, *Proceedings of ICICS 2001*, volume 2229 of *LNCS*, pages 352–363. Springer-Verlag, 2001.
10. C. Rackoff and D. Simon. Non-interactive zero-knowledge proof of knowledge and chosen ciphertext attack. In Joan Feigenbaum, editor, *Proceedings of CRYPTO '91*, volume 576 of *LNCS*, pages 433–444. Springer-Verlag, 1992.
11. Akihiro Yamamura and Taiichi Saito. Private information retrieval based on the subgroup membership problem. In V. Varadharajan and Y. Mu, editors, *Proceedings of ACISP 2001*, volume 2119 of *LNCS*, pages 206–220. Springer-Verlag, 2001.