

# Approximate-Deterministic Public Key Encryption from Hard Learning Problems

Yamin Liu<sup>1,2</sup>, Xianhui Lu<sup>1,2,3(✉)</sup>, Bao Li<sup>1,2,3</sup>, Wenpan Jing<sup>1,2</sup>,  
and Fuyang Fang<sup>1,2,3</sup>

<sup>1</sup> Data Assurance and Communication Security Research Center,  
Chinese Academy of Sciences, Beijing, China

<sup>2</sup> State Key Laboratory of Information Security,  
Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China  
{liuyamin, luxianhui, libao, jingwenpan, fangfuyang}@iie.ac.cn

<sup>3</sup> University of Chinese Academy of Sciences, Beijing, China

**Abstract.** We introduce the notion of approximate-deterministic public key encryption (A-DPKE), which extends the notion of deterministic public key encryption (DPKE) by allowing the encryption algorithm to be “slightly” randomized. However, a ciphertext convergence property is required for A-DPKE such that the ciphertexts of a message are gathering in a small metric space, while ciphertexts of different messages can be distinguished easily. Thus, A-DPKE maintains the convenience of DPKE in fast search and de-duplication on encrypted data, and encompasses new constructions. We present two simple constructions of A-DPKE, respectively from the learning parity with noise and the learning with errors assumptions.

**Keywords:** Deterministic public key encryption · Learning parity with noise · Learning with errors

## 1 Introduction

**Deterministic Public Key Encryption.** The provable security of deterministic public key encryption (DPKE) was initiated by Bellare, Boldyreva and O’Neill in 2007 [4]. Different from the widely accepted notion of probabilistic encryption [19], the encryption algorithm of DPKE does not require a fresh randomness; consequently, given a plaintext, its ciphertext is unique. Hence DPKE can serve as a candidate for efficiently searchable encryption, and supports de-duplication over encrypted data.

Though DPKE can not satisfy most security requirements of randomized public key encryption due to the deterministic encryption algorithm, Bellare, Boldyreva and O’Neill defined an as-strong-as-possible security notion for DPKE, called PRIV, over plaintext distributions with high min-entropy independent of the public key. More security definitions and constructions of DPKE were discussed in [5, 6, 13, 17, 33, 35]. Currently DPKE can be instantiated from various intractability assumptions, including lattice-related ones such as the learning with errors assumption (LWE).

**Hard Learning Problems.** Generally, hard learning problems, such as LWE and LPN (learning parity with noise), refer to learning a secret from several noisy linear equations. LWE was introduced by Regev in [32]. It states that recovering a secret  $\mathbf{s}$  giving  $(\mathbf{A}, \mathbf{b} = \mathbf{A}\mathbf{s} + \mathbf{e})$  is intractable, wherein  $\mathbf{A} \in \mathbb{Z}_q^{m \times n}$  and  $\mathbf{s} \in \mathbb{Z}_q^n$  are chosen at random, and  $\mathbf{e}$  is picked from an error distribution, for appropriate secret dimension  $n$ , number of samples  $m$ , and modulus  $q$ . The decisional version of LWE states that  $(\mathbf{A}, \mathbf{b} = \mathbf{A}\mathbf{s} + \mathbf{e})$  is computationally indistinguishable from the uniform pair  $(\mathbf{A}, \mathbf{u})$ , and is equivalent to the search version. Syntactically, LPN is LWE in the case of  $q = 2$  with the errors being picked from the Bernoulli distribution, however, LPN and LWE are different in many aspects.

The hardness of LWE is guaranteed by worst-case hard problems over lattices, as shown in a series of literatures [11, 29, 32], and LPN is essentially the hard problem of decoding a random binary linear code. In addition, both are believed to be intractable even for quantum algorithms. Thus, it is desirable to instantiate cryptographic primitives from them.

LWE is useful in the construction of various public key cryptographic primitives, such as chosen-ciphertext secure encryption [31], identity-based encryption [20], password-based authenticated key exchange [24], and in our interest, DPKE [35]. The low-noise version of LPN has also been used to construct secure public key encryption [1, 16, 21, 23, 34]. And recently, Yu and Zhang showed how to obtain several public key cryptographic primitives from constant-noise LPN [38], such as CPA/CCA secure encryption and oblivious transfer.

However, the syntax of hard learning problems seems incompatible with the definition of DPKE since it involves a randomly sampled error item, which is important to the intractability of the problems while causes a kind of nondeterminacy. Hence, current constructions of DPKE from LWE either take a detour from lossy trapdoor functions (LTDF) [6, 35], or use a deterministic variant of LWE called learning with rounding (LWR) [12]. The construction of LTDF from LWE are somewhat complicated [31]. The LWR-based constructions of DPKE [3, 37] are very simple, but to ensure the hardness of LWR, the modulus  $q$  should be large enough [3, 7]. Besides, as far as we know, currently there is no construction of DPKE or even LTDF from LPN, and it is believed that there is no “rounding version” of LPN [2].

Nevertheless, we try to address the problem in another way. Remember that using the LWE assumption to instantiate another important cryptographic primitive, the smooth projective hashing (SPH) [14], is also an open problem as stated in [30]. In 2009, Katz and Vaikuntanathan defined a variant of SPH called *approximate smooth projective hashing*, instantiated it with LWE, and obtained the first password-based authenticated key exchange protocol from a lattice-related assumption [24]. Thus, we believe that a similar solution should work for the case of DPKE.

## 1.1 Our Contributions

**Approximate-DPKE.** We extend the definition of DPKE to allow some sort of nondeterminacy while maintaining its advantages, by introducing the notion of *approximate-deterministic public key encryption* (A-DPKE). Compared

with DPKE, in A-DPKE the encryption algorithm is “slightly” randomized, thus there will be many ciphertexts corresponding to one message. However, these ciphertexts are not scattered in the ciphertext space, instead they are gathering in a small metric space. Moreover, ciphertexts of different plaintexts are distributed “far enough” that they will not mix up. We call the property *ciphertext convergence*, and with it A-DPKE preserves the advantages of DPKE in encrypted search and de-duplication, since the ciphertexts of a given message can be easily recognized without decryption.

A-DPKE can achieve the same security level of DPKE, namely, the PRIV-series of security definitions. However, though the encryption algorithm of A-DPKE is randomized, it cannot be as secure as probabilistic encryption due to the ciphertext convergence property: encryptions of the same message can be easily recognized while encryptions of different messages can be easily distinguished. It is a tradeoff between security and functionality just as in the case of DPKE.

Then we can bring out simple and natural instantiations of A-DPKE from hard learning problems.

**A-DPKE from LPN.** To the best of our knowledge, there is no constructions of DPKE from the LPN assumption (neither low-noise nor constant-noise) so far. However, by relaxing DPKE to A-DPKE, immediately we obtain a simple construction of A-DPKE from low-noise LPN, using the trapdoor generation techniques as in [23]. The secret key is a matrix  $\mathbf{T} \in \mathbb{Z}_2^{m \times m}$ , and the public key is a pair of matrices  $(\mathbf{A}, \mathbf{B} = \mathbf{T}\mathbf{A}) \in (\mathbb{Z}_2^{m \times n})^2$ . To encrypt a message  $\mathbf{m} \in \{0, 1\}^n$ , two ciphertext components are computed as  $\mathbf{c}_1 = \mathbf{A}\mathbf{m} + \mathbf{e}$ ,  $\mathbf{c}_2 = (\mathbf{B} + \mathbf{G})\mathbf{m} + \mathbf{T}\mathbf{e}$ , where  $\mathbf{e}$ ,  $\mathbf{T}$  are small errors, and  $\mathbf{G}$  is the generator matrix of an efficiently decodeable binary linear code. We can see that though the encryption is randomized, the Hamming distance of two ciphertexts of a message could be small if the error items are small. By choosing proper parameters, both the ciphertext convergence property and the security will hold.

**A-DPKE from LWE.** Further, we show a natural A-DPKE construction from LWE. The public key is simply a matrix  $\mathbf{A} \in \mathbb{Z}_q^{m \times n}$  generated with the trapdoor generation techniques from [28], and the secret key is the corresponding trapdoor  $\mathbf{R}$ . Then the encryption and decryption are simply the evaluation and inversion of the LWE function. Note that a message  $\mathbf{m} \in \{0, 1\}^n$  is encrypted as  $\mathbf{c} = \mathbf{A}\mathbf{m} + \mathbf{e}$ . By choosing the size of the error item  $\mathbf{e}$  properly, the ciphertext convergence property will hold. And the security is ensured by the hardness of LWE for high min-entropy secrets [3, 18]. Compared with the LWE-based DPKE via LTDF, our A-DPKE is simpler in structure; and compared with the LWR-based DPKE scheme, our LWE-based A-DPKE scheme can use smaller modulus  $q$ .

**Organization.** The rest of the paper is organized as follows. In Sect. 2 some notations and preliminaries about lattice is introduced. In Sect. 3 the definition of A-DPKE is given. In Sect. 4 and Sect. 5 the A-DPKE schemes from LPN and LWE are constructed respectively. Finally, Sect. 6 is the conclusion.

## 2 Preliminaries

### 2.1 Notations

We use bold lower-case characters to denote vectors, such as  $\mathbf{x}$ , and use bold upper-case letters to denote matrices, such as  $\mathbf{X}$ . If  $X$  is a set, then  $x \xleftarrow{\$} X$  denotes that  $x$  is chosen from  $X$  uniformly at random. If  $X$  is a distribution, then  $x \xleftarrow{\$} X$  denotes that  $x$  is randomly sampled according to  $X$ .

For a randomized algorithm  $\mathbf{A}$ ,  $x \xleftarrow{\$} \mathbf{A}(\cdot)$  denotes that  $x$  is assigned the output of  $\mathbf{A}$ . An algorithm is efficient if it runs in polynomial time in its input length. A function  $f(\lambda)$  is negligible if it decreases faster than any polynomial of the security parameter  $\lambda$ , and is denoted as  $f(\lambda) \leq \text{negl}(\lambda)$ .

The min-entropy of a random variable  $X$  is denoted as  $H_\infty(X) = -\log(\max_x P_X(x))$ , wherein  $P_X(x) = \Pr[X = x]$ .  $X$  is called a  $k$ -source if  $H_\infty(X) \geq k$ .

The statistical distance between two random variables  $X$  and  $Y$  is  $\Delta(X, Y) = \frac{1}{2} \sum_x |P_X(x) - P_Y(x)|$ .  $X$  and  $Y$  are statistically close if  $\Delta(X, Y) \leq \epsilon(\lambda)$ , and is denoted as  $X \stackrel{s}{\approx} Y$ .  $X$  and  $Y$  are computationally indistinguishable if no efficient algorithm can tell them apart given only oracle access, and is denoted as  $X \stackrel{c}{\approx} Y$ .

### 2.2 Lattices

A full-rank  $m$ -dimensional lattice  $\Lambda$  generated by a basis  $\mathbf{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_m\} \in \mathbb{Z}^{m \times m}$  is defined as

$$\Lambda = \mathcal{L}(\mathbf{B}) = \{\mathbf{B}\mathbf{x} : \mathbf{x} \in \mathbb{Z}^m\},$$

where  $\mathbf{b}_1, \dots, \mathbf{b}_m$  are linearly independent.

The length of lattice vectors is measured with norms. By default the Euclidean norm is used, i.e.,  $\|\mathbf{x}\|_2 = \sqrt{\sum x_i^2}$ , or solely denoted as  $\|\mathbf{x}\|$ . In some occasions in this work, the infinity norm is also used, i.e.,  $\|\mathbf{x}\|_\infty = \max x_i$ . Obviously, for an  $m$ -dimensional vector  $x$ , if  $\|\mathbf{x}\|_\infty \leq a$ , then  $\|\mathbf{x}\|_2 \leq \sqrt{m}a$ ; and if  $\|\mathbf{x}\|_\infty \geq a$ , then  $\|\mathbf{x}\|_2 \geq a$ .

The length of the shortest nonzero vector in a lattice  $\Lambda$  is denoted by  $\lambda_1(\Lambda)$ . Since lattice points are periodically arranged in every dimension, then  $\lambda_1(\Lambda)$  is the distance of two lattice points in the most “compact” dimension.

The LWE problem is essentially the bounded-distance decoding problem over a full-rank  $m$ -dimensional  $q$ -ary integer lattice  $\Lambda_q(\mathbf{A})$  generated by a random matrix  $\mathbf{A} \in \mathbb{Z}_q^{m \times n}$ :

$$\Lambda_q(\mathbf{A}) = \{y \in \mathbb{Z}^m : \exists \mathbf{s} \in \mathbb{Z}_q^n \text{ s.t. } y = \mathbf{A}\mathbf{s} \pmod{q}\}.$$

## 3 Approximate-DPKE: Definition and Security

Here we define approximate-DPKE. Compared with the original DPKE definition, the main difference is that the encryption algorithm of A-DPKE is randomized. And compared with the definition of randomized PKE, A-DPKE has the

additional property of ciphertext convergence, i.e., the ciphertexts of a message are distributed in a small metric space.

**Definition 1.** *An approximate-deterministic public key encryption scheme A-DPKE = (KG, ENC, DEC) consists of the following three algorithms:*

- *The probabilistic key generation algorithm:  $(pk, sk) \xleftarrow{\$} \text{KG}(1^\lambda)$ .*
- *The probabilistic encryption algorithm:  $c \xleftarrow{\$} \text{ENC}(pk, m; r)$ .*
- *The deterministic decryption algorithm:  $m \leftarrow \text{DEC}(sk, c)$ .*

*And we further require that the encryption scheme should satisfy a “ciphertext convergence” property, i.e., there are a function  $\text{dis}$  measuring the “distance” of ciphertexts, and a distance parameter  $t$ , fulfilling the following requirements:*

- *For arbitrary two ciphertexts  $c_1, c_2$  of a given plaintext  $m$ , there is  $\text{dis}(c_1, c_2) \leq t$ .*
- *For arbitrary two ciphertext-plaintext pairs  $(c, m)$  and  $(c', m')$ , there is  $\text{dis}(c, c') > t$  if  $m \neq m'$ .*

In the definition we explicitly contain the randomness  $r$  in the encryption algorithm. In the following we sometimes omit it in occasions that the choice of randomness is unimportant and just use  $\text{Enc}(pk, m)$ .

The correctness requirement of A-DPKE involves two aspect. One is trivially the decryption correctness, i.e., there should be  $\text{DEC}(sk, \text{ENC}(pk, m; r)) = m$ . The other is the ciphertext convergence property, wherein the choices of the metric function  $\text{dis}$  and parameter  $t$  depend on specific instantiations.

The definition of A-DPKE is a generalization of that of DPKE. Consider the metric function  $\text{dis}$  to be the Hamming distance, e.g., the numbers of bit-wise differences between two ciphertexts, and set the parameter  $t = 0$ , then a DPKE certainly satisfies the ciphertext convergence property.

As to the security requirement, it is clear that A-DPKE can achieve existing security requirements of DPKE, e.g., the PRIV security. The question is whether it can be semantically secure [19]. The answer is NO, but the consequence is not necessarily negative. On one side, though the encryption algorithm of A-DPKE is randomized, it still can not achieve semantic security due to the ciphertext convergence property. On the other side, with this property, A-DPKE preserves the advantages of DPKE in searchable encryption and de-duplication, since the ciphertexts of a certain message can be efficiently recognized without decryption, given  $\text{dis}$  and  $t$ .

In the following we give the definition of PRIV-IND security [5, 6] for A-DPKE, which requires that the encryptions of messages from two different high min-entropy distributions are indistinguishable.

**Definition 2 (PRIV-IND security for A-DPKE).** *An approximate-deterministic public-key encryption scheme  $\Pi = (\text{KG}, \text{Enc}, \text{Dec})$  is PRIV-IND secure if for any probabilistic polynomial time adversary  $\mathbf{A}$ , for any efficiently sampleable distributions  $\{M_\lambda^0\}_{\lambda \in \mathbb{N}}$  and  $\{M_\lambda^1\}_{\lambda \in \mathbb{N}}$  with sufficient min-entropy*

$H_\infty(M_\lambda^0) \geq k$  and  $H_\infty(M_\lambda^1) \geq k$ , there is  $(pk, \text{Enc}(pk, \mathbf{m}_0)) \stackrel{c}{\approx} (pk, \text{Enc}(pk, \mathbf{m}_1))$ , where  $(pk, sk) \stackrel{\$}{\leftarrow} \text{Gen}(1^\lambda)$ ,  $\mathbf{m}_0 \stackrel{\$}{\leftarrow} M_\lambda^0$  and  $\mathbf{m}_1 \stackrel{\$}{\leftarrow} M_\lambda^1$ .

In [37] Xie et al. defined a PRIV-INDr security for DPKE, which requires that the encryption is indistinguishable from uniform. It is clear that PRIV-INDr implies PRIV-IND. We also define the PRIV-INDr security for A-DPKE.

**Definition 3 (PRIV-INDr security for A-DPKE).** *An approximate-deterministic public-key encryption scheme  $\Pi = (\text{KG}, \text{Enc}, \text{Dec})$  is PRIV-INDr secure if for any probabilistic polynomial time adversary  $\mathbf{A}$ , for any efficiently sampleable distributions  $\{M_\lambda\}_{\lambda \in \mathbb{N}}$  with sufficient min-entropy  $H_\infty(M_\lambda) \geq k$ , there is  $(pk, \text{Enc}(pk, \mathbf{m})) \stackrel{c}{\approx} (pk, \mathbf{u})$ , where  $(pk, sk) \stackrel{\$}{\leftarrow} \text{Gen}(1^\lambda)$ ,  $\mathbf{m} \stackrel{\$}{\leftarrow} M_\lambda$  and  $\mathbf{u} \stackrel{\$}{\leftarrow} C_\lambda$ , where  $C_\lambda$  is the ciphertext space.*

Note that other forms of security definitions for DPKE can also be extended to the A-DPKE case naturally, such as PRIV with respect to hard-to-invert auxiliary information [13].

In the definition of PRIV security, the message blocks  $\mathbf{m}_0$  and  $\mathbf{m}_1$  contain several (possibly correlated) messages. If the block size is restricted to be one, then the security is called PRIV1 [4–6]. Full PRIV security in the standard model is considered to be elusive [36], and currently the only known approach to achieve it is the one proposed by Bellare and Hoang in [8], with the help of a newly introduced strong assumption UCE (universal computational extractor) [9]. Thus, in this work, we are satisfied with just the PRIV1 security.

## 4 A-DPKE from LPN

So far, there is no known constructions of DPKE from the learning parity with noise assumption. Now we propose an A-DPKE scheme under the LPN assumption, which depicts that the relaxation from deterministic to approximate-deterministic is worthwhile.

### 4.1 Coding Theory

In the LPN based A-DPKE construction, we will use a linear code as a building block. Thus some preliminaries about the coding theory are recalled below. The notations and definitions mainly come from [26] by Meurer.

For  $x \in [0, 1]$ , the  $q$ -ary entropy function is defined as  $H_q(x) = x \log_q(q-1) - x \log_q x - (1-x) \log_q(1-x)$ . In particular, when  $q = 2$ ,  $H(x) = x \log x - (1-x) \log(1-x)$ .

**Definition 4 (Linear Code).** *A linear code  $\mathcal{C}$  in a finite field  $\mathbb{Z}_q$  is a linear subspace of the linear space  $\mathbb{Z}_q^m$ . If the dimension of  $\mathcal{C}$  is  $n$ , then  $\mathcal{C}$  is called an  $[m, n]$ -code. The ratio  $R = \frac{n}{m}$  is called the information rate of  $\mathcal{C}$ .*

In this work, we use linear codes in  $\mathbb{Z}_2^m$ . Given a generator matrix  $\mathbf{A} \in \mathbb{Z}_2^{m \times n}$ , a code  $\mathcal{C}(\mathbf{A}) = \{\mathbf{c} = \mathbf{A}\mathbf{s} : \mathbf{s} \in \mathbb{Z}_2^n\}$  is specified. An important parameter of a code  $\mathcal{C}$  is the minimum distance  $d(\mathcal{C})$ , which is the minimum Hamming distance between two distinct codewords, i.e.,  $d(\mathcal{C}) = \min_{\mathbf{c}_1 \neq \mathbf{c}_2 \in \mathcal{C}} |\mathbf{c}_1 - \mathbf{c}_2| = \min_{\mathbf{c} \in \mathcal{C} \setminus \{\mathbf{0}\}} |\mathbf{c}|$ . With the relative Gilbert-Varshamov distance, a lower bound of  $d(\mathcal{C})$  can be estimated.

**Definition 5 (Relative Gilbert-Varshamov distance).** *Let  $0 < R < 1$ . The relative Gilbert-Varshamov distance  $D_{\text{GV}}(R) \in \mathbb{R}$  is the unique solution in  $0 \leq x \leq 1 - \frac{1}{q}$  of the equation  $H_q(x) = 1 - R$ .*

The following lemma from [26] shows the lower bound of a linear code  $\mathcal{C}$ .

**Lemma 1.** *Almost all linear codes meet the relative Gilbert-Varshamov distance, i.e., for almost all linear codes  $\mathcal{C}$  of rate  $R$  it holds  $d(\mathcal{C}) \geq \lfloor D_{\text{GV}}(R)m \rfloor$ .*

## 4.2 The LPN Assumption

Then we recall the LPN assumption, wherein the error item is sampled from the Bernoulli distribution  $\mathcal{B}_\rho$  with  $0 < \rho < 1/2$ , i.e.,  $\Pr[x = 1 : x \xleftarrow{\$} \mathcal{B}_\rho] = \rho$ .

**Definition 6 (Learning Parity with Noise).** *Let  $\lambda$  be the security parameter,  $n = n(\lambda), m = m(\lambda)$  be integers, and  $\rho \in (0, 1/2)$  be the Bernoulli parameter. The  $\text{LPN}_{n,m,\rho}$  assumption states that, if we choose  $\mathbf{A} \xleftarrow{\$} \mathbb{Z}_2^{m \times n}, \mathbf{s} \xleftarrow{\$} \mathbb{Z}_2^n, \mathbf{e} \xleftarrow{\$} \mathcal{B}_\rho^m, \mathbf{u} \xleftarrow{\$} \mathbb{Z}_2^m$ , then the following distributions are computationally indistinguishable:*

$$(\mathbf{A}, \mathbf{A}\mathbf{s} + \mathbf{e}) \stackrel{c}{\approx} (\mathbf{A}, \mathbf{u}).$$

In standard LPN, the Bernoulli parameter  $\rho$  is a constant such as  $1/10$ . However, for the purpose of constructing PKE schemes, we mainly use a low-noise variant of LPN contributed by Alekhnovich [1], in which  $\rho = \Theta(1/\sqrt{n})$ .

And we still need another variant of LPN called Knapsack LPN (KLPN), which is defined below.

**Definition 7 (Knapsack LPN).** *Let  $\lambda$  be the security parameter,  $n = n(\lambda), m = m(\lambda)$  be integers, and  $\rho \in (0, 1/2)$  be the Bernoulli parameter. The  $\text{KLPN}_{n,m,\rho}^m$  assumption states that, if we choose  $\mathbf{A} \xleftarrow{\$} \mathbb{Z}_2^{m \times (m-n)}, \mathbf{T} \xleftarrow{\$} \mathcal{B}_\rho^{m \times m}, \mathbf{u} \xleftarrow{\$} \mathbb{Z}_2^{m \times (m-n)}$ , then the following distributions are computationally indistinguishable:*

$$(\mathbf{A}, \mathbf{T}\mathbf{A}) \stackrel{c}{\approx} (\mathbf{A}, \mathbf{u}).$$

The equivalence of LPN and KLPN assumptions was stated in [23, 27] with the following lemma:

**Lemma 2 [23].** *For all algorithms  $\mathbf{B}$  there exists an algorithm  $\mathbf{A}$  that runs in roughly the same time as  $\mathbf{B}$  and  $\text{Adv}_{\text{LPN}_{n,m,\rho}}(\mathbf{A}) \geq \frac{1}{m} \text{Adv}_{\text{KLPN}_{n,m,\rho}^m}(\mathbf{B})$ .*

### 4.3 Construction 1: A-DPKE from LPN

Now we will describe the A-DPKE construction from LPN. Firstly we set the parameters used in the construction. Some choices of parameters are similar to those in [23].

- $\lambda$  is the security parameter,  $n(\lambda), m(\lambda)$  are integers, where  $n = \Theta(\lambda^2)$  and  $m \geq 2n$ . Besides,  $m > 1400$ . For reasonable choices of the security parameter, say  $\lambda \geq 80$ ,  $m > 1400$  can be trivially met.
- $0 < c < 1/4$  is a constant. And the Bernoulli parameter is  $\rho = \sqrt{c/m}$ .
- $\beta = 2\sqrt{cm}$  is a parameter used in the correctness proof of the construction.
- $\mathbf{G} \in \mathbb{Z}_2^{m \times n}$  is the generator-matrix of a binary linear error-correcting code  $\mathcal{C} : \mathbb{Z}_2^n \rightarrow \mathbb{Z}_2^m$  with an efficient decoding algorithm  $\text{Decode}_{\mathbf{G}}$  which corrects up to  $\alpha m$  errors with  $4c < \alpha \leq 0.05$ .

Now the A-DPKE based on LPN is given below:

- $\text{KG}(1^\lambda)$ : Choose  $\mathbf{T} \xleftarrow{\$} \mathcal{B}_\rho^{m \times m}$ ,  $\mathbf{A} \xleftarrow{\$} \mathbb{Z}_2^{m \times n}$ , and compute  $\mathbf{B} = \mathbf{T}\mathbf{A}$ . Set  $sk = \mathbf{T}$  and  $pk = (\mathbf{A}, \mathbf{B})$ .
- $\text{Enc}(pk, \mathbf{m}; (\mathbf{e}, \bar{\mathbf{T}}))$ : To encrypt a message  $\mathbf{m} \in \{0, 1\}^n$ , choose  $\mathbf{e} \xleftarrow{\$} \mathcal{B}_\rho^m$ ,  $\bar{\mathbf{T}} \xleftarrow{\$} \mathcal{B}_\rho^{m \times m}$ , and compute

$$\mathbf{c}_1 = \mathbf{A}\mathbf{m} + \mathbf{e}, \mathbf{c}_2 = (\mathbf{B} + \mathbf{G})\mathbf{m} + \bar{\mathbf{T}}\mathbf{e},$$

where  $\mathbf{G}$  is the generator matrix of a binary linear code defined above as part of the public parameter. Finally, set the ciphertext  $\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2)$ .

- $\text{Dec}(sk, \mathbf{C})$ : Parse  $\mathbf{C}$  as  $(\mathbf{c}_1, \mathbf{c}_2)$ . Compute  $\mathbf{y} = \mathbf{c}_2 - \mathbf{T}\mathbf{c}_1$  and set  $\mathbf{m} = \text{Decode}_{\mathbf{G}}(\mathbf{y})$ .

### 4.4 Correctness

To establish the correctness of **Construction 1** over the specified parameter settings, two lemmata from literatures are required. The first one is the Chernoff Bound for bounding the Hamming weight of a vector constituted by independent Bernoulli random variables, e.g., the weight of the error item in the first component of the ciphertext.

**Lemma 3 (Chernoff Bound).** For  $\mathbf{d} \xleftarrow{\$} \mathcal{B}_\rho^m$  and  $\delta > 0$ ,

$$\Pr[|\mathbf{d}| > (1 + \delta)\rho m] < e^{-\frac{\min(\delta, \delta^2)}{3}\rho m}.$$

In our case,  $\delta = 1$ . The other lemma is essentially from [23], bounding the Hamming weight of the error item in the second component of the ciphertext.

**Lemma 4 [23].** For  $\mathbf{e} \xleftarrow{\$} \mathcal{B}_\rho^m$  with  $|\mathbf{e}| \leq 2\rho m$ ,  $\mathbf{T} \xleftarrow{\$} \mathcal{B}_\rho^{m \times m}$ , and  $4c < \alpha < 1$ , there is

$$\Pr_{\mathbf{T}}[|\mathbf{T}\mathbf{e}| > \frac{\alpha}{2}m] < \text{negl}(\lambda).$$

Then we can establish the correctness of **Construction 1** as an A-DPKE.



**Theorem 1.** *Let  $\lambda$  be the security parameter,  $n, m, c, \rho, \beta, \alpha$  and the error correcting code  $\mathbf{G}$  be defined above. Choose the distance function  $\text{dis}$  such that  $\text{dis}(\mathbf{C}_1, \mathbf{C}_2) = |\mathbf{C}_1 - \mathbf{C}_2| = (|\mathbf{c}_{1,1} - \mathbf{c}_{1,2}|, |\mathbf{c}_{2,1} - \mathbf{c}_{2,2}|)$  denotes the Hamming distance of two ciphertexts  $\mathbf{C}_1 = (\mathbf{c}_{1,1}, \mathbf{c}_{1,2})$ ,  $\mathbf{C}_2 = (\mathbf{c}_{2,1}, \mathbf{c}_{2,2})$ , and set the parameter  $t = (t_1, t_2) = (2\beta, \alpha m)$ . Then the above construction is a correct A-DPKE scheme.*

*Proof.* The correctness includes the decryption correctness and the ciphertext convergence.

– Decryption correctness.

Given a ciphertext  $\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2)$ , the decryption algorithm computes

$$\mathbf{y} = \mathbf{c}_2 - \mathbf{T}\mathbf{c}_1 = \mathbf{G}\mathbf{m} + (\bar{\mathbf{T}} - \mathbf{T})\mathbf{e}.$$

To ensure that  $\text{Decode}_{\mathbf{G}}(\mathbf{y})$  correctly recovers  $\mathbf{m}$ , the Hamming weight of the error term  $(\bar{\mathbf{T}} - \mathbf{T})\mathbf{e}$  should be small, i.e.,  $|(\bar{\mathbf{T}} - \mathbf{T})\mathbf{e}| \leq \alpha m$ .

With the parameters  $\rho = \sqrt{c/m}$ ,  $\beta = 2\sqrt{cm} = 2\rho m$ , and the Chernoff Bound for  $\delta = 1$ , the Hamming weight of  $\mathbf{e}$  is bounded by  $\beta$  with overwhelming probability. That is,

$$\Pr_{\mathbf{e} \leftarrow \mathcal{B}_\rho^m} [|\mathbf{e}| > \beta] \leq e^{-\rho m/3} = 2^{-\Theta(\sqrt{m})}.$$

Then with the triangular inequality and Lemma 4 from [23], there is

$$|(\bar{\mathbf{T}} - \mathbf{T})\mathbf{e}| \leq |\bar{\mathbf{T}}\mathbf{e}| + |\mathbf{T}\mathbf{e}| \leq \alpha m,$$

with overwhelming probability. Thus,  $\text{Decode}_{\mathbf{G}}(\mathbf{y})$  will recover  $\mathbf{m}$  with overwhelming probability.

– Ciphertext convergence.

- Given a message  $\mathbf{m}$ , and its arbitrary two ciphertexts:

$$\mathbf{C}_1 = (\mathbf{c}_{1,1} = \mathbf{A}\mathbf{m} + \mathbf{e}_1, \mathbf{c}_{1,2} = (\mathbf{B} + \mathbf{G})\mathbf{m} + \bar{\mathbf{T}}_1\mathbf{e}_1),$$

$$\mathbf{C}_2 = (\mathbf{c}_{2,1} = \mathbf{A}\mathbf{m} + \mathbf{e}_2, \mathbf{c}_{2,2} = (\mathbf{B} + \mathbf{G})\mathbf{m} + \bar{\mathbf{T}}_2\mathbf{e}_2).$$

According to Lemmas 3 and 4, there is

$$\begin{aligned} \text{dist}(\mathbf{C}_1, \mathbf{C}_2) &= (|\mathbf{c}_{1,1} - \mathbf{c}_{1,2}|, |\mathbf{c}_{2,1} - \mathbf{c}_{2,2}|) \\ &= (|\mathbf{e}_1 - \mathbf{e}_2|, |\bar{\mathbf{T}}_1\mathbf{e}_1 - \bar{\mathbf{T}}_2\mathbf{e}_2|) \\ &\leq (2\beta, \alpha m), \end{aligned}$$

with overwhelming probability, i.e., the ciphertexts of the same message are close in Hamming distance.

- Given two different messages  $\mathbf{m}$  and  $\mathbf{m}'$ , and two ciphertexts of them:

$$\mathbf{C} = (\mathbf{c}_1 = \mathbf{A}\mathbf{m} + \mathbf{e}, \mathbf{c}_2 = (\mathbf{B} + \mathbf{G})\mathbf{m} + \bar{\mathbf{T}}\mathbf{e}),$$

$$\mathbf{C}' = (\mathbf{c}'_1 = \mathbf{A}\mathbf{m}' + \mathbf{e}', \mathbf{c}'_2 = (\mathbf{B} + \mathbf{G})\mathbf{m}' + \bar{\mathbf{T}}'\mathbf{e}'),$$

there is  $\text{dist}(\mathbf{C}, \mathbf{C}') = (|\mathbf{c}_1 - \mathbf{c}'_1|, |\mathbf{c}_2 - \mathbf{c}'_2|)$ .

Let us analyze the two components separately. Consider  $\mathbf{A}$  as the generator matrix of a linear code  $\mathcal{C}(\mathbf{A})$ , then  $|\mathbf{A}(\mathbf{m} - \mathbf{m}')|$  is not less than the minimum distance of  $\mathcal{C}(\mathbf{A})$ . With the triangular inequality and Lemma 1, there is

$$\begin{aligned} |\mathbf{c}_1 - \mathbf{c}'_1| &= |\mathbf{A}(\mathbf{m} - \mathbf{m}') + (\mathbf{e} - \mathbf{e}')| \\ &\geq |\mathbf{A}(\mathbf{m} - \mathbf{m}')| - |\mathbf{e} - \mathbf{e}'| \\ &\geq d(\mathcal{C}(\mathbf{A})) - 2\beta \\ &\geq \lfloor D_{\text{GV}}(\frac{n}{m})m \rfloor - 2\beta \\ &\geq D_{\text{GV}}(\frac{1}{2})m - 1 - 2\beta. \end{aligned}$$

By a routine calculation based on Definition 5 there is  $D_{\text{GV}}(\frac{1}{2}) > 0.11$ . Since  $m > 1400$  and  $c < 0.25$ , that is,  $c < 0.000179m$ , then  $2\beta = 4\sqrt{cm} < 0.0536m$ . Hence there is

$$\begin{aligned} |\mathbf{c}_1 - \mathbf{c}'_1| &\geq 0.11m - 1 - 0.0536m \\ &= 0.0564m - 1 \\ &> 2\beta = t_1. \end{aligned}$$

Similarly, view  $\mathbf{U} = \mathbf{B} + \mathbf{G}$  as the generator matrix of a linear code  $\mathcal{C}(\mathbf{U})$ , then there is

$$\begin{aligned} |\mathbf{c}_2 - \mathbf{c}'_2| &= |\mathbf{U}(\mathbf{m} - \mathbf{m}') + (\bar{\mathbf{T}}\mathbf{e} - \bar{\mathbf{T}}'\mathbf{e}')| \\ &\geq |\mathbf{U}(\mathbf{m} - \mathbf{m}')| - |\bar{\mathbf{T}}\mathbf{e} - \bar{\mathbf{T}}'\mathbf{e}'| \\ &\geq d(\mathcal{C}(\mathbf{U})) - \alpha m \\ &\geq \lfloor D_{\text{GV}}(\frac{n}{m})m \rfloor - \alpha m \\ &\geq D_{\text{GV}}(\frac{1}{2})m - 1 - \alpha m \\ &> 0.11m - 1 - 0.05m \\ &= 0.06m - 1 \\ &> 0.05m \geq \alpha m = t_2. \end{aligned}$$

Hence it holds that  $\text{dis}(\mathbf{C}, \mathbf{C}') > (2\beta, \alpha m)$ , i.e., ciphertexts of different messages are far enough in Hamming distance.  $\square$

## 4.5 Security

Now we can show the PRIV1-INDr security of **Construction 1**.

**Theorem 2.** *Let  $\lambda$  be the security parameter,  $n, m, c, \rho, \beta, \alpha$  and the error correcting code  $\mathbf{G}$  be defined above. If the LPN assumption holds, then the above construction is PRIV1-INDr secure for uniformly distributed messages.*

*Proof.* Since for  $(pk, sk) \xleftarrow{\$} \text{KG}(1^\lambda)$ ,  $\mathbf{m} \xleftarrow{\$} \{0, 1\}^n$ ,  $\mathbf{e} \xleftarrow{\$} \mathcal{B}_\rho^m$ ,  $\bar{\mathbf{T}} \xleftarrow{\$} \mathcal{B}_\rho^{m \times m}$ , there is

$$(pk, \text{Enc}(pk, \mathbf{m})) = ((\mathbf{A}, \mathbf{B}), (\mathbf{A}\mathbf{m} + \mathbf{e}, (\mathbf{B} + \mathbf{G})\mathbf{m} + \bar{\mathbf{T}}\mathbf{e})) \quad (1)$$

$$\stackrel{c}{\approx} ((\mathbf{A}, \mathbf{B}'), (\mathbf{A}\mathbf{m} + \mathbf{e}, (\mathbf{B}' + \mathbf{G})\mathbf{m} + \bar{\mathbf{T}}\mathbf{e})) \quad (2)$$

$$\stackrel{c}{\approx} ((\mathbf{A}, \mathbf{B}'), (\mathbf{A}\mathbf{m} + \mathbf{e}, \mathbf{U}\mathbf{m} + \bar{\mathbf{T}}\mathbf{e})) \quad (3)$$

$$\stackrel{c}{\approx} ((\mathbf{A}, \mathbf{B}'), (\mathbf{u}_1, \mathbf{u}_2)), \quad (4)$$

where  $\mathbf{B}', \mathbf{U} \xleftarrow{\$} \mathbb{Z}_2^{m \times n}$ ,  $\mathbf{u}_1, \mathbf{u}_2 \xleftarrow{\$} \mathbb{Z}_2^m$ . Step 1 is straightforward. Step 2 follows from the KLPN assumption. Step 3 is also natural since  $\mathbf{B}'$  is uniform. And step 4 follows from the LPN assumption.  $\square$

## 5 A-DPKE from LWE

### 5.1 The LWE Assumption

Next we will show a natural construction of A-DPKE from the learning with errors assumption. Firstly we recall the definition of (decisional) LWE.

**Definition 8 (Learning with Errors [32]).** *Let  $\lambda$  be the security parameter,  $n = n(\lambda), m = m(\lambda), q = q(\lambda)$  be integers, and  $\chi = \chi(\lambda)$  be a distribution over  $\mathbb{Z}_q$ . The  $\text{LWE}_{n,m,q,\chi}$  assumption states that, if we choose  $\mathbf{A} \xleftarrow{\$} \mathbb{Z}_q^{m \times n}$ ,  $\mathbf{s} \xleftarrow{\$} \mathbb{Z}_q^n$ ,  $\mathbf{e} \leftarrow \chi^m$ ,  $\mathbf{u} \xleftarrow{\$} \mathbb{Z}_q^m$ , then the following distributions are computationally indistinguishable:*

$$(\mathbf{A}, \mathbf{A}\mathbf{s} + \mathbf{e}) \stackrel{c}{\approx} (\mathbf{A}, \mathbf{u}).$$

Typically, the error distribution is the discrete Gaussian distribution over  $\mathbb{Z}_q$  with appropriate variance, or the uniform distribution over a small interval [15].

The equivalent computational version of LWE states that getting the secret  $\mathbf{s}$  from  $(\mathbf{A}, \mathbf{b} = \mathbf{A}\mathbf{s} + \mathbf{e})$  is hard. However, with the trapdoor generation technique from [28], the secret  $\mathbf{s}$  can be efficiently recovered.

**Lemma 5 [28].** *There is an efficient randomized algorithm  $\text{GenTrap}(1^n, 1^m, q)$  that, given any integers  $n \geq 1, q \geq 2$ , and sufficiently large  $m = O(n \log q)$ , outputs a parity-check matrix  $\mathbf{A} \in \mathbb{Z}_q^{m \times n}$  and a ‘trapdoor’  $\mathbf{R}$  such that the distribution of  $\mathbf{A}$  is  $\text{negl}(n)$ -far from uniform. Moreover, there are an efficient algorithm  $\text{Invert}$  that with overwhelming probability over all random choices, does the following:*

- For  $\mathbf{b} = \mathbf{A}\mathbf{s} + \mathbf{e}$ , where  $\mathbf{s} \in \mathbb{Z}_q^n$  is arbitrary and either  $\|\mathbf{e}\| < q/O(\sqrt{n \log q})$  or  $\mathbf{e} \leftarrow D_{\mathbb{Z}^m, \alpha q}$  for  $1/\alpha \geq \sqrt{n \log q} \cdot \omega(\sqrt{\log n})$ , the deterministic algorithm  $\text{Invert}(\mathbf{R}, \mathbf{A}, \mathbf{b})$  outputs  $\mathbf{s}$  and  $\mathbf{e}$ .

Goldwasser et al. proved that LWE is hard even for non-uniform secret  $\mathbf{s}$  with hard-to invert auxiliary information  $f(\mathbf{s})$ , provided that  $\mathbf{s}$  has high min-entropy and the modulus  $q$  is super-polynomial [18]. The size of the modulus  $q$  was improved to be polynomial by Alwen et al. in [3] (in the appendix of its full version) with the following definition and lemma.

**Definition 9 (LWE with Weak and Leaky Secrets [3]).** Let  $\lambda$  be the security parameter,  $n = n(\lambda)$ ,  $m = m(\lambda)$ ,  $q = q(\lambda)$  be integer parameters, and  $\chi$  be a distribution over  $\mathbb{Z}_q$ . Let  $\gamma = \gamma(\lambda) \in (0, q/2)$  be an integer and  $k = k(\lambda)$  be a real. The  $\text{LWE}_{n,m,q,\chi}^{\text{WL}(\gamma,k)}$  problem says that for any efficiently sampleable correlated random variables  $(\mathbf{s}, \text{aux})$ , where the support of  $\mathbf{s}$  is the integer interval  $[-\gamma, \gamma]^n$  and  $H_\infty(\mathbf{s}|\text{aux}) \geq k$ , the following distributions are computationally indistinguishable:

$$(\text{aux}, \mathbf{A}, \mathbf{A}\mathbf{s} + \mathbf{e}) \stackrel{c}{\approx} (\text{aux}, \mathbf{A}, \mathbf{u}),$$

where  $\mathbf{A} \xleftarrow{\$} \mathbb{Z}_q^{m \times n}$ ,  $\mathbf{u} \xleftarrow{\$} \mathbb{Z}_q^m$ ,  $\mathbf{e} \xleftarrow{\$} \chi^m$  are chosen randomly and independently of  $(\mathbf{s}, \text{aux})$ .

The lemma below states that the hardness of LWE for weak and leaky sources follows from that of the standard LWE.

**Lemma 6 [3].** Let  $k, l, m, n, \beta, \gamma, \sigma, q$  be integer parameters and  $\chi$  a distribution (all parameterized by  $\lambda$ ) such that  $\Pr_{x \leftarrow \chi}[|x| \geq \beta] \leq \text{negl}(\lambda)$  and  $\sigma \geq \beta\gamma nm$ . Let

$\Psi_\sigma$  be either:

- The discrete Gaussian distribution with standard deviation  $\sigma$ , or
- The uniform distribution over the integer interval  $[-\sigma, \sigma]$ .

Assuming that the  $\text{LWE}_{l,m,q,\chi}$  assumption holds, the weak and leaky  $\text{LWE}_{n,m,q,\Psi_\sigma}^{\text{WL}(\gamma,k)}$ -assumption holds if  $k \geq (l + \Omega(\lambda)) \log q$ .

In our construction, we choose  $\gamma = 1$ , and use binary secrets  $\mathbf{s} \in \{0, 1\}^n$ .

## 5.2 A-DPKE from LWE

### 5.3 Construction 2: A-DPKE from LWE

The A-DPKE construction from LWE is shown below. Firstly we list the parameter settings:

- $\lambda$  is the security parameter, and  $n(\lambda), m(\lambda), q(\lambda)$  are integers, with  $m \geq 2n \log q$ . For simplicity, we let  $q$  be prime.

- Let  $\Psi_\sigma$  be a suitable error distribution with  $\beta nm < \sigma < \min(\frac{q}{16}, \frac{q}{O(\sqrt{n \log q})})$ , where  $\beta$  is the parameter for another error distribution  $\chi_\beta$  with which the LWE assumption holds.

Then the encryption and decryption of A-DPKE are simply the evaluation and inversion of the LWE function, using the trapdoor generation technique in Lemma 5.

- $\text{KG}(1^\lambda)$ . Run  $(\mathbf{A}, \mathbf{R}) \xleftarrow{\$} \text{GenTrap}(1^n, 1^m, q)$ . Set  $pk = \mathbf{A}$  and  $sk = \mathbf{R}$ .
- $\text{Enc}(pk, \mathbf{m}; \mathbf{e})$ . To encrypt a message  $\mathbf{m} \in \{0, 1\}^n$ , compute  $\mathbf{c} = \mathbf{A}\mathbf{m} + \mathbf{e}$ , where  $\mathbf{e} \in \mathbb{Z}^m$  is randomly sampled according to the distribution  $\Psi_\sigma$ .
- $\text{Dec}(sk, \mathbf{c})$ . Run  $(\mathbf{m}, \mathbf{e}) \leftarrow \text{Invert}(\mathbf{R}, \mathbf{A}, \mathbf{c})$ , and output  $\mathbf{m}$ .

#### 5.4 Correctness

Intuitively, the encryption algorithm of **Construction 2** encodes a message  $\mathbf{m}$  to a point near the lattice point  $\mathbf{A}\mathbf{s}$  in the  $q$ -ary lattice  $\Lambda_q(\mathbf{A})$ , and the offset is the error size. Thus, to prove ciphertext convergence property, we need the following lemma bounding the length of the shortest nonzero vector of  $\Lambda_q(\mathbf{A})$ , in the form of infinity norm.

**Lemma 7** [20]. *Let  $n$  and  $q$  be positive integers with  $q$  prime, and let  $m \geq 2n \log q$ . Then for all but at most  $q^{-n}$  fraction of  $\mathbf{A} \in \mathbb{Z}_q^{m \times n}$ , we have  $\lambda_1^\infty(\Lambda_q(\mathbf{A})) \geq q/4$ .*

An immediate corollary explains the bound in the form of Euclidean norm.

**Corollary 1.** *Let  $n$  and  $q$  be positive integers with  $q$  prime, and let  $m \geq 2n \log q$ . Then for all but at most  $q^{-n}$  fraction of  $\mathbf{A} \in \mathbb{Z}_q^{m \times n}$ , we have  $\lambda_1^2(\Lambda_q(\mathbf{A})) \geq q/4$ .*

Now we can show the correctness of **Construction 2**.

**Theorem 3.** *Let  $\lambda$  be the security parameter,  $n = n(\lambda)$ ,  $m = m(\lambda)$ ,  $q = q(\lambda)$  be integers, with  $m \geq 2n \log q$  and  $q$  being prime. Let  $\Psi_\sigma$  be the error distribution with  $\beta nm < \sigma < \min(\frac{q}{16}, \frac{q}{O(\sqrt{n \log q})})$ , thus  $\|\mathbf{e}\| \leq \sigma$ . Choose the distance function  $\text{dis}$  such that  $\text{dis}(\mathbf{c}_1, \mathbf{c}_2) = \|\mathbf{c}_1 - \mathbf{c}_2\|$  denotes the distance of the two vectors  $\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{Z}_q^m$ , and set the parameter  $t = 2\sigma$ . Then the above construction is a correct A-DPKE scheme.*

*Proof.*

- The decryption correctness follows from Lemma 5.
- Ciphertext convergence.
  - Given a message  $\mathbf{m}$ , and its arbitrary two ciphertexts  $\mathbf{c}_1 = \mathbf{A}\mathbf{m} + \mathbf{e}_1, \mathbf{c}_2 = \mathbf{A}\mathbf{m} + \mathbf{e}_2$ , then there is

$$\begin{aligned}
 \text{dis}(\mathbf{c}_1, \mathbf{c}_2) &= \|\mathbf{c}_1 - \mathbf{c}_2\| \\
 &= \|\mathbf{e}_1 - \mathbf{e}_2\| \\
 &\leq \|\mathbf{e}_1\| + \|\mathbf{e}_2\| \\
 &\leq 2\sigma.
 \end{aligned}$$

It means that the ciphertexts of the same message are close in Euclidean distance.

- Given two different messages  $\mathbf{m}, \mathbf{m}'$ , and two ciphertexts of them,  $\mathbf{c} = \mathbf{A}\mathbf{m} + \mathbf{e}, \mathbf{c}' = \mathbf{A}\mathbf{m}' + \mathbf{e}'$ . With Lemma 7 and Corollary 1 there is

$$\begin{aligned}
\text{dis}(\mathbf{c}, \mathbf{c}') &= \|\mathbf{c} - \mathbf{c}'\| \\
&= \|(\mathbf{A}\mathbf{m} - \mathbf{A}\mathbf{m}') + (\mathbf{e} - \mathbf{e}')\| \\
&\geq \lambda_1^2(\Lambda_q(\mathbf{A})) - 2\sigma \\
&> q/4 - q/8 \\
&= 8/q > 2\sigma.
\end{aligned}$$

It means that the ciphertexts of different messages are far enough in Euclidean distance.  $\square$

## 5.5 Security

Now we show the PRIV1-IND security of **Construction 2**.

**Theorem 4.** *Let  $\lambda$  be the security parameter,  $n = n(\lambda) \geq \lambda, l = l(\lambda), m = m(\lambda), q = q(\lambda)$  be integers, and  $\chi$  be an efficiently sampleable distribution such that  $\Pr [|x| \geq \beta] \leq \text{negl}(\lambda)$  and  $\sigma \geq \beta nm$ . Define  $\Psi_\sigma$  as in Lemma 6 and choose  $\mathbf{e} \xleftarrow{\$} \chi$ . If the  $\text{LWE}_{l,m,q,\chi}$  assumption holds, then the above construction is PRIV1-IND secure for all  $k$ -sources where  $k \geq (l + \Omega(\lambda)) \log q$ .*

*Proof.* The parameters are chosen such that the  $\text{LWE}_{n,m,q,\Psi_\sigma}^{\text{WL}(1,k)}$ -assumption holds. Hence for any distributions  $M_\lambda^0, M_\lambda^1$  over  $\{0,1\}^n$  with  $H_\infty(M_\lambda^0) \geq k$  and  $H_\infty(M_\lambda^1) \geq k$ , there is

$$(pk, \text{Enc}(pk, \mathbf{m}_0; \mathbf{e}_0)) \stackrel{s}{\approx} (\mathbf{B}, \mathbf{B}\mathbf{m}_0 + \mathbf{e}_0) \quad (1)$$

$$\stackrel{c}{\approx} (\mathbf{B}, \mathbf{u}) \quad (2)$$

$$\stackrel{c}{\approx} (\mathbf{B}, \mathbf{B}\mathbf{m}_1 + \mathbf{e}_1) \quad (3)$$

$$\stackrel{s}{\approx} (pk, \text{Enc}(pk, \mathbf{m}_1; \mathbf{e}_1)), \quad (4)$$

wherein  $\mathbf{m}_0 \xleftarrow{\$} M_\lambda^0, \mathbf{m}_1 \xleftarrow{\$} M_\lambda^1, (pk, sk) \xleftarrow{\$} \text{Gen}(1^\lambda), \mathbf{e}_0, \mathbf{e}_1 \xleftarrow{\$} \Psi_\sigma, \mathbf{B} \leftarrow \mathbb{Z}_q^{m \times n}$ , and  $\mathbf{u} \xleftarrow{\$} \mathbb{Z}_q^m$ . Step 1 and Step 4 follow with Lemma 5, i.e., the trapdoor generation technique. Step 2 and Step 3 follow with Lemma 6, i.e., the LWE assumption with weak secret.  $\square$

*Remark 1.* Xie et al. proposed a very simple DPKE scheme which is basically the evaluation of inversion of the  $\text{LWR}$  function, by encrypting  $\mathbf{m}$  as  $\lfloor \mathbf{A}\mathbf{m} \rfloor_p$  where  $p \ll q$ , but the security analysis requires the modulus  $q$  to be super-polynomial [37]. Later, Alwen et al. improved the size of the modulus  $q$  to be polynomial, and the size of  $q$  is roughly  $q \geq 2\beta nm^2$ . In our A-DPKE scheme, there is roughly  $q \geq \beta nm^{\frac{3}{2}}$ , i.e., the modulus can be smaller.

*Remark 2.* Bellare et al. proved that with the trapdoor techniques in [28] the LWE function is a lossy trapdoor function for uniform input distributions, but they did not mention whether it is a secure DPKE [10] for high min-entropy message distributions.

*Remark 3.* In fact we can prove the construction is PRIV1-IND secure with respect to hard-to-invert auxiliary input [13], as long as the LWE with weak and leaky secrets assumption holds. We only show the “weak” secret aspect for simplicity.

## 6 Conclusion

In this work we proposed the notion of approximate-deterministic public key encryption by generalizing the original definition of DPKE. A-DPKE maintains the advantages of DPKE in applications such as searchable encryption and data de-duplication, while allows new constructions from quantum-resistant assumptions. We presented two simple constructions of A-DPKE from hard learning problems, e.g., LPN and LWE. The LWE based A-DPKE is as simple as the DPKE scheme from the LWR assumption, with smaller modulus. And we believe that the relaxation from deterministic to approximate-deterministic is meaningful since previously there is no construction of DPKE from LPN.

To make the new concept practical, it is desirable to instantiate A-DPKE with ring-based assumptions, such as ring-LPN [22] and ring-LWE [25]. However, in the current work we have not addressed the problem, and leave it for future work.

**Acknowledgments.** We are grateful to anonymous reviewers for their inspiring comments. Besides, we thank Yuanyuan Gao and Jingnan He for helpful discussions. Yamin Liu is supported by the National Natural Science Foundation of China (No. 61502480). Xianhui Lu is supported the by National Natural Science Foundation of China (No. 61572495, No. 61272534). Bao Li and Fuyang Fang are supported by the National Natural Science Foundation of China (No. 61379137) and the National Basic Research Program of China (973 project) (No. 2013CB338002).

## References

1. Alekhnovich, M.: More on average case vs approximation complexity. In: FOCS 2003, pp. 298–307. IEEE (2003)
2. Akavia, A., Bogdanov, A., Guo, S., Kamath, A., Rosen, A.: Candidate weak pseudorandom functions in  $AC0 \circ MOD2$ . In: ITCS 2014, pp. 251–259. ACM (2014)
3. Alwen, J., Krenn, S., Pietrzak, K., Wichs, D.: Learning with rounding, revisited - new reductions, properties and applications. In: Canetti, R., Garay, J.A. (eds.) CRYPTO 2013. LNCS, vol. 8042, pp. 57–74. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-40041-4\\_4](https://doi.org/10.1007/978-3-642-40041-4_4)
4. Bellare, M., Boldyreva, A., O’Neill, A.: Deterministic and efficiently searchable encryption. In: Menezes, A. (ed.) CRYPTO 2007. LNCS, vol. 4622, pp. 535–552. Springer, Heidelberg (2007). doi:[10.1007/978-3-540-74143-5\\_30](https://doi.org/10.1007/978-3-540-74143-5_30)

5. Bellare, M., Fischlin, M., O'Neill, A., Ristenpart, T.: Deterministic encryption: definitional equivalences and constructions without random oracles. In: Wagner, D. (ed.) CRYPTO 2008. LNCS, vol. 5157, pp. 360–378. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-85174-5\\_20](https://doi.org/10.1007/978-3-540-85174-5_20)
6. Boldyreva, A., Fehr, S., O'Neill, A.: On notions of security for deterministic encryption, and efficient constructions without random oracles. In: Wagner, D. (ed.) CRYPTO 2008. LNCS, vol. 5157, pp. 335–359. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-85174-5\\_19](https://doi.org/10.1007/978-3-540-85174-5_19)
7. Bogdanov, A., Guo, S., Masny, D., Richelson, S., Rosen, A.: On the hardness of learning with rounding over small modulus. ePrint Archive 2015/769 (2015)
8. Bellare, M., Hoang, V.T.: Resisting randomness subversion: fast deterministic and hedged public-key encryption in the standard model. In: Oswald, E., Fischlin, M. (eds.) EUROCRYPT 2015. LNCS, vol. 9057, pp. 627–656. Springer, Heidelberg (2015). doi:[10.1007/978-3-662-46803-6\\_21](https://doi.org/10.1007/978-3-662-46803-6_21)
9. Bellare, M., Hoang, V.T., Keelveedhi, S.: Instantiating random oracles via UCEs. In: Canetti, R., Garay, J.A. (eds.) CRYPTO 2013. LNCS, vol. 8043, pp. 398–415. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-40084-1\\_23](https://doi.org/10.1007/978-3-642-40084-1_23)
10. Bellare, M., Kiltz, E., Peikert, C., Waters, B.: Identity-based (lossy) trapdoor functions and applications. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012. LNCS, vol. 7237, pp. 228–245. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-29011-4\\_15](https://doi.org/10.1007/978-3-642-29011-4_15)
11. Brakerski, Z., Langlois, A., Peikert, C., Regev, O., Stehlé, D.: Classical hardness of learning with errors. In: STOC 2013, pp. 575–584. ACM (2013)
12. Banerjee, A., Peikert, C., Rosen, A.: Pseudorandom functions and lattices. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012. LNCS, vol. 7237, pp. 719–737. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-29011-4\\_42](https://doi.org/10.1007/978-3-642-29011-4_42)
13. Brakerski, Z., Segev, G.: Better security for deterministic public-key encryption: the auxiliary-input setting. In: Rogaway, P. (ed.) CRYPTO 2011. LNCS, vol. 6841, pp. 543–560. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-22792-9\\_31](https://doi.org/10.1007/978-3-642-22792-9_31)
14. Cramer, R., Shoup, V.: Universal hash proofs and a paradigm for adaptive chosen ciphertext secure public-key encryption. In: Knudsen, L.R. (ed.) EUROCRYPT 2002. LNCS, vol. 2332, pp. 45–64. Springer, Heidelberg (2002). doi:[10.1007/3-540-46035-7\\_4](https://doi.org/10.1007/3-540-46035-7_4)
15. Döttling, N., Müller-Quade, J.: Lossy codes and a new variant of the learning-with-errors problem. In: Johansson, T., Nguyen, P.Q. (eds.) EUROCRYPT 2013. LNCS, vol. 7881, pp. 18–34. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-38348-9\\_2](https://doi.org/10.1007/978-3-642-38348-9_2)
16. Döttling, N., Müller-Quade, J., Nascimento, A.C.A.: IND-CCA secure cryptography based on a variant of the LPN problem. In: Wang, X., Sako, K. (eds.) ASIACRYPT 2012. LNCS, vol. 7658, pp. 485–503. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-34961-4\\_30](https://doi.org/10.1007/978-3-642-34961-4_30)
17. Fuller, B., O'Neill, A., Reyzin, L.: A unified approach to deterministic encryption: new constructions and a connection to computational entropy. In: Cramer, R. (ed.) TCC 2012. LNCS, vol. 7194, pp. 582–599. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-28914-9\\_33](https://doi.org/10.1007/978-3-642-28914-9_33)
18. Goldwasser, S., Kalai, Y.T., Peikert, C., Vaikuntanathan, V.: Robustness of the learning with errors assumption. In: ICS 2010, pp. 230–240. Tsinghua University Press (2010)
19. Goldwasser, S., Micali, S.: Probabilistic encryption. J. Comput. Syst. Sci. **28**(2), 270–299 (1984)



20. Gentry, C., Peikert, C., Vaikuntanathan, V.: How to use a short basis: trapdoors for hard lattices and new cryptographic constructions. In: STOC 2008, pp. 197–206. ACM (2008)
21. Damgård, I., Park, S.: How practical is public-key encryption based on LPN? ePrint Archive, 2012/699 (2012)
22. Heyse, S., Kiltz, E., Lyubashevsky, V., Paar, C., Pietrzak, K.: Lapin: an efficient authentication protocol based on Ring-LPN. In: Canteaut, A. (ed.) FSE 2012. LNCS, vol. 7549, pp. 346–365. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-34047-5\\_20](https://doi.org/10.1007/978-3-642-34047-5_20)
23. Kiltz, E., Masny, D., Pietrzak, K.: Simple chosen-ciphertext security from low-noise LPN. In: Krawczyk, H. (ed.) PKC 2014. LNCS, vol. 8383, pp. 1–18. Springer, Heidelberg (2014). doi:[10.1007/978-3-642-54631-0\\_1](https://doi.org/10.1007/978-3-642-54631-0_1)
24. Katz, J., Vaikuntanathan, V.: Smooth projective hashing and password-based authenticated key exchange from lattices. In: Matsui, M. (ed.) ASIACRYPT 2009. LNCS, vol. 5912, pp. 636–652. Springer, Heidelberg (2009). doi:[10.1007/978-3-642-10366-7\\_37](https://doi.org/10.1007/978-3-642-10366-7_37)
25. Lyubashevsky, V., Peikert, C., Regev, O.: On ideal lattices and learning with errors over rings. In: Gilbert, H. (ed.) EUROCRYPT 2010. LNCS, vol. 6110, pp. 1–23. Springer, Heidelberg (2010). doi:[10.1007/978-3-642-13190-5\\_1](https://doi.org/10.1007/978-3-642-13190-5_1)
26. Meurer, A.: A coding-theoretic approach to cryptanalysis. Ph.D. dissertation thesis (2012)
27. Micciancio, D., Mol, P.: Pseudorandom knapsacks and the sample complexity of LWE search-to-decision reductions. In: Rogaway, P. (ed.) CRYPTO 2011. LNCS, vol. 6841, pp. 465–484. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-22792-9\\_26](https://doi.org/10.1007/978-3-642-22792-9_26)
28. Micciancio, D., Peikert, C.: Trapdoors for lattices: simpler, tighter, faster, smaller. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012. LNCS, vol. 7237, pp. 700–718. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-29011-4\\_41](https://doi.org/10.1007/978-3-642-29011-4_41)
29. Peikert, C.: Public-key cryptosystems from the worst-case shortest vector problem: extended abstract. In STOC 2009, pp. 333–342. ACM (2009)
30. Peikert, C., Vaikuntanathan, V., Waters, B.: A framework for efficient and composable oblivious transfer. In: Wagner, D. (ed.) CRYPTO 2008. LNCS, vol. 5157, pp. 554–571. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-85174-5\\_31](https://doi.org/10.1007/978-3-540-85174-5_31)
31. Peikert, C., Waters, B.: Lossy trapdoor functions and their applications. In: STOC 2008, pp. 187–196. ACM (2008)
32. Regev, O.: On lattices, learning with errors, random linear codes, and cryptography. In: STOC 2005, pp. 84–93. ACM (2005)
33. Raghunathan, A., Segev, G., Vadhan, S.: Deterministic public-key encryption for adaptively chosen plaintext distributions. In: Johansson, T., Nguyen, P.Q. (eds.) EUROCRYPT 2013. LNCS, vol. 7881, pp. 93–110. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-38348-9\\_6](https://doi.org/10.1007/978-3-642-38348-9_6)
34. Sun, X., Li, B., Lu, X.: Cramer-shoup like chosen ciphertext security from LPN. In: Lopez, J., Wu, Y. (eds.) ISPEC 2015. LNCS, vol. 9065, pp. 79–95. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-17533-1\\_6](https://doi.org/10.1007/978-3-319-17533-1_6)
35. Wee, H.: Dual projective hashing and its applications — lossy trapdoor functions and more. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012. LNCS, vol. 7237, pp. 246–262. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-29011-4\\_16](https://doi.org/10.1007/978-3-642-29011-4_16)
36. Wichs, D.: Barriers in cryptography with weak, correlated and leaky sources. In: ITCS 2013, pp. 111–126. ACM (2013)

37. Xie, X., Xue, R., Zhang, R.: Deterministic public key encryption and identity-based encryption from lattices in the auxiliary-input setting. In: Visconti, I., Prisco, R. (eds.) SCN 2012. LNCS, vol. 7485, pp. 1–18. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-32928-9\\_1](https://doi.org/10.1007/978-3-642-32928-9_1)
38. Yu, Y., Zhang, J.: Cryptography with auxiliary input and trapdoor from constant-noise LPN. In: Robshaw, M., Katz, J. (eds.) CRYPTO 2016. LNCS, vol. 9814, pp. 214–243. Springer, Heidelberg (2016). doi:[10.1007/978-3-662-53018-4\\_9](https://doi.org/10.1007/978-3-662-53018-4_9). ePrint Archive, 2016/514

Progress in Cryptology – INDOCRYPT 2016  
17th International Conference on Cryptology in India,  
Kolkata, India, December 11–14, 2016, Proceedings  
Dunkelman, O.; Sanadhya, S. (Eds.)  
2016, XVII, 429 p. 66 illus., Softcover  
ISBN: 978-3-319-49889-8