3-2010

Probabilistic public key encryption with equality test

Guomin YANG
Singapore Management University, gmyang@smu.edu.sg

Chik How TAN
National University of Singapore

Qiong HUANG
City University of Hong Kong

Duncan S. WONG
City University of Hong Kong

Probabilistic Public Key Encryption with Equality Test

Guomin Yang[1], Chik How Tan[1], Qiong Huang[2], and Duncan S. Wong[2]

[1] Temasek Laboratories
National University of Singapore
{tslyg,tsltch}@nus.edu.sg
[2] Department of Computer Science
City University of Hong Kong
duncan@cityu.edu.hk, csqhuang@student.cityu.edu.hk

Abstract. We present a (probabilistic) public key encryption (PKE) scheme such that when being implemented in a bilinear group, anyone is able to check whether two ciphertexts are encryptions of the same message. Interestingly, bilinear map operations are not required in key generation, encryption or decryption procedures of the PKE scheme, but is only required when people want to do an equality test (on the encrypted messages) between two ciphertexts that may be generated using different public keys. We show that our PKE scheme can be used in different applications such as searchable encryption and partitioning encrypted data. Moreover, we show that when being implemented in a non-bilinear group, the security of our PKE scheme can be strengthened from One-Way CCA to a weak form of IND-CCA.

Keywords: Public Key Encryption, Adaptive Chosen Ciphertext Attacks, Ciphertext Comparability, Searchable Encryption, Bilinear Map.

1 Introduction

Consider an outsourced database, data are stored in encrypted form. In order to maintain a good data structure, or extract some statistical information of the data, data may need to be partitioned. However, classical encryption schemes are not suitable for this purpose, since given a pile of ciphertexts, no one is able to tell the relationships among encrypted messages without knowing the decryption keys.

Searchable encryption (SE) schemes, introduced by Boneh et al. [7], and intensively studied in [3,1,21] may be one candidate to solve the problem. Informally speaking, in a searchable encryption scheme, a Tag T_M can be generated with respect to a message M and a key pair (PK, SK) (given T_M and PK, one should not be able to derive the message M). There is also a function Test' such that $\mathsf{Test}'(T_M, C)$ returns 1 if and only if the ciphertext C is an encryption of M under public key PK. So using the tag, one should be able to categorize the ciphertexts according to the encrypted messages. However, this method has one

J. Pieprzyk (Ed.): CT-RSA 2010, LNCS 5985, pp. 119–131, 2010.

shortcoming: the same message encrypted under different public keys cannot be categorized into one cluster.

Another possible approach is to use *deterministic* encryption schemes, which has been studied in recent years [3,5,6]. A PKE scheme is deterministic if the encryption algorithm is deterministic, namely given the same public key and the message, the encryption algorithm always outputs the same ciphertext. However, deterministic encryption suffers from the same problem that searchable encryption does for our purpose.

In this paper, we formalize the notion of Ciphertext Comparability for public key encryption schemes. We introduce a (probabilistic) public key encryption scheme such that when being implemented in a bilinear group its ciphertexts are publicly comparable. That is, given two ciphertexts C_1 and C_2 generated under public keys PK and PK', respectively, there is a function Test such that $\mathsf{Test}(C_1, C_2)$ returns 1 if and only if C_1 and C_2 are encryptions of the same message, no matter $PK = PK'$ or not. Our encryption scheme itself does not invoke any bilinear map operation in key generation, encryption or decryption procedures, only the Test function does. Further more, we show that when being implemented in a non-bilinear group, our PKE scheme can achieve a higher level of confidentiality, but at the cost of losing ciphertext comparability.

Related Work. Since the introduction of public key cryptography due to the seminal paper of Diffie and Hellman [14], public key encryption schemes are in the center of modern cryptography. Different types of PKE schemes with different security goals have been constructed. For a long time, people were searching for PKE schemes providing strong confidentiality, namely indistinguishability or semantic security under chosen-ciphertext attacks ($\mathsf{IND\text{-}CCA2}$) [24,15]. Nowadays there are many PKE schemes (e.g. [11,12,23,22,9,2,10,19,18,20]) achieving $\mathsf{IND\text{-}CCA2}$ security. A historical survey on PKE can be found in [13].

In [7], Boneh et al. presented the notion of public key encryption with keyword search (PEKS) and several constructions that achieve semantic security. In a PEKS scheme, Alice, with a key pair (pk, sk), can provide Bob with a trapdoor T_W, which is computed as a function of her secret key sk and any keyword W of her choice. Using T_W, it is possible to check, using a function Test', whether an arbitrary given ciphertext c is an encryption of W or not. The consistency condition is that $\mathsf{Test}'(T_W, c)$ returns 1 if and only if c is an encryption of W. Otherwise, the trapdoor T_W should not give any information about the real encrypted message W' (besides $W' \neq W$). Later, a general connection between PEKS and (anonymous) IBE was given by Abdalla et al. [1]. Recently, Hofheinz and Weinreb [20] presented a searchable public key encryption with decryption (PEKSD) in the standard model.

In [3], Bellare et al. initiated the notion of deterministic public key encryption where the encryption algorithm is deterministic. Deterministic encryption has been shown to be a useful tool in many applications, e.g. fast (i.e. logarithmic time) searching on encrypted data. This topic is further studied in [5,6].

As discussed in the introduction, if we are concerning encryptions generated under one public key, both searchable encryption and deterministic encryption

can be used to do equality tests among ciphertexts. However, we are interested in a more general multi-key setting.

This work is different from the *verifiable* encryption considered by Camenisch and Shoup in [8]. In a verifiable encryption scheme, the encrypter, who knows the plaintext, is able to prove (in zero-knowledge) to a third party that the message "encapsulated" in the ciphertext satisfies some property, e.g. the plaintext is the discrete log of an element with respect to a base in a group. So after generating the ciphertext, additional work is required from the encrypter in order to conduct the proof. Our work does not require any extra effort from the encrypters after generating the ciphertexts, and equality test can be performed just using the ciphertexts.

Our Contributions. In this paper, we give the notion of ciphertext comparability for public key encryption schemes, and present such a scheme. Our scheme allows anyone to compare two ciphertexts and check if they are encryptions of the same message, even though the ciphertexts may be generated using different public keys. We show that when being implemented in a bilinear group, our encryption scheme is one-way under chosen ciphertext attacks (as we will see shortly, it is impossible to achieve IND-ATK type of security for encryption schemes with ciphertext comparability) in the random oracle model. We show that PKE schemes with ciphertext comparability can be used in many applications, such as constructing searchable encryption, and partitioning encrypted data.

We then analyze the security of our encryption scheme when being implemented in a non-bilinear group. We show that under the DDH assumption our scheme achieves *Weak* Indistinguishability under Chosen Ciphertext Attacks (W-IND-CCA2) in the random oracle model.

Paper Organization. In the next section, we give the definitions and security models for public key encryption schemes with ciphertext comparability. Then we give our construction of a PKE scheme with ciphertext comparability in Sec. 3 and show some of its applications in Sec. 4. In Sec. 5, we give the definition of W-IND-ATK and prove that when being implemented in a non-bilinear group, our PKE scheme can achieve W-IND-CCA2 security.

2 Definitions

A public key encryption scheme $\Pi = (\mathcal{G}, \mathcal{E}, \mathcal{D})$ consists of a triple of algorithms. The key generation algorithm $\mathcal{G}$ takes a security parameter $k \in \mathbb{N}$ and outputs a public/private key pair (pk, sk). The encryption algorithm $\mathcal{E}$ takes a message m and the public key pk, and outputs a ciphertext c. The decryption algorithm $\mathcal{D}$ takes sk and c as input, and outputs m or $\perp$ (which indicates decryption failure). The correctness requirement is that $\forall k \in \mathbb{N}$ and $\forall m \in \mathsf{PtSp}(k)$, $(pk, sk) \leftarrow \mathcal{G}(1^k)$, $m \leftarrow \mathcal{D}(sk, \mathcal{E}(pk, m))$ where $\mathsf{PtSp}(k)$ is the plaintext space associated to Π.

We say that Π has *Ciphertext Comparability* with error ϵ for some function $\epsilon(\cdot)$ if there exists an efficiently computable deterministic function $\mathsf{Test}(\cdot,\cdot)$ such that for every k we have

1. Perfect Consistency: for every $x \in \mathsf{PtSp}(k)$

$$\Pr\left[\begin{array}{l}(pk,sk)\leftarrow\mathcal{G}(1^k),(pk',sk')\leftarrow\mathcal{G}(1^k),C\leftarrow\mathcal{E}(pk,x)\\ C'\leftarrow\mathcal{E}(pk',x)\text{: }\mathsf{Test}(C,C')=1\end{array}\right]=1$$

2. Soundness: for every polynomial time algorithm $\mathcal{M}$

$$\Pr\left[\begin{array}{l}(C,C',sk,sk')\leftarrow\mathcal{M}(1^k),x\leftarrow\mathcal{D}(sk,C),x'\leftarrow\mathcal{D}(sk',C'):\\ x\neq\perp\wedge x'\neq\perp\wedge x\neq x'\wedge\mathsf{Test}(C,C')=1\end{array}\right]\leq\epsilon(k)$$

In the above definition, consistency ensures that encryptions (even under different public keys) of the same message can be recognized. Soundness measures the probability of false-hits (i.e. $\mathsf{Test}(C,C')=1$ but C and C' are encryptions of different messages).

Sanity Check. In our definition for ciphertext comparability, the Test function does not perform any sanity check on the ciphertexts, namely, we don't specify the output of Test when its input cannot be decrypted. We only require that when the ciphertexts are real encryptions of messages, the function works properly. On the other hand, checking whether a ciphertext is in the correct form without using the private key is another problem out of the scope of this paper.

In the following, we review the classical notions of privacy for public key encryption schemes, namely, indistinguishability under chosen plaintext and chosen ciphertext attacks [17,24,15].

Definition 1 (IND-ATK [4]). *Let $\Pi=(\mathcal{G},\mathcal{E},\mathcal{D})$ be a public key encryption scheme and let $\mathcal{A}=(\mathcal{A}_1,\mathcal{A}_2)$ be a polynomial-time adversary. For $atk\in\{cpa,cca1,cca2\}$ and $k\in\mathbb{N}$ let*

$$\mathrm{Adv}^{\mathrm{ind-atk}}_{\mathcal{A},\Pi}\stackrel{\mathrm{def}}{=}\Pr\left[\begin{array}{l}(pk,sk)\leftarrow\mathcal{G}(1^k),(x_0,x_1,\delta)\leftarrow\mathcal{A}_1^{\mathcal{O}_1}(pk),b\leftarrow\{0,1\},\\ y\leftarrow\mathcal{E}(pk,x_b),b'\leftarrow\mathcal{A}_2^{\mathcal{O}_2}(pk,x_0,x_1,\delta,y):b'=b\end{array}\right]-\frac{1}{2}$$

where $x_0\neq x_1\wedge|x_0|=|x_1|$ and

$$\begin{array}{llll}\textit{If } atk=cpa & \textit{then} & \mathcal{O}_1(\cdot)=\varepsilon & \textit{and} \quad \mathcal{O}_2(\cdot)=\varepsilon\\ \textit{If } atk=cca1 & \textit{then} & \mathcal{O}_1(\cdot)=\mathcal{D}_{sk}(\cdot) & \textit{and} \quad \mathcal{O}_2(\cdot)=\varepsilon\\ \textit{If } atk=cca2 & \textit{then} & \mathcal{O}_1(\cdot)=\mathcal{D}_{sk}(\cdot) & \textit{and} \quad \mathcal{O}_2(\cdot)=\mathcal{D}_{sk}(\cdot)\end{array}$$

In the case of CCA2, we insist that $\mathcal{A}_2$ does not ask its oracle for decrypting y. We say that Π is secure in the sense of IND-ATK if $\mathrm{Adv}^{\mathrm{ind-atk}}_{\mathcal{A},\Pi}$ is negligible for any $\mathcal{A}$.

Unfortunately, indistinguishability based security notions are not applicable to PKE schemes with ciphertext comparability. Given the challenge ciphertext y and plaintexts x_0,x_1, an adversary $\mathcal{A}$ can compute another ciphertext

$y' = \mathcal{E}(pk, x_1)$, and then return $\mathsf{Test}(y, y')$ as her guess of the value b in the IND-ATK games. The advantage of the adversary $\mathcal{A}$ is

$$\begin{aligned}\mathrm{Adv}^{\mathrm{ind-atk}}_{\mathcal{A},\Pi} &= \frac{1}{2}\Pr[\mathsf{Test}(y, y') = 1 | b = 1] + \frac{1}{2}\Pr[\mathsf{Test}(y, y') = 0 | b = 0] - \frac{1}{2} \\ &= \frac{1}{2} + \frac{1}{2}\Pr[\mathsf{Test}(y, y') = 0 | b = 0] - \frac{1}{2} \\ &= \frac{1}{2}(1 - \Pr[\mathsf{Test}(y, y') = 1 | b = 0]) \\ &\geq \frac{1}{2}(1 - \epsilon(k))\end{aligned}$$

As ciphertext comparability and indistinguishability are irreconcilable, we go back to the one-way definition of privacy for public key encryption schemes.

Definition 2 (OW-ATK). *Let $\Pi = (\mathcal{G}, \mathcal{E}, \mathcal{D})$ be a public key encryption scheme and let $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ be a polynomial-time adversary. For $atk \in \{cpa, cca1, cca2\}$ and $k \in \mathbb{N}$ let*

$$\mathrm{Adv}^{\mathrm{ow-atk}}_{\mathcal{A},\Pi} \stackrel{\mathrm{def}}{=} \Pr\left[\begin{array}{l}(pk, sk) \leftarrow \mathcal{G}(1^k), \delta \leftarrow \mathcal{A}_1^{\mathcal{O}_1}(pk), x \leftarrow \mathsf{PtSp}(k) \\ y \leftarrow \mathcal{E}(pk, x), x' \leftarrow \mathcal{A}_2^{\mathcal{O}_2}(pk, \delta, y) : x' = x\end{array}\right]$$

where

If $atk = cpa$	*then*	$\mathcal{O}_1(\cdot) = \varepsilon$	*and*	$\mathcal{O}_2(\cdot) = \varepsilon$
If $atk = cca1$	*then*	$\mathcal{O}_1(\cdot) = \mathcal{D}_{sk}(\cdot)$	*and*	$\mathcal{O}_2(\cdot) = \varepsilon$
If $atk = cca2$	*then*	$\mathcal{O}_1(\cdot) = \mathcal{D}_{sk}(\cdot)$	*and*	$\mathcal{O}_2(\cdot) = \mathcal{D}_{sk}(\cdot)$

In the case of CCA2, we insist that $\mathcal{A}_2$ does not ask its oracle to decrypt y. We say that Π is secure in the sense of OW-ATK if $\mathrm{Adv}^{\mathrm{ow-atk}}_{\mathcal{A},\Pi}$ is negligible for any $\mathcal{A}$.

A Simpler Definition of OW-CCA2. The following theorem states that the OW-CCA2 definition can be simplified.

Theorem 1. *In the case of OW-CCA2, the definition with $\mathcal{O}_1 = \mathcal{D}_{sk}(\cdot)$ (denoted Def1) is equivalent to that with $\mathcal{O}_1 = \varepsilon$ (denoted Def2).*

Proof. 1. Def1 $\Rightarrow$ Def2: trivial.

2. Def2 $\Rightarrow$ Def1: we prove that if a PKE scheme Π is secure under definition Def2, it is also secure under definition Def1. First of all, an obvious fact is that if Π is secure under definition Def2 then $|\mathsf{PtSp}(k)| > p(k)$ for any polynomial p.

The proof is by contradiction. Suppose there exists a polynomial-time adversary $\mathcal{A}$ that breaks Π in Def1 with a non-negligible advantage, we construct another polynomial-time adversary $\mathcal{B}$ that breaks Π in Def2 also with a non-negligible advantage.

$\mathcal{B}$ is given (pk, y), where $(pk, sk) \leftarrow \mathcal{G}(1^k), x \leftarrow \mathsf{PtSp}(k), y \leftarrow \mathcal{E}(pk, x)$. $\mathcal{B}$ simulates the experiment of Def1 as follows: $\mathcal{B}$ gives pk to $\mathcal{A}$ as the public key, and simulates $\mathcal{O}_1 = \mathcal{D}_{sk}(\cdot)$ by asking its own decryption oracle. If $\mathcal{O}_1$ is queried

with y, $\mathcal{B}$ makes a random guess on x and aborts the game. Otherwise, $\mathcal{B}$ gives y to $\mathcal{A}$ after $\mathcal{A}$ finishes to query $\mathcal{O}_1$, and continues to simulate $\mathcal{O}_2 = \mathcal{D}_{sk}(\cdot)$ in $\mathsf{Def1}$ by using its own decryption oracle. Finally, $\mathcal{B}$ outputs whatever $\mathcal{A}$ outputs.

Denote E the event $\mathcal{A}$ queries $\mathcal{O}_1$ with y. If E does not occur, then the simulation is perfect. Since $|\mathsf{PtSp}(k)| > p(k)$ for any polynomial p, for any $\tilde{y}$ $\mathcal{A}$ queried to $\mathcal{O}_1$, the probability that $\mathcal{D}_{sk}(\tilde{y}) = x$ is negligible as x is randomly selected from $\mathsf{PtSp}(k)$. So the probability that E occurs is negligible. □

The above theorem shows that the simplification does not weaken the definition of security. On the other hand, it helps simplify the security proofs.

3 PKE with Ciphertext Comparability in Bilinear Groups

Our construction is based on the Computational Diffie-Hellman (CDH) assumption in bilinear groups. Let $\mathbb{G}_1, \mathbb{G}_2$ denote two groups of prime order q, and $e : \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$ a bilinear map between them. The map satisfies the following properties:

1. Bilinear: For any $U, V \in \mathbb{G}_1$, and $a, b \in \mathbb{Z}_q$, we have $e(U^a, V^b) = e(U, V)^{ab}$;
2. Non-degenerate: If g is a generator of $\mathbb{G}_1$, then $e(g, g)$ is a generator of $\mathbb{G}_2$;
3. Computable: there exists an efficient algorithm to compute $e(U, V)$ for any $U, V \in \mathbb{G}_1$.

Computational Diffie-Hellman (CDH) Problem: Fix a generator g of $\mathbb{G}_1$. The CDH problem is as follows: given g, g^a, g^b as input where a, b are randomly selected from $\mathbb{Z}_q$, compute g^{ab}. We say that CDH is intractable if all polynomial time algorithms have a negligible advantage in solving CDH.

We build a public key encryption scheme with ciphertext comparability in a bilinear group where CDH is intractable. Our construction uses a hash function $H : \mathbb{G}_1^3 \to \{0,1\}^{k+\ell}$, where k and ℓ are security parameters such that elements of $\mathbb{G}_1$ are represented in k bits and elements of $\mathbb{Z}_q$ are represented in ℓ bits. Our PKE with ciphertext comparability works as follows:

- $\mathcal{G}(1^k)$: Select $x \leftarrow \mathbb{Z}_q^*$ and compute $y = g^x$. Set $pk = y$ and $sk = x$.
- $\mathcal{E}(pk, m)$: To encrypt a plaintext $m \in \mathbb{G}_1^*$ ($\stackrel{\mathsf{def}}{=} \mathbb{G}_1 \backslash \{1\}$), select $r \leftarrow \mathbb{Z}_q^*$, compute $U = g^r$, $V = m^r$, $W = H(U, V, y^r) \oplus m \| r$. The ciphertext is $C = (U, V, W)$.
- $\mathcal{D}(sk, C)$: To decrypt a ciphertext $C = (U, V, W)$, compute $m \| r \leftarrow H(U, V, U^x) \oplus W$. If $(m \in \mathbb{G}_1^* \wedge r \in \mathbb{Z}_q^* \wedge U = g^r \wedge V = m^r)$, return m; otherwise, return $\perp$.
- $\mathsf{Test}(C_1, C_2)$: Given two ciphertexts $C_1 = (U_1, V_1, W_1)$ and $C_2 = (U_2, V_2, W_2)$, if $e(U_1, V_2) = e(U_2, V_1)$, return 1; otherwise, return 0.

Theorem 2. *The above PKE scheme has perfect consistency and perfect soundness.*

Proof. The proof is straightforward, as follows:

1. Perfect Consistency. It is easy to see that for any $(pk, sk) \leftarrow \mathcal{G}(1^k)$, $(pk', sk') \leftarrow \mathcal{G}(1^k)$ and $C \leftarrow \mathcal{E}(pk, m), C' \leftarrow \mathcal{E}(pk', m)$ where $C = (g^r, m^r, W)$, $C' = (g^{r'}, m^{r'}, W')$, we have

$$e(g^r, m^{r'}) = e(g^{r'}, m^r) = e(g, m)^{rr'}$$

for any $m \in \mathbb{G}_1^*$ and $(r, r') \in \mathbb{Z}_q^{*2}$.

2. Perfect Soundness. Given two ciphertexts $C = (g^r, m^r, W), C' = (g^{r'}, m'^{r'}, W')$, we have

$$e(g^r, m'^{r'}) = e(g, m')^{rr'}, e(g^{r'}, m^r) = e(g, m)^{rr'}$$

then it must be true that $e(g, m')^{rr'} \neq e(g, m)^{rr'}$ for any $m \neq m'$ and $(r, r') \in \mathbb{Z}_q^{*2}$. □

Theorem 3. *The PKE scheme above with message space $\mathbb{G}_1^*$ is OW-CCA2 secure in the random oracle model assuming CDH is intractable.*

Proof. Let $\mathcal{A}$ be a PPT adversary attacking the OW-CCA2 security of the above PKE scheme. Suppose that $\mathcal{A}$ runs in time t and makes at most q_H hash queries and q_D decryption queries. Let $\mathsf{Adv}_{\mathcal{A}}^{\mathsf{OW\text{-}CCA2}}(t, q_H, q_D)$ denote the advantage of $\mathcal{A}$ in the OW-CCA2 experiment. We first consider the original game:

Game $\mathbf{G_0}$

1. $x \leftarrow \mathbb{Z}_q^*$, $y = g^x$
2. $m \leftarrow \mathbb{G}_1^*$, $r \leftarrow \mathbb{Z}_q^*$, $U^* = g^r$, $V^* = m^r$, $W^* = H(U^*, V^*, y^r) \oplus (m \| r)$
3. $m' \leftarrow \mathcal{A}^{\mathcal{O}_H, \mathcal{O}_2}(y, U^*, V^*, W^*)$, where the oracles work as follows.
   - $\mathcal{O}_H$: On input a triple $(U, V, Y) \in \mathbb{G}_1^3$, a *compatible* random value is returned, where by '*compatible*' we mean that if the same input is asked multiple times, the same answer will be returned. Note that a query to this oracle is also issued when computing the challenge ciphertext or simulating the decryption oracle.
   - $\mathcal{O}_2$: On input a ciphertext (U, V, W), it runs the decryption algorithm $\mathcal{D}$ to decrypt it using the secret key x.

Let $\mathbf{X}_0$ be the event that $m' = m$ in **Game $\mathbf{G_0}$**. In the following, let $\mathbf{X}_i$ be the event that $m' = m$ in **Game $\mathbf{G_i}$** for $i = 1, 2, \cdots$. $\mathcal{A}$'s winning probability in **Game $\mathbf{G_i}$** is $\Pr[\mathbf{X}_i]$. Next we modify **Game $\mathbf{G_0}$** and obtain the following game.

Game $\mathbf{G_1}$

1. $x \leftarrow \mathbb{Z}_q^*$, $y = g^x$, $T = \emptyset$
2. $m \leftarrow \mathbb{G}_1^*$, $r \leftarrow \mathbb{Z}_q^*$, $U^* = g^r$, $V^* = m^r$, $R^* \leftarrow \{0,1\}^{k+\ell}$, $W^* = R^* \oplus (m \| r)$, $T = T \cup \{(U^*, V^*, (U^*)^x, R^*)\}$
3. $m' \leftarrow \mathcal{A}^{\mathcal{O}_H, \mathcal{O}_2}(y, U^*, V^*, W^*)$, where the oracles work as follows.
   - $\mathcal{O}_H$: On input $(U, V, Y) \in \mathbb{G}_1^3$, if there is an entry (U, V, Y, h) in the hash table T, h is returned; otherwise, a random value h is selected and returned, and (U, V, Y, h) is added into T.

- $\mathcal{O}_2$: On input a ciphertext (U, V, W), a hash query on (U, V, U^x) is issued. Suppose the answer is $h \in \{0,1\}^{k+\ell}$. Then $m\|r$ is computed as $h \oplus W$, and the validity check on whether $U = g^r$ and $V = m^r$ is performed. If the check fails, $\perp$ is returned; otherwise, m is returned.

Due to the idealness of the random oracle, **Game $\mathbf{G_1}$** is identical to **Game $\mathbf{G_0}$**. Thus $\Pr[\mathbf{X}_1] = \Pr[\mathbf{X}_0]$. In the next game, we further modify the simulation in an indistinguishable way.

Game $\mathbf{G_2}$

1. $x \leftarrow \mathbb{Z}_q^*$, $y = g^x$, $T = \emptyset$
2. $m \leftarrow \mathbb{G}_1^*$, $r \leftarrow \mathbb{Z}_q^*$, $U^* = g^r$, $V^* = m^r$, $W^* \leftarrow \{0,1\}^{k+\ell}$, $T = T \cup \{(U^*, V^*, (U^*)^x, W^* \oplus (m\|r))\}$
3. $m' \leftarrow \mathcal{A}^{\mathcal{O}_H, \mathcal{O}_2}(y, U^*, V^*, W^*)$, where the oracles work as follows.
   - $\mathcal{O}_H$: It is simulated in the same way as that in **Game $\mathbf{G_1}$** except that if $\mathcal{A}$ asks $(U^*, \cdot, (U^*)^x)$, the game is aborted. Let this event be $\mathbf{E}$.
   - $\mathcal{O}_2$: The same as that in **Game $\mathbf{G_1}$** except that if $\mathcal{A}$ asks for decryption of (U^*, V^*, W') where $W' \neq W^*$, $\perp$ is returned.

The challenge ciphertext generated in this game is identically distributed to that in **Game $\mathbf{G_1}$**, as W^* is a random value in both **Game $\mathbf{G_1}$** and **Game $\mathbf{G_2}$**. Also, the simulation of $\mathcal{O}_2$ is perfect since W^* is uniquely determined by U^* and V^*. Therefore, if event $\mathbf{E}$ does not occur, **Game $\mathbf{G_2}$** is identical to **Game $\mathbf{G_1}$**. Next, we show that event $\mathbf{E}$ occurs with negligible probability.

Lemma 1. *Event* $\mathbf{E}$ *happens in* ***Game*** $\mathbf{G_2}$ *with negligible probability if CDH is intractable.*

Proof. Suppose that $\Pr[\mathbf{E}]$ is non-negligible. We construct a PPT algorithm $\mathcal{B}$ to break the CDH assumption. Given a tuple $(g, g^a, g^c) \in \mathbb{G}_1^3$, $\mathcal{B}$ randomly selects $\alpha \leftarrow \mathbb{Z}_q^*$ and $R^* \leftarrow \{0,1\}^{k+\ell}$, then sets the public key $y = g^a$ and $U^* = g^c$, and computes $m = g^\alpha$. It then adds $(U^*, V^* = (U^*)^\alpha, \top, \top)$ into table T which is initially empty, where $\top$ represents that the value is unknown. $\mathcal{B}$ invokes adversary $\mathcal{A}$ on input (y, U^*, V^*, R^*), where the challenge ciphertext (U^*, V^*, R^*) has the same distribution as that in **Game $\mathbf{G_2}$**. The oracles for $\mathcal{A}$ are simulated as follows.

- $\mathcal{O}_H$: $\mathcal{B}$ simulates the oracle as described in **Game $\mathbf{G_1}$**, except that if $\mathcal{A}$ makes a query on $(U^*, \cdot, Z)$, $\mathcal{B}$ checks if $e(g, Z) = e(y, U^*)$. If the equation holds, $\mathcal{B}$ outputs Z and aborts the game.
- $\mathcal{O}_2$: On input (U, V, W), if the input is $U = U^*$, $V = V^*$ and $W \neq R^*$, $\mathcal{B}$ returns $\perp$. Otherwise, $\mathcal{B}$ searches T for an entry of the form $(U, V, \cdot, \cdot)$. For each item (U, V, Y, h), $\mathcal{B}$ computes $m\|r = h \oplus W$ and proceeds as follows.
  1. If $U = g^r, V = m^r$ and $Y = y^r$, m is returned;
  2. Otherwise, $\mathcal{B}$ continues to search T for the next entry of the form $(U, V, \cdot, \cdot)$.

  If nothing is returned to $\mathcal{A}$ in the above loop for all entries $(U, V, \cdot, \cdot)$ in T, $\mathcal{B}$ returns $\perp$.

Denote $\mathbf{E}'$ the event that $\mathcal{A}$ queries $\mathcal{O}_H$ on input $(U^*, \cdot, g^{ac})$. At the end of the simulation, if $\mathbf{E}'$ does not occur, $\mathcal{B}$ aborts with failure.

(*Analysis*): We first show that the decryption queries are simulated indistinguishably from **Game $\mathbf{G_2}$**. We separate all the decryption queries into two types:

1. Type 1: (U, V, U^a) has been queried to $\mathcal{O}_H$ before a decryption query (U, V, W) is issued. In this case, W is uniquely determined after (U, V, U^a) is queried to $\mathcal{O}_H$. So the decryption oracle is simulated perfectly.
2. Type 2: (U, V, U^a) has never been queried to $\mathcal{O}_H$ when a decryption query (U, V, W) is issued. In this case, $\perp$ is returned by the decryption oracle. The simulation fails if (U, V, W) is a valid ciphertext. However, due to the idealness of the random oracle, this happens with probability $1/2^{k+\ell}$.

Denote $\mathbf{E}_2$ the event that a valid ciphertext is rejected in the simulation. Then we have

$$\Pr[\mathbf{E}_2] \leq \frac{q_D}{2^{k+\ell}}.$$

If $\mathbf{E}_2$ does not happen, then the simulation is identical to **Game $\mathbf{G_2}$**, so $\Pr[\mathbf{E}'|\neg\mathbf{E}_2] = \Pr[\mathbf{E}]$. Then we have

$$\begin{aligned}\Pr[\mathbf{E}'] &= \Pr[\mathbf{E}'|\mathbf{E}_2]\Pr[\mathbf{E}_2] + \Pr[\mathbf{E}'|\neg\mathbf{E}_2]\Pr[\neg\mathbf{E}_2] \\ &\geq \Pr[\mathbf{E}'|\neg\mathbf{E}_2]\Pr[\neg\mathbf{E}_2] \\ &= \Pr[\mathbf{E}](1 - \Pr[\mathbf{E}_2]) \\ &\geq \Pr[\mathbf{E}] - \Pr[\mathbf{E}_2]\end{aligned}$$

Therefore,

$$\mathsf{Adv}_{\mathcal{B}}^{\mathsf{CDH}} \geq \Pr[\mathbf{E}] - \frac{q_D}{2^{k+\ell}}$$

which is non-negligible. This completes the proof of Lemma 1. □

Since **Game $\mathbf{G_1}$** and **Game $\mathbf{G_2}$** are the same if event $\mathbf{E}$ does not occur, we have,

$$|\Pr[\mathbf{X}_1] - \Pr[\mathbf{X}_2]| \leq \Pr[\mathbf{E}].$$

Lemma 2. $\Pr[\mathbf{X}_2]$ *is negligible under the CDH assumption.*

Proof. Suppose that $\Pr[\mathbf{X}_2]$ is non-negligible. We construct a PPT algorithm $\mathcal{B}$ to break the CDH assumption. Given a tuple $(g, \hat{U} = g^r, \hat{V} = m^r) \in \mathbb{G}_1^3$, where $r \leftarrow \mathbb{Z}_q$ and $m \leftarrow \mathbb{G}_1^*$, $\mathcal{B}$'s goal is to compute m. $\mathcal{B}$ randomly selects $x \leftarrow \mathbb{Z}_q^*$ and sets the public key $y = g^x$. It then adds $(U^* = \hat{U}, V^* = \hat{V}, (U^*)^x, \top)$ into table T which is initially empty, where $\top$ represents that the value is unknown. $\mathcal{B}$ randomly selects $R^* \leftarrow \{0,1\}^{k+\ell}$ and invokes adversary $\mathcal{A}$ on input (y, U^*, V^*, R^*). $\mathcal{B}$ simulates the game by following the description of **Game $\mathbf{G_2}$**. Finally $\mathcal{B}$ outputs whatever $\mathcal{A}$ outputs. So we have

$$\Pr[\mathbf{X}_2] \leq \mathsf{Adv}_{\mathcal{B}}^{\mathsf{CDH}}.$$

□

Therefore, we have

$$\begin{aligned}\mathsf{Adv}^{\mathsf{OW\text{-}CCA2}}_{\mathcal{A}}(t, q_H, q_D) &= \Pr[\mathbf{X}_0] \\ &= \Pr[\mathbf{X}_1] \\ &\leq \Pr[\mathbf{X}_2] + \mathsf{Adv}^{\mathsf{CDH}} + \frac{q_D}{2^{k+\ell}} \\ &\leq 2\mathsf{Adv}^{\mathsf{CDH}} + \frac{q_D}{2^{k+\ell}}.\end{aligned}$$

This completes the proof of Theorem 3. □

4 Variants and Applications

Encrypting Long Messages. In our PKE scheme above, we assume that messages are elements of group $\mathbb{G}_1^*$. To encrypt long messages, we can use a collision resistant hash function $H' : \{0,1\}^* \to \mathbb{G}_1^*$ and a pseudo-random bit generator PRG [16]. We modify the scheme as follows:

- $\mathcal{G}(1^k)$: Unchanged.
- $\mathcal{E}(pk, m)$: To encrypt a plaintext $M \in \{0,1\}^*$, compute $m \leftarrow H'(M)$. Select $r \leftarrow \mathbb{Z}_q^*$, compute $U \leftarrow g^r$, $V \leftarrow m^r$, $K \leftarrow H(U, V, pk^r)$ and $W \leftarrow \mathsf{PRG}(K) \oplus M \| r$. The ciphertext is $C = (U, V, W)$.
- $\mathcal{D}(sk, C)$: To decrypt a ciphertext $C = (U, V, W)$, compute $K \leftarrow H(U, V, U^{sk})$, $M\|r \leftarrow \mathsf{PRG}(K) \oplus W$, and $m \leftarrow H'(M)$. If $(r \in \mathbb{Z}_q^* \wedge U = g^r \wedge V = m^r)$, return M; otherwise, return $\perp$.
- $\mathsf{Test}(C_1, C_2)$: Unchanged

It follows that the above (hybrid) encryption scheme also has ciphertext comparability. The perfect consistency is still maintained, however, the soundness is no longer perfect, but is bounded by the collision probability of H'.

Theorem 4. *The modified PKE scheme above is OW-CCA2 secure assuming H, H' are random oracles,* PRG *is a secure pseudo-random bit generator, and CDH is intractable.*

The proof essentially follows that of Theorem 3, we can replace the pseudo-random bit string with a truly random string when generating the challenge ciphertext. Since the adversary does not know the random seed of the PRG (due to the CDH assumption and H is a random oracle), the difference between the games is negligible provided PRG is a secure pseudo-random bit generator.

Searchable Encryption. A PKE scheme with ciphertext comparability is naturally searchable. To generate a tag T_M for message M, one can simply encrypt M under any valid public key to generate a ciphertext C, and set $T_M = C$. Then using the Test function, anyone is able to search encryptions of the message M, even if they are generated using different public keys.

Partitioning Encrypted Data. In applications such as outsourced databases, by using a PKE scheme with ciphertext comparability, the database administrator is able to do a partition of the database according to the encrypted messages

without any help from the message owners. These schemes may also be useful in other similar applications such as collection and categorization of confidential data through an agent.

5 Weak IND-CCA2 vs Ciphertext Comparability

In Sec. 2, we have shown that ciphertext comparability and indistinguishability are irreconcilable. In this section, we are interested in the following question: if we don't need ciphertext comparability, or when being implemented in a non-bilinear group, what kind of security level can our PKE scheme in Sec. 3 achieve? The first security model we'd like to try is of course IND-CCA. Unfortunately, our scheme is even not IND-CPA secure, as shown below.

Theorem 5. *The PKE scheme in Sec. 3 with message space $\mathbb{G}_1^*$ is not IND-CPA secure.*

Proof. We construct a PPT adversary $\mathcal{A}$ as follows. Given public key y, $\mathcal{A}$ computes $m_0 = g^{r_0}$ and $m_1 = g^{r_1}$ for any two distinct r_0 and r_1 chosen arbitrarily from $\mathbb{Z}_q^*$. $\mathcal{A}$ sends (m_0, m_1) to the game simulator. After receiving the challenge ciphertext $c^* = (U, V, W)$, $\mathcal{A}$ checks if $V = U^{r_0}$. If yes, $\mathcal{A}$ returns 0; otherwise $\mathcal{A}$ returns 1. The probability that $\mathcal{A}$ guesses correctly the value of b is 1. □

The above attack demonstrates the advantage the adversary can get from selecting the challenge plaintexts. In the next, we define a different set of indistinguishability games where the adversary has no such power.

Definition 3 (W-IND-ATK). *Let $\Pi = (\mathcal{G}, \mathcal{E}, \mathcal{D})$ be a public key encryption scheme and let $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ be a polynomial-time adversary. For $atk \in \{cpa, cca1, cca2\}$ and $k \in \mathbb{N}$ let*

$$\mathrm{Adv}^{\mathrm{w-ind-atk}}_{\mathcal{A},\Pi} \stackrel{\mathrm{def}}{=} \Pr\left[\begin{array}{l}(pk, sk) \leftarrow \mathcal{G}(1^k), \delta \leftarrow \mathcal{A}_1^{\mathcal{O}_1}(pk),\\ (x_0, x_1) \leftarrow \mathsf{PtSp}(k), b \leftarrow \{0,1\}, y \leftarrow \mathcal{E}(pk, x_b),\\ b' \leftarrow \mathcal{A}_2^{\mathcal{O}_2}(pk, x_0, x_1, \delta, y) : b' = b\end{array}\right] - \frac{1}{2}$$

where $x_0 \neq x_1 \wedge |x_0| = |x_1|$ and

If $atk = cpa$	*then*	$\mathcal{O}_1(\cdot) = \varepsilon$	*and*	$\mathcal{O}_2(\cdot) = \varepsilon$
If $atk = cca1$	*then*	$\mathcal{O}_1(\cdot) = \mathcal{D}_{sk}(\cdot)$	*and*	$\mathcal{O}_2(\cdot) = \varepsilon$
If $atk = cca2$	*then*	$\mathcal{O}_1(\cdot) = \mathcal{D}_{sk}(\cdot)$	*and*	$\mathcal{O}_2(\cdot) = \mathcal{D}_{sk}(\cdot)$

In the case of CCA2, we insist that $\mathcal{A}_2$ does not ask its oracle for decrypting y. We say that Π is secure in the sense of W-IND-ATK if $\mathrm{Adv}^{\mathrm{w-ind-atk}}_{\mathcal{A},\Pi}$ is negligible for any $\mathcal{A}$.

W-IND-CCA2 Security of Our PKE. Interestingly, we can show that when being implemented in a non-bilinear group, our PKE scheme given in Sec. 3 can achieve W-IND-CCA2 security under the DDH assumption which is described below.

Decisional Diffie-Hellman (DDH) Problem. Fix a generator g of $\mathbb{G}_1$. The DDH assumption claims that $\{g, g^a, g^b, Z\}$ and $\{g, g^a, g^b, g^{ab}\}$ are computationally indistinguishable where a, b are randomly selected from $\mathbb{Z}_q$ and Z is a random element of $\mathbb{G}_1$.

Theorem 6. *The PKE scheme in Sec. 3 with message space $\mathbb{G}_1^*$ is W-IND-CCA2 secure in the random oracle model under the DDH assumption.*

The proof is by contradiction. Suppose there exists an adversary who can break the encryption scheme, we plant the DDH problem $(g, m, U = g^r, V = Z)$ into the challenge ciphertext to the adversary, and simulate the decryption oracle in a similar way as in the proof of Theorem 3. Then depending on $Z = m^r$ (i.e. Z is in the "right" form) or $Z \leftarrow \mathbb{G}_1$ (Z is independent of m), the adversary would have different probability in winning the game, so we can use the adversary to solve the DDH problem. The detailed proof is deferred to the full version of the paper.

Acknowledgement. We would like to thank the anonymous reviewers for their comments and suggestions.

References

1. Abdalla, M., Bellare, M., Catalano, D., Kiltz, E., Kohno, T., Lange, T., Malone-Lee, J., Neven, G., Paillier, P., Shi, H.: Searchable encryption revisited: Consistency properties, relation to anonymous IBE, and extensions. J. Cryptology 21(3), 350–391 (2008)
2. Abe, M., Gennaro, R., Kurosawa, K., Shoup, V.: Tag-KEM/DEM: A new framework for hybrid encryption and a new analysis of Kurosawa-Desmedt KEM. In: Cramer, R. (ed.) EUROCRYPT 2005. LNCS, vol. 3494, pp. 128–146. Springer, Heidelberg (2005)
3. Bellare, M., Boldyreva, A., O'Neill, A.: Deterministic and efficiently searchable encryption. In: Menezes, A. (ed.) CRYPTO 2007. LNCS, vol. 4622, pp. 535–552. Springer, Heidelberg (2007)
4. Bellare, M., Desai, A., Pointcheval, D., Rogaway, P.: Relations among notions of security for public-key encryption schemes. In: Krawczyk, H. (ed.) CRYPTO 1998. LNCS, vol. 1462, pp. 26–45. Springer, Heidelberg (1998)
5. Bellare, M., Fischlin, M., O'Neill, A., Ristenpart, T.: Deterministic encryption: Definitional equivalences and constructions without random oracles. In: Wagner, D. (ed.) CRYPTO 2008. LNCS, vol. 5157, pp. 360–378. Springer, Heidelberg (2008)
6. Boldyreva, A., Fehr, S., O'Neill, A.: On notions of security for deterministic encryption, and efficient constructions without random oracles. In: Wagner, D. (ed.) CRYPTO 2008. LNCS, vol. 5157, pp. 335–359. Springer, Heidelberg (2008)
7. Boneh, D., Crescenzo, G.D., Ostrovsky, R., Persiano, G.: Public key encryption with keyword search. In: Cachin, C., Camenisch, J.L. (eds.) EUROCRYPT 2004. LNCS, vol. 3027, pp. 506–522. Springer, Heidelberg (2004)
8. Camenisch, J., Shoup, V.: Practical verifiable encryption and decryption of discrete logarithms. In: Boneh, D. (ed.) CRYPTO 2003. LNCS, vol. 2729, pp. 126–144. Springer, Heidelberg (2003)

9. Canetti, R., Halevi, S., Katz, J.: Chosen-ciphertext security from identity-based encryption. In: Cachin, C., Camenisch, J.L. (eds.) EUROCRYPT 2004. LNCS, vol. 3027, pp. 207–222. Springer, Heidelberg (2004)
10. Canetti, R., Halevi, S., Katz, J.: A forward-secure public-key encryption scheme. J. Cryptology 20(3), 265–294 (2007)
11. Cramer, R., Shoup, V.: A practical public key cryptosystem provably secure against adaptive chosen ciphertext attack. In: Krawczyk, H. (ed.) CRYPTO 1998. LNCS, vol. 1462, pp. 13–25. Springer, Heidelberg (1998)
12. Cramer, R., Shoup, V.: Universal hash proofs and a paradigm for adaptive chosen ciphertext secure public-key encryption. In: Knudsen, L.R. (ed.) EUROCRYPT 2002. LNCS, vol. 2332, pp. 45–64. Springer, Heidelberg (2002)
13. Dent, A.W.: A brief history of provably-secure public-key encryption. In: Vaudenay, S. (ed.) AFRICACRYPT 2008. LNCS, vol. 5023, pp. 357–370. Springer, Heidelberg (2008)
14. Diffie, W., Hellman, M.E.: New directions in cryptography. IEEE Transactions on Information Theory 22, 644–654 (1978)
15. Dolev, D., Dwork, C., Naor, M.: Nonmalleable cryptography. SIAM J. Comput. 30(2), 391–437 (2000)
16. Goldreich, O.: Foundations of Cryptography: Basic Tools. Cambridge University Press, Cambridge (2001)
17. Goldwasser, S., Micali, S.: Probabilistic encryption. J. Comput. Syst. Sci. 28(2), 270–299 (1984)
18. Hanaoka, G., Kurosawa, K.: Efficient chosen ciphertext secure public key encryption under the computational diffie-hellman assumption. In: Pieprzyk, J. (ed.) ASIACRYPT 2008. LNCS, vol. 5350, pp. 308–325. Springer, Heidelberg (2008)
19. Hofheinz, D., Kiltz, E.: Secure hybrid encryption from weakened key encapsulation. In: Menezes, A. (ed.) CRYPTO 2007. LNCS, vol. 4622, pp. 553–571. Springer, Heidelberg (2007)
20. Hofheinz, D., Kiltz, E.: Practical chosen ciphertext secure encryption from factoring. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 313–332. Springer, Heidelberg (2009)
21. Hofheinz, D., Weinreb, E.: Searchable encryption with decryption in the standard model. Cryptology ePrint Archive, Report 2008/423 (2008), `http://eprint.iacr.org/`
22. Kurosawa, K., Desmedt, Y.: A new paradigm of hybrid encryption scheme. In: Franklin, M. (ed.) CRYPTO 2004. LNCS, vol. 3152, pp. 426–442. Springer, Heidelberg (2004)
23. Lucks, S.: A variant of the cramer-shoup cryptosystem for groups of unknown order. In: Zheng, Y. (ed.) ASIACRYPT 2002. LNCS, vol. 2501, pp. 27–45. Springer, Heidelberg (2002)
24. Rackoff, C., Simon, D.R.: Non-interactive zero-knowledge proof of knowledge and chosen ciphertext attack. In: Feigenbaum, J. (ed.) CRYPTO 1991. LNCS, vol. 576, pp. 433–444. Springer, Heidelberg (1992)