A New Security Definition for Public Key Encryption Schemes and Its Applications

Guomin Yang, Duncan S. Wong, Qiong Huang, and Xiaotie Deng

Department of Computer Science
City University of Hong Kong
Hong Kong, China
{csyanggm,duncan,deng}@cs.cityu.edu.hk
csqionghuang@cityu.edu.hk

Abstract. The strongest security definition for public key encryption (PKE) schemes is indistinguishability against adaptive chosen ciphertext attacks (IND-CCA). A practical IND-CCA secure PKE scheme in the standard model is well-known to be difficult to construct given the fact that there are only a few such kind of PKE schemes available. From another perspective, we observe that for a large class of PKE-based applications, although IND-CCA security is sufficient, it is not a necessary requirement. Examples are Key Encapsulation Mechanism (KEM), MT-authenticator, providing pseudorandomness with a-priori information, and so on. This observation leads us to propose a slightly weaker version of IND-CCA, which requires ciphertexts of two *randomly* selected messages are indistinguishable under chosen ciphertext attacks. Under this new security notion, we show that highly efficient schemes proven secure in the standard model can be built in a straightforward way. We also demonstrate that such a security definition is already sufficient for the applications above.

Keywords: Public Key Encryption, Adaptive Chosen Ciphertext Attacks, Standard Model

1 Introduction

Design and analysis of public key encryption (PKE) schemes are among the most important tasks for cryptographers, and an appropriate security definition is a prerequisite before any work. If the definition is too weak, the resulting schemes are not able to be widely deployed; on the other hand, if the definition is too strong, the design work may become very difficult, and the schemes may be very complicated and inefficient.

The commonly agreed (and also strongest) security definition for PKE schemes is Indistinguishability under Chosen Ciphertext Attacks (or IND-CCA) [17, 25]. It ensures that the ciphertexts of any two different messages[1] *selected* by any polynomial time adversary are indistinguishable even if the adversary has access to a decryption oracle. Such a definition of security allows the encryption scheme to be safely deployed in the widest range of applications. Many practical PKE schemes

[1] Hereinafter we simply assume that all the messages have equal length.

have been proven to be IND-CCA in the random oracle model, for example, RSA-OAEP[18], OAEP+[30], SAEP[8]. However, security in the random oracle model does not rule out all the possible attacks in the real world [6, 11]. On the other hand, only few efficient IND-CCA schemes are constructed in the standard model. Some IND-CCA schemes were proposed by following the Naor-Yung paradigm [22, 23, 26, 27], but they are quite impractical as they rely on generic non-interactive zero knowledge proofs. The first provably secure and practical PKE scheme in the standard model was proposed by Cramer and Shoup[13], later, two fairly practical schemes were also proposed by the same authors [14], and these are the only known efficient PKE schemes in the standard model[2].

In this paper, we reconsider the notion of IND-CCA. After reviewing some applications of PKE, we find that IND-CCA is a sufficient but not necessary condition. For example, in the construction of hybrid encryption schemes, we can use Key Encapsulation Mechanism (KEM) to replace PKE in order to achieve high efficiency [2, 20, 21, 29]. In another example, when we use PKE for authentication purpose [4], we only need to ensure the confidentiality of a random challenge. Hence, we consider a variant of the IND-CCA definition for PKE schemes, which we call Weak INDistinguishability under Chosen Ciphertext Attacks (or simply, WIND-CCA). In this definition, we require that the ciphertexts of two messages *drawn* independently and randomly from the message space are indistinguishable under chosen ciphertext attacks. An alternative definition that is similar to semantic security can also be derived. To see the difference, consider the following example: we construct the encryption algorithm in such a way that if the input is the first element (we denote it by m_0) in lexical order in the message space, we append a '0' at the end of the ciphertext, otherwise, we append a '1'. Obviously, the ciphertext of m_0 is distinguishable from that of any other element in the message space, but such an encryption scheme may be still WIND-CCA if the message space is sufficiently large. We will see later that such a modification in security definition allows highly efficient PKE schemes to be built.

Our Contributions. We first define a new type of indistinguishability against adaptive chosen ciphertext attacks for public key encryption schemes. This new definition (WIND-CCA) is strictly weaker than the conventional IND-CCA security. With this new notion, we hope more efficient PKE schemes can be built (in the standard model) in the future.

Then we propose an efficient PKE scheme that is WIND-CCA in the standard model. Different from the Naor-Yung paradigm, we use an efficient non-interactive zero knowledge proof system for proving the equivalence of discrete logarithms. Compared with the schemes by Cramer and Shoup [13, 14], our new scheme has a slightly longer ciphertext but a smaller key size.

Furthermore, we give several applications where a WIND-CCA PKE scheme can be applied. In particular, we show a generic way to construct CCA secure KEM

[2] In this paper, when we say PKE schemes, we do not refer to hybrid encryption schemes. In fact, we concentrate on the asymmetric primitives for building hybrid encryption schemes.

(and hence CCA secure hybrid encryption schemes). Compared with existing KEM schemes, the derived KEM is more flexible and has a wider range of applications. Some other applications are also presented, such as providing pseudorandomness with a-priori information, constructing an encryption based MT-authenticator and so on.

Organisation: In the next section we review the underlying assumptions, and the notions of IND-CCA PKE schemes and IND-CCA KEM schemes. The definition of WIND-CCA PKE schemes is also given here. We then give our concrete construction in Sec. 3, along with the security proof. In Sec. 4, 5 and 6 we discuss the applications of WIND-CCA PKE schemes in constructing KEM schemes, providing pseudorandomness with a-priori information and building encryption-based MT-authenticators, respectively.

2 Preliminaries

We denote the selection of a random element e from a set S by $e \xleftarrow{R} S$, and let $|S|$ denote the cardinality of S. If $A(\cdot,\cdot)$ is a probabilistic algorithm, $y \leftarrow A(x_1, x_2)$ denotes the experiment of running A with input x_1, x_2 and output y. In the rest of this paper, we only consider probabilistic polynomial time (PPT) algorithms.

2.1 Intractable Assumptions

Let k be a security parameter. Let $p = 2q+1$ be a randomly chosen k-bit safe prime. Let Q be the group of squares modulo p, and $g \in Q \setminus \{1\}$.

Decisional Diffie-Hellman (DDH) Assumption [28]: For $a, b, c \xleftarrow{R} Z_q^*$, $\langle g, g^a, g^b, g^{ab}\rangle$ and $\langle g, g^a, g^b, g^c\rangle$ are computationally indistinguishable.

The Knowledge of Exponent Assumption (KEA) [1, 7, 15]: Given a pair $g, \hat{g} = g^a$ with unknown discrete log a, the only way to efficiently come up with another pair $A, \hat{A} = A^a$ is by raising g and g^a to some power b (i.e. $A = g^b, \hat{A} = \hat{g}^b$).

2.2 Public Key Encryption Schemes

A public key encryption scheme consists of a tuple of algorithms PKE $= (\mathcal{G}, \mathcal{E}, \mathcal{D})$. The key generation algorithm takes the security parameter as input and generates a key pair $(pk, sk) \leftarrow \mathcal{G}(1^k)$ where pk is the public key and sk is the private key. The encryption algorithm takes a message m and pk as input and outputs a ciphertext $c \leftarrow \mathcal{E}(pk, m)$. The decryption algorithm is a (probabilistic or deterministic) algorithm which takes the private key and a ciphertext as input and outputs $m \leftarrow \mathcal{D}(sk, c)$ or rejects the ciphtertext by outputting a special symbol '$\perp$'.

Indistinguishability against Chosen Ciphertext Attacks (IND-CCA) is defined by the following game.

$\text{Game}_{\mathcal{A}}^{\text{IND-CCA}}$:

$(pk, sk) \leftarrow \mathcal{G}(1^k)$
$(m_0, m_1, \sigma) \leftarrow \mathcal{A}_1^{\mathcal{O}_{sk}}(pk)$
$b \overset{R}{\leftarrow} \{0,1\}, c \leftarrow \mathcal{E}(pk, m_b)$
$b' \leftarrow \mathcal{A}_2^{\mathcal{O}_{sk}}(pk, \sigma, c)$

In the above game, $\mathcal{A}$ is separated into two stages and σ denotes the internal state of $\mathcal{A}$ after the 'find' stage. $\mathcal{O}_{sk}$ denotes a decryption oracle with respect to the private key sk, the only restriction is that in the 'guess' stage, the adversary cannot query the decryption oracle with input c. $\mathcal{A}$ wins the game if $b' = b$, and the advantage of $\mathcal{A}$ is defined as follows:

$$\mathrm{Adv}_{\mathcal{A}}^{\mathrm{IND-CCA}} = \Pr(b' = b) - \frac{1}{2}$$

A PKE scheme is secure under chosen ciphertext attacks if $\mathrm{Adv}_{\mathcal{A}}^{\mathrm{IND-CCA}}$ is negligible for any PPT adversary $\mathcal{A}$.

The definition of IND-CPA is similar except that $\mathcal{A}$ does not have access to a decryption oracle. These definitions require that the ciphertexts of any two distinct plaintexts *selected* by the adversary are indistinguishable. However, as stated in the introduction, in a lot of applications, we only use PKE schemes to protect randomly selected messages, so we introduce the following weaker version of IND-CCA.

$\mathrm{Game}_{\mathcal{A}}^{\mathrm{WIND-CCA}}$:

$(pk, sk) \leftarrow \mathcal{G}(1^k)$
$(m_0, m_1) \overset{R}{\leftarrow} \mathcal{M}(pk)$
$b \overset{R}{\leftarrow} \{0,1\}, c \leftarrow \mathcal{E}(pk, m_b)$
$b' \leftarrow \mathcal{A}^{\mathcal{O}_{sk}}(pk, m_0, m_1, c)$

In the above game, $\mathcal{M}(pk)$ denotes the message space with respect to pk. As in the traditional IND-CCA game, $\mathcal{A}$ has access to a decryption oracle $\mathcal{O}_{sk}$ except that $\mathcal{A}$ cannot query $\mathcal{O}_{sk}$ with input c[3]. The advantage of an adversary $\mathcal{A}$ is defined analogously to the original IND-CCA game, and a PKE scheme is WIND-CCA if $\mathrm{Adv}_{\mathcal{A}}^{\mathrm{WIND-CCA}}$ is negligible for any PPT adversary $\mathcal{A}$. The definition of WIND-CPA is defined analogously except that $\mathcal{A}$ does not have access to the decryption oracle.

2.3 Key Encapsulation Mechanism

A KEM is a key derivation and encapsulation algorithm. It takes a public key as input and outputs a symmetric key K and a ciphertext ψ of the keying material of K, and the private key owner decrypts ψ and derives K. KEM and its variants are combined with symmetric key encryption schemes to produce hybrid encryption schemes [2, 20, 21, 29]. In the following we describe the syntax and security definition for KEM.

A KEM consists of a tuple of algorithms KEM = (KEM.G, KEM.E, KEM.D).

[3] It is easy to see that allowing the adversary to access the decryption oracle before receiving the challenge ciphertext does not give the adversary any additional power, since the the adversary can simulate this process after receiving the challenge ciphertext.

1. The key generation algorithm takes the security parameter as input and generates a key pair $(pk, sk) \leftarrow \text{KEM.G}(1^k)$
2. The encryption algorithm takes the public key as input and outputs a symmetric key K and a ciphertext ψ of the keying material of K, $(K, \psi) \leftarrow \text{KEM.E}(pk)$
3. The decryption algorithm takes the private key and the ciphertext as input and generates a symmetric key $K = \text{KEM.D}(sk, \psi)$ or rejects the ciphertext by outputting '$\perp$'.

The security definition of KEM has the same flavor with that of session key security in key exchange protocols [5, 9]. It is defined by the following game[4].

$\text{Game}_{\mathcal{A}}^{\text{KEM.CCA}}$:

$(pk, sk) \leftarrow \text{KEM.G}(1^k)$
$(K_1, \psi) \leftarrow \text{KEM.E}(pk)$, $K_0 \stackrel{R}{\leftarrow} \mathcal{K}$
$b \stackrel{R}{\leftarrow} \{0, 1\}$
$b' \leftarrow \mathcal{A}^{\mathcal{O}_{sk}}(pk, \psi, K_b)$

where $\mathcal{K}$ denotes the symmetric key space. The advantage of the adversary $\mathcal{A}$ is defined as

$$\text{Adv}_{\mathcal{A}}^{\text{KEM.CCA}} = \Pr(b' = b) - \frac{1}{2}$$

We observe that two techniques are used in the KEM constructions[5] by Shoup [29] and Kurosawa-Desmedt [21]:

i) Since generating the symmetric key and encrypting the keying material are done in one step, they can use the same random coins in both processes. However, this combination limits the flexibility and range of applications of KEM. For example, if the symmetric key is generated from some other party or process, the randomness in sampling the (random) key may not be able be recovered (e.g. $K = g^r$), and we cannot rule out this possibility in real applications.
ii) The key derivation process relies on the public key. This will again limit the flexibility, consider that we want to use KEM to transport the same symmetric key to more than one receivers, if we use Shoup's KEM, then we need to find r_1 and r_2 such that $h_1^{r_1} = h_2^{r_2}$ where h_i is part of $user_i$'s public key. This is impossible since the sender does not know the secret exponents corresponding to h_1 and h_2.

Jumping ahead, we will see a construction that can avoid these problems in Sec. 4.

[4] Due to the same reason as footnote 3, we let the adversary query the decryption oracle after receiving the challenge ciphertext.

[5] For the 'PKE + hash' based Tag-KEM construction by Abe et al.[2], they do not follow this approach, but they use a fully IND-CCA PKE scheme. In their generic construction of hybrid encryption, they also adopt these techniques in the Tag-KEM part.

2.4 Non-interactive Zero Knowledge Proof

In this section, we review a non-interactive zero-knowledge proof (NIZK) system which is important in our construction of a WIND-CCA PKE scheme. Non-interactive zero-knowledge proof is an essential tool in the Naor-Yung paradigm, but generic NIZK proofs are quite inefficient due to the complex NP reductions. However, there are efficient NIZK proofs for some specific problems, for example, proving the equivalence of two discrete logarithms [12, 16].

In [16], Damgård et al. developed a method that compiles all known discrete-log based Σ-protocols (3-move public-coin protocols) into NIZK arguments. In the following, we review the compilation of a Σ-protocol $\mathcal{P}_{eqdlog}$ for proving the equivalence of two discrete logarithms.

Prover P and verifier V get common input $x = (p, q, g, g', h_1, h_2)$ where p, q, g are the parameters in the DDH assumption, g' is a random element of Q and $h_1 = g^w, h_2 = g'^w$ for some $w \in Z_q^*$. P gets w as private input.

1. P chooses $r \overset{R}{\leftarrow} Z_q^*$ and sends $a = (a_1, a_2)$ to V, where $a_1 = g^r, a_2 = g'^r$;
2. V chooses $e \overset{R}{\leftarrow} Z_q^*$ and sends it to P;
3. P sends $z = r + ew$ to V who checks that $g^z = a_1 h_1^e, g'^z = a_2 h_2^e$.

Let $R_{eqdlog} = \{(x, w)\}$ be the relation as specified above, and $L_R = \{x | \exists\ w : (x, w) \in R_{eqdlog}\}$. The above Σ-protocol has the following properties:

1. **Special Soundness.** From $x \in L_R$ and two accepting conversations (a, e, z), (a, e', z') where $e \neq e'$, w can be efficiently computed.
2. **Special Honest Verifier Zero Knowledge.** There exists a PPT simulator which on input x and e outputs a conversation (a, e, z) whose distribution is (statistically) indistinguishable from that of conversations between P and V for the given statement $x \in L_R$ and challenge e.

Under the assumption that Paillier cryptosystem [24] is 2-harder than the discrete logarithm problem (informally, $\mathcal{H}$ is 2-harder than $\mathcal{G}$ means even if there exists an algorithm $\mathcal{A}$ that completely breaks $\mathcal{G}$ on instances of size k, no algorithm running in comparable time can break $\mathcal{H}$ on instances of size $2k$ or larger), Damgård et al. compiled the above Σ-protocol to an NIZK proof in the registered public key model [3]. The compilation makes use of the homomorphic property of the Paillier encryption.

In the key setup phase for the verifier, a public/private key pair (pk, sk) of the Paillier cryptosystem is generated, also a challenge e as V would do in the Σ-protocol is chosen, and set c to be a random encryption of e under pk. The public key is $\hat{pk} = (pk, c)$ and the private key is $\hat{sk} = (sk, e)$.

Since the Σ-protocol is with linear answer and Paillier encryption is homomorphic, it is possible to execute the prover's side given only an encryption of the challenge. The compilation is reviewed below:

Protocol Compile($\mathcal{P}_{eqdlog}$):

1. Given an instance x, w to prove, P gets V's public key (pk, c) and computes the first message a according to $\mathcal{P}_{eqdlog}$. Let the final message be of form $u + ev$, P computes $\xi \leftarrow \text{randomize}(E_{pk}^{Paillier}(u)c^v)$, which can be done by multiplying $E_{pk}^{Paillier}(u)c^v$ with a random encryption of 0. P sends x, π to V where $\pi = (a, \xi)$.
2. On input (x, π), V gets $z \leftarrow D_{sk}^{Paillier}(\xi)$ and verifies $(x; a, e, z)$ by running the verifier of $\mathcal{P}_{eqdlog}$.

Damgård et al. showed that the above NIZK protocol is statistically zero knowledge in the registered public key model. Namely, there exists a PPT algorithm M, such that for all instances $(x, w) \in R_{eqdlog}$, the following two ensembles are statistically close:

Verifier's key pair, real proof:

$$\{(\hat{pk}, \hat{sk}, \pi) \mid (\hat{pk}, \hat{sk}) \leftarrow \text{KeySetup}(1^k); \pi \leftarrow P(1^k, x, w, \hat{pk})\}$$

Verifier's key pair, simulated proof:

$$\{(\hat{pk}, \hat{sk}, \pi) \mid (\hat{pk}, \hat{sk}) \leftarrow \text{KeySetup}(1^k); \pi \leftarrow M(1^k, x, \hat{sk}, \hat{pk})\}$$

Remark: In order to prove the above protocol is zero-knowledge, the simulator has to interact with V in order to emulate the view of V in real life. In particular, M has to know the private key of V.

3 A WIND-CCA (but not IND-CCA) PKE Scheme

In this section, we show how to construct a PKE scheme that is WIND-CCA secure. We first present our construction, and then show that this scheme is not semantically secure, thus separate WIND-CCA from IND-CCA.

The Construction: Let p, q, Q, g be defined as in the DDH assumption, we assume that messages are (or can be encoded as) elements of Q.

Key Generation: Run the key setup algorithm of the NIZK proof system in Sec. 2.4 and get $(\hat{pk}, \hat{sk})$; randomly choose $s \stackrel{R}{\leftarrow} Z_q^*$, and compute $h = g^s$. The public key is $(\hat{pk}, h)$ and the private key is $(\hat{sk}, s)$.

Encryption: Given a message $m \in Q$, randomly select $r \stackrel{R}{\leftarrow} Z_q^*$ and computes $u_1 = g^r, u_2 = m^r, \lambda = mh^r$. Then generate an NIZK proof $\pi = P(1^k, x = (g, u_1, m, u_2), r, \hat{pk})$ as described in Sec. 2.4. The ciphertext is (u_1, u_2, λ, π).

Decryption: Given a ciphertext u_1, u_2, λ, π, calculate $m = \lambda / u_1^s$, and $d \leftarrow V(1^k, \hat{sk}, x = (g, u_1, m, u_2), \pi)$ as described in Sec. 2.4. If $d =$ accept, return m, otherwise, return $\perp$. (Notice that only the private key owner can verify the statement/proof pair (x, π)).

Theorem 1. *The above PKE scheme is* not *IND-CPA.*

Proof. It's easily observed that the ciphertext of g is distinguishable from that of any other element in Q. □

Theorem 2. *The above PKE scheme is WIND-CCA.*

Proof. The proof is by contradiction, if there exists an adversary $\mathcal{A}$ that breaks the above scheme in the WIND-CCA game with a non-negligible probability, we construct another adversary $\mathcal{B}$ that can break the DDH assumption with non-negligible probability.

$\mathcal{B}$ is given $g, g_1 = g^r, g_2 = m, g_3 = m^z$ where $r \overset{R}{\leftarrow} Z_q^*, m \overset{R}{\leftarrow} Q$ and $z = r$ or $z \overset{R}{\leftarrow} Z_q^*$. $\mathcal{B}$ generates a public/private key pair $(\hat{pk}, h)/(\hat{sk}, s)$ by running the key generation algorithm ($\hat{pk}$ is of the form $(pk^{Paillier}, c)$ and $\hat{sk}$ is of the form $(sk^{Paillier}, e)$ where $e \overset{R}{\leftarrow} Z_q^*$ and $c \leftarrow E_{sk}^{Paillier}(e)$). Then $\mathcal{B}$ generates a ciphertext $\gamma = (g_1, g_2 g_1^s, g_3, \pi)$ where π is generated by running the simulation algorithm M with input $(x = (g, g_1, g_2, g_3), \hat{sk}, \hat{pk})$ (i.e. $\pi = (a_1, a_2, E_{sk}^{Paillier}(\varpi))$ where $\varpi \overset{R}{\leftarrow} Z_q^*$, $a_1 = g^{\varpi}/g_1^e$ and $a_2 = g_2^{\varpi}/g_3^e$). $\mathcal{B}$ tosses a coin b and sets $m_b = g_2$, $\mathcal{B}$ also selects $m_{1-b} \overset{R}{\leftarrow} Q$ and runs $\mathcal{A}$ with input $(\hat{pk}, h), m_b, m_{1-b}, \gamma$. Up to now (before $\mathcal{A}$ issues any decryption queries), we can get that

Claim 1 *The challenge ciphertext created by $\mathcal{B}$ reveals no information about b.*

Proof. We separate the case that $z = r$ from the case that z is a random element of Q.

1. $z = r$: Due to the DDH assumption, g_3 and $g_2 g_1^s$ reveals no information about b. The fact that g_3 does not revel b follows from that (g^r, g, m^r, m) and (g^r, g, m^r, U) are computationally indistinguishable (here U denotes a random element of Q); and $g_2 g_1^s$ does not reveal b follows from that (g, g_1, g^s, g_1^s) and (g, g_1, g^s, U) are computationally indistinguishable. The fact that π reveals no information about b follows from that Paillier's cryptosystem is semantically secure under the decisional composite residuosity assumption
2. $z \overset{R}{\leftarrow} Z_q^*$: By following the same reason as above, $g_2 g_1^s$ and π do not reveal b. And g_3 is independent of b. □

After giving the challenge ciphertext to $\mathcal{A}$, $\mathcal{B}$ answers $\mathcal{A}$'s decryption queries by following the normal decryption procedures[6]. Finally, if $\mathcal{A}$ outputs $b' = b$, $\mathcal{B}$ outputs $z = r$, otherwise $\mathcal{B}$ outputs $z \overset{R}{\leftarrow} Z_q^*$.

If $\mathcal{B}$'s input is a DH tuple: the game is essentially the same as the real game except that the proof π in the challenge ciphertext is generated by the simulation algorithm M. Since the challenge ciphertext to $\mathcal{A}$ is statistically close to a valid ciphertext, $\mathcal{A}$ has a non-negligible advantage over $\frac{1}{2}$ to guess correctly.

[6] Here we require the stronger result of [16], that is the NIZK protocol is unboundedly sound, see [16] for details.

If $\mathcal{B}$'s input is not a DH tuple: the ciphertext $\mathcal{A}$ has received is an invalid ciphertext. Suppose $\mathcal{A}$ asks a decryption query with input u'_1, u'_2, e', π'. From the soundness of the NIZK proof, it follows that $\mathcal{B}$ (with negligible error) will only decrypt ciphertexts in correct form (i.e. $u'_1 = g^{r'}, u'_2 = m'^{r'}$ where $m' = e'/u'^s_1$). For those ciphertexts in the correct form, we have the following claim:

Claim 2 *Under the knowledge of exponent assumption, $\mathcal{A}$ can produce a ciphertext u'_1, u'_2, e', π' in the correct form only if $\mathcal{A}$ knows the message $m' = e'/u'^s_1$.*

Proof. Suppose $\mathcal{A}$ gives $\mathcal{B}$ a ciphertext $u'_1 = g^{r'}, u'_2 = m'^{r'}, e', \pi'$ where $m' = e'/u'^s_1$. We consider the following cases (these are just mental experiments):

1. $\mathcal{A}$ knows r'. Then $\mathcal{A}$ can get $m' = e'/h^{r'}$ by herself.
2. $\mathcal{A}$ does not know r' and $u'_2 = u'_1$. This implies $m' = g$ and $\mathcal{A}$ can get this information by herself.
3. $\mathcal{A}$ does not know r' and $u'_2 \neq u'_1$. If $\mathcal{A}$ does not know m', then $\mathcal{A}$ and $\mathcal{B}$ constitute an algorithm that breaks the knowledge of exponent assumption. By contradiction, $\mathcal{A}$ must know the message m'. □

By the above claim, it follows that if the ciphertext is in correct form, $\mathcal{A}$ must know the corresponding plaintext. So we can conclude that the decryption queries essentially do not provide any help to $\mathcal{A}$. Hence, due to the DDH assumption, g_1^s is essentially a one-time pad, and $\mathcal{A}$ has only negligible advantage over $\frac{1}{2}$ to guess correctly. □

4 Constructing Hybrid Encryption Schemes

KEM is an essential tool of existing approaches to constructing hybrid encryption schemes. In the next, we give a construction of KEM based on WIND-CCA PKE schemes.

Let (pk, sk) be the key pair of the PKE scheme. As in [21, 29], let $H : \mathcal{M}_{pk} \to \{0,1\}^k$ be a hash function that essentially preserves the entropy of the input distribution, which can be constructed without any intractability assumptions. Without loss of generality, we assume that $|\mathcal{M}_{pk}| > 2^k$. Then a CCA secure KEM can be constructed as follows:

KEM.ENC(pk):	KEM.DEC(sk, ψ'):
Select $m \xleftarrow{R} \mathcal{M}_{pk}$ Calculate $\psi \leftarrow \mathcal{E}(pk, m)$ and $K = H(m)$ Return (K, ψ)	Calculate $m' \leftarrow \mathcal{D}(sk, \psi')$ If $m' = \perp$, return $\perp$ Else, calculate $K' = H(m')$ Return K'

Theorem 3. *If there exists a PPT algorithm that breaks the* CCA *security of the above* KEM *scheme, there exists another PPT algorithm that can break the* WIND-CCA *security of the* PKE *scheme.*

The proof is given in Appendix A.

From the result of Shoup [29], we can construct a hybrid encryption scheme that is IND-CCA by combining the above KEM with the symmetric key encryption scheme in [29]. We can also derive a CCA-secure Tag-KEM based on the above KEM and a secure message authentication code (MAC) [2]. It's easy to see that the above KEM construction does not have those limitations discussed in Sec. 2.3.

5 Pseudorandomness with A-Priori Information

In [10], Canetti proposed a new cryptographic primitive which is named *Oracle Hashing*. It can provide pseudorandomness with a-priori information. However, in order to prove the security of the oracle hashing, Canetti raised a new assumption that is related to the DDH assumption (we call it *Strong DDH assumption*). He proved that the construction $H(x, r) = r \| r^x$ is a strong oracle hashing under the *Strong DDH assumption*. First we recall this assumption, as below:

Strong Decisional Diffie-Hellman (SDDH) Assumption [10]: Let p, q, g be the parameters as in the DDH assumption. For any non-invertible function f and for $a, b, c \stackrel{R}{\leftarrow} Z_q^*$, $\langle f(a), g^b, g^{ab} \rangle$ and $\langle f(a), g^b, g^c \rangle$ are computationally indistinguishable.

Canetti showed a very simple way to construct an IND-CPA hybrid encryption scheme by using trapdoor one-way permutation and oracle hashing. In other words,

$$\text{SDDH} + \text{Trapdoor One-way Permutation} \rightarrow \text{IND-CPA}$$

However, it would be better to replace the SDDH assumption by a commonly used one, and we find that WIND-CPA PKE schemes can help. In the next, we will show that

$$\text{DDH} + \text{WIND-CPA} \rightarrow \text{IND-CPA}$$

In other words, besides the generic transformation from WIND-CCA to IND-CCA hybrid encryption, there is another generic transformation from WIND-CPA to IND-CPA hybrid encryption. We start by defining the following variance of the DDH assumption:

Definition 1 (DDH with A-Priori Information). *Let p, q, g be defined as in the DDH assumption. Let $PKE = (\mathcal{G}, \mathcal{E}, \mathcal{D})$ denote an asymmetric key encryption scheme. Define $T_R = \langle y, g, g^a, g^b, g^c \rangle$ and $T_D = \langle y, g, g^a, g^b, g^{ab} \rangle$. For any PPT algorithm $\mathcal{A}$, define*

$$P_{\mathcal{A}}^R = \Pr\left(\mathcal{A}(T_R) = 1 \mid (pk, sk) \leftarrow \mathcal{G}(1^k);\ a, b, c \stackrel{R}{\leftarrow} Z_q^*;\ y \leftarrow \mathcal{E}(pk, a)\right)$$

$$P_{\mathcal{A}}^D = \Pr\left(\mathcal{A}(T_D) = 1 \mid (pk, sk) \leftarrow \mathcal{G}(1^k);\ a, b \stackrel{R}{\leftarrow} Z_q^*;\ y \leftarrow \mathcal{E}(pk, a)\right)$$

where the probabilities are taken over the random choices of a, b, c and the random coins of $\mathcal{G}, \mathcal{E}, \mathcal{A}$. We say T_R and T_D are computationally indistinguishable *with respect to PKE if for any positive polynomial $q(\cdot)$ and all sufficiently large k,*

$$|P_{\mathcal{A}}^{D} - P_{\mathcal{A}}^{R}| < \frac{1}{q(k)}$$

We say T_R and T_D are computationally indistinguishable under chosen ciphertext attacks (CI-CCA) *with respect to PKE if $\mathcal{A}$ has access to a decryption oracle $\mathcal{O}_{sk}$ with the restriction that $\mathcal{A}$ cannot ask the decryption of y.*

In the next, we prove that $\langle \mathcal{E}(pk, x), H(x, r) \rangle$ and $\langle \mathcal{E}(pk, x), H(y, r) \rangle$ are computationally indistinguishable (CI-CCA respectively) if the encryption scheme is WIND-CPA (WIND-CCA respectively). Hence, by following the same proof in [10], it follows that $\mathcal{E}(pk, x), r, r^x \oplus m$ is IND-CPA if $\mathcal{E}(\cdot, \cdot)$ is WIND-CPA.

Lemma 1. *For any $PKE = (\mathcal{G}, \mathcal{E}, \mathcal{D})$ that is* WIND-CPA (WIND-CCA *respectively), T_R and T_D are computationally indistinguishable* (CI-CCA *respectively*) *under the DDH assumption.*

The proof is given in Appendix B. Here we only prove the part corresponding to CCA security, the part corresponding to CPA security essentially follows the same proof and can be obtained by simply removing the decryption oracle.

Pseudorandomness with A-Priori Information: From the above lemma, we can derive the following theorem by using Hoefding Inequality [19].

Theorem 4. *For any encryption scheme $(\mathcal{G}, \mathcal{E}, \mathcal{D})$ that is* WIND-CPA (WIND-CCA *respectively), $\langle \mathcal{E}(pk, x), H(x, r) \rangle$ and $\langle \mathcal{E}(pk, x), H(y, r) \rangle$ are computationally indistinguishable* (CI-CCA *respectively*) *where $x, y \stackrel{R}{\leftarrow} Z_q^*$, and $r \stackrel{R}{\leftarrow} Q$.*

The proof is given in Appendix C. Again, we only prove the part corresponding to CCA security, and the part corresponding to CPA security can be obtained straightforwardly.

6 Constructing an Encryption Based MT-authenticator

The concept of MT-authenticator was introduced by Bellare et al. [4] in constructing authenticated key exchange (AKE) protocols. An MT-authenticator is a protocol 'emulating' the ideal message transfer functionality. By replacing every message transfer with an MT-authenticator, the AKE protocol is immune to any impersonation attacks. So authentication and session key security can be separated in the design of AKE protocols. Later, the work was further extended by Canetti and Krawczyk [9].

In [4], several MT-authenticators are constructed, and one of them is based on a PKE scheme and an MAC scheme, which is called 'Encryption-based MT-authenticator'.

$$P_i \leftarrow P_j : ENC_{P_i}(N_j)$$
$$P_i \rightarrow P_j : m,\ MAC_{N_j}(m, P_j)$$

Encryption-based MT-authenticator

In this protocol, $N_j \stackrel{R}{\leftarrow} \{0,1\}^k$ is a random challenge. ENC_{P_i} denotes the public key encryption function of P_i and MAC_{N_j} denotes an MAC under key N_j. Bellare et al. proved that the above protocol is an MT-authenticator if the PKE scheme is IND-CCA and the MAC scheme is unforgeable against chosen message attacks. However, it's easy to observe that in the security proof we merely need to guarantee that the ciphertexts of two randomly selected message are indistinguishable, thus a WIND-CCA secure PKE scheme would suffice. The proof essentially follows the original one in [4] and thus is omitted here.

7 Conclusion

In the face of the difficulty in constructing efficient IND-CCA PKE schemes in the standard model, and the fact that IND-CCA is a sufficient but not necessary condition for many PKE-based applications, we define a new type of indistinguishability against adaptive chosen ciphertext attacks (or WIND-CCA in short). We show that efficient PKE schemes fulfilling this new security definition can be easily constructed in the standard model. We also demonstrate that WIND-CCA schemes are sufficient for many PKE-based applications, such as constructing more flexible CCA secure KEM (and hence CCA secure hybrid encryption schemes), providing pseudorandomness with a priori information, constructing MT-authenticators, and so on. With this new security notion, we hope more practical PKE schemes can be built in the future.

References

1. Masayuki Abe and Serge Fehr. Perfect NIZK with adaptive soundness. In *TCC*, pages 118–136, 2007.
2. Masayuki Abe, Rosario Gennaro, Kaoru Kurosawa, and Victor Shoup. Tag-KEM/DEM: A new framework for hybrid encryption and a new analysis of Kurosawa-Desmedt KEM. In *EUROCRYPT*, pages 128–146, 2005.
3. Boaz Barak, Ran Canetti, Jesper Buus Nielsen, and Rafael Pass. Universally composable protocols with relaxed set-up assumptions. In *FOCS*, pages 186–195, 2004.
4. M. Bellare, R. Canetti, and H. Krawczyk. A modular approach to the design and analysis of authentication and key exchange protocols. In *Proc. 30th ACM Symp. on Theory of Computing*, pages 419–428. ACM, May 1998.
5. M. Bellare and P. Rogaway. Entity authentication and key distribution. In *Proc. CRYPTO 93*, pages 232–249. Springer-Verlag, 1994. LNCS 773.
6. Mihir Bellare, Alexandra Boldyreva, and Adriana Palacio. An uninstantiable random-oracle-model scheme for a hybrid-encryption problem. In *EUROCRYPT*, pages 171–188, 2004.
7. Mihir Bellare and Adriana Palacio. The knowledge-of-exponent assumptions and 3-round zero-knowledge protocols. In *CRYPTO*, pages 273–289, 2004.
8. Dan Boneh. Simplified OAEP for the RSA and Rabin functions. In *CRYPTO*, pages 275–291, 2001.

9. R. Canetti and H. Krawczyk. Analysis of key-exchange protocols and their use for building secure channels. In *Proc. EUROCRYPT 2001*, pages 453–474. Springer-Verlag, 2001. LNCS 2045. `http://eprint.iacr.org/2001/040/`.
10. Ran Canetti. Towards realizing random oracles: Hash functions that hide all partial information. In *Advances in Cryptology - CRYPTO97*, pages 455–469. Springer-Verlag, 1997. LNCS 1294.
11. Ran Canetti, Oded Goldreich, and Shai Halevi. The random oracle methodology, revisited. *J. ACM*, 51(4):557–594, 2004.
12. David Chaum and Torben P. Pedersen. Wallet databases with observers. In *CRYPTO*, pages 89–105, 1992.
13. Ronald Cramer and Victor Shoup. A practical public key cryptosystem provably secure against adaptive chosen ciphertext attack. In *CRYPTO*, pages 13–25, 1998.
14. Ronald Cramer and Victor Shoup. Universal hash proofs and a paradigm for adaptive chosen ciphertext secure public-key encryption. In *EUROCRYPT*, pages 45–64, 2002.
15. Ivan Damgård. Towards practical public key systems secure against chosen ciphertext attacks. In *CRYPTO*, pages 445–456, 1991.
16. Ivan Damgård, Nelly Fazio, and Antonio Nicolosi. Non-interactive zero-knowledge from homomorphic encryption. In *TCC*, pages 41–59, 2006.
17. Danny Dolev, Cynthia Dwork, and Moni Naor. Nonmalleable cryptography. *SIAM J. Comput.*, 30(2):391–437, 2000.
18. Eiichiro Fujisaki, Tatsuaki Okamoto, David Pointcheval, and Jacques Stern. RSA-OAEP is secure under the RSA assumption. In *CRYPTO*, pages 260–274, 2001.
19. O. Goldreich. *Foundations of Cryptography: Basic Tools*. Cambridge University Press, 2001.
20. Eike Kiltz. Chosen-ciphertext secure key-encapsulation based on gap hashed diffie-hellman. In *Public Key Cryptography*, pages 282–297, 2007.
21. Kaoru Kurosawa and Yvo Desmedt. A new paradigm of hybrid encryption scheme. In *CRYPTO*, pages 426–442, 2004.
22. Yehuda Lindell. A simpler construction of CCA2-Secure public-key encryption under general assumptions. In *EUROCRYPT*, pages 241–254, 2003.
23. Moni Naor and Moti Yung. Public-key cryptosystems provably secure against chosen ciphertext attacks. In *STOC*, pages 427–437, 1990.
24. Pascal Paillier. Public-key cryptosystems based on composite degree residuosity classes. In *EUROCRYPT*, pages 223–238, 1999.
25. Charles Rackoff and Daniel R. Simon. Non-interactive zero-knowledge proof of knowledge and chosen ciphertext attack. In *CRYPTO*, pages 433–444, 1991.
26. Amit Sahai. Non-malleable non-interactive zero knowledge and adaptive chosen-ciphertext security. In *FOCS*, pages 543–553, 1999.
27. Alfredo De Santis, Giovanni Di Crescenzo, Rafail Ostrovsky, Giuseppe Persiano, and Amit Sahai. Robust non-interactive zero knowledge. In *CRYPTO*, pages 566–598, 2001.
28. Victor Shoup. Lower bounds for discrete logarithms and related problems. In *EUROCRYPT*, pages 256–266, 1997.
29. Victor Shoup. Using hash functions as a hedge against chosen ciphertext attack. In *EUROCRYPT*, pages 275–288, 2000.
30. Victor Shoup. OAEP reconsidered. In *CRYPTO*, pages 239–259, 2001.

A Proof of Theorem 3

Proof. Let $PKE = (\mathcal{G}, \mathcal{E}, \mathcal{D})$ be a WIND-CCA public key encryption scheme, and $(pk, sk) \leftarrow \mathcal{G}(1^k)$. Let $m \overset{R}{\leftarrow} \mathcal{M}(pk)$, $K = H(m)$ and $\psi \leftarrow \mathcal{E}(pk, m)$. Let K' be a symmetric key chosen randomly from the key space. Assume that there exists a PPT algorithm $\mathcal{A}$ such that

$$\Pr\left(\mathcal{A}^{\mathcal{O}_{sk}}(K, \psi) = 1\right) - \Pr\left(\mathcal{A}^{\mathcal{O}_{sk}}(K', \psi) = 1\right) \geq \epsilon$$

We then construct another adversary $\mathcal{B}$ which breaks PKE in the WIND-CCA game. $\mathcal{B}$ is given $m_0, m_1 \stackrel{R}{\leftarrow} \mathcal{M}(pk)$ and $c \leftarrow \mathcal{E}(pk, m_b)$ where b denotes the random coin of the challenger. $\mathcal{B}$ first calculates $K_0 = H(m_0)$, $K_1 = H(m_1)$ and tosses a coin b'. $\mathcal{B}$ then runs $\mathcal{A}$ on input $K_{b'}, c$. $\mathcal{B}$ answers $\mathcal{A}$'s decryption queries using its own decryption oracle. Finally, if $\mathcal{A}$ outputs 1, $\mathcal{B}$ outputs b'; otherwise it outputs $1 - b'$. It is easy to see that

$$\Pr\left(\mathcal{B} \text{ guesses b correctly}\right) \geq \frac{1}{2} + \frac{\epsilon}{2}$$

□

B Proof of Lemma 1

Proof. The theorem is proved by a hybrid argument. First, we define $T_0 = \langle \mathcal{E}(pk, m), g, g^a, g^b, g^{ab} \rangle$ and $T_1 = \langle \mathcal{E}(pk, m), g, g^a, g^b, g^c \rangle$ where $m, a, b, c \stackrel{R}{\leftarrow} Z_q^*$. Then we have the following claims:

Claim 3 *T_0 and T_D are* CI-CCA.

Proof. Suppose there is a PPT algorithm $\mathcal{A}$ such that

$$\Pr\left(\mathcal{A}^{\mathcal{O}_{sk}}(T_D) = 1\right) - \Pr\left(\mathcal{A}^{\mathcal{O}_{sk}}(T_0) = 1\right) \geq \delta$$

we construct an algorithm $\mathcal{B}$ that breaks the encryption scheme in the WIND-CCA game. $\mathcal{B}$ is given m, a, y, where m and a denote the two randomly selected messages and y denote the challenging ciphertext. After receiving the challenge y, $\mathcal{B}$ chooses $b \stackrel{R}{\leftarrow} Z_q^*$ and runs $\mathcal{A}$ on input y, g, g^a, g^b, g^{ab}. It answers $\mathcal{A}$'s decryption queries by asking its own decryption oracle. Finally, $\mathcal{B}$ outputs what $\mathcal{A}$ outputs. It is obvious that $\mathcal{B}$ succeeds with probability at least $\frac{1}{2} + \frac{\delta}{2}$. □

Similarly, we can prove the following claim:

Claim 4 *T_1 and T_R are* CI-CCA.

Now assume that there is an algorithm $\mathcal{D}$ such that

$$\Pr\left(\mathcal{D}^{\mathcal{O}_{sk}}(T_D) = 1\right) - \Pr\left(\mathcal{D}^{\mathcal{O}_{sk}}(T_R) = 1\right) \geq \epsilon$$

Then we construct another algorithm $\mathcal{D}'$ that breaks the DDH assumption. $\mathcal{D}'$ is given $T = (g, g^a, g^b, g^z)$ where $z = ab$ or $z \stackrel{R}{\leftarrow} Z_q^*$. $\mathcal{D}'$ first runs the key generation algorithm of the public key encryption scheme to generate the encryption/decryption key pair (pk, sk). Then $\mathcal{D}'$ selects $m \stackrel{R}{\leftarrow} Z_q^*$, and runs $\mathcal{D}$ on input $(\mathcal{E}(pk, m), T)$. It answers $\mathcal{D}$'s decryption queries using the decryption key. Finally, $\mathcal{D}'$ outputs what $\mathcal{D}$ outputs.

1. If $z = ab$, we have that $(\mathcal{E}(pk, m), T) = T_0$. According to Claim 3, $\mathcal{D}$ outputs 1 with probability $\Pr(\mathcal{D}^{\mathcal{O}_D}(T_D) = 1) + \epsilon_1$ where ϵ_1 is negligible.

2. If $z \xleftarrow{R} Z_q^*$, we have that $(\mathcal{E}(pk, m), T) = T_1$. According to Claim 4, $\mathcal{D}$ outputs 1 with probability $\Pr(\mathcal{D}^{\mathcal{O}_D}(T_R) = 1) + \epsilon_2$ where ϵ_2 is negligible.

   Therefore, we can get that

$$\Pr\left(\mathcal{D}'(g, g^a, g^b, g^{ab}) = 1\right) - \Pr\left(\mathcal{D}'(g, g^a, g^b, g^z) = 1\right) \geq \epsilon + \epsilon_1 - \epsilon_2$$

□

C Proof of Theorem 4

Hoefding Inequality: Let $X_1, X_2, ..., X_n$ be n independent random variables with the same probability distribution, each ranging over the real interval $[a, b]$, and let μ denote the expected value of each of these variables. Then, for every $\epsilon > 0$,

$$\Pr\left(\left|\frac{\Sigma_{i=1}^n X_i}{n} - \mu\right| > \epsilon\right) < 2e^{-\frac{2\epsilon^2}{(b-a)^2}n}$$

Proof. Assume that there exists a distinguisher $\mathcal{D}$ and a polynomial ρ such that $\forall x, y \xleftarrow{R} Z_q^*, r \xleftarrow{R} Q$

$$\Pr\left(\mathcal{D}^{\mathcal{O}_{sk}}(\mathcal{E}(pk, x), r, r^x) = 1\right) - \Pr\left(\mathcal{D}^{\mathcal{O}_{sk}}(\mathcal{E}(pk, x), r, r^y) = 1\right) \geq \frac{1}{\rho} \tag{1}$$

Then we construct an distinguisher $\mathcal{D}'$ that distinguishes between $(\mathcal{E}(pk, a), g, g^a, g^b, g^{ab})$ and $(\mathcal{E}(pk, a), g, g^a, g^b, g^c)$. First, we define

$$\hat{P}_1 = \Pr(\mathcal{D}^{\mathcal{O}_{sk}}(\mathcal{E}(pk, x), r, r^x) = 1)$$

$$\hat{P}_2 = \Pr(\mathcal{D}^{\mathcal{O}_{sk}}(\mathcal{E}(pk, x), r, r^y) = 1)$$

Given $(\mathcal{E}(pk, a), g, g^a, g^b, g^z)$ where $z = ab$ or $z \xleftarrow{R} Z_q^*$, $\mathcal{D}'$ runs as follows:

For $i = 1$ to ρ^3

1. Choose $r_i, r_i' \xleftarrow{R} Z_q^*$
2. Run algorithm $\mathcal{D}$ with input $(\mathcal{E}(pk, a), g^{r_i}, (g^a)^{r_i})$ and get $P_1^i = \mathcal{D}^{\mathcal{O}_{sk}}(\mathcal{E}(pk, a), g^{r_i}, (g^a)^{r_i})$. $\mathcal{D}'$ answers $\mathcal{D}$'s decryption queries using its own oracle.
3. Run algorithm $\mathcal{D}$ again with input $(\mathcal{E}(pk, a), (g^b)^{r_i'}, (g^z)^{r_i'})$ and get $P_2^i = \mathcal{D}^{\mathcal{O}_{sk}}(\mathcal{E}(pk, a), (g^b)^{r_i'}, (g^z)^{r_i'})$. $\mathcal{D}'$ answers $\mathcal{D}$'s decryption queries using its own oracle.

Then $\mathcal{D}'$ computes $P_1 = \Sigma_{i=1}^{\rho^3} P_1^i$ and $P_2 = \Sigma_{i=1}^{\rho^3} P_2^i$. If $|P_1 - P_2| > \frac{\rho^2}{2}$, it outputs 0; otherwise, it outputs 1.

1. If $z = ab$, both P_1 and P_2 are the sums of ρ^3 independent samples from a distribution over $\{0, 1\}$ with mean $\rho^3 \hat{P}_1$. By Hoefding Inequality, it follows that $\Pr(|P_1 - P_2| > \frac{\rho^2}{2}) < \exp(-\rho)$.

2. If $z \stackrel{R}{\leftarrow} Z_q^*$, P_1 is the sum of ρ^3 independent samples from a distribution over $\{0,1\}$ with mean $\rho^3\hat{P}_1$ (denoted by μ_1), and P_2 is the sum of ρ^3 independent samples from a distribution over $\{0,1\}$ with mean $\rho^3\hat{P}_2$ (denoted by μ_2). By equation (1), it follows that $\mu_1 - \mu_2 \geq \rho^2$. Then by Hoefding Inequality again, it follows that $\Pr(|P_1 - \mu_1| > \frac{\rho^2}{4}) < \exp(-\rho)$, and $\Pr(|P_2 - \mu_2| > \frac{\rho^2}{4}) < \exp(-\rho)$. Hence, we have that with overwhelming probability, $|P_1 - P_2| > \frac{\rho^2}{2}$ holds. □