Computers and Electrical Engineering 38 (2012) 975–985

Contents lists available at SciVerse ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

# Public key encryption without random oracle made truly practical [☆,☆☆]

Puwen Wei [a,b,*], Xiaoyun Wang [c], Yuliang Zheng [d]

[a] School of Mathematics, Shandong University, Jinan 250100, China
[b] Key Laboratory of Cryptologic Technology and Information Security, Ministry of Education, Jinan 250100, China
[c] Center for Advanced Study, Tsinghua University, Beijing 100084, China
[d] Department of Software and Information Systems, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

ARTICLE INFO

*Article history:*
Received 15 November 2010
Received in revised form 1 February 2012
Accepted 1 February 2012
Available online 23 February 2012

ABSTRACT

In this paper, we report our success in identifying an efficient public key encryption scheme whose formal security proof does not require a random oracle. Specifically, we focus our attention on a universal hash based public key encryption scheme proposed by Zheng and Seberry at Crypto'92. Although Zheng and Seberry's encryption scheme is very simple and efficient, its reductionist security proof has not been provided. We show how to tweak the Zheng–Seberry scheme so that the resultant scheme not only preserves the efficiency of the original scheme but also admits provable security against adaptive chosen ciphertext attack without random oracle. For the security proof, our first attempt is based on a strong assumption called the oracle Diffie–Hellman$^+$ assumption. This is followed by a more challenging proof that employs a weaker assumption called the adaptive decisional Diffie–Hellman assumption, which is in alignment with adaptively secure assumptions advocated by Pandey, Pass and Vaikuntanathan.

## 1. Introduction

The notion of chosen ciphertext security was introduced by Naor and Yung [1]. Rackoff and Simon [2] provided a stronger notion called indistinguishability under adaptive chosen ciphertext attack (IND-CCA2), which is equivalent to the notion of non-malleability [3]. Adaptive chosen ciphertext security has since become a standard notion for the security of public key encryption.

A significant number of efforts have been devoted by researchers to the construction of public key encryption that is secure against adaptive chosen ciphertext attack. Some of the research outcomes of these efforts were based on non-interactive zero-knowledge proofs [3], which were not quite practical in real world applications. To construct an efficient encryption scheme, many encryption techniques have been proposed in the so-called random oracle model [4–6]. The random oracle model, however, is one of the most controversial issues in cryptography. A notable argument against the random oracle model was made by Canetti et al. [7] who demonstrated that there existed cryptographic schemes that were secure in the random oracle model but insecure for any instantiation of a random oracle. Recently, Leurent and Nguyen [8] showed that instantiations of full domain hash functions (random oracles) proposed in the literature are insecure. They also advocated to assess carefully the impact of potential flaws in random oracle instantiations on a system that relies on such instantiations.

[☆] An extended abstract appears in the Proceedings of the 11th International Conference on Information and Communications Security, LNCS 5927, pp. 107–120, Springer-Verlag, 2009. The first author was supported by the National Natural Science Foundation of China under Grant No. 61103237; part of the third author's work was done while visiting Shandong University on a Changjiang Scholars program sponsored by the Chinese Ministry of Education and Li Ka Shing Foundation in Hong Kong.

[☆☆] Reviews processed and approved for publication by Editor-in-Chief Dr. Manu Malek.

[*] Corresponding author at: School of Mathematics, Shandong University, Jinan 250100, China. Fax: +86 531 88363873.
*E-mail addresses:* pwei@sdu.edu.cn (P. Wei), xiaoyunwang@tsinghua.edu.cn (X. Wang), yzheng@uncc.edu (Y. Zheng).

0045-7906/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compeleceng.2012.02.001

To address the concern over random oracles, an obvious approach is to design a public key encryption scheme that does not rely on a random oracle for its security against adaptive chosen ciphertext attack. The often cited encryption scheme proposed by Cramer and Shoup [9] represents the first concrete result in this line of research. A multiple number of techniques have since been proposed and studied by many researchers. Most of these techniques, however, share a common drawback that impedes their possible adoption in practice, that is, they generally require at least a few times more computation than their random oracle based counterparts.

Given the superiority in computational efficiency of random oracle based encryption, it is a shared view among most researchers that alternative encryption techniques without random oracles will not be able to win over practitioners unless these alternatives afford a computational speed comparable to that enjoyed by random oracle based techniques.

Aside from computational efficiency, another major advantage of random oracle based schemes [4–6] lies in its simplicity. To preserve the simplicity while not relying on a random oracle for security proofs, new computational assumptions have been examined. One such effort was made by Pandey et al. [10] who introduced a few complexity theoretical hardness assumptions that abstracted out concrete properties of a random oracle. Based on these assumptions, they were able to solve a number of open problems, including the construction of a non-interactive concurrently non-malleable string commitment. Their results point to an interesting approach towards designing efficient and provably secure cryptographic schemes without random oracles. We note that although these assumptions are stronger than traditional cryptographic hardness assumptions, they seem quite reasonable and it is conceivable that, like many other assumptions in the field such as the decisional Diffie–Hellman assumption (DDH), this type of new assumptions may gain wider acceptance after further screening by peers in the field.

### 1.1. Our contribution

The goal of this paper is to search for a public key encryption scheme that (1) does not rely on a random oracle for its adaptive chosen ciphertext security, and (2) is truly practical in that it requires no more exponentiations of large integers than does a comparable random oracle based scheme. To achieve our goal, our first attempt is to prove Zheng and Seberry's encryption scheme based on the oracle Diffie–Hellman assumption$^+$ (ODH$^+$). However, ODH$^+$ is shown to be a very strong assumption. Hence, in order to use a more reasonable assumption, we examine a variant of Pandey et al.'s assumption [10], called the adaptive DDH assumption. Based on the adaptive DDH assumption, a modified version of Zheng and Seberry's encryption scheme proposed in [11] is proved to be adaptive chosen ciphertext secure without a random oracle.

Zheng and Seberry [11] proposed three simple methods for immunizing public key cryptosystems against chosen ciphertext attacks. The nature of the three methods is the same. They immunize a public key cryptosystem by appending to each ciphertext a tag that is correlated to the message to be encrypted. Soldera et al. [12] showed a potential weakness of the first scheme, denoted by Zheng–Seberry$_{1wh}$, in some special circumstances. Based on the gap Diffie–Hellman assumption (GDH), Baek and Zheng [13] provided a security proof for the slightly modified version of Zheng–Seberry$_{1wh}$, in the random oracle model, leaving as an open problem proofs for the other two schemes. The focus of this paper is to modify the second scheme in [11], denoted by Zheng–Seberry$_{uh}$, so that the resultant scheme is adaptive chosen ciphertext secure (see Section 5). The scheme Zheng–Seberry$_{uh}$ is worth studying for the following reasons: First, the scheme immunizes public key encryption against adaptive chosen ciphertext attacks with the help of a universal hash function. This allows the scheme to steer clear of a one-way hash function with non-standard output size, whereby successfully averting potential risks recently discovered in [8]. Second, the input length of a plaintext can be arbitrary, while the overhead of the corresponding ciphertext is a constant. As a result, the ratio between the length of the ciphertext and that of the plaintext can be close to 1 as the length of the plaintext increases.

### 1.2. Related work

Hybrid encryption, which is also known as the KEM–DEM approach [11], applies a public key cryptosystem to encapsulate the key of a symmetric cryptosystem (KEM) and the symmetric cryptosystem is subsequently used to conceal data (DEM). Cramer and Shoup first generalized the notion in their work [14,15]. Kurosawa and Desmedt [16] later presented a more efficient hybrid encryption scheme by using a KEM which is not necessarily adaptive chosen ciphertext secure. More recently, Kiltz et al. [17] improved on the Kurosawa–Desmedt technique and proposed a new approach to designing adaptive chosen ciphertext secure hybrid encryption schemes without a random oracle. Compared with Kiltz et al.'s concrete scheme which relies on the DDH assumption and AE-OT[1] secure symmetric encryption, our modified Zheng–Seberry$_{uh}$ scheme is conceptually much simpler and relies only on the adaptive DDH assumption. More important, this newly modified scheme requires significantly less computation time than Kiltz et al.'s.

Another important progress was made by Hofheinz and Kiltz [18] recently. They proposed a new public key encryption scheme based on factoring. Their scheme requires only roughly two exponentiations in encryption and roughly one exponentiation in decryption. (Here, "roughly" two or one exponentiation means two or one full exponentiation and additional

[1] According to [17], a symmetric cipher is AE-OT secure if it satisfies (one-time) ciphertext indistinguishability (IND-OT) and (one-time) ciphertext integrity (INT-OT).

exponentiations with small exponents.) While for the encryption schemes based on discrete logarithm, DHIES [19] is one of the most efficient schemes without random oracle.

Compared with DHIES which relies on the oracle Diffie–Hellman (ODH) assumption together with the security of symmetric encryption and a message authentication code (MAC), our modified scheme relies on the adaptive DDH assumption only and preserves the computational efficiency of Zheng–Seberry$_{uh}$. However, it is fair to say that our modified Zheng–Seberry scheme and DHIES are comparable, each having its own pros and cons in practice. With DHIES, all three assumptions on symmetric encryption, MAC and ODH are responsible for the security of DHIES and it is relatively easy to select practical candidates to instantiate functions underlying the assumptions. With our modified Zheng–Seberry scheme, the adaptive DDH assumption which is solely responsible for the security of the scheme is slightly stronger than the ODH assumption required by DHIES.

## 2. Preliminaries

*Notation and definition:* $|X|$ denotes the length of a binary string $X$ or the size of (or number of elements in) a set $X$. $x \overset{R}{\leftarrow} X$ denotes picking an element $x$ from $X$ uniformly at random. $y \leftarrow A(x)$ denotes the experiment of running an algorithm $A$ on input $x$ and outputting $y$. PPT denotes probabilistic polynomial time. $x \| y$ denotes the concatenation of strings $x$ and $y$. A function $\mu : \mathbb{N} \to \mathbb{R}$ is called negligible in $n$ if for every positive polynomial $p(\cdot)$ and all sufficiently large $n$'s, we have $\mu(n)<1/p(n)$.

*Decisional Diffie–Hellman assumption:* Let $\mathbb{G}$ be an Abelian group with prime order $q$. $g \in \mathbb{G}$ is a generator of $\mathbb{G}$. The decisional Diffie–Hellman assumption states that, for any PPT algorithm $A$, there exists a negligible function $\mu$ such that for all sufficiently large $|q|$

$$\left|\Pr\left[a,b \overset{R}{\leftarrow} \mathbb{Z}_q^* : A(q,g,g^a,g^b,g^{ab}) = 1\right] - \Pr\left[a,b,c \overset{R}{\leftarrow} \mathbb{Z}_q^* : A(q,g,g^a,g^b,g^c) = 1\right]\right| \leqslant \mu(|q|)$$

where the probability is taken over the random choice of $q$, $g$, $a$, $b$, $c$ and the coin-tosses of $A$.

*Universal hashing [20]:* A family of functions $H$: $\{0,1\}^P \to \{0,1\}^l$ is a universal family of hash functions if, for every $x_1 \neq x_2 \in \{0,1\}^P$ and every $y_1, y_2 \in \{0,1\}^l$, the number of functions in $H$ mapping $x_1$ to $y_1$ and $x_2$ to $y_2$ is precisely $|H|/2^{2l}$, where $|H|$ denotes the number of functions in $H$. Simply speaking, if $h$ is chosen uniformly from the universal class of hash functions $H$, $h(x_1)$ and $h(x_2)$ are distributed, uniformly and independently of each other, over $\{0,1\}^l \times \{0,1\}^l$.

For the security proof in this paper, we need the following lemma whose proof can be found in [21].

**Lemma 1** [21]. *Let $S_1$,$S_2$, and $S_3$ be events defined on a probability space such that $Pr[S_1 \wedge \neg S_3] = Pr[S_2 \wedge \neg S_3]$. Then we have $|Pr[S_1] - Pr[S_2]| \leqslant Pr[S_3]$.*

## 3. New assumptions

In this section, we give the definitions of adaptive DDH and other related assumptions. First, we recall the definition of an adaptive one-to-one one-way function introduced in [10]. In the definition, an adversary picks an index $tag^*$ and is given $y^* = f_{tag^*}(x^*)$ for a random $x^*$ in the domain of $f_{tag^*}(x)$. The aim of the adversary is to compute $x^*$. The difference between the traditional definition for an one-way function and the one in [10] is that the adversary in [10] has access to a "magic oracle" that on input $(tag,y)$ with $tag \neq tag^*$, returns $f_{tag}^{-1}(y)$. The security requirement is that the adversary can compute $x^*$ with a negligible probability only, even if the adversary can get help from the "magic oracle".

**Definition 1** (*Family of adaptive one-to-one one-way functions [10]*). A family of injective one-way functions $\mathcal{F} = \{f_{tag} : D_{tag} \to \{0,1\}^*\}_{tag \in \{0,1\}^n}$ is called adaptively secure if

- There is an efficient randomized domain sampler $D$, which on input $tag \in \{0,1\}^n$, outputs a random element in $D_{tag}$. There is a deterministic polynomial algorithm $M$ such that for all $tag \in \{0,1\}^n$ and for all $x \in D_{tag}$, $M(tag,x) = f_{tag}(x)$.
- Let $\mathcal{O}_{tag}(\cdot,\cdot)$ denote an oracle that, on input $tag'$ and $y$, outputs $f_{tag'}^{-1}(y)$ if $tag' \neq tag$ with $|tag'| = |tag|$, and $\perp$ otherwise. The family $\mathcal{F}$ is adaptively secure if, for any probabilistic polynomial time adversary $A$ which has access to the oracle $\mathcal{O}_{tag}(\cdot,\cdot)$, there exists a negligible function $\mu$ such that for all $n$, and for all tags $tag \in \{0,1\}^n$,

$$\Pr\left[x \leftarrow D_{tag} : A^{\mathcal{O}_{\sqcup\dashv\}}(\cdot,\cdot)}(tag, f_{tag}(x)) = x\right] \leqslant \mu(n)$$

where the probability is over the random choice of $x$ and the coin tosses of $A$.

Similarly to the definition of adaptive one-way function, the definition of adaptive pseudorandom generator $G_{tag}$ requires that the adversary cannot tell the output of $G_{tag^*}$ from a random string, even if the adversary can get help from a magic oracle that, on input $(tag,y)$ with $tag \neq tag^*$, returns 0 or 1 depending on whether $y$ is in the range of $G_{tag}$ or not.

**Definition 2** (*Adaptive PRG [10]*). Let a family of functions $\mathcal{G} = \{G_{tag} : \{0,1\}^n \to \{0,1\}^{s(n)}\}_{tag \in \{0,1\}^n}$ be a pseudorandom generator (PRG). And let $\mathcal{O}_{tag}(\cdot,\cdot)$ denote an oracle that, on input $(tag', y)$ such that $tag' \neq tag, |tag'| = |tag|$, outputs 1 if $y$ is in the range of $G_{tag'}$, and 0 otherwise.

We say that $\mathcal{G}$ is an adaptively secure PRG if, for any probability polynomial-time adversary $A$ which has access to the oracle $\mathcal{O}_{tag}(\cdot,\cdot)$, there exists a negligible function $\mu$ such that for all $n$ and for all tags $tag \in \{0,1\}^n$,

$$|\Pr[y \leftarrow G_{tag}(U_n) : A^{\mathcal{O}_{tag}(\cdot,\cdot)}(y) = 1] - \Pr[y \leftarrow U_{s(n)} : A^{\mathcal{O}_{tag}(\cdot,\cdot)}(y) = 1]| \leqslant \mu(n)$$

where the probability is over the random choice of $y$ and the coin-tosses of the adversary $A$.

Combining the above two definitions, we have a definition for a variant of the adaptive PRG. The variant is similar to Definition 2 except that, the adversary $A$ has some auxiliary information $f_{tag}(x)$ on a seed $x$ and interacts with the oracle $\mathcal{O}_{tag}(\cdot,\cdot,\cdot)$.

**Definition 3** (*Auxiliary adaptive PRG*). Let $\mathcal{G} = \{G_{tag} : \{0,1\}^n \to \{0,1\}^{s(n)}\}_{tag \in \{0,1\}^n}$ be a pseudorandom generator (PRG). And let $\mathcal{O}_{tag}(\cdot,\cdot,\cdot)$ denote an oracle that, on input $(tag', f_{tag'}(x), y)$ such that $tag' \neq tag$, $|tag'| = |tag|$, outputs the seed $x$ if $y = G_{tag'}(x)$ and $x$ is consistent with its auxiliary information $f_{tag'}(x)$; $\mathcal{O}_{tag}(\cdot,\cdot,\cdot)$ outputs $\perp$ otherwise.

We say that the PRG $\mathcal{G}$ is adaptively secure if, for any probability polynomial-time adversary $A$ which has the auxiliary information $f_{tag}(x)$ on the seed $x$ and has access to the oracle $\mathcal{O}_{tag}(\cdot,\cdot,\cdot)$, there exists a negligible function $\mu$ such that for all $n$ and for all tags $tag \in \{0,1\}^n$,

$$\left|Adv_A^{real} - Adv_A^{rand}\right| \leqslant \mu(n)$$

where $Adv_A^{real}$ denotes $\Pr[x \leftarrow U_n : A^{\mathcal{O}_{\sqcup\dashv\}}(\cdot,\cdot,\cdot)}(f_{tag}(x), G_{tag}(x)) = 1], Adv_A^{rand}$ denotes $\Pr[y \leftarrow U_{s(n)} : A^{\mathcal{O}_{tag}(\cdot,\cdot,\cdot)}(f_{tag}(x), y) = 1]$ and the probability is over the random choice of $y$ and $x$, and the coin-tosses of $A$.

Definition 3 is a combination of Definitions 2 and 1 in that the auxiliary information on $x$ in Definition 3 can be replaced by a one-way function $f(x)$ and the inversion oracle $\mathcal{O}_{tag}(\cdot,\cdot,\cdot)$ plays the role of $\mathcal{O}_{tag}(\cdot,\cdot)$ in Definition 1. In addition, Definition 3 also implies that the adversary cannot invert the one-way function $f(x)$ even with the help from $\mathcal{O}_{tag}(\cdot,\cdot,\cdot)$. A candidate construction for the auxiliary adaptive PRG, based on AES, is defined by $G_{tag}(x) = AES_x(tag\|0)\|AES_x(tag\|1)$.

From Definition 3 and the specific number theoretic assumption DDH, we derive a definition for the adaptive DDH assumption.

Let $\mathbb{G}$ be a group with prime order $q$. $g \in \mathbb{G}$ is the generator. $G_{tag}(\cdot) : \mathbb{G} \to \{0,1\}^*$ is a pseudorandom generator. $G_{tag}(\cdot)_{[i,\ldots,j]}$ denotes the substring from the $i$-th bit to the $j$-th bit of the output of $G_{tag}(\cdot)$.

**Definition 4** (*Adaptive DDH assumption*). Given $\{g, g^a, g^b, G_{g^a,g^b}(g^c)_{[1,\ldots,P+W]}\}$, it is computationally infeasible for any PPT distinguisher $D$ to tell whether $c = ab$, even if $D$ has access to an oracle $\mathcal{O}_{g^a,g^b}(\cdot,\cdot,\cdot)$, where $P$ and $W$ are polynomials in a security parameter. The oracle $\mathcal{O}_{g^a,g^b}(\cdot,\cdot,\cdot)$, on input $(g^{a'}, g^{b'}, G_{g^{a'},g^{b'}}(g^{c'})_{[P+1,\ldots,P+W]})$, outputs $g^{a'b'}$ if its input satisfies:

- $(g^{a'}, g^{b'}, G_{g^{a'},g^{b'}}(g^{c'})_{[P+1,\ldots,P+W]}) \neq (g^a, g^b, G_{g^a,g^b}(g^c)_{[P+1,\ldots,P+W]})$
- $G_{g^{a'},g^{b'}}(g^{a'b'})_{[P+1,\ldots,P+W]} = G_{g^{a'},g^{b'}}(g^{c'})_{[P+1,\ldots,P+W]}$

Otherwise, the oracle outputs $\perp$.

The definition implies that for any PPT $D$, there is a negligible function $\mu$ such that

$$\left|\Pr_{a,b,c \xleftarrow{R} Z_q}[D^{\mathcal{O}_{g^a,g^b}(\cdot,\cdot,\cdot)}(S) = 1] - \Pr_{a,b \xleftarrow{R} Z_q}[D^{\mathcal{O}_{g^a,g^b}(\cdot,\cdot,\cdot)}(S') = 1]\right| \leqslant \mu(n)$$

where $S = \left(g, g^a, g^b, G_{g^a,g^b}(g^c)_{[1,\ldots,P+W]}\right)$, $S' = (g, g^a, g^b, G_{g^a,g^b}(g^{ab})_{[1,\ldots,P+W]})$, and $n$ is the security parameter.

A quadruple $\{g, g^a, g^b, G_{g^a,g^b}(g^c)_{[1,\ldots,P+W]}\}$ satisfying $c = ab$ is called an adaptive DDH quadruple.

**Remark.** Comparing with Definition 3, $(g^a, g^b)$ is not only a tag but also represents some auxiliary information on $g^{ab}$. Note that it is not required that the length of the substring of $G_{g^{a'},g^{b'}}(g^{c'})$ in the adversary's query be equal to that of $G_{g^a,g^b}(g^c)_{[1,\ldots,P+W]}$. However, the length of $G_{g^{a'},g^{b'}}(g^{c'})_{[P+1,\ldots,P+W]}$, that is $W$, should be large enough to guarantee that the adversary can guess a "right" query with a negligible probability only. Intuitively, that means, in almost all cases, the oracle does not offer any "useful" help to the adversary. But that does not mean the adversary cannot provide the right query with a non-negligible probability. In fact, the adversary can randomly pick $a'$, $b'$ and generate the "right" query $(g^{a'}, g^{b'}, G_{g^{a'},g^{b'}}(g^{a'b'})_{[P+1,\ldots,P+W]})$ by himself. Although the oracle's answer to such a query does not carry any useful information to the adversary, it is important for simulation purposes in a security proof, which will be explained later.

### 3.1. Relationships to other assumptions

*HDH, ODH and SDH assumptions:* Abdalla, Bellare and Rogaway introduce three related notions, which are the hash Diffie–Hellman assumption (HDH), the oracle Diffie–Hellman (ODH) assumption and the strong Diffie–Hellman assumption (SDH)

[22,19]. The HDH assumption states that it is hard to tell $h(g^{ab})$ from a random string, where $h$ is a cryptographic hash function, even if you know $g^a$ and $g^b$. That is, it is hard to tell $\{g^a, g^b, h(g^{ab})\}$ from $\{g^a, g^b, U_n\}$, where $n$ is the output length of $h(\cdot)$. HDH is much weaker than DDH, but stronger than CDH. The difference between HDH and ODH is that the adversary can have access to the oracle $\mathcal{O}_b(X)$, which computes $\mathcal{O}_b(X) = h(X^b)$. As long as $\mathcal{O}_b(X)$ is not queried at $g^a$, $\mathcal{O}_b(X)$ seems to be useless to the adversary. Under the ODH assumption, they proved the adaptive chosen ciphertext security of their encryption scheme DHIES. It seems that the ODH assumption and the adaptive DDH assumption are similar in flavor. But the adversary's power in the adaptive DDH assumption is much more restricted, as the adversary can get the help of the oracle only if it can produce a useful and "right" query, which happens with only a negligible probability.

*Non-malleable pseudorandom generator:* In order to prove the security ($NM-CPA) of OAEP without random oracle, Boldyreva and Fischlin [23] fully instantiated OAEP by assuming special properties of the two pseudorandom generators $G$ and $H$ in OAEP. To be more precise, $G$ is a near-collision resistant trapdoor pseudorandom generator, which can recover the preimage $s$ of $G(s)$ according to the $k$ least significant bits of $G(s)$; $H$ is a non-malleable pseudorandom generator. Our adaptive DDH assumption is closely related to their assumption. To some extent, the adaptive DDH combines the above properties of $G$ and $H$, and takes advantage of concrete algebraic structures to replace the random oracle.

## 4. Original Zheng–Seberry$_{uh}$ encryption scheme (immunizing with universal hash function)

Assume that $H$: $\{0,1\}^* \to \{0,1\}^l$ is a family of universal hash functions. Each function in $H$ is specified by a string of exactly $Q$ bits. Denote by $h_s$ the function in $H$ that is specified by a string $s \in \{0,1\}^Q$. $L$ denotes an encryption label, which consists of public data. In addition, $m$ denotes a plaintext to be encrypted. Zheng–Seberry$_{uh}$ scheme is described in Table 1.

Note that there are two minor differences between Zheng–Seberry$_{uh}$ in Table 1 and the scheme B in [11]. The first difference is that a public label $L$ is employed in Table 1. Using such a label is a widely adopted practice and does not affect the security proof. The second difference is that in Table 1, the universal hash value is encrypted together with a message, which allows the use of a broader range of universal hash functions that may not necessarily hide all the information on a message.

### 4.1. Security proof of the original Zheng–Seberry$_{uh}$ scheme

In order to prove the adaptive chosen ciphertext security of Zheng–Seberry$_{uh}$ scheme, our first attempt is to use a modified ODH assumption, called ODH$^+$. To define ODH$^+$, a small adjustment is made on ODH. That is, we replace the hash function in ODH with a pseudorandom generator.

**Definition 5** (*Oracle Diffie–Hellman$^+$ or ODH$^+$ assumption*). Let $\mathbb{G}$ be a group with order $q$ and $g$ be a generator of $\mathbb{G}$, $G : \mathbb{G} \to \{0,1\}^{P+Q}$ be a pseudorandom generator, $A$ be an adversary, $k$ be a security parameter, $P$ and $Q$ be polynomials in $k$. Consider the following two experiments.

| | |
|---|---|
| • Experiment $Exp^{ODH^+-real}$ | • Experiment $Exp^{ODH^+-rand}$ |
| – $u \overset{R}{\leftarrow} \mathbb{Z}_q^*$, $U \leftarrow g^u$; | – $u \overset{R}{\leftarrow} \mathbb{Z}_q^*, U \leftarrow g^u$; |
| $v \overset{R}{\leftarrow} \mathbb{Z}_q^*, V \leftarrow g^v$; $W \leftarrow G(g^{uv})$ | $v \overset{R}{\leftarrow} \mathbb{Z}_q^*, V \leftarrow g^v$; $W \overset{R}{\leftarrow} \{0,1\}^{P+Q}$ |
| – $\mathcal{O}_{g^v}(X) \overset{def}{=} G(X^v)$ | – $\mathcal{O}_{g^v}(X) \overset{def}{=} G(X^v)$ |
| – $b \leftarrow A^{\mathcal{O}_{g^v}(\cdot)}(U, V, W)$ | – $b \leftarrow A^{\mathcal{O}_{g^v}(\cdot)}(U, V, W)$ |
| – Return $b$ | – Return $b$ |

Let the advantage of $A$ be

$$Adv^{ODH^+} = |\Pr[Exp^{ODH^+-real} = 1] - \Pr[Exp^{ODH^+-rand} = 1]|$$

**Table 1**
The original Zheng–Seberry$_{uh}$ scheme.

| *Zheng–Seberry$_{uh}$ scheme* | |
|---|---|
| *Public parameters:* A label $L$, the universal class of hash functions $H$: $\{0,1\}^* \to \{0,1\}^l$, the group $\mathbb{G}$, the generator $g$ of $\mathbb{G}$ with order $q$, and the pseudorandom generator $G : \mathbb{G} \to \{0, 1\}^*$. | |
| *Key generation:* Choose $x_A$ uniformly at random from $\mathbb{Z}_q^*$ and compute $y_A = g^{x_A}$. The public key is $y_A$ and the private key is $x_A$. | |
| **Encyrption** $E_{uhf}(y_A, m, L)$ | **Decryption** $D_{uhf}(x_A, c_1, c_2, L)$ |
| 1. $x \overset{R}{\leftarrow} \mathbb{Z}_q^*, r = y_A^x$ | 1. $r' = c_1^{x_A}, z' = G(r')_{[1,\ldots,P]},\ s' = G(r')_{[P+1,\ldots,P+Q]}$ |
| 2. $z = G(r)_{[1,\ldots,P]}$, $s = G(r)_{[P+1,\ldots,P+Q]}$ | 2. $m' \| t' = c_2 \oplus z'$, where $m' = (c_2 \oplus z')_{[1,\ldots,P-l]}$, $t' = (c_2 \oplus z')_{[P-l+1,\ldots,P]}$ |
| 3. $c_1 = g^x$, $c_2 = z \oplus (m \| t)$, where $t = h_s(m \| L)$ | 3. If $t' = h_{s'}(m' \| L)$, then output $m'$; otherwise output $\perp$ |
| Output the ciphertext $(c_1, c_2)$ | |

where $A$ is not allowed to call $\mathcal{O}_{g^v}(\cdot)$ on $g^u$. ODH$^+$ states that $Adv^{ODH^+} \leqslant \mu(k)$, where $\mu(k)$ is a negligible function.

A quadruple $\{g, g^u, g^v, G(g^c)_{[1,\ldots,P+Q]}\}$ satisfying $c = uv$ is called an ODH$^+$ quadruple.

However, it will become clear later that the ODH$^+$ assumption appears to be somewhat "too strong". For the ODH$^+$ assumption to hold, it will be necessary to assume more about the pseudorandom generator $G(\cdot)$ than what we would like it to have. To that end, we will modify the Zheng–Seberry$_{uh}$ scheme so that its security can be proven under a weaker, more reasonable assumption than the ODH$^+$ assumption. Before we achieve this final goal in Section 5, let us first proceed to prove the security of the Zheng–Seberry$_{uh}$ scheme under the ODH$^+$ assumption.

**Theorem 1.** *Under the ODH$^+$ assumption, Zheng–Seberry$_{uh}$ scheme is secure against adaptive chosen ciphertext attack.*

**Proof.** The main idea of the security proof is to construct three adaptive chosen ciphertext attack games, which are denoted by Game 1, Game 2 and Game 3, so that the adversary's views in these games are indistinguishable.

More specifically, Game 1 is a real run of a standard adaptive chosen ciphertext attack game and Game 3 is similar to Game 1, except that pseudorandom strings $(s^*, z^*)$ used in a target ciphertext are replaced with truly random strings. It turns out that in Game 3, as the choice of $\beta \in \{0,1\}$ is independent of the distribution of the ciphertext, the adversary can correctly guess $\beta$ with probability 1/2 only. Thus if we can show that the adversary's views in Game 1 and Game 3 are indistinguishable, then we know that the adversary can win the real adaptive chosen ciphertext attack game with a negligible advantage only. To that end, we introduce a new game, called Game 2, that can be considered a hybrid of Game 1 and Game 3. In Game 2, $(s^*, z^*)$ are generated by the pseudorandom generator but the seed is chosen randomly from $\mathbb{G}$. With the help of Game 2, we can proceed to show that both $|\Pr[Game\,1] - \Pr[Game\,2]|$ and $|\Pr[Game\,3] - \Pr[Game\,2]|$ are negligible, where $\Pr[Game\,i]$ denotes the probability that the adversary wins Game $i$, for $1 \leqslant i \leqslant 3$. From these facts, the theorem will be finally proven. We note that the public key and private key of the cryptosystem, the coin tosses of the adversary, and the bit $\beta$ all maintain identical values across the three games. Therefore, indistinguishability between Game 2 and Game 3 relies on the pseudorandomness of $G(\cdot)$. However, indistinguishability between Game 2 and Game 1 is not so obvious. We follow the proof method in [9] to measure the difference between Game 2 and Game 1. More specifically, if, for some adversary, $|\Pr[Game\,1] - \Pr[Game\,2]|$ is non-negligible, we construct an experiment to test the ODH$^+$ quadruple using the adversary. The proof relies only on properties of a universal hash function and the ODH$^+$ assumption.

Now we are ready to describe in detail the three games, namely Game 1, Game 2 and Game 3.

*Game 1:*
Game 1 is a real run of a standard adaptive chosen ciphertext attack game. First, a challenger runs the public parameter generation algorithm and the key generation algorithm to obtain the public parameters and the public/private key pair $(y_A, x_A)$. Then, the challenger gives the public parameters and the public key to an adversary. In Phase 1 and Phase 2 of the standard game, the challenger can answer the adversary's decryption query using his private key. After the adversary submits a pair of plaintexts $(m_0, m_1)$ in the challenge phase, the challenger creates a target ciphertext as follows: $c^* = (c_1^*, c_2^*) = (g^{x^*}, z^* \oplus (m_\beta \| t^*))$, where $r^* = y_A^{x^*}, s^* = G(r^*)_{[P+1,\ldots,P+Q]}, z^* = G(r^*)_{[1,\ldots,P]}, t^* = h_{s^*}(m_\beta \| L)$, and $\beta \overset{R}{\leftarrow} \{0,1\}$.
*Game 2:*
Game 2 is similar to Game 1 except that the target ciphertext $c^*$ is modified to $c_+^{**} = (g^{x^*}, z^{**} \oplus (m_\beta \| t^{**}))$, where $r^{**} \overset{R}{\leftarrow} \mathbb{G}$, $s^{**} = G(r^{**})_{[P+1,\ldots,P+Q]}$, $z^{**} = G(r^{**})_{[1,\ldots,P]}$, $t^{**} = h_{s^{**}}(m_\beta \| L)$. A further difference between Game 2 and Game 1 lies in the generation of the pseudorandom string. In Game 1, $(s^*, z^*)$ is generated by the seed $y_A^{x^*}$, whereas in Game 2, $(s^{**}, z^{**})$ is generated by a random element $r^{**} \in \mathbb{G}$.
*Game 3:*
Game 3 is similar to Game 2 except that the target ciphertext is modified to $c_+^* = (g^{x^*}, u_3 \oplus (m_\beta \| t_+^*))$, where $t_+^* = h_{u_2}(m_\beta \| L), u_2 \overset{R}{\leftarrow} \{0,1\}^Q$, and $u_3 \overset{R}{\leftarrow} \{0,1\}^P$. Since the distribution of $c_+^*$ is independent of the choice of $\beta$, the probability that the adversary can guess $\beta$ correctly in Game 3 is 1/2.

Next, we show that $|\Pr[Game\,1] - Pr[Game\,2]| \leqslant \mu(k)$, where $\mu(k)$ is a negligible function. Assume for contradiction that there exists a polynomial $p(k)$ such that, for infinitely many $k$'s, $|\Pr[Game\,1] - \Pr[Game\,2]| \geqslant 1/p(k)$, which means there exists an adversary $B$ for Game 1 and Game 2 such that $|\Pr[Game\,1] - \Pr[Game\,2]|$ is non-negligible. We show how to construct a PPT algorithm $A$ to break the ODH$^+$ assumption using $B$, by explicitly constructing an experiment of statistical test for the ODH$^+$ problem.

Given $\{g, g^a, g^b, G(g^c)_{[1,\ldots,P+Q]}\}$, $A$ sets $y_A = g^a$ and simulates the adaptive chosen ciphertext attack game for the adversary $B$ in the following experiment.

*Experiment:* $A$ sets the target ciphertext to

$$(g^b, G(g^c)_{[1,\ldots,P]} \oplus (m_\beta \| h_{G(g^c)_{[P+1,\ldots,P+Q]}}(m_\beta \| L))).$$

Notice that, in Game 1, Game 2 and Game 3, the challenger $A$ can perfectly decrypt the adversary's decryption query, since the challenger can use the private key. In the experiment, the challenger $A$, however, can only use the oracle $\mathcal{O}_{y_A}(\cdot)$ to answer the decryption query. Specifically, when the challenger receives a decryption query $(c_1, c_2)$, he will try to decrypt it as follows

1. If $c_1 \neq c_1^*$, the challenger queries the oracle $\mathcal{O}_{y_A}(\cdot)$ with $c_1$. $\mathcal{O}_{y_A}(\cdot)$ returns $G(c_1^a)_{[1,\ldots,P+Q]}$ as an answer. The challenger computes $m\|t = c_2 \oplus G(c_1^a)_{[1,\ldots,P]}$ and $s = G(c_1^a)_{[P+1,\ldots,P+Q]}$, and checks whether $t = h_s(m\|L)$. If $t = h_s(m\|L)$, the challenger returns $m$ as a plaintext. Otherwise, the challenger outputs $\perp$.
2. If $c_1 = c_1^*$, the challenger cannot get help from the oracle $\mathcal{O}_{y_A}(\cdot)$ and outputs $\perp$. In this case, Pr[*Bad*] denotes the probability that $(c_1,c_2)$ is a valid ciphertext such that $c_{2[1,\ldots P-l]} \neq c^*_{2[1,\ldots P-l]}$ and $c_1 = c_1^*$. In fact, Pr[*Bad*] is negligible, for the following reason. In order to form a valid ciphertext, the adversary needs to find a $c_2$ satisfying

$$(c_2 \oplus z^*)_{[P-l+1,\ldots,P]} = h_{s^*}((c_2 \oplus z^*)_{[1,\ldots,P-l]}\|L)$$
$$(c_2^* \oplus z^*)_{[P-l+1,\ldots,P]} = h_{s^*}\left((c_2^* \oplus z^*)_{[1,\ldots,P-l]}\|L\right)$$

According to the definition of the universal hash functions, if $h$ is chosen uniformly from the universal class $H$, for every $c_2, c_2^* \in \{0,1\}^P$ with $c_2 \neq c_2^*$, it holds that $c_{2[P-l+1,\ldots,P]}$ and $c^*_{2[P-l+1,\ldots,P]}$ are uniformly and independently distributed over $\{0,1\}^l \times \{0,1\}^l$. That is, the adversary can find such a $c_2$ only with a negligible probability $1/2^l$. Otherwise, it would imply that $h$ is not chosen uniformly from $H$, that is, the pseudorandom string $s$ could be distinguished from a random string by an efficient algorithm with a non-negligible advantage. This is a contradiction.

Let Pr[*Exp*] denote the probability that the adversary B wins the above game in the experiment. The following claims, Claims 1 and 2, explain how to identify a ODH$^+$ quadruple according to the adversary's performance in the above experiment.

**Claim 1.**

*If $\{g, g^a, g^b, G(g^c)_{[1,\ldots,P+Q]}\}$ is an ODH$^+$ quadruple, then |Pr[Game 1] – Pr[Exp]| is negligible and |Pr[Game 2] – Pr[Exp]| is non-negligible.*

To show that Claim 1 holds, we first note that if $\{g, g^a, g^b, G(g^c)_{[1,\ldots,P+Q]}\}$ is an ODH$^+$ quadruple and the event *Bad* does not happen, then the experiment perfectly simulates Game 1 and the adversary's views in the experiment and Game 1 are identical. Hence, we have

$$\Pr[Game\ 1 \wedge \neg Bad] = \Pr[Exp \wedge \neg Bad]$$

Applying Lemma 1, we have $|\Pr[Game\,1] - \Pr[Exp]| \leqslant \Pr[Bad]$, where $\Pr[Bad]$ is negligible. On the other hand, since $|\Pr[Game\,1] - \Pr[Game\,2]| \geqslant 1/p(k)$, we have

$$|\Pr[Game\ 1] - \Pr[Exp]| + |\Pr[Game\ 2] - \Pr[Exp]| \geqslant |\Pr[Game\ 1] - \Pr[Game\ 2]| \geqslant 1/p(k)$$

which implies $|\Pr[Game\,2] - \Pr[Exp]| \geqslant 1/p(k) - \Pr[Bad]$. Therefore, $|\Pr[Game\,2] - \Pr[Exp]|$ is non-negligible, from which Claim 1 follows.

Using a similar argument to the correctness of Claim 1, we have the following Claim 2.

**Claim 2.**

*If $\{g, g^a, g^b, G(g^c)_{[1,\ldots,P+Q]}\}$ is not an ODH$^+$ quadruple, then|Pr[Game 2] – Pr[Exp]| is negligible and |Pr[Game 1] – Pr[Exp]| is non-negligible.*

Summing up Claims 1 and 2, the ODH$^+$ assumption can be compromised by observing the behavior of the adversary. Specifically, if $|\Pr[Game\,1] - \Pr[Exp]|$ is negligible, then $|\Pr[Game\,2] - \Pr[Exp]|$ must be non-negligible. In this case, $\{g, g^a, g^b, G(g^c)_{[1,\ldots,P+Q]}\}$ must be an ODH$^+$ quadruple. Likewise, if $|\Pr[Game\,1] - \Pr[Exp]|$ is non-negligible, then $|\Pr[Game\,2] - \Pr[Exp]|$ must be negligible. In this case, $\{g, g^a, g^b, G(g^c)_{[1,\ldots,P+Q]}\}$ must not be an ODH$^+$ quadruple. These lead to the following claim:

**Claim 3.**

*|Pr[Game 1] – Pr[Game 2]| ⩽ μ(k) holds under the ODH$^+$ assumption, where μ(k) is a negligible function.*

Finally, it remains to show that $|\Pr[Game\,3] - \Pr[Game\,2]|$ is negligible. Since the only difference between Game 3 and Game 2 is the target ciphertext, the adversary's view of Game 3 is indistinguishable from his view of Game 2 if $G$ is a secure pseudorandom generator. Otherwise it were not the case, Game 2 would serve as an efficient algorithm to distinguish the output distribution of $G$ from the uniform distribution. Hence, we obtain the following claim:

**Claim 4.**

*|Pr[Game 3] – Pr[Game 2]| ⩽ μ′(k) if G is a secure pseudorandom generator, where μ′(k) is a negligible function.*

From Claims 3 and 4, we have

$$\begin{aligned} |\Pr[Game\ 1] - 1/2| &= |\Pr[Game\ 1] - \Pr[Game\ 3]| \leqslant |\Pr[Game\ 1] - \Pr[Game\ 2]| + |\Pr[Game\ 2] - \Pr[Game\ 3]| \\ &\leqslant \mu(k) + \mu'(k) \end{aligned} \tag{1}$$

where $\mu(k) + \mu'(k)$ is a negligible function.

That is, the adversary can win the standard adaptive chosen ciphertext attack game with only a negligible advantage. This completes the proof of Theorem 1. □

*Problems with the ODH⁺ assumption:* As discussed in [19], if there is no oracle $\mathcal{O}_{g^v}(\cdot)$ in the ODH assumption, the correctness of ODH can be easily achieved under the DDH assumption. Because $g^{uv}$ is already indistinguishable from a random group element, one only needs to convert the random group element $g^{uv}$ to a random string of proper length. According to the leftover hash lemma [24], the application of the universal hash function suffices [25]. However, when taking into account the help from the oracle $\mathcal{O}_{g^v}(\cdot)$, using only the universal hash could be dangerous due to self-reducibility. Hence, Abdalla et al. suggest to use an one-way cryptographic hash function, such as SHA-1, instead of a universal hash function. In our ODH$^+$ assumption, we use a pseudorandom generator $G(g^{uv}) = G'(h'(g^{uv}))$ which consists of a pseudorandom generator $G'$ and a universal hash function $h'$. On one hand, the universal hash $h'$ guarantees that the $h'(g^{uv})$ is a (pseudo)random string and can be used as a seed of $G(\cdot)$. On the other hand, the pseudorandon generator $G'$ guarantees the one-wayness and pseudorandomness of $G$.

However, the ODH assumption implicitly assumes that the one-way cryptographic hash function can eliminate dependency or correlations among different outputs and inputs of the hash function. Therefore the ODH$^+$ assumption also implicitly requires the pseudorandom generator have such a strong property. But the problem is that an ordinary pseudorandom generator does not necessarily guarantee to eliminate such dependency. We further notice that, in the ODH$^+$ assumption, the adversary can always get answers from the oracle and the pseudorandom generator should be strong enough to prevent the adversary getting useful information from these answers. To mitigate the problem, one method is to weaken the requirement for the pseudorandom generator by restricting the behavior of the adversary. As discussed in the previous section, the adaptive DDH assumption is in fact a weaker version of ODH$^+$, which does restrict the adversary's query while providing help for the simulator. This leads us to arriving at the main goal of this work, that is to modify the Zheng–Seberry$_{uh}$ scheme and prove its security under the weaker adaptive DDH assumption.

## 5. Modified Zheng–Seberry$_{uh}$ scheme

### 5.1. Description of the modified Zheng–Seberry$_{uh}$ scheme

Our major modification to the Zheng–Seberry$_{uh}$ scheme is to increase the output length of the pseudorandom generator by $W$ bits. These additional $W$ bits play the role of a tag for an ephemeral key $y_A^x$ and will be sent to a recipient as part of a ciphertext. In practice, in order to minimize the impact of these additional bits on the efficiency of the scheme, $W$ should be chosen to be as short as practical. For a security level of $2^{80}$, we suggest $W \geqslant 160$.

Additionally, the pseudorandom generator $G(\cdot)$ is required to be a adaptively secure pseudorandom generator $G_{tag}(\cdot)$, where $tag = (y_A, c_1)$. The modified Zheng–Seberry$_{uh}$ scheme is described in Table 2.

### 5.2. Security proof of the modified Zheng–Seberry$_{uh}$ scheme

**Theorem 2.** *Assuming the adaptive DDH assumption holds, the modified Zheng–Seberry$_{uh}$ scheme is secure against adaptive chosen ciphertext attacks.*

**Proof.** The main idea of the proof is similar to that for the Zheng–Seberry$_{uh}$ scheme under the ODH$^+$ assumption (Section 4). Our aim here is to prove that $|\Pr[Game\,1] - 1/2|$ is negligible, by showing that both $|\Pr[Game\,1] - \Pr[Game\,2]|$ and $|\Pr[Game\,3] - \Pr[Game\,2]|$ are negligible. More details follow.

*Game 1:*
Game 1 is the standard adaptive chosen ciphertext attack game and the target ciphertext is $c^* = (c_1^*, c_2^*) = (g^{x^*}, z^* \oplus (m_\beta \| t^* \| 0^W))$, where $t^* = h_{s^*}(m_\beta \| L)$.

**Table 2**
The modified Zheng–Seberry$_{uh}$ scheme.

*Modified Zheng–Seberry$_{uh}$ scheme*

*Public parameters:* A label $L$, a universal class of hash functions $H$: $\{0,1\}^* \to \{0,1\}^l$, a group $\mathbb{G}$, a generator $g$ of $\mathbb{G}$ with order $q$, and an adaptively secure pseudorandom generator $G_{tag} : \mathbb{G} \to \{0,1\}^*$.

*Key generation:* Choose $x_A$ randomly in $\mathbb{Z}_q^*$ and compute $y_A = g^{x_A}$. The public key is $y_A$ and the private key is $x_A$.

| **Encyrption** $E_{uhf}(y_A, m, L)$ | **Decryption** $D_{uhf}(x_A, y_A, c_1, c_2, L)$ |
|---|---|
| 1. $x \overset{R}{\leftarrow} \mathbb{Z}_q^*, r = y_A^x$ | 1. $r' = c_1^{x_A}, s' = G_{tag}(r')_{[1,\ldots,Q]}, z' = G_{tag}(r')_{[Q+1,\ldots,Q+P+W]}$, |
| 2. $c_1 = g^x$. Let $tag = (y_A, c_1)$. | 2. $m' \| t' = (c_2 \oplus z')_{[1,\ldots,P]}$, where $m' = (c_2 \oplus z')_{[1,\ldots,P-l]}$, $t' = (c_2 \oplus z')_{[P-l+1,\ldots,P]}$ |
| 3. $s = G_{tag}(r)_{[1,\ldots,Q]}, t = h_s(m \| L)$ | 3. if $h_{s'}(m' \| L) = t'$ and $z'_{[P+1,\ldots,P+W]} = c_{2\,[P+1,\ldots,P+W]}$, then output $m'$ as a plaintext; otherwise output $\perp$. |
| 4. $z = G_{tag}(r)_{[Q+1,\ldots,Q+P+W]}, c_2 = z \oplus (m \| t \| 0^W)$ | |
| Output the ciphertext $(c_1, c_2)$ | |

*Game 2:*
Game 2 is similar to Game 1 except that the target ciphertext is modified to $c_+^{**} = (g^{x^*}, z^{**} \oplus (m_\beta \| t^{**} \| 0^W))$, where $s^{**} = G_{y_A, g^{x^*}}(r^{**})_{[1,\ldots,Q]}, z^{**} = G_{y_A, g^{x^*}}(r^{**})_{[Q+1,\ldots,Q+P+W]}, r^{**} \overset{R}{\leftarrow} \mathbb{G}, t^{**} = h_{s^{**}}(m_\beta \| L)$.
*Game 3:*
Game 3 is similar to Game 2 except that the target ciphertext is modified to $c_+^* = \left(g^{x^*}, u_3 \oplus \left(m_\beta \| t_+^* \| 0^W\right)\right)$, where $u_2 \overset{R}{\leftarrow} \{0,1\}^Q, u_3 \overset{R}{\leftarrow} \{0,1\}^{P+W}, t_+^* = h_{u_2}(m_\beta \| L)$.
As discussed in the proof of Claim 3, if $|\Pr[\mathit{Game}\ 1] - \Pr[\mathit{Game}\ 2]|$ is a non-negligible, then there exists an adversary $B$ which can distinguish between Game 1 and Game 2 with non-negligible advantage. We can then construct an algorithm $A$ to break the adaptive DDH assumption using $B$, resulting in a contradiction.

Given $\{g, g^a, g^b, G_{g^a, g^b}(g^c)_{[1,\ldots,Q+P+W]}\}, A$ sets $y_A = g^a$ and simulates the adaptive chosen ciphertext attack game for the adversary $B$ in the following experiment.

*Experiment:* $A$ sets the target ciphertext $(c_1^*, c_2^*)$ to

$$\left(g^b, G_{g^a, g^b}(g^c)_{[Q+1,\ldots,Q+P+W]} \oplus (m_\beta \| h_{G_{g^a, g^b}(g^c)_{[1,\ldots,Q]}}(m_\beta \| L) \| 0^W)\right)$$

and uses the oracle $\mathcal{O}_{g^a, g^b}(\cdot,\cdot,\cdot)$ to answer the decryption query. Notice that, $A$ is a bit different from the one constructed in the proof of Theorem 1 in that the oracle $\mathcal{O}_{g^a, g^b}(\cdot,\cdot,\cdot)$ would output $\perp$ if the challenger does not propose the "right" query. More precisely, when the challenger receives the decryption query $(c_1, c_2)$, he computes $T = c_{2[P+1,\ldots,P+W]}$ and decrypts as follows

1. If $c_1 \neq c_1^*$, the challenger makes the query $(g^a, c_1, T)$ to the oracle $\mathcal{O}_{g^a, g^b}(\cdot,\cdot,\cdot)$.
   - If the oracle returns the answer $r$, the challenger can compute

     $$m \| t = c_{2[1,\ldots,P]} \oplus G_{g^a, g^b}(r)_{[Q+1,\ldots,Q+P]}$$
     $$s = G_{g^a, g^b}(r)_{[1,\ldots,Q]}$$

     and check whether $t = h_s(m \| L)$. If $t = h_s(m \| L)$, the challenger returns the plaintext $m$. Otherwise, the challenger outputs $\perp$.
   - If the oracle outputs $\perp$ which means the $(g, g^a, c_1, T)$ is not a adaptive DDH quadruple and the corresponding ciphertext is not valid, then the challenger outputs $\perp$.
2. If $c_1 = c_1^*, c_2 \neq c_2^*, T = T^*$, where $T^* = c^*_{2[P+1,\ldots,P+W]}$, the challenger cannot get help from the oracle $\mathcal{O}_{g^a, g^b}(\cdot,\cdot,\cdot)$ and outputs $\perp$. Let $\Pr[\mathit{Bad}]$ denote the probability that $(c_1, c_2)$ is a valid ciphertext such that $c_1 = c_1^*, c_{2[1,\ldots,P-l]} \neq c^*_{2[1,\ldots,P-l]}, T = T^*$. $\Pr[\mathit{Bad}]$ is negligible, because the adversary needs to find a $c_2$ satisfying

   $$(c_2 \oplus z^*)_{[P-l+1,\ldots,P]} = h_{s^*}((c_2 \oplus z^*)_{[1,\ldots,P-l]} \| L)$$
   $$(c_2^* \oplus z^*)_{[P-l+1,\ldots,P]} = h_{s^*}\left((c_2^* \oplus z^*)_{[1,\ldots,P-l]} \| L\right)$$

   If $h$ is chosen uniformly from the universal class $H$, for every $c_{2[1,\ldots,P]}, c^*_{2[1,\ldots,P]} \in \{0,1\}^P$ with $c_{2[1,\ldots,P]} \neq c^*_{2[1,\ldots,P]}, c_{2[P-l+1,\ldots,P]}$ and $c^*_{2[P-l+1,\ldots,P]}$ are uniformly and independently distributed over $\{0,1\}^l \times \{0,1\}^l$. Therefore, the adversary can find such a $c_2$ only with negligible probability. Otherwise, it would imply that $h$ is not chosen uniformly from $H$, that is, the pseudorandom string $s$ could be distinguished from a random string by an efficient algorithm with a non-negligible advantage.
3. Otherwise, the challenger outputs $\perp$.

Finally, Claims 5 and 6 show that, if $|\Pr[\mathit{Game}\ 1] - \Pr[\mathit{Game}\ 2]|$ is non-negligible, then whether $\{g, g^a, g^b, G_{g^a, g^b}(g^c)_{[1,\ldots,P+Q+W]}\}$ is an adaptive DDH quadruple or not can be decided with a non-negligible advantage. □

**Claim 5.** *If* $\{g, g^a, g^b, G_{g^a, g^b}(g^c)_{[1,\ldots,P+Q+W]}\}$ *is an adaptive DDH quadruple, then |Pr[Game 1] – Pr[Exp]| is negligible and |Pr[Game 2] – Pr[Exp]| is non-negligible.*

**Claim 6.** *If* $\{g, g^a, g^b, G_{g^a, g^b}(g^c)_{[1,\ldots,P+Q+W]}\}$ *is not an adaptive DDH quadruple, then |Pr[Game 2] – Pr[Exp]| is negligible and |Pr[Game 1] – Pr[Exp]| is non-negligible.*

The correctness of Claims 5 and 6 can be shown in a similar way to that for the correctness of Claims 1 and 2. Combining Claims 5 and 6, we have the following Claim 7:

**Claim 7.** *|Pr[Game 1] – Pr[Game 2]| is negligible if the adaptive DDH assumption holds.*

Since the only difference between Game 3 and Game 2 is the target ciphertext, we have

**Claim 8.** *|Pr[Game 3] – Pr[Game 2]| is negligible if* $G_{tag}$ *is a secure pseudorandom generator.*

From Claims 7, 8 and the inequality (1), we conclude that $|\Pr[\mathit{Game}\ 1] - 1/2|$ is negligible. That is the adversary cannot win the standard adaptive chosen ciphertext attack game with a non-negligible advantage. The proof of Theorem 2 is completed.

**Table 3**
"Trapdoor permutation$^+$" denotes trapdoor permutations that are uninvertible with access to a $H$-inverting oracle. "One-way hash$^+$" denotes adaptively secure perfectly one-way hash. SPD-OW denotes set partial domain one-wayness. "SKE" denotes secure symmetric encryption. "MAC" denotes secure message authentication code. "Enc Exp" ("Dec Exp") denotes the number of exponentiations or double exponentiations in encryption (decryption).

| | Enc Exp | Dec Exp | Assumption | RO |
|---|---|---|---|---|
| Modified Zheng–Seberry$_{uh}$ (this paper) | 2 | 1 | Adaptive DDH | No |
| Zheng–Seberry$_{uh}$ (this paper) | 2 | 1 | ODH$^+$ | No |
| Cramer–Shoup [9] | 4 | 3 | DDH | No |
| Kurosawa–Desmedt [16] | 3 | 1 | DDH, SKE | No |
| Hofheinz–Kiltz [29] | 3 | 1 | DDH | No |
| Hofheinz–Kiltz [18] | Roughly 2 | Roughly 1 | Rabin's trapdoor OWP | No |
| DHIES [19] | 2 | 1 | ODH, SKE, MAC | No |
| Pandey–Pass–Vaikuntanathan [10] | – | – | Trapdoor permutation$^+$, one-way hash$^+$ | No |
| Zheng–Seberry$_{1wh}$ [13] | 2 | 1 | GDH | Yes |
| OAEP [30] | 1 | 1 | SPD-OW | Yes |
| Bellare–Rogaway [6] | – | – | Trapdoor OWP | Yes |

## 6. Instantiation

First we note that an $\epsilon$-AXU hash function [27] can be used in place of a universal hash function. One may also use an efficient universal hash function family proposed by Bernstein [28]. Such a substitution requires only minor revisions to be made in the security proofs. Specifically, in Case 2 of the experiment for the security proof of the modified Zheng–Seberry$_{uh}$ scheme, the probability that the adversary can find $c_2$ satisfying $(c_2 \oplus z^*)_{[P-l+1,\ldots,P]} = h_{s^*}((c_2 \oplus z^*)_{[1,\ldots,P-l]} \| L)$ and $\left(c_2^* \oplus z^*\right)_{[P-l+1,\ldots,P]} = h_{s^*}\left(\left(c_2^* \oplus z^*\right)_{[1,\ldots,P-l]} \| L\right)$ needs to be changed. To instantiate the adaptively secure pseudorandom generator $G_{tag}(\cdot)$, we can use the HMAC-based key derivation function (KDF) [26], which follows the extract-then-expand paradigm.

## 7. Comparison

For the modified Zheng–Seberry$_{uh}$ scheme, the length of a ciphertext is $|m| + |p| + 320$, where $|p|$ denotes the binary length of a element in $\mathbb{G}$. (We recall that for the original Zheng–Seberry$_{uh}$ scheme, it is $|m| + |p| + 160$.) Thanks to the use of a pseudorandom generator and a universal hash function, no limit needs to be placed on the length of a plaintext. For a long plaintext $m$, the ratio between the lengths of a ciphertext and a plaintext, $\alpha = \frac{|m|+|p|+320}{|m|}$, approaches to 1.

This advantage makes our modified Zheng–Seberry$_{uh}$ scheme more preferable to other schemes such as those proposed in [9,10]. The ratio $\alpha$ in [9] is 4, that is the length of a ciphertext is always 4 times as long as a plaintext. The scheme in [10] which is also based on adaptive security assumptions has five elements in a ciphertext, although the output length of the hash function can be modified to fit the length of a plaintext. Table 3 shows a comparison of the Zheng–Seberry$_{uh}$ and modified Zheng–Seberry$_{uh}$ schemes with a few of the relevant encryption schemes.[2]

## 8. Concluding remarks

We have demonstrated how to modify a universal hash based public key encryption scheme by Zheng and Seberry so that the resultant scheme not only preserves the efficiency of the original scheme but also admits provable security against adaptive chosen ciphertext attack without random oracle. This represents the first public key encryption scheme that is practical *in a true sense* while not relying for its security on a random oracle. A further advantage of the scheme lies in its flexibility to encrypt messages of any length. We also compare the modified Zheng–Seberry scheme with related encryption schemes in terms of efficiency and underlying assumptions, supporting our conclusion that the modified Zheng–Seberry is preferable to its competitors.

The schemes investigated in this work are based on discrete logarithms in a subgroup. Naturally, these schemes have their equivalents on elliptic curves. A possible interesting area for further research is to investigate whether similar results can be obtained with schemes built on other computationally hard problems, such as the integer factorization problem.

## References

[1] Naor M, Yung M. Public-key cryptosystems provably secure against chosen cipher-text attacks. In: ACM symposium on theory of computing. ACM Press; 1990. p. 14–6.

[2] Rackoff C, Simon D. Non-interactive zero-knowledge proof of knowledge and chosen ciphertext attack. In: Advances in cryptology – CRYPTO 1991. LNCS, vol. 576. Springer-Verlag; 1991. p. 433–44.

[3] Dolev D, Dwork C, Naor M. Non-malleable cryptography. SIAM J Comput 2000;30(2):391–437.

[2] In [19], MAC is strong unforgeable against chosen message attack [19] and SKE is secure against IND-CPA attack (indistinguishability against chosen plaintext attack). In [16], SKE is $\epsilon$-rejection secure which is information-theoretical assumption but Gennaro and Shoup [31] showed that a MAC unforgeable against one-more forgery and indistinguishable symmetric ciphers are enough.

[4] Bellare M, Rogaway P. Optimal asymmetric encryption – how to encrypt with RSA. In: Advances in cryptology – EUROCRYPT 1993. LNCS, vol. 950. Springer-Verlag; 1994. p. 92–111.

[5] Fujisaki E, Okamoto T. Secure integration of asymmetric and symmetric encryption schemes. In: Advances in cryptology – CRYPTO 1999. LNCS, vol. 1666. Springer-Verlag; 1999. p. 537–54.

[6] Bellare M, Rogaway P. Random oracles are practical: a paradigm for designing efficient protocols. In: Proceedings of the first ACM conference on computer and communications security. The Association for Computing Machinery; 1993. p. 62–73.

[7] Canetti R, Goldreich O, Halevi S. The random oracle methodology, revisited. In: 30th annual ACM symposium on theory of computing. ACM Press; 1998. p. 23–6.

[8] Leurent G, Nguyen PQ. How risky is the random oracle model. In: Advances in cryptology – CRYPTO 2009. LNCS, vol. 5677. Springer-Verlag; 2009. p. 445–64.

[9] Cramer R, Shoup V. A practical public key cryptosystem provably secure against adaptive chosen ciphertext attack. In: Advances in cryptology – CRYPTO 1998. LNCS, vol. 1462. Springer-Verlag; 1998. p. 13–25.

[10] Pandey O, Pass R, Vaikuntanathan V. Adaptive one-way functions and applications. In: Advances in cryptology – CRYPTO 2008. LNCS, vol. 5157. Springer-Verlag; 2008. p. 57–74.

[11] Zheng Y, Seberry J. Immunizing public key cryptosystems against chosen ciphertext attacks. IEEE journal on selected areas in communications, 1993. The extended abstract of this paper appears in advances in cryptology – CRYPTO 1992;11(5):715–724.

[12] Soldera D, Seberry J, Qu C. The analysis of Zheng–Seberry scheme. In: Proceedings of the 7th Australian conference on information security and privacy. LNCS, vol. 2384. Springer-Verlag; 2002. p. 159–68.

[13] Baek J, Zheng Y. Zheng and Seberry's public key encryption scheme revisited. Int J Inform Security 2003;2(1):37–44.

[14] Shoup V. Using hash functions as a hedge against chosen ciphertext attack. In: Advances in cryptology – EUROCRYPT 2000. LNCS, vol. 1807. Springer-Verlag; 2000. p. 275–88.

[15] Cramer R, Shoup V. Universal hash proofs and a paradigm for adaptive chosen ciphertext secure public-key encryption. In: Advances in cryptology – EUROCRYPT 2002. LNCS, vol. 2332. Springer-Verlag; 2002. p. 45–64.

[16] Kurosawa K, Desmedt Y. A new paradigm of hybrid encryption scheme. In: Advances in cryptology – CRYPTO 2004. LNCS, vol. 3152. Springer-Verlag; 2004. p. 426–42.

[17] Kiltz E, Pietrzak K, Stam M, Yung M. A new randomness extraction paradigm for hybrid encryption. In: Advances in cryptology – EUROCRYPT 2009. LNCS, vol. 5479. Springer-Verlag; 2009. p. 590–609.

[18] Hofheinz D, Kiltz E. Practical chosen ciphertext secure encryption from factoring. In: Advances in cryptology – EUROCRYPT 2009. LNCS, vol. 5479. Springer-Verlag; 2009. p. 313–32.

[19] Abdalla M, Bellare M, Rogaway P. The oracle Diffie–Hellman assumptions and an analysis of DHIES. In: Topics in cryptology – CT-RSA 2001. LNCS, vol. 2020. Springer-Verlag; 2001. p. 143–58.

[20] Carter JL, Wegman MN. Universal classes of hash functions. J Comput Syst Sci 1979;18(2):143–54.

[21] Shoup V. OAEP reconsidered. In: Advances in cryptology – CRYPTO 2001. LNCS, vol. 2139. Springer-Verlag; 2001. p. 239–59.

[22] Bellare M, Rogaway P. Minimizing the use of random oracles in authenticated encryption schemes. In: Information and Communications Security. LNCS, vol. 1334. Springer-Verlag; 1997. p. 1–16.

[23] Boldyreva A, Fischlin M. On the security of OAEP. In: Advances in cryptology – ASIACRYPT 2006. LNCS, vol. 4284. Springer-Verlag; 2006. p. 210–25.

[24] Impagliazzo R, Levin LA, Luby M. Pseudo-random generation from one-way functions. In: ACM symposium on theory of computing. ACM Press; 1989. p. 12–24.

[25] Naor M, Reingold O. Number-theoretic constructions of efficient pseudo-random functions. In: 38th annual symposium on foundations of computer science. ACM Press; 1997. p. 458–67.

[26] Krawczyk H. On extract-then-expand key derivation functions and an HMAC-based KDF; 2008. Available from: http://www.ee.technion.ac.il/hugo/kdf/.

[27] den Boer B. A simple and key-economical unconditional authentication scheme. J Comput Security 1993;2(1):65–71.

[28] Bernstein DJ. Polynomial evaluation and message authentication; 2007. Available from: http://cr.yp.to/papers.html#pema.

[29] Hofheinz D, Kiltz E. Secure hybrid encryption from weakened key encapsulation. In: Advances in cryptology – CRYPTO 2007. LNCS, vol. 4622. Springer-Verlag; 2007. p. 553–71.

[30] Fujisaki E, Okamoto T, Pointcheval D, Stern J. RSA-OAEP is secure under the RSA assumption. In: Advances in cryptology – CRYPTO 2001. LNCS, vol. 2139. Springer-Verlag; 2001. p. 260–74.

[31] Gennaro R, Shoup V. A note on an encryption scheme of Kurosawa and Desmedt; 2005. Available from: http://www.shoup.net/papers/kdnote.pdf.

**Puwen Wei** received his Ph.D. degree in mathematics from Shandong University, China, in 2009. From 2008 to 2009, he was an exchange visitor at the Department of Software and Information Systems, University of North Carolina at Charlotte, USA. Currently, he is a lecturer of Shandong University. His research interests are in the field of public key cryptography.

**Xiaoyun Wang** received her Ph.D. degree in mathematics from Shandong University, China, in 1993. She is currently a professor of the Institute for Advanced Study, Tsinghua University, China. Her research interests include cryptography and secure computing in number theory and abstract algebra.

**Yuliang Zheng** received his Ph.D. degree in electrical and computer engineering from Yokohama National University, Japan, in 1991. He is currently a professor of Software and Information Systems, University of North Carolina at Charlotte, USA. His research interests include cryptography, network security, and the protection of critical infrastructures.