IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON INFORMATION SECURITY SOLUTIONS FOR TELEMEDICINE APPLICATIONS

Received April 10, 2018, accepted May 17, 2018, date of publication May 24, 2018, date of current version June 29, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2840119

A Practical Public Key Encryption Scheme Based on Learning Parity With Noise

ZHIMIN YU[1], CHONG-ZHI GAO[2,3], ZHENGJUN JING[1], BRIJ BHOOSHAN GUPTA[4], AND QIURU CAI[1]

[1]School of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, China
[2]School of Computer Science and Educational Software, Guangzhou University, Guangzhou 510000, China
[3]State Key Laboratory of Cryptology, Beijing 100878, China
[4]Department of Computer Engineering, National Institute of Technology at Kurukshetra, Kurukshetra 136119, India

Corresponding author: Chong-zhi Gao (czgao@gzhu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61672270, Grant 61602216, and Grant 61702236, in part by the Changzhou Applied Basic Research Guidance Project under Grant 2016365, in part by the Changzhou Science and Technology Program under Grant CJ20179027, and in part by the State Key Laboratory of Cryptology, China.

ABSTRACT To protect cyber security and privacy, it is critical to design security and practical public key encryption schemes. Today, big data and cloud computing bring not only unprecedented opportunities but also fundamental security challenges. Big data faces many security risks in the collection, storage, and use of data and brings serious problems regarding the disclosure of private user data. It is challenging to achieve security and privacy protection in the big data environment. Thus, to meet the growing demand of public key encryption in this environment, we proposed a single-bit public key encryption scheme based on a variant of learning parity with noise (LPN) and extended it to a multi-bit public key encryption scheme. We proved the correctness and chosen plaintext attack security of the proposed method. Our schemes solved encoding error rate problems of the existing public key schemes based on LPN, and the encoding error rate in our schemes is negligible.

INDEX TERMS CPA, encoding error ratio, encryption, LPN, public key encryption.

I. INTRODUCTION

With the development and application of big data and cloud computing technology, the large data environment has put forward higher requirements for data encryption, and the design of a practical and secure public key encryption scheme has important practical significance. Considering data security in the big data environment, many valuable schemes have been put forward [1]–[3]. They have been shown to be useful in applications such as protecting the privacy in machine learning [4], [5], and protecting security in cloud computing [6], [7]. The main classical public key schemes were designed based on a number of difficult number theory problems, such as large number factorization and discrete logarithms [8]–[11]. However, many traditional number theory assumptions on which the above schemes are based can be solved by quantum algorithms [12]. That is, in the era of quantum computing, these public key encryption schemes have been broken. Therefore, in the post quantum era, new public key encryption schemes based on new difficult problems need to be designed and implemented [13], [14] for the new computing environments and applications.

In 2003, Boneh and Silverberg defined the concept of ideal multilinear mapping and demonstrated its application scenarios [15]. However, until 2013, Garg, Gentry and Halevi (GGH) proposed the first realistic multilinear mapping based on ideal lattice [16], with its security based on the multi-level Diffie-Hellman computation and decision problem (GCDH/GDDH). Many new schemes have been designed based on the GGH scheme [17], [18]. Recently, the GGH scheme was proved to be insecure [19], and new multilinear mapping construction is being explored.

Regev proposed LWE (Learning with Error) based on lattice theory [20], which has been widely used in public key cryptosystem design and applications of data encryption in cloud computing [21]–[28]. Although LWE issues can resist quantum attacks, the public key size in schemes designed based on LWE is too large, and the reduction of this size is a public problem.

If we design public key schemes based on the variety of LPN that is the special case of LWE in F_2, the size of public key is small. There is a randomly selected open n-dimensional vector $\mathbf{a} \in Z_2^n$ and a randomly selected private n-dimensional vector $\mathbf{s} \in Z_2^n$ in an LPN (Learning parity with noise) problem. An attacker can get a sample set $(\mathbf{a}, \langle \mathbf{a}, \mathbf{s} \rangle + e)$, where $e \leftarrow Ber_\tau$. Ber_τ represents the Bernoulli distribution that is discrete 0, 1 probability distribution, and the probability of an occurrence of 1 is $0 < \tau < 1$. The parameter in the standard LPN problem is $0 < \tau < 0.5$, which is essentially the noise rate. On this basis, if the attacker is able to distinguish between the sampling element and the random element $(\mathbf{a}, r)$, $r \xleftarrow{\$} Z_2$, the attacker can solve the DLPN problem (decisional LPN).

To date, there are two kinds of non-trivial solving methods for LPN problems. One kind of method intends to exhaust all possible noise vectors, and the other solves the LPN problem based on the Blum-Kalai-Wasserman (BKW) algorithm [29]. The original BKW algorithm has sub index time complexity $2^{O(n/\log n)}$ with sampling times $2^{O(n/\log n)}$. Lyubashevsky gives a BKW algorithm variant that requires higher time complexity $2^{O(n/\log\log n)}$ but with sampling times $n^{1+\varepsilon}$ [30]. Recently, Kirchner [31] also proposed an improved algorithm with less running time. Although there are many solving algorithms for a variety of LPN problems, there are no polynomial time algorithms or quantum algorithms.

The creation and calculation of LPN instances are very simple, but it is very difficult to solve the DLPN problem. Therefore, it is very attractive to design cryptographic applications based on LPN. The LPN problem has been widely used in symmetric encryption [8], [32]–[35], but there has been little progress in the design of the public key scheme. In 2003, Alekhnovich proposed a public-key encryption scheme based on a decisional LPN problem [36].

In this scheme, the noise ratio is $\tau \approx 1/\sqrt{n}$ instead of a constant defined in a standard LPN problem. Subsequently, Damgård et al. proposed not only a public key encryption scheme based on decisional LPN problem but also a public-key encryption scheme based on a ring-LPN problem [37]. Damgård et al. proved the security of these schemes. Meanwhile, these schemes are practical. Damgård et al. compared some practical public key encryption algorithms such as RSA for computational efficiency, public key size and ciphertext. Although the RSA algorithm does not have an anti-quantum offensive, the performance comparison is meaningful.

However, non-negligible encoding error exists in all existing public key schemes based on an LPN variant [36], [37]. To solve this problem, we designed a new public-key encryption scheme. First, our issue will extend the LPN variant to a matrix LPN problem, and a new public key encryption scheme will be proposed based on an LPN variant. There are two advantages to the proposed scheme. First, we maintain the largest advantages of LPN, which are rapid instance generation, and rapid and efficient encryption and decryption computing; second, we solve the encoding error problem of existing public key encryption schemes. There are two vectors in Damgård's scheme $\mathbf{f}, \mathbf{e} \leftarrow Ber_\tau^n$. The correctness of the scheme relies on the fact that the inner product $\mathbf{f}^T\mathbf{e}$ will be zero with the greater probability $\Pr(\mathbf{f}^T\mathbf{e} = 0) = 1/2 + (1-2\tau^2)^n/2$ if the parameter is selected carefully. As this probability is greater but not negligible, there is an encoding error in the decryption. Damgård chose parameters to ensure the decryption error rate is less than 25% and chose the ciphertext expansion as 5. However, all the five bits of decoding error probability are still $(1/4)^5 = \left(1/2^{10}\right)$. Meanwhile, Damgård chose a small noise rate $\tau = \Theta\left(1/\sqrt{n}\right)$ to meet this condition. Obviously, if τ is too small, the attacker will crack the scheme easily.

Our contributions include the following: Firstly, we reduce the DLPN variety problem with $\mathbf{S} \leftarrow Ber_\tau^{n\times n}$ to the normal DLPN problem. So, our schemes are under the normal DLPN assumption. Secondly, we construct a new single-bit public key encryption algorithm in which a plaintext bit will be converted to a bit-vector involved in cryptographic operations. When a ciphertext is decrypted, if the hamming weight of the n dimensional vector is less than $n/2$, the plaintext bit is 0, and vice versa, the plaintext bit is 1. The probability of the hamming weight exceeding expectations will exponentially decay rapidly to a value that is negligible; thus, decryption error probability is negligible. Thirdly, we extend the single-bit scheme to the multi-bit public key encryption algorithm.

In our single-bit and multi-bit schemes, even if we choose a larger parameter $\tau = 1/\sqrt{n}$, it can also ensure that the decryption error can be ignored. Therefore, under the promise of security, the size of the public key is smaller than in Damgård's scheme. Meanwhile, total encryption and decryption time of our algorithms is greatly reduced.

The remainder of this paper is organized as follows. In section 2, preliminary knowledge will be given. In section 3, we propose a single-bit and a multi-bit public key encryption scheme. Then, in section 4, we give the comparison between our scheme and the existing scheme. The conclusion is given in section 5.

II. PRELIMINARIES

We first introduce the notation used in this paper and present the definition of the LPN problem [38].

A. NOTATION

We will completely work in the field $\mathbf{GF}_2$. For a vector $\mathbf{u} \in \mathbb{Z}_2^k$, the i-th entry of column vector $\mathbf{u}$ will be denoted by u_i. The i-th column vector of matrix $\mathbf{U}$ will be denoted by $\mathbf{u}_i$. $x \leftarrow D$ means that x is drawn from distribution D. Assuming $\mathbf{A}$ be n order matrix, $\mathbf{A}^T$ denotes the transpose of $\mathbf{A}$ and $\mathbf{A}^{-1}$ denotes the inverse matrix of $\mathbf{A}$. A probability $\varepsilon(n)$ is said to be negligible if $\varepsilon(n) \leq 1/p(n)$ for an arbitrarily large enough integer n. A Bernoulli distribution with parameter τ will be denoted by Ber_τ. Ber_τ^k denotes the distribution of vectors $\mathbf{a} \in \mathbb{Z}_2^k$ where each entry of the vector is drawn independently from Ber_τ. $Bin_{n,\tau}$ denotes the binomial distribution with n trials, each with success probability τ. $X \sim Bin_{n,\tau}$

TABLE 1. Mathematical expectation of h(d) when $m = 0$.

n	$n/2$	$\tau = 1/\sqrt{n}$	$E(\mathrm{h}(\mathbf{r}^T\mathbf{E}))$	$E(\mathrm{h}(\mathbf{e}_1^T\mathbf{S}))$	$E(\mathrm{h}(\mathbf{e}_2^T))$	$E(\mathrm{h}(\mathbf{d}))$
9000	4500	0.010541	514	514	95	≈1123
21000	10500	0.006901	1276	1276	145	≈2697
29000	14500	0.005872	1795	1795	166	≈3755
80000	40000	0.003536	5131	5131	282	≈10544
145000	72500	0.002626	9431	9431	381	≈19243

denotes that X is drawn from distribution $Bin_{n,\tau}$. For a vector $\mathbf{a} \in \mathbb{Z}_2^k$, its hamming weight is the number of ones in $\mathbf{a}$. A function $\mathrm{h}(\mathbf{a})$ calculates the hamming weight of $\mathbf{a} \in \mathbb{Z}_2^k$. Let $\mathrm{I} = (1, 1, \ldots, 1) \in \mathbb{Z}_2^n$.

B. DECISIONAL LPN PROBLEM

Definition 1 (Decisional LPN Problem): Given parameters $n \leftarrow \mathbb{N}$, $\tau \in \mathbb{R}$, $\tau = \Theta(1/\sqrt{n})$ and randomly selected matrix $\mathbf{A} \leftarrow \mathbb{Z}_2^{n\times n}$, $\mathbf{S} \leftarrow \mathbb{Z}_2^{n\times n}$. An attacker can obtain a sample set $(\mathbf{A}, \mathbf{AS} + \mathbf{E})$, where $\mathbf{E} \leftarrow Ber_\tau^{n\times n}$. If the attacker can distinguish between a new sample $(\mathbf{A}, \mathbf{AS} + \mathbf{E})$ and $(\mathbf{A}, \mathbf{R})$, $\mathbf{R} \xleftarrow{\$} \mathbb{Z}_2^{n\times n}$ with non-negligible probability after obtaining enough samples; then, the attacker is able to solve the decisional LPN (DLPN) problem.

Definition 2 (Decisional LPN Assumption): The probability of any probabilistic polynomial time (PPT) attacker to solve the decisional LPN problem with parameters (n, τ) is negligible. Alekhnovich defined the noise ratio $\tau = \Theta(1/\sqrt{n})$ [36].

III. PUBLIC KEY ENCRYPTION SCHEME BASED ON DLPN

In this section, we first give a single-bit public key encryption scheme based DLPN, and then we prove the correctness and security of the scheme. Second, we extend a single-bit scheme to the multi-bit public key encryption scheme and prove its correctness and security.

A. SINGLE-BIT PUBLIC KEY ENCRYPTION SCHEME

1) CONSTRUCTION OF THE SCHEME

A single-bit public key encryption scheme includes three PPT algorithms (KeyGen, Enc, Dec) following these steps:

(1) The key generation algorithm KeyGen$(1^n, \tau)$ takes as input an integer n and noise rate τ. Choose matrix $\mathbf{A} \leftarrow \mathbb{Z}_2^{n\times n}$ randomly and choose $\mathbf{S} \leftarrow Ber_\tau^{n\times n}$, $\mathbf{E} \leftarrow Ber_\tau^{n\times n}$. Compute $\mathbf{B} = \mathbf{AS}+\mathbf{E}$. It returns a public key $pk = (\mathbf{A}, \mathbf{B})$ and a private key $sk = (\mathbf{S})$.

(2) The encryption algorithm Enc(pk, m) takes as input the public key pk and message $m \in Z_2$. Compute $\mathbf{c}_1 = \mathbf{r}^T\mathbf{A} + \mathbf{e}_1^T$, $\mathbf{c}_2 = \mathbf{r}^T\mathbf{B} + \mathbf{e}_2^T + m\mathrm{I}$. It returns a ciphertext $\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2)$.

(3) The decryption algorithm Dec$(sk,\mathbf{c})$ takes the private key sk and a ciphertext $\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2)$ as input. Compute $\mathbf{d} = \mathbf{c}_1 \times \mathbf{S} + \mathbf{c}_2$. If $\mathrm{h}(\mathbf{d}) \ll n/2$, it returns $m = 0$, else it returns $m = 1$.

2) CORRECTNESS

Before giving proof of the correctness of the scheme, we introduce lemma 3, whose proof can be found in [37].

Lemma 3 ([37] Lemma 2.5): Let $X \sim Bin_{n,\tau}$. Then, the probability that X is even is $1/2 + (1 - 2\tau^2)^n/2$.

Lemma 4: The probability of decryption error of the single-bit public key encryption scheme is negligible.

Proof: Because $\mathbf{d} = \mathbf{c}_1 \times \mathbf{S} + \mathbf{c}_2$ and $\mathbf{d} = (\mathbf{r}^T\mathbf{A} + \mathbf{e}_1^T) \times \mathbf{S} + \mathbf{r}^T\mathbf{B} + \mathbf{e}_2^T + m\mathrm{I}$, substituting $\mathbf{B} = \mathbf{AS} + \mathbf{E}$ into the above equations, we can get $\mathbf{d} = \mathbf{r}^T\mathbf{E} + \mathbf{e}_1^T\mathbf{S} + \mathbf{e}_2^T + m\mathrm{I}$. As we know $\mathbf{r} \leftarrow Ber_\tau^n$, $\mathbf{E} \leftarrow Ber_\tau^{n\times n}$, each entry of $\mathbf{c} = \mathbf{r}^T\mathbf{E}$ is $c_i = \sum_{j=1}^{n} r_j e_{i,j}$. From Lemma 3, the probability that c_i is 0 will be $1/2 + (1 - 2\tau^2)^n/2$. Similarly, the probability that each entry of $\mathbf{e}_1^T\mathbf{S}$ is 0 will be $1/2 + (1 - 2\tau^2)^n/2$. Lastly, because $\mathbf{e}_2^T \leftarrow Ber_\tau^n$, we can reach the following conclusions: $\mathrm{h}(\mathbf{r}^T\mathbf{E} + \mathbf{e}_1^T\mathbf{S} + \mathbf{e}_2^T) \le \mathrm{h}(\mathbf{r}^T\mathbf{E}) + \mathrm{h}(\mathbf{e}_1^T\mathbf{S}) + \mathrm{h}(\mathbf{e}_2^T)$, $\mathrm{h}(\mathbf{r}^T\mathbf{E}) + \mathrm{h}(\mathbf{e}_1^T\mathbf{S}) + \mathrm{h}(\mathbf{e}_2^T) \approx n(1 - (1 - 2\tau^2)^n + \tau)$.

According to parameters selected in the scheme, $\mathrm{h}(\mathbf{r}^T\mathbf{E} + \mathbf{e}_1^T\mathbf{S} + \mathbf{e}_2^T) \ll n/2$ can be met. Obviously, if plaintext is 1, it is equivalent to do the inverse operation on $\mathbf{e}_2^T$, and if the plaintext is 0, $\mathbf{e}_2^T$ remains unchanged. So, if $m = 0$, then $\mathrm{h}(\mathbf{r}^T\mathbf{E} + \mathbf{e}_1^T\mathbf{S} + \mathbf{e}_2^T) \ll n/2$, on the contrary, there must be $\mathrm{h}(\mathbf{r}^T\mathbf{E} + \mathbf{e}_1^T\mathbf{S} + \mathbf{e}_2^T) \gg n/2$. □

The function $\mathrm{h}(\mathbf{r}^T\mathbf{E} + \mathbf{e}_1^T\mathbf{S} + \mathbf{e}_2^T)$ takes as input different integer n and larger noise rate $\tau = 1/\sqrt{n}$. We give in Table 1 the mathematical expectation of $\mathrm{h}(\mathbf{r}^T\mathbf{E}+\mathbf{e}_1^T\mathbf{S}+\mathbf{e}_2^T)$ when the plaintext $m = 0$.

3) SECURITY PROOF

Although we sample private key $\mathbf{S} \leftarrow Ber_\tau^{n\times n}$ instead of $\mathbb{Z}_2^{n\times n}$, its security is still based on the DLPN. Therefore, before given a security proof, we introduce lemma 5.

Lemma 5: Choose $\mathbf{A} \in \mathbb{Z}_2^{n\times n}$, $\mathbf{S} \leftarrow Ber_\tau^{n\times n}$ and $\mathbf{E} \leftarrow Ber_\tau^{n\times n}$ randomly. Compute $\mathbf{B} = \mathbf{AS}+\mathbf{E}$. Under the assumption f DLPN, it is indistinguishable between $(\mathbf{A}, \mathbf{B})$ and $\mathbb{Z}_2^{n\times n} \times \mathbb{Z}_2^{n\times n}$ sampled from uniform distribution.

Proof: Given a set of LPN sample $(\mathbf{A}_i, \mathbf{B}_i = \mathbf{A}_i\mathbf{S} + \mathbf{E}_i)$, where $\mathbf{A}_i \leftarrow \mathbb{Z}_2^{n\times n}$, $\mathbf{S} \leftarrow \mathbb{Z}_2^{n\times n}$ and $\mathbf{E}_i \leftarrow Ber_\tau^{n\times n}$. Without loss of generality, if we assume $\mathbf{A}_1^{-1} \in \mathbb{Z}_2^{n\times n}$, then

$$\begin{aligned}(\mathbf{A}_i\mathbf{A}_1^{-1}, \mathbf{B}_i - \mathbf{A}_i\mathbf{A}_1^{-1}\mathbf{B}_1) &= (\mathbf{A}_i\mathbf{A}_1^{-1}, \mathbf{B}_i' = \mathbf{E}_i - \mathbf{A}_i\mathbf{A}_1^{-1}\mathbf{E}_1)\\ &= (\mathbf{A}_i', \mathbf{B}_i' = \mathbf{A}_i'\mathbf{E}_1 + \mathbf{E}_i),\end{aligned}$$

where $\mathbf{A}_i' = \mathbf{A}_i\mathbf{A}_1^{-1}$. If $(\mathbf{A}_i, \mathbf{B}_i)$ is the LPN sample selected according to the definition 2 instead of uniform distribution $\mathbb{Z}_2^{n\times n} \times \mathbb{Z}_2^{n\times n}$, then $(\mathbf{A}_i', \mathbf{B}_i')$ meets the definition in section 3.1.1. When there is a PPT algorithm that can distinguish $(\mathbf{A}_i', \mathbf{B}_i')$ and a uniform distribution $\mathbb{Z}_2^{n\times n} \times \mathbb{Z}_2^{n\times n}$, then this algorithm can distinguish $(\mathbf{A}_i, \mathbf{B}_i)$ and a uniform distribution $\mathbb{Z}_2^{n\times n} \times \mathbb{Z}_2^{n\times n}$.

Therefore, the DLPN variety problem with $\mathbf{S} \leftarrow Ber_\tau^{n\times n}$ can be reduced to the normal DLPN problem.

Theorem 6 (Security): Under the DLPN assumption, the single-bit public key scheme is secure against a chosen plaintext attack.

Proof: Suppose the single-bit public key scheme defined in section 3.1.1 has parameters n, τ and public key $pk = (\mathbf{A}, \mathbf{B})$. Let $\mathbf{R} \in \mathbb{Z}_2^{n\times 2n}$ be

$$r_{i,j} = \begin{cases} a_{i,j} & 1 \le i \le n, 1 \le j \le n \\ b_{i,j} & 1 \le i \le n, n < j \le 2n. \end{cases}$$

Obviously, $\mathbf{R}$ has the same distribution as $pk = (\mathbf{A}, \mathbf{B})$.

If plaintext is $m = 0$, the ciphertext is $(\mathbf{r}^T\mathbf{A} + \mathbf{e}_1^T, \mathbf{r}^T\mathbf{B} + \mathbf{e}_2^T)$, which can be written as $\mathbf{r}^T\mathbf{R} + \mathbf{e}^{*T}$, where $\mathbf{e}^* \in \mathbb{Z}_2^{2n}$, $e_i^* = \begin{cases} (e_1)_i, & 1 \le i \le n \\ (e_2)_i, & n < i \le 2n \end{cases}$. According DLPN assumptions, it is indistinguishable between $(\mathbf{R}, \mathbf{r}^T\mathbf{R} + \mathbf{e}^{*T})$ and $(\mathbf{S}, \mathbf{r}'^T\mathbf{S} + \mathbf{e}'^T)$, where $\mathbf{S} \in Z^{n\times 2n}$, $\mathbf{r}' \leftarrow Ber_\tau^n$ and $\mathbf{e}' \leftarrow Ber_\tau^{2n}$. Furthermore, $(\mathbf{S}, \mathbf{r}'^T\mathbf{S} + \mathbf{e}'^T)$ and $(\mathbf{S}, \mathbf{r}^{*T})$ also cannot be distinguished, in which $\mathbf{r}^*$ is randomly chosen from $\mathbb{Z}^{2n}$.

If the plaintext is $m = 1$, $\mathbf{e}_2^T + m\mathrm{I}$ does not change the distribution of $\mathbf{e}_2^T$ and only makes a negated operation to $\mathbf{e}_2^T$. Hence, a ciphertext is indistinguishable from random digits.

B. MULTI-BIT PUBLIC KEY ENCRYPTION SCHEME

1) CONSTRUCTION OF THE SCHEME

A multi-bit public key encryption scheme includes three PPT algorithms (KeyGen, Enc, Dec) following these steps:

(1) The key generation algorithm KeyGen$(1^n, \tau)$ takes as input an integer n and noise rate τ. Choose matrix $\mathbf{A} \leftarrow \mathbb{Z}_2^{n\times n}$ randomly and choose $\mathbf{S} \leftarrow Ber_\tau^{n\times n}$, $\mathbf{E} \leftarrow Ber_\tau^{n\times n}$. Compute $\mathbf{B} = \mathbf{AS} + \mathbf{E}$. It returns a public key $pk = (\mathbf{A}, \mathbf{B})$ and private key $sk = (\mathbf{S})$.

(2) The encryption algorithm Enc$(pk, \mathbf{m})$ takes as input the public key pk and message $\mathbf{m} \in \mathbb{Z}_2^n$. First, convert $\mathbf{m}$ to a square matrix $\mathbf{M}^* \in \mathbb{Z}_2^{n\times n}$, if $m_i = 0$, each entry of the *i-th* column of $\mathbf{M}^*$ are 0, and vice versa, each entry of the *i-th* column are 1, e.g., $\mathbf{m} = (1, 1, 0, 0)^{\mathrm{T}}$, then

$$\mathbf{M}^* = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

Choose $\mathbf{R}, \mathbf{E}_1, \mathbf{E}_2 \leftarrow Ber_\tau^{n\times n}$, compute $\mathbf{C}_1 = \mathbf{RA} + \mathbf{E}_1$, $\mathbf{C}_2 = \mathbf{RB} + \mathbf{E}_2 + \mathbf{M}^*$. It returns a ciphertext $\mathbf{C} = (\mathbf{C}_1, \mathbf{C}_2)$.

(3) The decryption algorithm Dec$(sk, \mathbf{c})$ takes as input the private key sk and a ciphertext $\mathbf{C} = (\mathbf{C}_1, \mathbf{C}_2)$. Compute $\mathbf{D} = \mathbf{C}_1 \times \mathbf{S} + \mathbf{C}_2$. If hamming weight of the $i - th$ column of $\mathbf{D}$ is $\mathrm{h}(\mathbf{d}) < n/2$, then $m_i = 0$, $m_i = 1$. At last it returns $\mathbf{m}$.

2) CORRECTNESS

Lemma 7: The probability of decryption error of the multi-bit public key encryption scheme is negligible.

Proof: It is very easy to verify that $\mathbf{D} = \mathbf{C}_1 \times \mathbf{S} + \mathbf{C}_2 = (\mathbf{RA} + \mathbf{E}_1) \times \mathbf{S} + \mathbf{RB} + \mathbf{E}_2 + \mathbf{M}^*$, substituting $\mathbf{B} = \mathbf{AS} + \mathbf{E}$ into the above equations, we get $\mathbf{D} = \mathbf{RE} + \mathbf{E}_1\mathbf{S} + \mathbf{E}_2 + \mathbf{M}^*$, where $\mathbf{R}, \mathbf{E}_1, \mathbf{E}_2 \leftarrow Ber_\tau^{n\times n}$. According to Lemma 3, the hamming weight of each column of $\mathbf{RE}$ and $\mathbf{E}_1\mathbf{S}$ is $n\left(1/2 - (1 - 2\tau^2)^n/2\right)$. When each entry of the column vector $\mathbf{m}_i^*$ is zero, the hamming weight $\mathrm{h}(\mathbf{d}_i)$ is at most $n(1 - (1 - 2\tau^2)^n + \tau)$. According to the parameters selected in the scheme, $\mathrm{h}(\mathbf{d}_i) \ll n/2$ can be met. Obviously, if the plaintext is $m_i = 0$, it is equivalent to doing an inverse operation on $\mathbf{e}_{2_i}^T$; if the plaintext is 0, $\mathbf{e}_2^T$ remains unchanged. Therefore, if $m = 0$, then $\mathrm{h}(\mathbf{r}^T\mathbf{E} + \mathbf{e}_1^T\mathbf{S} + \mathbf{e}_2^T) \ll n/2$; in contrast, there must be $\mathrm{h}(\mathbf{r}^T\mathbf{E} + \mathbf{e}_1^T\mathbf{S} + \mathbf{e}_2^T) \gg n/2$ if $m_i = 1$. □

3) SECURITY PROOF

Theorem 8 (Security): Under the DLPN assumption, the multi-bit public key scheme is secure against the chosen plaintext attack.

Proof: Suppose the multi-bit public key scheme defined in section 3.2.1 has parameters n, τ and public key $pk = (\mathbf{A}, \mathbf{B})$. Let

$$\mathbf{Q} \in \mathbb{Z}_2^{n\times 2n}, \quad q_{i,j} = \begin{cases} a_{i,j} & 1 \le i \le n, 1 \le j \le n \\ b_{i,j} & 1 \le i \le n, n < j \le 2n. \end{cases}$$

Obviously, $\mathbf{Q}$ has the same distribution as $pk = (\mathbf{A}, \mathbf{B})$.

If each entry of plaintext is $m_i = 0$, the ciphertext is $(\mathbf{RA} + \mathbf{E}_1, \mathbf{RB} + \mathbf{E}_2)$, which can be written as $\mathbf{RQ} + \mathbf{E}^*$, where $\mathbf{E}^* \in \mathbb{Z}_2^{n\times 2n}$,

$$(\mathbf{e}^*)_{i,j} = \begin{cases} (\mathbf{e}_1)_{i,j} & 1 \le i \le n, \ 1 \le j \le n \\ (\mathbf{e}_2)_{i,j} & 1 \le i \le n, \ n < j \le 2n. \end{cases}$$

According to the DLPN assumptions, it is indistinguishable between $(\mathbf{Q}, \mathbf{RQ} + \mathbf{E}^*)$ and $(\mathbf{S}, \mathbf{R}'\mathbf{S} + \mathbf{E}')$, where $\mathbf{S} \in \mathbb{Z}^{n\times 2n}$, $\mathbf{R}' \leftarrow Ber_\tau^{n\times n}$ and $\mathbf{E}' \leftarrow Ber_\tau^{n\times 2n}$. $(\mathbf{S}, \mathbf{R}'\mathbf{S} + \mathbf{E}^*)$ and $(\mathbf{S}, \mathbf{R}^*)$ can also not be distinguished, in which $\mathbf{R}^*$ is randomly chosen from $\mathbb{Z}_2^{n\times 2n}$.

If a plaintext entry is $m_i = 1$, $(\mathbf{e}_2)_i^T + m\mathrm{I}$ does not change the distribution of $(\mathbf{e}_2)_i^T$ and simply makes a negated operation to $(\mathbf{e}_2)_i^T$. Hence, the ciphertext is indistinguishable from random.

IV. PERFORMANCE ANALYSIS

We choose for 80-, 112-, and 128-bit security, respectively, $n = 9000, 21000$ and 29000, which are suitable and correspond to the security levels of 1024-, 2048-, 3072-bit RSA [23]. Table 2 lists the comparison between our schemes and the Damgård schemes in computational efficiency. All calculations in the schemes based on LPN are on all fields F_2. Therefore, the multiplication and addition have the same overhead. Thus, the computational times in the table are the sum of the multiplication and addition results.

TABLE 2. Comparison between our scheme and Damgård's scheme in size of public key and ciphertext.

Scheme	Size of public key (bit)	Size of ciphertext (bit)	Encoding error
Damgård's single-bit	$2n^2+2n$	$n+1$	have
Our single-bit	$2n^2$	$2n$	no
Damgård's multi-bit	$4n^2$	$2n$	have
Our multi-bit	$2n^2$	$2n^2$	no

TABLE 3. Comparison with Damgård Scheme and RSA public key encryption scheme.

	Time per encryption (ms)			Time per decryption		
Security level (bits)	80	112	128	80	112	128
RSA scheme(not padding)	0.010	0.030	0.060	0.140	0.940	2.890
Damgård's multi-bit	25.80	128.40	241.70	0.052	0.098	0.128
Our multi-bit scheme	15.60	45.30	102.10	0.11	0.221	0.258

Our scheme has the same public key size as in the Damgård scheme. Although our scheme increases slightly in ciphertext size and computational overhead, the decryption error can be negligible.

We compare the performance of our multi-bit scheme with RSA(not padding) and Damgård's scheme in implementation for various security levels as shown in Table 3. The implementation was written in C++ and made use of the NTL library for some mathematical operations. We can see that the encryption in our scheme is slower than in RSA and the decryption in our scheme is faster than in RSA. We get the opposite result when compared with Damgård's multi-bit scheme.

The limitation of our approach is that it does not meet the stronger CCA security. Overcoming this shortcoming is one of our future research directions.

V. CONCLUSIONS

In the post quantum era, the design of public key cryptography under the DLPN assumption is an important research direction. Such schemes have many advantages such as shorter public key and ciphertext, faster encryption and decryption. But the existing scheme is still having the problem of decryption error, which is not satisfactory.

Based on the LPN variants problem, we proposed a single-bit and a multi-bit public key encryption scheme. Our scheme solved the decryption error problem of the existing public key encryption schemes based on DLPN. Compared to existing schemes, there is an increase in only a small amount of ciphertext space and computing overhead in our scheme. Our scheme not only is able to withstand quantum attack but also provides strong practical security at the same time. In the future, we will design a public key scheme based DLPN with high security, smaller public key and ciphertext size, and smaller computational overhead. Furthermore, designing public key cryptography that satisfies CCA security is also one of our future work.

REFERENCES

[1] X. Sun, B. Li, X. Lu, and F. Fang, "CCA secure public key encryption scheme based on LWE without Gaussian sampling," in *Proc. Int. Conf. Inf. Secur. Cryptol.*, vol. 9589, 2015, pp. 361–378.

[2] J. Xu, L. Wei, Y. Zhang, A. Wang, F. Zhou, and C.-Z. Gao, "Dynamic fully homomorphic encryption-based Merkle tree for lightweight streaming authenticated data structures," *J. Netw. Comput. Appl.*, vol. 107, pp. 113–124, Apr. 2018.

[3] Z. Liu, Y. Huang, J. Li, X. Cheng, and C. Shen, "DivORAM: Towards a practical oblivious RAM with variable block size," *Inf. Sci.*, vol. 447, pp. 1–11, Jun. 2018.

[4] T. Li, J. Li, Z. Liu, P. Li, and C. Jia, "Differentially private Naive Bayes learning over multiple data sources," *Inf. Sci.*, vol. 444, pp. 89–104, May 2018.

[5] C.-Z. Gao, Q. Cheng, P. He, W. Susilo, and J. Li, "Privacy-preserving Naive Bayes classifiers secure against the substitution-then-comparison attack," *Inf. Sci.*, vol. 444, pp. 72–88, May 2018.

[6] J. Li, J. Li, X. Chen, C. Jia, and W. Lou, "Identity-based encryption with outsourced revocation in cloud computing," *IEEE Trans. Comput.*, vol. 64, no. 2, pp. 425–437, Feb. 2015.

[7] P. Li *et al.*, "Multi-key privacy-preserving deep learning in cloud computing," *Future Generat. Comput. Syst.*, vol. 74, pp. 76–85, Sep. 2017.

[8] B. Applebaum, D. Cash, C. Peikert, and A. Sahai, "Fast cryptographic primitives and circular-secure encryption based on hard learning problems," in *Advances in Cryptology—CRYPTO* (Lecture Notes in Computer Science), vol. 5677. Berlin, Germany: Springer, 2009, pp. 595–618.

[9] G. Liu, H. Li, and L. Yang, "A topology preserving method of evolving contours based on sparsity constraint for object segmentation," *IEEE Access*, vol. 5, pp. 19971–19982, 2017.

[10] L. Yang, Y. Xiang, and D. Peng, "Precoding-based blind separation of MIMO FIR mixtures," *IEEE Access*, vol. 5, pp. 12417–12427, 2017.

[11] N. Koblitz, A. Menezes, and S. Vanstone, "The state of elliptic curve cryptography," *Des., Codes Cryptogr.*, vol. 19, nos. 2–3, pp. 173–193, 2000.

[12] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM J. Comput.*, vol. 26, no. 5, pp. 1484–1509, 1997.

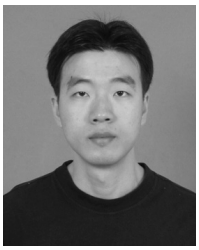
[13] Z. Huang, S. Liu, X. Mao, K. Chen, and J. Li, "Insight of the protection for data security under selective opening attacks," *Inf. Sci.*, vols. 412–413, pp. 223–241, Oct. 2017.

[14] Q. Lin, H. Yan, Z. Huang, W. Chen, J. Shen, and Y. Tang, "An ID-based linearly homomorphic signature scheme and its application in blockchain," *IEEE Access*, vol. 6, pp. 20632–20640, 2018, doi: 10.1109/ACCESS.2018.2809426.

[15] D. Boneh and A. Silverberg, "Applications of multilinear forms to cryptography," *J. Contemp. Math.*, vol. 324, no. 1, pp. 71–90, 2003.

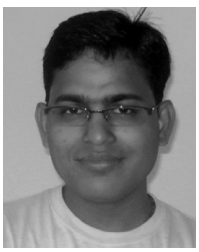
[16] S. Garg, C. Gentry, and S. Halevi, "Candidate multilinear maps from ideal lattices," in *Proc. Annu. Int. Conf. Theory Appl. Cryptograph. Techn.*, vol. 7881, 2013, pp. 1–17.
[17] J.-S. Coron, T. Lepoint, and M. Tibouchi, "Practical multilinear maps over the integers," in *Advances in Cryptology—CRYPTO* (Lecture Notes in Computer Science), vol. 8042. Berlin, Germany: Springer, 2013, pp. 476–493.
[18] S. Hohenberger, A. Sahai, and B. Waters, "Full domain hash from (leveled) multilinear maps and identity-based aggregate signatures," Cryptol. ePrint Arch., Int. Assoc. Cryptol. Res., Tech. Rep. 2013/434, Jul. 2013. [Online]. Available: http://eprint.iacr.org/2013/434.pdf
[19] Y. Hu and H. Jia, "Cryptanalysis of GGH map," Cryptol. ePrint Arch., Int. Assoc. Cryptol. Res., Tech. Rep., Feb. 2016. [Online]. Available: http://eprint.iacr.org/2015/301.pdf
[20] O. Regev, "On lattices, learning with errors, random linear codes, and cryptography," *J. ACM*, vol. 56, no. 6, 2009, Art. no. 34.
[21] C. Gentry, C. Peikert, and V. Vaikuntanathan, "Trapdoors for hard lattices and new cryptographic constructions," in *Proc. 40th Annu. ACM Symp. Theory Comput.*, 2008, pp. 197–206
[22] Z. Brakerski and V. Vaikuntanathan, "Efficient fully homomorphic encryption from (standard) LWE," in *Proc. IEEE Symp. Found. Comput. Sci.*, Oct. 2011, pp. 97–106.
[23] J. Li, Y. Li, X. Chen, P. Lee, and W. Lou, "A hybrid cloud approach for secure authorized deduplication," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 5, pp. 1206–1216 May 2015.
[24] D. Cabarcas, F. Göpfert, and P. Weiden, "Provably secure LWE encryption with smallish uniform noise and secret," *J. ACM*, vol. 2014, pp. 33–42, Jun. 2014.
[25] W. Chen, H. Lei, and K. Qi, "Lattice-based linearly homomorphic signatures in the standard model," *Theor. Comput. Sci.*, vol. 634, pp. 47–54, Jun. 2016.
[26] R. Lindner and C. Peikert, "Better key sizes (and attacks) for LWE-based encryption," in *Proc. Cryptographers' Track RSA Conf.*, vol. 6558, 2011, pp. 319–339.
[27] S. Garg, C. Gentry, S. Halevi, M. Raykova, A. Sahai, and B. Waters, "Candidate indistinguishability obfuscation and functional encryption for all circuits," in *Proc. Annu. IEEE Symp. Found. Comput. Sci.*, Oct. 2013, pp. 40–49.
[28] Q. Lin, J. Li, Z. Huang, W. Chen, and A. J. Shen, "Short linearly homomorphic proxy signature scheme," *IEEE Access*, vol. 6, pp. 12966–12972, 2018.
[29] A. Blum, A. Kalai, and H. Wasserman, "Noise-tolerant learning, the parity problem, and the statistical query model," *J. ACM*, vol. 50, no. 4, pp. 506–519, 2003.
[30] V. Lyubashevsky, "The parity problem in the presence of noise, decoding random linear codes, and the subset sum problem," in *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques* (Lecture Notes in Computer Science), vol. 3624. Berlin, Germany: Springer, 2005, pp. 378–389.
[31] P. Kirchner, "Improved generalized birthday attack," Cryptol. ePrint Arch., Int. Assoc. Cryptol. Res., Tech. Rep., Jun. 2016. [Online]. Available: http://eprint.iacr.org/2011/377.pdf
[32] J. N. Hopper and M. Blum, "Secure human identification protocols," in *Proc. Int. Conf. Theory Appl. Cryptol. Inf. Secur.*, vol. 2248, 2001, pp. 52–66.
[33] A. Juels and A. Stephen, "Authenticating pervasive devices with human protocols," in *Proc. Annu. Int. Cryptol. Conf.*, vol. 3621, 2005, pp. 293–308.
[34] H. Gilbert, M. J. B. Robshaw, and Y. Seurin, "How to encrypt with the LPN problem," in *Automata, Languages and Programming—ICALP* (Lecture Notes in Computer Science), vol. 5126. Berlin, Germany: Springer, 2008, pp. 679–690.
[35] J. Katz, J. S. Shin, and A. Smith, "Parallel and concurrent security of the HB and HB$^+$ protocols," *J. Cryptol.*, vol. 23, no. 3 pp. 402–421, 2010.
[36] M. Alekhnovich, "More on average case vs approximation complexity," in *Proc. 44th IEEE Symp. Found. Comput. Sci.*, Oct. 2003, pp. 298–307.
[37] I. Damgård and S. Park, "How practical is public-key encryption based on LPN and ring-LPN?" Cryptol. ePrint Arch., Int. Assoc. Cryptol. Res., Tech. Rep., Jun. 2016. [Online]. Available: http://eprint.iacr.org/2012/699.pdf
[38] A. Blum, M. Furst, M. Kearns, and R. J. Lipton, "Cryptographic primitives based on hard learning problems," in *Proc. Annu. Int. Cryptol. Conf.*, vol. 773, 2001, pp. 278–291.

ZHIMIN YU was born in Meihekou, China, in 1973. He received the B.S. and M.S. degrees in computer engineering from Tongji University, Shanghai, China, in 1996 and 2004, respectively. He is currently a Lecturer with the School of Computer Engineering, Jiangsu University of Technology, China. His research interests include cryptology and information security.



CHONG-ZHI GAO received the Ph.D. degree in applied mathematics from Sun Yat-sen University in 2004. He is currently a Professor with the School of Computer Science, Guangzhou University. His research interests include cryptography and privacy in machine learning.

ZHENGJUN JING was born in Danyang, China, in 1978. He received the Ph.D. degree in information and security from the Nanjing University of Posts and Telecommunications in 2015. Since 2016, he has been an Associate Professor with the Department of Computer Engineering, Jiangsu University of Technology. His interests are in the cryptanalysis and design of cryptography.



BRIJ BHOOSHAN GUPTA received the Ph.D. degree in information and cyber security from IIT Roorkee, Roorkee, India. He is currently an Assistant Professor with the Department of Computer Engineering, National Institute of Technology at Kurukshetra, Kurukshetra, India. His research interests include information security, cyber security, cloud computing, Web security, intrusion detection, and phishing.

QIURU CAI was born in Qinhuangdao, China, in 1972. She received the B.S. degree in computer engineering from Northeastern University, Shenyang, China, in 1996, and the M.S. degree in computer application from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2008. She is currently a Lecturer with the School of Computer Engineering, Jiangsu University of Technology, China. Her research interests include cryptology and information security.