SECURITY AND COMMUNICATION NETWORKS
Security Comm. Networks (2010)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/sec.274

RESEARCH ARTICLE

An efficient and provably secure public key encryption scheme based on coding theory

Rongxing Lu[1], Xiaodong Lin[2], Xiaohui Liang[1] and Xuemin (Sherman) Shen[1]*

[1] Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada
[2] Faculty of Business and Information Technology, University of Ontario Institute of Technology, Oshawa, Ontario, L1H 7K4, Canada

ABSTRACT

Although coding-based public key encryption schemes such as McEliece and Niederreiter cryptosystems have been well studied, it is not a trivial task to design an *efficient* coding-based cryptosystem with semantic security against adaptive chosen ciphertext attacks (IND-CCA2). To tackle this challenging issue, in this paper, we first propose an *efficient* IND-CCA2-secure public key encryption scheme based on coding theory. We then use the provable security technique to formally prove the security of the proposed scheme is tightly related to the syndrome decoding (SD) problem in the random oracle model. Compared with the previously reported schemes, the proposed scheme is merited with simple construction and fast encryption speed. Copyright © 2010 John Wiley & Sons, Ltd.

KEYWORDS

coding-based cryptography; public key encryption; semantic security; chosen-ciphertext attacks; syndrom decoding problem

***Correspondence**

Xuemin (Sherman) Shen, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada.
E-mail: xshen@bbcr.uwaterloo.ca

1. INTRODUCTION

Post-quantum cryptography (PQC), as a popular cryptography terminology aiming at providing 'Post-Quantum' alternative to the currently existing number theory cryptography [1–3], has obtained great attention in recent years. In essence, PQC overlaps many existing cryptography branches including coding-based cryptography [4,5], lattice-based cryptography, hash-based cryptography, and multivariate-quadratic-equations cryptography [6]. However, driven largely by the possible invention of a large quantum computer in the near future, PQC becomes a new *buzz word* in cryptography communities and these non-number theory cryptography branches, especially the coding-based cryptography, have brought renewed attention.

In coding-based cryptography, there are two well-known public key encryption schemes, namely McEliece and Niederreiter cryptosystems [4,5]. McEliece cryptosystem was first proposed in 1978 [4], which represents the first public key encryption scheme based on linear error-correcting codes. Compared with the classical RSA cryptosystem [7], the McEliece cryptosystem has two advantages: (i) the speeds of both encryption and decryption algorithms are faster; and (ii) with the increase of the key size, the security level also grows much faster. Niederreiter cryptosystem [5] is a dual encryption scheme proposed in 1986, which is not only ten times faster than the McEliece cryptosytem in terms of encryption speed, but also equivalent to the McEliece cryptosystem in terms of security. Following these two seminal works, over the past years, many efforts have been put in coding-based cryptography [8–14]. For example, Stern has proposed a coding-based zero knowledge identification scheme in 1993 [14]; Courtois, Finiasz, and Sendrier have presented the first practical coding-based signature scheme in 2001 [10]. More recently, how to reduce the public key size and how to secure the parameter choice in coding-based cryptography are also deeply explored [15–19].

The *semantic security* (a.k.a indistinguishability) against adaptive chosen ciphertext attacks (IND-CCA2) is the strongest known notion of security for the public key encryption schemes. However, in coding-based cryptography, 'IND-CCA2' has not been widely discussed. To the best of our knowledge, only a few papers have touched this research issue [20–22]. Because McEliece cryptosystem has some special architecture, some general IND-CCA2 conversions [23,24], though they achieve IND-CCA2 versions of McEliece cryptosystem, may incur some redundancy. Therefore, Kobta and Imai have pro-

posed two specific conversions to reduce the redundancy [20]. Recently, Nojima *et al*. [21] have studied the semantic for the McEliece cryptosystem without random oracles. However, they only achieve the semantic security against the chosen plaintext attacks and the tight reductions are also questionable, especially for the Niederreiter cryptosystem. In Reference [22], Dowsley *et al*. have also discussed the CCA2 secure public key encryption scheme based on the McEliece assumption in the standard model, but their scheme needs some special constructions. Therefore, how to design an *efficient* and IND-CCA2 secure coding-based cryptosystem with/without random oracles is still worth of investigation.

In this paper, we propose an *efficient* IND-CCA2 secure public key encryption scheme based on coding theory. Concretely, we design our scheme based on the syndrome decoding (SD) problem, and use the provable security technique to get a tight reduction in the random oracle model [25]. Compared with Niederreiter cryptosystem, only two additional hash operations are required in the proposed scheme. Thus, our scheme achieves fast encryption speed.

The remainder of this paper is organized as follows. In Section 2, we formalize the definition of public key encryption and the corresponding security model. In Section 3, we review the coding theory and the complexity assumption, the base of our proposed scheme. In Section 4, we present our efficient public key encryption scheme based on coding theory, following by its formal security proof and parameter selection in Section 5 and Section 6, respectively. Finally, we draw our conclusions in Section 7.

2. DEFINITION AND SECURITY MODEL

2.1. Notation

Let $\mathbb{N} = \{1, 2, 3, \ldots\}$ denote the set of natural numbers, and $k \in \mathbb{N}$ be a security parameter. An event is said to be negligible if it happens with probability less than the inverse of any polynomial in k. If $n \in \mathcal{N}$, then 0^n denotes the string of n zeros. If x, y are strings, then $|x|$ denotes the length of x, $[x]_n$ denotes the n least significant bits of x, $[x]^n$ denotes the n most significant bits of x, and $x \oplus y$ denotes the bit XOR if $|x| = |y|$, while if $\mathcal{S}$ is a finite set, then $|\mathcal{S}|$ is its cardinality, and $s \xleftarrow{R} \mathcal{S}$ indicates the process of selecting s uniformly and at random in $\mathcal{S}$. If $\mathcal{A}$ is a randomized algorithm, then $y \longleftarrow \mathcal{A}(x_1, x_2, \cdots)$ denotes the processing of $\mathcal{A}$ on inputs $x_1, x_2, \cdots$, and letting y denote its output.

2.2. Definition

In general, a public key encryption scheme PKE = (Setup, Kgen, Enc, Dec) consists of four algorithms:

- The randomized setup algorithm Setup takes a security parameter κ as input, and returns the system public parameters *params* in a polynomial time of κ; we write $params \xleftarrow{R} \text{Setup}(\kappa)$.
- The randomized key generation algorithm Kgen takes the system public parameters params as input, and returns a pair (pk, sk) consisting of a public key and a corresponding private key in a polynomial time of κ, we write $(pk, sk) \xleftarrow{R} \text{Kgen}(params)$.
- The randomized encryption algorithm Enc takes a public key pk, a random number r, and a plaintext M as input, and returns a ciphertext C in a polynomial time of κ; we write $C \longleftarrow \text{Enc}(pk, r, M)$.
- The deterministic decryption algorithm Dec takes the private key sk and a ciphertext C as input, and returns the corresponding plaintext M or a special symbol $\perp$ indicating that the ciphertext was invalid in a polynomial time of κ; we write $x \leftarrow \text{Dec}(sk, C)$, where $x \in \{M, \perp\}$.

All algorithms should satisfy the standard consistency constraint of public key encryption, i.e., for any message M, $\text{Dec}(sk, C = \text{Enc}(pk, r, M)) = M$.

2.3. Security model

We recall the standard notion of security of public key encryption schemes in terms of indistinguishability [26]. Concretely, we consider the security notion for a public key encryption scheme is indistinguishable against the adaptive chosen ciphertext attacks, call it the 'IND-CCA2' security model for brevity.

Definition 1. *(IND-CCA2) Let k and t be integers and ε a real number in* $[0, 1]$*, and PKE a secure public key encryption scheme with the security parameter k. Let $\mathcal{A}$ be an IND-CCA2 adversary, which is allowed to access the decryption oracle O_D (and some random oracles $O_{H_1}, O_{H_2}, \cdots$, in the random oracle model), against the indistinguishability of PKE. We consider the following random experiment*:

Experiment $\mathbf{Exp}^{\text{ind-cca2}}_{\text{PKE},\mathcal{A}}(k)$

$$params \xleftarrow{R} \text{Setup}(k)$$

$$(pk, sk) \xleftarrow{R} \text{Kgen}(params)$$

$$(M_0, M_1, \text{state}) \longleftarrow \mathcal{A}^{\mathcal{O}_D(\cdot, \mathcal{O}_{H_1}, \mathcal{O}_{H_2}, \cdots)}(params, pk)$$

$$b \xleftarrow{R} \{0, 1\},\ C_b \longleftarrow \text{Enc}(pk, r, M_b)$$

$$b' \longleftarrow \mathcal{A}^{\mathcal{O}_D(\cdot, \mathcal{O}_{H_1}, \mathcal{O}_{H_2}, \cdots)}(params, pk, C_b, \text{state})$$

if $b = b'$ then return $b^* \leftarrow 1$ else $b^* \leftarrow 0$
return b^*

We define the success probability of $\mathcal{A}$ *via*

$$\begin{aligned}\mathbf{Succ}^{\text{ind-cca2}}_{\text{PKE},\mathcal{A}}(k) &= 2\Pr\left[\mathbf{Exp}^{\text{ind-cca2}}_{\text{PKE},\mathcal{A}}(k)\right]-1\\ &= 2\Pr\left[b=b'\right]-1\end{aligned}$$

PKE is said to be (k, t, ε)-IND-CCA2 secure, if no adversary $\mathcal{A}$ running in time t has a success $\mathbf{Succ}^{\text{ind-cca2}}_{\text{PKE},\mathcal{A}}(k) \geq \varepsilon$.

3. CODING THEORY AND COMPLEXITY ASSUMPTION

Let $\mathbb{F}_2$ be the finite field with 2 elements $\{0, 1\}$, $k\in\mathcal{N}$ be a security parameter, and $\mathcal{C}$ denote an $[n, k]$-binary linear code of length n and dimension k, i.e., a subspace of dimension k of the vector space $\mathbb{F}_2^n$. Elements of $\mathbb{F}_2^n$ are called *words*, and elements of $\mathcal{C}$ are called *codewords*. An $[n, k]$-binary linear code is usually given in the form of a $(n-k)\times n$ binary matrix $\mathbf{H}$, lines of which form a basis of the code. We call the *syndrome* of a word $x = (x_1, x_2, \cdots, x_n)\in\mathbb{F}_2^n$ is the quantity $s = \mathbf{H}x^T$ computed by

$$\begin{aligned}s &= (s_1, s_2, \cdots, s_{n-k}) = \mathbf{H}x^T\\ &= \begin{pmatrix} \mathrm{h}_{11} & h_{12} & .. & h_{1n}\\ h_{21} & h_{22} & .. & h_{2n}\\ \vdots & \vdots & \vdots & \vdots\\ h_{(n-k)1} & h_{(n-k)2} & .. & h_{(n-k)n}\end{pmatrix}\cdot\begin{pmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}\end{aligned}$$

If the quantity $s = 0$, i.e., $\mathbf{H}x^T = 0$, the word $x = (x_1, x_2, \cdots, x_n)\in\mathbb{F}_2^n$ is a codeword. The *Hamming weight* of a word $x = (x_1, x_2, \cdots, x_n)\in\mathbb{F}_2^n$ is referred to the number of its non-zero positions, denoted as $hw(x)$; the *Hamming distance* between two words $x = (x_1, x_2, \cdots, x_n)\in\mathbb{F}_2^n$ and $y = (y_1, y_2, \cdots, y_n)\in\mathbb{F}_2^n$ is the number of positions where they differ, and denoted as $d_H(x, y)$; and the minimal distance of an $[n, k]$-binary linear code $\mathcal{C}$ is defined by $d = \min_{x,y\in\mathcal{C}} d_H(x, y)$. Then, the $[n, k]$-binary linear code $\mathcal{C}$ is called $[n, k, d]$ code. All $[n, k, d]$ codes satisfy the Singleton bound which states that $d \leq n-k+1$ [27]. A binary linear $[n, k, d]$ code is ensured to exist as long as

$$\sum_{j=0}^{d-2}\binom{n-1}{j} < 2^{n-k}$$

This is called the *Gilbert-Varshamov* (GV) bound. Note that, random binary codes are known to meet the GV bound, in the sense that the above inequality comes very close to being an equality [28], and no available family of binary codes can be decoded in subexponential time up to the GV bound [27].

Syndrome Decoding Problem [27]*:* Let $t \leq \lfloor\frac{d-1}{2}\rfloor$, we know that, for any syndrome $s\in\mathbb{F}_2^{n-k}$, there exists at most one word $x\in\mathbb{F}_2^n$ such that $hw(x) \leq t$ and $\mathbf{H}x^T = s$. A syndrome $s\in\mathbb{F}_2^{n-k}$ is said to be t-decodable in the $[n, k, d]$-binary linear code $\mathcal{C}$ defined by $\mathbf{H}$ if there exists such a word x. The SD problem is stated as follows: given an $(n-k)\times n$ binary matrix $\mathbf{H}$ and a syndrome $s\in\mathbb{F}_2^{n-k}$, compute a word $x\in\mathbb{F}_2^n$ such that $hw(x) \leq t$ and $\mathrm{H}x^{\mathrm{T}} = s$. Note that, to ensure the hardness of SD problem, the parameters should be carefully chosen [27,29,30].

Definition 2. *(SD Assumption) Let $\mathcal{C}$ be an $[n, k, d]$-binary linear code defined by a $(n-k)\times n$ binary matrix $\mathbf{H}$ with the minimal distance d, and $t \leq \lfloor\frac{d-1}{2}\rfloor$. An adversary that takes an input of a syndrome $s\in\mathbb{F}_2^{n-k}$, returns a word $s\in\mathbb{F}_2^{n-k}$. We consider the following random experiment on SD problem.*

Experiment $\mathbf{Exp}^{\text{SD}}_{\mathcal{A}}$

$$x\in\mathbb{F}_2^n \leftarrow \mathcal{A}\left(\mathbf{H}, s\in\mathbb{F}_2^{n-k}\right)$$
$$if\ hw(x) \leq t \text{ and } \mathbf{H}x^T = s$$

then $b \leftarrow 1$ *else* $b \leftarrow 0$

return b

We define the corresponding success probability of $\mathcal{A}$ in solving the SD problem *via*

$$\mathbf{Succ}^{\text{SD}}_{\mathcal{A}} = \Pr\left[\mathbf{Exp}^{\text{SD}}_{\mathcal{A}} = 1\right]$$

Let $\tau\in\mathbb{N}$ and $\varepsilon\in[0, 1]$. We call SD to be (τ, ε)-secure if no polynomial algorithm $\mathcal{A}$ running in time τ has success $\mathbf{Succ}^{\text{SD}}_{\mathcal{A}} \geq \varepsilon$.

Parameters of Goppa Codes: Goppa codes are subfield subcodes of particular alternant codes. For given integers $m, t\in\mathbb{N}$, binary Goppa codes are of length $n = 2^m$ and with the dimension of $k = n-mt$. Let $\mathcal{F}_{m,t}$ denote the family of such Goppa codes, then we have $|\mathcal{F}_{m,t}| = \frac{2^{tm}}{t}$. Since their algebraic structure can be efficiently *hidden* and provide a good t-decoding algorithm, $\mathcal{F}_{m,t}$ are good candidates for constructing efficient cryptographic algorithms. In the next section, we will use the Goppa codes to designed our efficient and provably secure public key encryption scheme.

4. PROPOSED PUBLIC KEY ENCRYPTION SCHEME BASED ON CODING THEORY

In this section, we present our public key encryption PKE scheme based on coding theory, which can be regarded as the CCA2 version of Niederreiter cryptosystem [5] and mainly consists of four algorithms, namely Setup, Kgen, Enc, and Dec, as shown in Figure 1.

Setup. Given the security parameter κ, four integers $(m, t, k_0, k_1)\in\mathcal{N}$ are chosen such that the t-decoding in a Goppa code of length $n = k_0 + k_1 = 2^m$, of dimension $k = 2^m-mt$ has complexity at least 2^κ [10]. In addition, two secure cryptographic hash functions $\mathcal{H}_1, \mathcal{H}_2$ are also chosen, where $\mathcal{H}_1 : \{0, 1\}^* \rightarrow \{0, 1\}^{k_0+k_1}$ and $\mathcal{H}_2 : \{0, 1\}^* \rightarrow \{0, 1\}^{k_0+k_1}$. In the end, the system parameters $params = (m, t, k_0, k_1, \mathcal{H}_1, \mathcal{H}_2)$ are published.

Kgen. Given the system parameters $params = (m, t, k_0, k_1, \mathcal{H}_1, \mathcal{H}_2)$, choose a random binary Goppa code $\mathcal{C}_0$ from the Goppa code family $\mathcal{G}_{m,t}$. Let $\mathbf{H}_0$ be a parity check matrix of $\mathcal{C}_0$ and $\Delta_{\mathbf{H}_0}$ be a t-decoding algorithm in $\mathcal{C}_0$. In addition, a random non-singular $mt\times mt$ binary matrix $\mathbf{U}$ and a random permutation matrix $\mathbf{P}$ of size $2^m\times 2^m$ are

SETUP	KGEN
SETUP(κ) $\rightarrow$system parameters $params = (m, t, k_0, k_1, \mathcal{H}_1, \mathcal{H}_2)$	system parameters $params \rightarrow$ public key $\mathbf{H} = \mathbf{U}\mathbf{H}_0\mathbf{P}$ private key $(\Delta_{\mathbf{H}_0}, \mathbf{U}, \mathbf{P})$
ENC	DEC
$M \in \{0,1\}^{k_1}, r \in \{0,1\}^{k_0+k_1}$ $x = (0^{k_0}\|\|M) \oplus \mathcal{H}_1(r) \in \{0,1\}^{k_0+k_1}$ $\alpha = \mathbf{H}(x \oplus \beta)^T \in \mathbb{F}_2^{mt}, \beta = r \oplus \mathcal{H}_2(x)$ $C = (\alpha, \beta)$	$y' = \mathbf{P}y^T = \Delta_{\mathbf{H}_0}(\mathbf{U}^{-1}\alpha), y = y'\mathbf{P}^{-1}$ $x = y \oplus \beta = (0^{k_0}\|\|M) \oplus \mathcal{H}_1(r)$ $\beta \oplus \mathcal{H}_2(x) = r$ $x \oplus \mathcal{H}_1(r) = 0^{k_0}\|\|M$

Figure 1. Proposed public key encryption scheme based on coding theory.

also chosen. Set the private key $sk = (\Delta_{\mathbf{H}_0}, \mathbf{U}, \mathbf{P})$ and the corresponding public key pk as $\mathbf{H} = \mathbf{U}\mathbf{H}_0\mathbf{P}$.

Enc. Given a message $M \in \{0,1\}^{k_1}$ and the public key $\mathbf{H}$, choose a random number $r \in \{0,1\}^{k_0+k_1}$, and execute the following steps:

- compute $x \in \{0,1\}^{k_0+k_1}$, where $x = (0^{k_0}||M) \oplus \mathcal{H}_1(r)$,
- compute (α, β) such that $\alpha = \mathbf{H}(x \oplus \beta)^T \in \mathbb{F}_2^{mt}$, $B = r \oplus \mathcal{H}_2(x)$,
- set the ciphertext $C = (\alpha, \beta)$.

Dec. Given a ciphertext $C = (\alpha, \beta)$ and the private key $(\Delta_{\mathbf{H}_0}, \mathbf{U}, \mathbf{P})$, the following steps are executed:

- compute $y' = \mathbf{P}y^T = \Delta_{\mathbf{H}_0}(\mathbf{U}^{-1}\alpha)$,
- compute $y = y'\mathbf{P}^{-1}$, $x = y \oplus B$, and $\beta \oplus \mathcal{H}_2(x) = r$,
- compute $x \oplus H_1(r)$, if $[x \oplus \mathcal{H}_1(r)]^{k_0}$ is 0^{k_0}, parse $x \oplus H_1(r)$ as $0^{k_0}||M$, i.e., $M = [x \oplus H_1(r)]_{k_1}$; otherwise output $\perp$ indicating an invalid ciphertext.

Compared with Niederreiter cryptosystem [5], only two additional hash operations are required. As a result, the encryption speed of the proposed scheme is as fast as Niederreiter cryptosystem.

5. SECURITY PROOF

In this section, we prove that the proposed scheme is IND-CCA2-secure in the random oracle model, where the hash functions $\mathcal{H}_1$ and $\mathcal{H}_2$ are modelled as random oracles [25].

Theorem 1. *Let $\mathcal{A}$ be an adversary against the proposed PKE scheme in the random oracle model, where the hash functions $\mathcal{H}_1$ and $\mathcal{H}_2$ behave as random oracles. Assume that $\mathcal{A}$ has the success probability $\mathbf{Succ}^{\text{ind-cca2}}_{\text{PKE},\mathcal{A}} \geq \varepsilon$ to break the indistinguishability of the ciphertext $C = (\alpha, \beta)$ within the running time τ, after $q_{\mathcal{H}_1}, q_{\mathcal{H}_2}$ and q_D queries to the random oracles $\mathcal{O}_{\mathcal{H}_1}, \mathcal{O}_{\mathcal{H}_2}$ and the decryption oracle $\mathcal{O}_D$, respectively. Then, there exist $\varepsilon' \in [0,1]$ and $\tau' \in \mathbb{N}$ as follows*

$$\varepsilon' = \mathbf{Succ}^{\text{SD}}_{\mathcal{A}}(\tau') \geq \frac{\varepsilon}{2} - \frac{2q_D^2 + q_D \cdot (q_{H_1} + q_{H_2}) + q_{H_1}}{2^{k_0+k_1}} - \frac{q_D(q_D + q_{H_1})}{2^{k_1}}, \quad \tau' \leq \tau + \Theta(.) \tag{1}$$

such that the SD problem can be solved with probability ε' within time τ', where $\Theta(.)$ is the time complexity for the simulation.

Proof. We define a sequence of games $\mathbf{Game}_0, \mathbf{Game}_1, \cdots$ of modified attacks starting from the actual adversary $\mathcal{A}$ [31,32]. All the games operate on the same underlying probability space: the system parameters $params = (m, t, k_0, k_1, \mathcal{H}_1, \mathcal{H}_2)$ and public key $\mathbf{H}$, the coin tosses of $\mathcal{A}$. Let $(\mathbf{H}, s^* \in \mathbb{F}_2^{mt})$ be a random instance of SD problem, we will use these incremental games to reduce the SD instance to the adversary $\mathcal{A}$ against the IND-CCA2 security of the ciphertext $C = (\alpha, \beta)$ in the proposed PKE scheme.

$\mathbf{Game}_0$: This is the real attack game. In the game, the adversary $\mathcal{A}$ is fed with the system parameters $params = (m, t, k_0, k_1, \mathcal{H}_1, \mathcal{H}_2)$ and public key $\mathbf{H} = \mathbf{U}\mathbf{H}_0\mathbf{P}$. In the first phase, the adversary $\mathcal{A}$ can access to the random oracles $\mathcal{O}_{\mathcal{H}_1}, \mathcal{O}_{\mathcal{H}_2}$ and the decryption oracle $\mathcal{O}_D$ for any input. At some point, the adversary $\mathcal{A}$ chooses a pair of messages $(M_0, M_1) \in \{0,1\}^{k_1}$. Then, we randomly choose a bit $b \in \{0,1\}$ and produce the message $M^\star = M_b$'s ciphertext $C^\star = (\alpha^\star, \beta^\star)$ as the challenge to the adversary $\mathcal{A}$. The challenge comes from the public key $\mathbf{H}$ and one random number $r^\star \in \{0,1\}^{k_0+k_1}$, and $\alpha^\star = \mathbf{H}(x^\star \oplus \beta^\star)^T$, $\beta^\star = r^\star \oplus \mathcal{H}_2(x^\star)$ with $x^\star = (0^{l_0}||M^\star) \oplus \mathcal{H}_1(r^\star)$. In the second stage, the adversary $\mathcal{A}$ is still allowed to access to the random oracles $\mathcal{O}_{\mathcal{H}_1}, \mathcal{O}_{\mathcal{H}_2}$ and the decryption oracle $\mathcal{O}_D$ for any input, except the challenge $C^\star$ to $\mathcal{O}_{\text{D}}$. Finally, the adversary $\mathcal{A}$ outputs a bit $b' \in \{0,1\}$. In any $\mathbf{Game}_j$, we denote by $\mathbf{Guess}_j$ the event $b = b'$. Then, we have

$$\begin{aligned} \varepsilon \leq \mathbf{Succ}^{\text{ind-cca2}}_{\text{PKE},\mathcal{A}} &= 2\Pr[b = b'] - 1 \\ &= 2\Pr[\mathbf{Guess}_0] - 1 \\ \Pr[\mathbf{Guess}_0] &\geq \frac{\varepsilon}{2} + \frac{1}{2} \end{aligned} \tag{2}$$

$\mathbf{Game}_1$: In this game, we simulate the random oracles $\mathcal{O}_{\mathcal{H}_1}, \mathcal{O}_{\mathcal{H}_2}$, and the decryption oracle $\mathcal{O}_D$, by maintaining the lists $\mathcal{H}_1$-List, $\mathcal{H}_2$-List and $\mathcal{D}$-List to deal with the identical queries. In addition, we also simulate the way that the challenge $C^\star$ is generated as the challenger would do. The detailed simulation in this game is described in Figure 2. Because the distribution of $(params, \mathbf{H})$ is unchanged in the eye of the adversary $\mathcal{A}$, the simulation is perfect, and we have

$$\Pr[\mathbf{Guess}_1] = \Pr[\mathbf{Guess}_0] \tag{3}$$

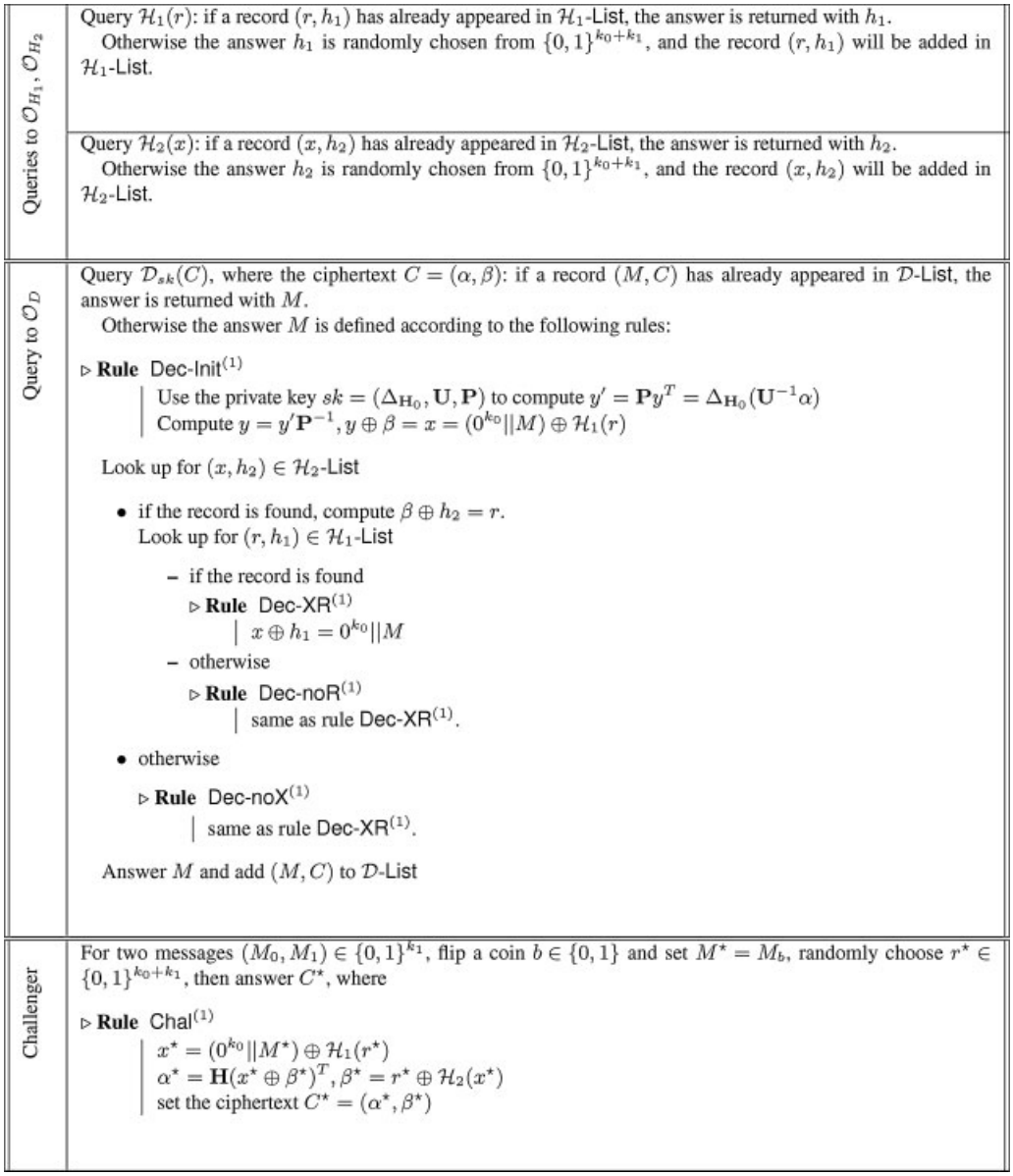

Figure 2. Formal simulation of the IND-CCA2 game against the proposed PKE based on coding theory.

$\mathbf{Game}_2$: In this game, we modify the simulation of the decryption oracle $\mathcal{O}_D$ by outputting a random message $M \in \{0,1\}^{k_1}$ when the ciphertext $C = (\alpha, \beta)$ has not been 'correctly' encrypted.

▷ **Rule** Dec-noX(2)

$h_2 \xleftarrow{R} \{0,1\}^{k_0+k_1}$, and set $r = h_2 \oplus \beta$
compute $h_1 = \mathcal{H}_1(r)$, choose a random message $M \xleftarrow{R} \{0,1\}^{k_1}$
compute $x = (0^{k_0}||M) \oplus h_1$
add (x, h_2) in $\mathcal{H}_2$-List

The two games $\mathbf{Game}_2$ and $\mathbf{Game}_1$ are perfectly indistinguishable unless x is already in $\mathcal{O}_{H_2}$. Because h_1 is queried from $\mathcal{O}_{H_1}$ and behaves uniformly, we can consider $x \in \{0,1\}^{k_0+k_1}$ a uniform random variable as well. So, the probability that x has already been queried to $\mathcal{O}_{H_2}$ is bounded to $(q_D + q_{H_2})/2^{k_0+k_1}$, then,

$$|\Pr[\mathbf{Guess}_2] - \Pr[\mathbf{Guess}_1]| \leq \frac{q_D(q_D + q_{H_2})}{2^{k_0+k_1}} \qquad (4)$$

$\mathbf{Game}_3$: In this game, we modify the simulation of the decryption oracle $\mathcal{O}_D$ without resorting to the random oracle $\mathcal{O}_{H_1}$.

▷ **Rule** Dec-noX$^{(3)}$

$h_2 \xleftarrow{R} \{0,1\}^{k_0+k_1}$, and set $r = h_2 \ominus \beta$
$h_1 \xleftarrow{R} \{0,1\}^{k_0+k_1}$, choose a random message
$M \xleftarrow{R} \{0,1\}^{k_1}$
compute $x = (0^{k_0}||M) \oplus h_1$
add (r, h_1) in $\mathcal{H}_1$-List and (x, h_2) in $\mathcal{H}_2$-List

The two games $\mathbf{Game}_3$ and $\mathbf{Game}_2$ are perfectly indistinguishable unless r is already in $\mathcal{O}_{H_1}$. Because h_2 is randomly chosen, we consider $r \in \{0,1\}^{k_0+k_1}$ as a uniform random variable, So, the probability that r has been queried to $\mathcal{O}_{H_1}$ is bounded to $(q_D + q_{H_1})/2^{k_0+k_1}$, then,

$$|\Pr[\mathbf{Guess}_3] - \Pr[\mathbf{Guess}_2]| \leq \frac{q_D(q_D + q_{H_1})}{2^{k_0+k_1}} \quad (5)$$

$\mathbf{Game}_4$: In this game, we modify the rule Dec-noR in the decryption oracle $\mathcal{O}_D$ simulation without resorting to the random oracle $\mathcal{O}_{H_1}$.

▷ **Rule** Dec-noR$^{(4)}$

set $r = h_2 \oplus \beta$
choose a random message $M \xleftarrow{R} \{0,1\}^{k_1}$
compute $h_1 = (0^{k_0}||M) \oplus x$
add (r, h_1) in $\mathcal{H}_1$-List

The two games $\mathbf{Game}_4$ and $\mathbf{Game}_3$ are perfectly indistinguishable unless (r, h_1) is already in $\mathcal{O}_{H_1}$. Because $[h_1]^{k_0}$ is known to the adversary $\mathcal{A}$ due to $h_1 = (0^{k_0}||M) \oplus x$, we consider $[h_1]_{k_1} \in \{0,1\}^{k_1}$ as a uniform random variable, then the probability that r has been queried to $\mathcal{O}_{H_1}$ is bounded to $(q_D + q_{H_1})/2^{k_1}$, then,

$$|\Pr[\mathbf{Guess}_4] - \Pr[\mathbf{Guess}_3]| \leq \frac{q_D(q_D + q_{H_1})}{2^{k_1}} \quad (6)$$

$\mathbf{Game}_5$: In this game, we modify the rule Dec-Init in the decryption oracle $\mathcal{O}_D$ simulation.

▷ **Rule** Dec-Init$^{(5)}$

look up (x, h_2) in $\mathcal{H}_2$-List such that $\alpha = \mathbf{H}(x \ominus \beta)^T$
if (x, h_2) is not found, set $x = \perp, h_2 = \perp$

The two games $\mathbf{Game}_5$ and $\mathbf{Game}_4$ are perfectly indistinguishable. If (x, h_2) is found in $\mathcal{H}_2$-List, the answer of the decryption oracle $\mathcal{O}_D$ is the same as that in $\mathbf{Game}_1$. If (x, h_2) is not found, i.e., $x = \perp$, and $h_2 = \perp$, the answer of the decryption oracle $\mathcal{O}_D$ is returning a random message $M \in \{0,1\}^{k_1}$ as that in $\mathbf{Game}_2$. Therefore, we have

$$\Pr[\mathbf{Guess}_5] = \Pr[\mathbf{Guess}_4] \quad (7)$$

$\mathbf{Game}_6$: In this game, we manufacture the challenge $C^\star = (\alpha^\star, \beta^\star)$ by first choosing the random value of $r^\star$ ahead of time.

▷ **Rule** Chal$^{(6)}$

two random values $(r^+, h_1^+) \in \{0,1\}^n$ are chosen ahead of time, then
$r^\star = r^+, h_1^\star = h_1^+$
$x^\star = (0^{k_0}||M^\star) \oplus h_1^+$
$\alpha^\star = \mathbf{H}(x^\star \oplus \beta^\star)^T, \beta^\star = r^+ \oplus \mathcal{H}_2(x^\star)$
set the ciphertext $C^\star = (\alpha^\star, \beta^\star)$

The two games $\mathbf{Game}_6$ and $\mathbf{Game}_5$ are perfectly indistinguishable unless $r^\star$ has been asked for $\mathcal{H}_1$. We define this event $\mathbf{AskH}_1^{(6)}$, then we have

$$|\Pr[\mathbf{Guess}_6] - \Pr[\mathbf{Guess}_5]| \leq \Pr[\mathbf{AskH}_1^{(6)}] \quad (8)$$

In this game, h_1^+ is only used in $x^\star$, but does not appear in the computation since $H_1(r^+)$ is not defined to be h_1^+. Then, the distribution of $C^\star = (\alpha^\star, \beta^\star)$ doesn't depend on b. As a result, we have

$$\Pr[\mathbf{Guess}_6] = \frac{1}{2} \quad (9)$$

$\mathbf{Game}_7$: In this game, instead of defining $x^\star$ from $h_1^\star$, we randomly choose $x^\star$ firstly and define $h_1^\star$ from $x^\star$. Because $x^\star$ is randomly chosen, we give a random answer for the question $x^\star$ to H_2.

▷ **Rule** Chal$^{(7)}$

three random values $(r^+, x^+, h_2^+) \in \{0,1\}^{k_0+k_1}$ are chosen ahead of time, then
$r^\star = r^+, x^\star = x^+, h_2^\star = h_2^+$
$h_1^\star = (0^{l_0}||M^\star) \oplus x^+$
$\alpha^\star = \mathbf{H}(x^+ \oplus \beta^\star)^T, \beta^\star = r^+ \ominus h_2^+$
set the ciphertext $C^\star = (\alpha^\star, \beta^\star)$

The two games $\mathbf{Game}_7$ and $\mathbf{Game}_6$ are perfectly indistinguishable unless $x^\star$ has been asked for $\mathcal{H}_2$. We define this event $\mathbf{AskH}_2^{(7)}$, then we have

$$\left|\Pr[\mathbf{AskH}_1^{(7)}] - \Pr[\mathbf{AskH}_1^{(6)}]\right| \leq \Pr[\mathbf{AskH}_2^{(7)}] \quad (10)$$

In this game, $h_1^\star = (0^{k_0}||M^\star) \oplus x^+$ is uniformly distributed, and independently of the view of the adversary $\mathcal{A}$, since x^+ hasn't been revealed. Therefore, we have

$$\Pr[\mathbf{AskH}_1^{(7)}] = \frac{q_{H_1}}{2^{k_0+k_1}} \quad (11)$$

$\mathbf{Game}_8$: In this game, instead of defining $\beta^\star$ from $h_2^\star$, we randomly choose $\beta^\star$ and then we define h_2^+ from $\beta^\star$.

three random values $(r^+, x^+, \beta^+) \in \{0,1\}^{k_0+k_1}$ are chosen ahead of time, then
$r^\star = r^+, x^\star = x^+, \beta^\star = \beta^+$
$h_1^\star = (0^{k_0}||M^\star) \oplus x^+$
$h_2^\star = \beta^+ \oplus r^+$
$\alpha^\star = \mathbf{H}(x^+ \oplus \beta^+)^T$
set the ciphertext $C^\star = (\alpha^\star, \beta^\star)$

Table 1. The sizes of plaintext/ciphertext and public key under the typical McEliece/Niederreiter parameters.

(m, t)	$n=2^m$	$k=n-mt$	Plaintext size M	Ciphertext size C	Public key size $\boldsymbol{H}$
(10, 50)	1024 bits	524 bits	$k_1 < 1024$ bits	1524 bits	62.5 kbytes
(11, 32)	2048 bits	1696 bits	$k_1 < 2048$ bits	2400 bits	88 kbytes
(12, 41)	4096 bits	3604 bits	$k_1 < 4096$ bits	4588 bits	246 kbytes

In this game, the distribution of $C^\star = (\alpha^\star, \beta^\star)$ is unchanged. Therefore, we have

$$\Pr[\mathbf{AskH}_2^{(8)}] = \Pr[\mathbf{AskH}_2^{(7)}] \tag{12}$$

$\mathbf{Game}_9$: In this game, we embed the SD challenge $(\mathbf{H}, s^* \in \mathbb{F}_2^{mt})$ in the game by setting $\alpha^\star = s^*$.

▷ **Rule** $\mathsf{Chal}^{(9)}$

random value $\beta^+ \in \{0,1\}^{k_0+k_1}$ is chosen ahead of time, then
$\alpha^\star = s^*, \beta^\star = \beta^+,$
set the ciphertext $C^\star = (\alpha^\star, \beta^\star)$

Clearly, the distribution of $C^\star = (\alpha^\star, \beta^\star)$ is still unchanged. Therefore, we have

$$\Pr[\mathbf{AskH}_2^{(9)}] = \Pr[\mathbf{AskH}_2^{(8)}] \tag{13}$$

In this game, when the event $\mathbf{AskH}_2^{(9)}$ takes place, i.e., there exists an $x^+ \in \{0,1\}^n$ such that $\alpha^\star = s^* = \mathbf{H}(x^+ \oplus \beta^+)^T$ has been queried to $\mathcal{O}_{H_2}$. Then, such an $x^* = x^+ \oplus \beta^+ \in \{0,1\}^n$ is just the SD challenge. As a result, we have

$$\Pr[\mathbf{AskH}_2^{(9)}] \leq \mathbf{Succ}_{\mathcal{A}}^{\mathrm{SD}}(\tau') = \varepsilon' \tag{14}$$

Summarizing all the above cases, we have

$$\varepsilon' = \mathbf{Succ}_{\mathcal{A}}^{\mathrm{SD}}(\tau') \geq \frac{\varepsilon}{2} - \frac{2q_D^2 + q_D \cdot (q_{H_1} + q_{H_2}) + q_{H_1}}{2^{k_0+k_1}} - \frac{q_D(q_D + q_{H_1})}{2^{k_1}} \tag{15}$$

and the running time $\tau' \leq \tau + \Theta(\cdot)$, where $\Theta(\cdot)$ is the time complexity for the simulation. This completes the proof.

6. SELECTION OF PARAMETERS

Parameter selection is imperative for the security of coding-based cryptography. If the parameters are not properly chosen, a coding-based system could suffer from threatening attacks based on either Information Set Decoding (ISD) or Generalized Birthday Algorithm (GBA) [18,33]. Since the decryption of the proposed scheme requires knowing one and only one solution for an SD problem, the GBA-based attacks can be ruled out [18]. Therefore, to resist the possible ISD-based attacks, some typical parameters used in McEliece be chosen for the proposed scheme. Then, the sizes of the public key, plaintext, and ciphertext can be calcuated, as shown in Table 1.

From the table, we can see that, though the sizes of the public keys are relatively large, the construction of the proposed scheme makes the speed of CCA2-secure encryption almost as fast as that of McEliece/Nederreiter cryptosystems. Furthermore, the recent works by Bender et al. [15] and Misoczki and Barreto [19] can be used to reduce the sizes of public key coding-based cryptography, which can make the coding-based cryptosystems more practical.

7. CONCLUSIONS

In this paper, we have proposed an efficient public key encryption scheme based on coding theory, and formally shown its IND-CCA2 security in the random oracle model. Since the size of the public key $\mathbf{H}$ in the proposed scheme is relatively large, our future work will focus on reducing the key size [15,19].

ACKNOWLEDGEMENTS

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) Strategic Projects of Canada.

REFERENCES

1. Lu R, Lin X, Cao Z, Shao J, Liang X. New (t,n) threshold directed signature scheme with provable security. *Information Sciences* 2008; **178**(3): 756–765.
2. Lin X, Lu R, Ho PH, Shen X, Cao Z. Tua: a novel compromise-resilient authentication architecture for wireless mesh networks. *IEEE Transactions on Wireless Communications* 2008; **7**(4): 1389–1399.
3. Lu R, Cao Z. Efficient remote user authentication scheme using smart card. *Computer Networks* 2005; **49**(4): 535–540.
4. McEliece RJ. A public-key cryptosystem based on algebraic coding theory. *DSN Progress Report 114–116*, Jet Propulsion Laboratory, Pasadena, CA, 1978.
5. Niederreiter H. Knapsack-type cryptosystems and algebraic coding theory. *Problems of Control and Information Theory* 1986; **15**(2): 159–166.
6. Bernstein DJ, Buchmann J, Dahmen E. Post-Quantum Cryptography. Springer: Berlin, Heidelberg, 2009.

7. Rivest RL, Shamir A, Adleman L. A method for obtaining digital signatures and public-key cryptosystems. *Communications of ACM* 1978; **21**(2): 120–126.
8. Canteaut A, Sendrier N. Cryptanalysis of the original mceliece cryptosystem. In ASIACRYPT'98, Vol. LNCS 1514. Springer-Verlag: Berlin, 1998; 187–199.
9. Augot D, Finiasz M, Sendrier N. A family of fast syndrome based cryptographic hash function. In MYCRYPT'05, Vol. LNCS 3715. Springer-Verlag: Berlin, 2005; 64–83.
10. Courtois N, Finiasz M, Sendrier N. How to achieve a mceliece-based digital signature scheme. In ASIACRYPT'01, Vol. LNCS 2248. Springer-Verlag: Berlin, 2001; 157–174.
11. Gaborit P, Girault M. Lightweight code-based identification and signature. *IEEE ISIT'07* Nice, France, July 2007; 191–195.
12. Gaborit P, Laudaroux C, Sendrier N. Synd: a very fast code-based cipher stream with a security reduction. IEEE ISIT'07 Nice, France, July 2007; 186–190.
13. Melchor CA, Cayrel PL, Gaborit P. A new efficient threshold ring signature scheme based on coding theory. In Post-Quantum Cryptography, Vol. LNCS 5299. Springer-Verlag: Berlin, 2008; 1–16.
14. Stern J. A new identification scheme based on syndrome decoding. In CRYPTO'93, Vol. LNCS 773. Springer-Verlag: Berlin, 1993; 13–21.
15. Berger T, Cayrel PL, Gaborit P, Otmani A. Reducing key length of the mceliece cryptosystem. In AFRICACRYPT'09, Vol. LNCS 5580. Springer-Verlag, 2009; 77–97.
16. Bernstein D, Lange T, Peters C. Attacking and defending the mceliece cryptosystem. In Post-Quantum Cryptography, Vol. LNCS 5299. Springer-Verlag: Berlin, 2008; 31–46.
17. Bernstein D, Lange T, Peters C, van Tilborg H. Explicit bounds for generic decoding algorithms for code-based cryptography. *Preproceedings of WCC 2009*; 168–180.
18. Finiasz M, Sendrier N. Security bounds for the design of code-based cryptosystems. In ASIACRYPT'09, Vol. LNCS 5912. Springer-Verlag: Berlin, 2009; 88–105.
19. Misoczki R, Barreto P. Compact mceliece keys from goppa codes. In SAC 2009, Vol. LNCS 5867. Springer-Verlag: Berlin, 2009; 376–392.
20. Kobara K, Imai H. Semantically secure mceliece public-key cryptosystems -conversions for mceliece pkc. In PKC'01, Vol. LNCS 1992. Springer-Verlag: Berlin, 2001; 19–35.
21. Nojima R, Imai H, Kobara K, Morozov K. Semantic security for the mceliece cryptosystem without random oracles. *Designs, Codes and Cryptography* 2008; **49** (1–3): 289–305.
22. Dowsley R, Muler-Quade J, Nascimento ACA. A CCA2 secure public key encryption scheme based on the mceliece assumptions in the standard model. In CT-RSA'09, Vol. LNCS 5473. Springer-Verlag: Berlin, 2009; 240–251.
23. Fujisaki E, Okamoto T. Secure intergration of asymmetric and symmetric encryption schemes. In CRYPTO'99, Vol. LNCS 1666. Springer-Verlag: Berlin, 1999; 537–554.
24. Pointcheval D. Chosen-ciphertext security for any one-way cryptosystem. In PKC'00, Vol. LNCS 1751. Springer-Verlag: Berlin, 2000; 129–146.
25. Bellar M, Rogaway P. Random oracles are practical: a paradigm for designing efficient protocols. In CCS'93. ACM, Fairfax: Virginia, US, 1993; 92–111.
26. Goldwasser S, Micali S. Probabilistic encryption. *Journal of Computer System Science* 1984; **28**: 270–299.
27. Barreto PSLM, Misoczki R. A new one-time signature scheme from syndrome decoding. *Cryptology ePrint Archive, Report 2010/017*, 2010. Available at: eprint.iacr.org
28. MacWilliams FJ, Sloane NJA. The Theory of Error-Correcting Codes. North-Holland, Amsterdam: New York, Oxford, 1978.
29. Chabaud F, Stern J. The cryptographic security of the syndrome decoding problem for rank distance codes. In ASIACRYPT'96, Vol. LNCS 1163. Springer-Verlag: Berlin, 1996; 368–381.
30. Berlekamp ER, McEliece RJ, van Tilborg HCA. On the inherent intractability of certain coding problems. *IEEE Transactions on Information Theory* 1978; **24**: 384–386.
31. Shoup V. OAEP reconsidered. *Journal of Cryptology* 2002; **15**(4): 223–249.
32. Phan DH, Pointcheval D. Chosen-ciphertext security without redundancy. In ASIACRYPT'03, Vol. LNCS 2894. Springer-Verlag: Berlin, 2003; 1–18.
33. Overbeck R, Sendrier N. Code-based cryptography. In Post-Quantum Cryptography. Springer-Verlag: Berlin, 2009; 95–145.