Deterministic Public Key Encryption and Identity-Based Encryption from Lattices in the Auxiliary-Input Setting[⋆]

Xiang Xie[1], Rui Xue[2], Rui Zhang[2]

[1] Institute of Software, Chinese Academy of Sciences
[2] The State Key Laboratory of Information Security
Institute of Information Engineering, Chinese Academy of Sciences
xiexiang@is.iscas.ac.cn, {xuerui,r-zhang}@iie.ac.cn

Abstract. Deterministic public key encryption (D-PKE) provides an alternative to randomized public key encryption in various scenarios (e.g. search on encrypted data) where the latter exhibits inherent drawbacks. In CRYPTO'11, Brakerski and Segev formalized a framework for studying the security of deterministic public key encryption schemes with respect to auxiliary inputs. A trivial requirement is that the plaintext should not be efficiently recoverable from the auxiliary inputs.

In this paper, we present an efficient deterministic public key encryption scheme in the auxiliary-input setting from lattices. The public key size, ciphertext size and ciphertext expansion factor are improved compared with the scheme proposed by Brakerski and Segev. Our scheme is also secure even in the multi-user setting where related messages may be encrypted under multiple public keys. In addition, the security of our scheme is based on the hardness of the learning with errors (LWE) problem which remains hard even for quantum algorithms.

Furthermore, we consider deterministic identity-based public key encryption (D-IBE) in the auxiliary-input setting. The only known D-IBE scheme (without considering auxiliary inputs) in the standard model was proposed by Bellare et al. in EUROCRYPT'12. However, this scheme is only secure in the selective security setting, and Bellare et al. identified it as an open problem to construct adaptively secure D-IBE schemes. The second contribution of this work is to propose a D-IBE scheme from lattices that is adaptively secure.

Keywords: deterministic (identity-based) public key encryption, auxiliary inputs, lattices

[⋆] This is the full version of the paper accepted by SCN 2012.

1 Introduction

The fundamental notion of *semantic security* for public key encryption schemes was introduced by Goldwasser and Micali [16]. While semantic security provides strong privacy guarantees, it inherently requires a *randomized* encryption algorithm. Unfortunately, randomized encryption only allows linear time search [1,10] on outsourced databases, which is prohibitive when the databases are terabytes in size. Further, randomized encryption necessarily expand the length of the plaintext, which may be undesirable in some applications such as legacy code or in-place encryption.

Bellare, Bolyreva, and O'Neill [6] initiated the study of *deterministic* public key encryption schemes that were oriented to search on encrypted data. Clearly, in this setting, no meaningful notion of security can be achieved if the plaintext space is small. Therefore, Bellare et al. [6] required security to hold only when the plaintexts are drawn from a high min-entropic distribution. Very recently, Brakerski and Segev [11] introduced a framework for modeling the security of deterministic encryption schemes with respect to auxiliary inputs. This framework is a generalization of the one formalized by Bellare et al. [6] (and further in [7,9,18]) to the auxiliary-input setting, in which an adversary possibly obtains additional information that is related to encrypted plaintext, and might even fully determine the encrypted plaintext information theoretically. An immediate consequence of having a deterministic encryption algorithm is that no meaningful notion of security can be satisfied if the plaintext can be recovered from the adversary's auxiliary information. Therefore, their framework focuses on the case of *hard-to-invert* auxiliary inputs. Brakerski and Segev [11] proposed two schemes satisfy this notion of security. However, these two schemes have large public key size, ciphertext size and ciphertext expansion factor. One result of this work is to propose a new scheme from lattices with improved public key size, ciphertext size and ciphertext expansion factor.

A deterministic identity-based encryption (D-IBE) scheme is an identity-based encryption [22] scheme with deterministic encryption algorithm. Bellare et al. [8] extended the security definition under high min-entropy into the identity-based setting. D-IBE allows efficiently searchable identity-based encryption of database entries while maintaining the maximal possible privacy, bringing the key-management benefits of the identity-based setting. Bellare et al. proposed a D-IBE scheme by first constructing identity-based lossy trapdoor functions (IB-LTDFs). Due to the inherent limitation of IB-LTDFs, their scheme only achieves selective security, and in fact, it has been identified as an open problem to construct adaptively secure D-IBE schemes [8].

1.1 Our Contributions

In this work, we propose a D-PKE scheme in the auxiliary-input setting from lattices in the standard model. The security of our scheme is based on the hardness of the LWE problem, which is known to be as hard as worst-case lattice problems [21,19]. The public key size, ciphertext size and ciphertext expansion factor are better than the scheme in [11], while the private key size is almost the same. The computations involved in encryption of our scheme are matrix-vector multiplication and followed by a rounding step. Matrix-vector multiplication can be implemented very fast in parallel, and rounding operations can also be computed by small low-depth arithmetic circuits. Therefore, the encryption can be implemented very fast. In addition, our scheme is secure even in the multi-user setting (as in [11]) where related messages may be encrypted under multiple public keys. In this setting we obtain security, with respect to auxiliary inputs, for any polynomial number of messages and users as long as the messages are related by invertible linear transformations.

Furthermore, we extend the security definition in the auxiliary-input setting to D-IBE, and propose a D-IBE scheme in the standard model. The only known (selectively secure) D-IBE scheme (not under the auxiliary-input setting) in the standard model was proposed by Bellare, Kiltz, Peikert and Waters [8], based on IB-LTDFs.

Our D-IBE scheme is the first adaptively secure one in the auxiliary-input setting. In Appendix A, we also give a more efficient selectively secure D-IBE scheme in the auxiliary-input setting whose ciphertext size and ciphertext expansion factor are comparable to our D-PKE scheme. All our schemes are secure with respect to auxiliary inputs that are sub-exponentially hard to invert.

1.2 Overview of Our Approach

A crucial technique hurdle is that the hardness of the LWE problem depends essentially on adding *random, independent* errors to every output of a mod-q "parity" function. Actually, without any error, parity functions are trivially easy to learn. Fortunately, Banerjee, Peikert and Rosen [5] introduced a "derandomized" LWE problem, i.e., generating the errors efficiently and deterministically, while preserving hardness.

The $\mathrm{LWE}_{q,n,m,\alpha}$ assumption says that for any $m = \mathrm{poly}(n)$, modulus q and error rate α: The pairs $(\mathbf{A}, \mathbf{A}^t\mathbf{s}+\mathbf{e})$, for random matrix $\mathbf{A} \leftarrow \mathbb{Z}_q^{n\times m}$, random vector $\mathbf{s} \leftarrow \mathbb{Z}_q^n$, and "small" random error terms $\mathbf{e} \leftarrow \mathbb{Z}^m$ of magnitude $\approx \alpha q$, are indistinguishable from $(\mathbf{A}, \mathbf{u})$, where $\mathbf{u}$ is uniformly random in $\mathbb{Z}_q^m$. The derandomization technique for LWE in [5] is very simple: instead of adding a small random error term to the vector $\mathbf{A}^t\mathbf{s} \in \mathbb{Z}_q^m$. They deterministically *round* it to the nearest element of a sufficiently "coarse" subgroup $\mathbb{Z}_p^m$ where $p \ll q$. In other words, the "error term" comes solely from deterministically rounding $\mathbf{A}^t\mathbf{s}$ to a relatively nearby value. Denoting the rounding operation as $\lfloor \mathbf{A}^t\mathbf{s} \rceil_p \in \mathbb{Z}_p^m$, Banerjee et al. call the problem of distinguishing $(\mathbf{A}, \lfloor \mathbf{A}^t\mathbf{s} \rceil_p)$ from uniform random samples the *learn with rounding* ($\mathrm{LWR}_{q,p,n,m}$) problem. In [5], Banerjee et al. show that the $\mathrm{LWR}_{q,p,n,m}$ is at least as hard as $\mathrm{LWE}_{q,n,m,\alpha}$ for an error rate α proportional to $1/p$, and super-polynomial q ($q \gg p$).

In order to make our D-PKE scheme secure in the auxiliary-input setting, it seems that we need more than the pseudorandomness of $\mathrm{LWR}_{q,p,n,m}$ with uniformly random secret. We hope the $\mathrm{LWR}_{q,p,n,m}$ samples still to be uniformly random even given some auxiliary information of the secret. That is, we want $(\mathbf{A}, \lfloor \mathbf{A}^t\mathbf{s} \rceil_p, f(\mathbf{s})) \approx (\mathbf{A}, \mathbf{u}, f(\mathbf{s}))$ for any hard-to-invert function f. Analogous result of LWE problem was shown in [15], namely $(\mathbf{A}, \mathbf{A}^t\mathbf{s} + \mathbf{e}, f(\mathbf{s})) \approx (\mathbf{A}, \mathbf{u}, f(\mathbf{s}))$ for properly chosen parameters. We briefly explain this statement. LWE assumption implies that $\mathbf{A}^t$ can be substituted by $\mathbf{Z} = \mathbf{B} \cdot \mathbf{C} + \mathbf{E}$, where $\mathbf{B} \leftarrow \mathbb{Z}_q^{m\times d}$, $\mathbf{C} \leftarrow \mathbb{Z}_q^{d\times n}$, and $\mathbf{E} \in \mathbb{Z}^{m\times n}$ is the error matrix (d is determined by the function f). Considering the distribution $(\mathbf{B}, \mathbf{C}, \mathbf{E}, \mathbf{BCs} + \mathbf{Es} + \mathbf{e}, f(\mathbf{s}))$. If $\mathbf{s}$ is sampled from "small" subgroup in $\mathbb{Z}_q^n$ such as $\{0, 1\}^n$, $\mathbf{Es}$ is "small". For sufficiently "large" $\mathbf{e}$, the distribution of $\mathbf{e}$ statistically hides $\mathbf{Es}$. Then we only need to consider the distribution $(\mathbf{B}, \mathbf{C}, \mathbf{E}, \mathbf{BCs} + \mathbf{e}, f(\mathbf{s}))$. According to the generalized Goldreich-Levin theorem of Dodis et al. in [13], we know that the distributions of $(\mathbf{Cs}, f(\mathbf{s}))$ and $(\mathbf{u}, f(\mathbf{s}))$ are statistically close. Applying LWE again, we obtain the above statement.

Randomized IBE schemes from lattices have been proposed in [14,12,2,3,17]. We adopt some of the techniques in [2] to construct our D-IBE. A non-trivial problem is how to use the artificial abort technique. The artificial abort technique in [2] does not work here, because that method only works on polynomial q. But, to guarantee the security, here we need q to be super-polynomial. We solve this problem by extending the technique first appeared in [23].

1.3 Related Work

Deterministic public key encryption for high min-entropic messages was introduced by Bellare, Boldyreva and O'Neill [6] who formalized a definitional framework, which was later refined and extended in [7,9,18]. Bellare et at. [6] presented two constructions in the random oracle model: The first relies on any semantically secure public key encryption scheme; whereas the second relies on the RSA function. Constructions in the standard model were then presented in [7,9], based on trapdoor permutations with (almost) uniformly plaintext space [7], and lossy trapdoor functions [9]. However these constructions fall short in the multi-message setting, where arbitrarily related messages are encrypted under the same public key. O'Neill [18] made a step forwards addressing this problem.

Deterministic public key encryption for auxiliary inputs was proposed by Brakerski and Segev [11]. In the auxiliary-input setting, Brakerski and Segev [11] proposed two constructions in the standard model. The first one is based on d-linear assumptions. This scheme is also secure in the multi-user setting, which solved an open problem in [6]. The second one is based on a rather general class of subgroup indistinguishability assumptions. These two schemes are secure with respect to auxiliary inputs that are sub-exponentially hard to invert.

Deterministic identity-based public key encryption was introduced by Bellare, Kiltz, Peikert and Waters [8]. Bellare et al. aimed to construct identity-based lossy trapdoor functions (IB-LTDFs), which is an extension of lossy trapdoor functions [20]. They built a selectively secure D-IBE as an application of IB-LTDFs. Bellare et al. gave two constructions of IB-LTDFs, while only the one based on Decision Linear Diffie-Hellman assumption can be used to get D-IBE schemes[3]. Since the inherent limitations of IB-LTDFs, it's hard to be directly used to construct adaptively secure D-IBE schemes.

2 Preliminaries

For an integer m, we denote $[m]$ as an integer set $\{1,...,m\}$. We use bold capital letters to denote matrices, and bold lowercase letters to denote vectors. The notation $\mathbf{A}^t$ denotes the transpose of the matrix $\mathbf{A}$. When we say a matrix defined over $\mathbb{Z}_q$ has full rank, we mean that it has full rank modulo q. If $\mathbf{A}_1$ is an $n \times m$ matrix and $\mathbf{A}_2$ is an $n \times m'$ matrix, then $[\mathbf{A}_1|\mathbf{A}_2]$ denotes the $n \times (m+m')$ matrix formed by concatenating $\mathbf{A}_1$ and $\mathbf{A}_2$. If $\mathbf{x}_1$ is a vector of length m and $\mathbf{x}_2$ is of length m', then we let $[\mathbf{x}_1|\mathbf{x}_2]$ denote the length $m+m'$ vector formed by concatenating $\mathbf{x}_1$ and $\mathbf{x}_2$. When doing matrix-vector multiplication, we always view vectors as column vectors.

A function negl(λ) is *negligible*, if it vanishes faster than the inverse of any polynomial in λ. The *statistical distance* between two distributions X, Y over some finite or countable set S is defined as $\Delta(X,Y) = \frac{1}{2}\sum_{s\in S}\big|\Pr[X=s]-\Pr[Y=s]\big|$. X and Y are statistically indistinguishable if $\Delta(X,Y)$ is negligible.

For any integer modulus $q \geq 2$, $\mathbb{Z}_q$ denotes the quotient ring of integer modulo q, and we represent $\mathbb{Z}_q$ by the numbers $\{-\lfloor\frac{q-1}{2}\rfloor, ..., \lceil\frac{q-1}{2}\rceil\}$. We define a "rounding" function $\lfloor\cdot\rceil_p : \mathbb{Z}_q \to \mathbb{Z}_p$, where $q \geq p \geq 2$, as $\lfloor x\rceil_p = \lfloor(p/q)\cdot x\rceil \mod p$. We extend $\lfloor\cdot\rceil_p$ component-wise to vectors and matrices over $\mathbb{Z}_q$.

2.1 Lattices

A full-rank m-dimensional integer lattice $\Lambda \subseteq \mathbb{Z}^m$ is a discrete additive subgroup whose linear span is $\mathbb{R}^m$. Every lattice is generated as the $\mathbb{Z}$-linear combination of some basis of linearly independent vectors $\mathbf{B} = \{\mathbf{b}_1, ..., \mathbf{b}_m\} \subset \mathbb{Z}^m$, i.e.,$\Lambda = \{\sum_{i=1}^m z_i\mathbf{b}_i : z_i \in \mathbb{Z}\}$. In this work we deal exclusively with "q-ary" lattices. For a matrix $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$, define the integer lattice

$$\Lambda^{\perp}(\mathbf{A}) = \{\mathbf{z} \in \mathbb{Z}^m : \mathbf{A}\mathbf{z} = 0 \mod q\}.$$

[3] The other identity-based lossy trapdoor function is based on LWE assumption.

Let $\mathbf{S} = \{\mathbf{s}_1, ..., \mathbf{s}_k\}$ be a set of vectors in $\mathbb{R}^m$. We use $\widetilde{\mathbf{S}} = \{\widetilde{\mathbf{s}_1}, ..., \widetilde{\mathbf{s}_k}\}$ to denote the Gram-Schmidt orthogonalization of the vectors $\mathbf{s}_1, ..., \mathbf{s}_k$. We use $\|\mathbf{S}\|$ to denote the length of the longest vector in $\mathbf{S}$, and $\|\mathbf{S}\|_\infty$ to denote the largest magnitude of the entries in $\mathbf{S}$. For a real-valued matrix $\mathbf{R}$, we let $s_1(\mathbf{R})$ denote the largest singular value of $\mathbf{R}$, i.e. $s_1(\mathbf{R})$=$\max_{\|\mathbf{u}\|=1}\|\mathbf{Ru}\|$.

Let Λ be a discrete subset of $\mathbb{Z}^m$. For any vector $\mathbf{c} \in \mathbb{R}^m$ and any positive parameter $\sigma \in \mathbb{R}_{>0}$, let $\rho_{\sigma,\mathbf{c}}(\mathbf{x}) = \exp(-\pi\|\mathbf{x} - \mathbf{c}\|^2/\sigma^2)$ be the Gaussian function on $\mathbb{R}^m$ with center $\mathbf{c}$ and parameter σ. Let $\rho_{\sigma,\mathbf{c}}(\Lambda) = \sum_{\mathbf{x}\in\Lambda} \rho_{\sigma,\mathbf{c}}(\mathbf{x})$ be the discrete integral of $\rho_{\sigma,\mathbf{c}}$ over Λ, and let $\mathcal{D}_{\Lambda,\sigma,\mathbf{c}}$ be the discrete Gaussian distribution over Λ with center $\mathbf{c}$ and parameter σ. Specifically, for all $\mathbf{y} \in \Lambda$, we have $\mathcal{D}_{\Lambda,\sigma,\mathbf{c}}(\mathbf{y}) = \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{y})}{\rho_{\sigma,\mathbf{c}}(\Lambda)}$. For notional convenience, $\rho_{\sigma,\mathbf{0}}$ and $\mathcal{D}_{\Lambda,\sigma,\mathbf{0}}$ are abbreviated as ρ_σ and $\mathcal{D}_{\Lambda,\sigma}$, respectively.

We recall the learning with errors (LWE) problem, a classic hard problem on lattices defined by Regev [21]. The (decisional) learning with errors problem LWE$_{q,n,m,\alpha}$, in dimension n with error rate $\alpha \in (0, 1)$, stated in matrix form, is: given an input $(\mathbf{A}, \mathbf{b})$, where $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ for any m=poly(n) is uniformly random and $\mathbf{b} \in \mathbb{Z}_q^m$ is either of the form $\mathbf{b} = \mathbf{A}^t\mathbf{s} + \mathbf{e} \mod q$ for uniform $\mathbf{s} \in \mathbb{Z}_q^n$ and $\mathbf{e} \leftarrow \mathcal{D}_{\mathbb{Z}^m,\alpha q}$ or is uniformly random (and independent of $\mathbf{A}$), distinguish which is the case, with non-negligible advantage. It is known that when $\alpha q \geq 2\sqrt{n}$, this decision problem is at least as hard as approximating several problems on n-dimensional lattices in the *worst-case* to within $\widetilde{O}(n/\alpha)$ factors with a quantum computer [21] or on a classical computer for a subset of these problems [19]. In the following, we list some useful facts that make our constructions work.

Lemma 1 ([17] Lemma 2.11). *Let $x \leftarrow \mathcal{D}_{\mathbb{Z},r}$ with $r > 0$, then with overwhelming probability, $|x| \leq r\sqrt{n}$.*

Lemma 2 ([4] Lemma 2.1). *Let q, n, m be positive integers with $q \geq 2$ be prime, and $m \geq n \lg q + \omega(\lg \lambda)$. Let $\mathbf{A} \leftarrow \mathbb{Z}_q^{n\times m}$ and $\mathbf{R} \leftarrow \{-1, 1\}^{m\times m}$. Then $(\mathbf{A}, \mathbf{AR})$ is statistically close to uniform.*

Lemma 3 ([2] Lemma 15). *Let $\mathbf{R}$ be a $k \times m$ matrix chosen at random from $\{-1, 1\}^{k\times m}$. Then with overwhelming probability, $s_1(\mathbf{R}) \leq 12 \cdot \sqrt{k + m}$.*

Lemma 4 ([4] Lemma 3.5). *Let q, n, m be positive integers with $q \geq 2$ and $m \geq 6n \lg q$. There is a probabilistic polynomial-time algorithm* `TrapGen`(q, n, m) *that outputs a pair $(\mathbf{A}, \mathbf{T}) \in \mathbb{Z}_q^{n\times m} \times \mathbb{Z}^{m\times m}$ such that $\mathbf{A}$ is statistically close to uniform in $\mathbb{Z}_q^{n\times m}$ and $\mathbf{T}$ is a basis for $\Lambda^\perp(\mathbf{A})$, satisfying $\|\mathbf{T}\|_\infty \leq O(n \lg q)$ and $\|\widetilde{\mathbf{T}}\| \leq O(\sqrt{n \lg q})$ (Alwen and Peikert assert that the constant hidden in the first $O(\cdot)$ is no more than 20).*

Lemma 5 ([2] Theorem 17). *Let $q > 2, m > n$, $\mathbf{A}, \mathbf{B} \in \mathbb{Z}_q^{n\times m}$, $\mathbf{T_A}$ be a basis of $\Lambda^\perp(\mathbf{A})$, and $\sigma \geq \|\widetilde{\mathbf{T_A}}\| \cdot \omega(\sqrt{\log m})$. There exists an efficient randomized algorithm* `SampleLeft` *that, takes as inputs $\mathbf{A}, \mathbf{B}, \mathbf{T_A}, \sigma$, and outputs a basis $\mathbf{S}$ of $\Lambda^\perp(\mathbf{U})$ for $\mathbf{U} = [\mathbf{A}|\mathbf{B}]$ with $\|\mathbf{S}\| \leq O(\sigma \cdot m)$ whose distribution depends on $\mathbf{U}, \sigma$.*

Lemma 6 ([2] Theorem 18). *Let $q > 2, m > n$, $\mathbf{A}, \mathbf{B} \in \mathbb{Z}_q^{n\times m}$, $\mathbf{B}$ be full rank, $\mathbf{R} \in \{-1, 1\}^{m\times m}$, $\mathbf{T_B}$ be a basis of $\Lambda^\perp(\mathbf{B})$, and $\sigma \geq \|\widetilde{\mathbf{T_B}}\| \cdot s_1(\mathbf{R}) \cdot \omega(\sqrt{\log m})$. There exists an efficient randomized algorithm* `SampleRight` *that, takes as inputs $\mathbf{A}, \mathbf{R}, \mathbf{B}, \mathbf{T_B}, \sigma$, and outputs a basis $\mathbf{S}$ of $\Lambda^\perp(\mathbf{U})$ for $\mathbf{U} = [\mathbf{A}|\mathbf{AR} + \mathbf{B}]$ with $\|\mathbf{S}\| \leq O(\sigma \cdot m)$ whose distribution depends on $\mathbf{U}, \sigma$. Note that this algorithm still works if we replace $\mathbf{B}$ with $k\mathbf{B}$ or $\mathbf{CB}$, where $k \in \mathbb{Z}_q$ is coprime with q and $\mathbf{C} \in \mathbb{Z}_q^{n\times n}$ is full-rank.*

We consider any auxiliary input $f(x)$ from which it is hard to recover the input x. We say that a function f is ϵ-hard-to-invert with respect to a distribution $\mathcal{D}$, if for every efficient algorithm $\mathcal{A}$

it holds that $\Pr[\mathcal{A}(f(x)) = x] \le \epsilon$ over the choice of $x \leftarrow \mathcal{D}$ and the internal coin tosses of $\mathcal{A}$. We describe a useful statement as follows which is crucial to our constructions.

Lemma 7 ([15] Theorem 5). *Let $k \lg t > \lg q + \omega(\lg \lambda)$, $t = poly(\lambda)$. Let $\mathcal{D}$ be any distribution over $\mathbb{Z}_t^n$ and $f : \mathbb{Z}_t^n \to \{0,1\}^*$ be any (possibly randomized) function that is $2^{-k \lg t}$-hard-to-invert with respect to $\mathcal{D}$. For any super-polynomial $q = q(\lambda)$, and any $m = poly(n)$, any $\alpha, \beta \in (0,1)$ such that $\alpha/\beta = negl(\lambda)$.*

$$(\mathbf{A}, \mathbf{A}^t\mathbf{s} + \mathbf{e}, f(\mathbf{s})) \approx (\mathbf{A}, \mathbf{u}, f(\mathbf{s})),$$

where $\mathbf{A} \leftarrow \mathbb{Z}_q^{n \times m}$, $\mathbf{s} \leftarrow \mathcal{D} \subseteq \mathbb{Z}_t^n$ and $\mathbf{u} \leftarrow \mathbb{Z}_q^m$ are uniformly random and $\mathbf{e} \leftarrow \mathcal{D}_{\mathbb{Z},\beta q}^m$. Assuming the $\mathrm{LWE}_{q,d,m,\alpha}$ *assumption, where $d \triangleq \frac{k \lg t - \omega(\lg \lambda)}{\lg q}$.*

For the case of simplicity, we denote the $\mathbf{Adv}_{LWE_{q,n,m,\beta,f}}(\lambda)$ as the advantage of any efficient distinguisher of the above two distribution in Lemma 7. According to Lemma 7, we know that $\mathbf{Adv}_{LWE_{q,n,m,\beta,f}}(\lambda)$ is negligible in λ. Assuming the $\mathrm{LWE}_{q,d,m,\alpha}$ assumption, where $d \triangleq \frac{k \lg t - \omega(\lg \lambda)}{\lg q}$.

2.2 Security Definition

In this section, we describe the security notions introduced in [11]. Brakerski and Segev [11] formalized three security notions with respect to auxiliary inputs, and proved that all these three are equivalent. Brakerski and Segev [11] also showed that for the case of blockwise-hard-to-invert (see [11] for a definition of blockwise-hard-to-invert function) auxiliary inputs, encrypting a single message is equivalent to encrypting multiple messages. For the case of simplicity, in this paper, we only consider the case of a single message. In the single message case, hard-to-invert function and the blockwise-hard-to-invert function are equivalent. Furthermore, we slightly extend the notion in [11]. We require the ciphertext is indistinguishable from uniformly random elements in the ciphertext space. This property implies the strong PRIV1-IND notion defined in [11] and recipient anonymity.

A deterministic public key encryption scheme consists of three algorithms: (`KeyGen`, `Enc`, `Dec`). The probabilistic `KeyGen` algorithm produces a secret key and a corresponding public key. The deterministic `Enc` algorithm uses the public key to map plaintexts into ciphertexts. The deterministic `Dec` algorithm uses the secret key to recover plaintexts from ciphertexts.

Definition 1. *A deterministic public key encryption scheme* D-PKE=(`KeyGen`,`Enc`,`Dec`) *is* PRIV1-INDr-secure *with respect to ϵ-hard-to-invert auxiliary inputs if for any probabilistic polynomial-time algorithm $\mathcal{A}$, for any efficiently sampleable distributions $\mathcal{M}$, and any efficiently computable $\mathcal{F} = \{f\}$ that is ϵ-hard-to-invert with respect to $\mathcal{M}$ such that the advantage of $\mathcal{A}$ in the following game is negligible.*

$$\mathbf{Adv}_{\text{D-PKE},\mathcal{A},\mathcal{F}}^{PRIV1\text{-}INDr}(\lambda) = \Big| \Pr[(pk, sk) \leftarrow \mathtt{KeyGen}(\lambda); b \leftarrow \{0,1\}; m \leftarrow \mathcal{M}; f \leftarrow \mathcal{F};$$
$$c_0^* = \mathtt{Enc}(pk, m); c_1^* \leftarrow \mathcal{C}; b' \leftarrow \mathcal{A}(pk, c_b^*, f(m)) : b = b'] - 1/2 \Big|.$$

Where $\mathcal{C}$ is the ciphertext space. The probability is taken over the choices of $m \leftarrow \mathcal{M}$, $(pk, sk) \leftarrow$ `KeyGen`(λ)*, and over the internal coin tosses of $\mathcal{A}$.*

The multi-user setting of deterministic public key encryption is a straightforward extension of the above definition. Namely, for any efficient adversary $\mathcal{A}$, given polynomial many encryptions of the related messages under multiple public keys and auxiliary information of these message, can not distinguish them from uniformly random elements in the ciphertext space with the same auxiliary information.

A deterministic identity-based public key encryption consists of four algorithms: ($\texttt{IBE.Setup}$, $\texttt{IBE.KGen}$, $\texttt{IBE.Enc}$, $\texttt{IBE.Dec}$). The probabilistic $\texttt{IBE.Setup}$ algorithm generates public parameters, denoted by PP, and a master key MSK. The possibly probabilistic $\texttt{IBE.KGen}$ algorithm uses the master key to extract a private key sk_{id} corresponding to a given identity id. The deterministic $\texttt{IBE.Enc}$ algorithm encrypts messages for a given identity. The deterministic $\texttt{IBE.Dec}$ algorithm decrypts ciphertexts using the private key.

Definition 2. *A deterministic identity-based public key encryption scheme* D-IBE=($\texttt{IBE.Setup}$*,* $\texttt{IBE.KGen}$*,* $\texttt{IBE.Enc}$*,* $\texttt{IBE.Dec}$) *is* PRIV1-ID-INDr-secure *with respect to ϵ-hard-to-invert auxiliary inputs if for any probabilistic polynomial-time algorithm $\mathcal{A}$, for any efficiently sampleable distribution $\mathcal{M}$, and any efficiently computable $\mathcal{F} = \{f\}$ that is ϵ-hard-to-invert with respect to $\mathcal{M}$, such that the advantage of $\mathcal{A}$ in the following game is negligible.*

$$\mathbf{Adv}^{PRIV1\text{-}ID\text{-}INDr}_{\text{D-IBE},\mathcal{A},\mathcal{F}}(\lambda) = \Big| \Pr[(PP, MSK) \leftarrow \texttt{IBE.Setup}(\lambda); id^* \leftarrow \mathcal{A}^{\texttt{IBE.KGen}(\cdot)}(PP);$$
$$b \leftarrow \{0,1\}; m \leftarrow \mathcal{M}; f \leftarrow \mathcal{F}; c_0^* = \texttt{IBE.Enc}(PP, id^*, m); c_1^* \leftarrow \mathcal{C};$$
$$b' \leftarrow \mathcal{A}^{\texttt{IBE.KGen}(\cdot)}(PP, c_b^*, f(m)) : b = b'] - 1/2 \Big|.$$

Where $\mathcal{C}$ is the ciphertext space, and oracle $\texttt{IBE.KGen}(\cdot)$ *on input id generates a private key sk_{id} for the identity id with the restriction that $\mathcal{A}$ is not allowed to query id^*. The probability is taken over the choices of $m \leftarrow \mathcal{M}$, $(PP, MSK) \leftarrow$* $\texttt{IBE.Setup}(\lambda)$*, $sk_{id} \leftarrow$* $\texttt{IBE.KGen}(PP, id, MSK)$*, and over the internal coin tosses of $\mathcal{A}$.*

3 The D-PKE Scheme

In this section, we propose a deterministic public key encryption scheme in the auxiliary-input setting. Before going to the concrete scheme, we first give a useful lemma, i.e. a trapdoor to invert the rounding function.

Lemma 8. *Let p, q, n, m be positive integers with $q \geq p \geq 2$. Let $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ be full-rank, and $\mathbf{T}$ be a basis of $\Lambda^{\perp}(\mathbf{A})$ with $\|\mathbf{T}\|_\infty < p/m$. Given $\mathbf{c} = \lfloor \mathbf{A}^t\mathbf{x} \rceil_p$, where $\mathbf{x} \in \mathbb{Z}_t^n$ with $t \leq q$, there is a polynomial-time algorithm* $\texttt{Invert}(\mathbf{c}, \mathbf{A}, \mathbf{T})$ *that outputs $\mathbf{x}$.*[4]

Proof. Given $\mathbf{c} = \lfloor \mathbf{A}^t\mathbf{x} \rceil_p$, rewrite it into $\mathbf{c} = (p/q)\mathbf{A}^t\mathbf{x} + \mathbf{e} + p\mathbf{v}$, where $\mathbf{e} \in \mathbb{R}^m$ is an "error" vector with $\|\mathbf{e}\|_\infty \leq 1/2$, and $\mathbf{v} \in \mathbb{Z}^m$. Then compute $\mathbf{T}^t\mathbf{c} = (p/q)(\mathbf{AT})^t\mathbf{x} + \mathbf{T}^t\mathbf{e} + p\mathbf{T}^t\mathbf{v}$. Since $\mathbf{T}$ is a basis of $\Lambda^{\perp}(\mathbf{A})$, we have $\mathbf{T}^t\mathbf{c} = p\mathbf{v}' + \mathbf{T}^t\mathbf{e} + p\mathbf{T}^t\mathbf{v} = \mathbf{T}^t\mathbf{e} + p\mathbf{w}$, for some $\mathbf{v}', \mathbf{w} \in \mathbb{Z}^m$. Since $\mathbf{T}^t\mathbf{c}$ and $p\mathbf{w}$ are integer vectors, then $\mathbf{T}^t\mathbf{e}$ is an integer vector as well. Therefore, $\mathbf{T}^t\mathbf{c} = \mathbf{T}^t\mathbf{e} \mod p$. By the hypothesis of $\mathbf{T}$, we know $\|\mathbf{T}^t\mathbf{e}\|_\infty \leq 1/2 \cdot m \cdot \|\mathbf{T}\|_\infty < p/2$. Then we get that $\mathbf{T}^t\mathbf{e} \mod p = \mathbf{T}^t\mathbf{e}$, and obtain $\mathbf{e}$, since $\mathbf{T}$ is invertible in $\mathbb{R}$. We next compute $(q/p)(\mathbf{c} - \mathbf{e}) = \mathbf{A}^t\mathbf{x} + q\mathbf{v}$, then, $(q/p)(\mathbf{c} - \mathbf{e}) \mod q = \mathbf{A}^t\mathbf{x}$. Since $\mathbf{A}$ is full-rank modulo q, $\mathbf{x}$ can be recovered by Gaussian elimination. □

The D-PKE scheme is described as follows. Set the parameters p, q, n, m as specified in Sec. 3.1.

- **Key Generation.** Algorithm $\texttt{KeyGen}(\lambda)$ takes as input a security parameter λ. It uses the algorithm from Lemma 4 to generate a (nearly) uniform matrix and a trapdoor, i.e., $(\mathbf{A}, \mathbf{T}) \leftarrow \texttt{TrapGen}(q, n, m)$. It outputs $pk = \mathbf{A} \in \mathbb{Z}_q^{n\times m}$ and $sk = \mathbf{T} \in \mathbb{Z}^{m\times m}$.

[4] The strong trapdoor presented in [17] can be used here.

- **Encryption.** Algorithm $\mathtt{Enc}(pk, \mathbf{m})$ takes as input a public key $pk = \mathbf{A}$ and a message $\mathbf{m} \in \mathbb{Z}_t^n (\subset \mathbb{Z}_q^n)$. It outputs a ciphertext $\mathbf{c} = \lfloor \mathbf{A}^t \mathbf{m} \rceil_p \in \mathbb{Z}_p^m$.
- **Decryption.** Algorithm $\mathtt{Dec}(sk, \mathbf{c})$ takes as input a secret key $sk = \mathbf{T}$ and a ciphertext $\mathbf{c} \in \mathbb{Z}_p^m$. It first computes $\mathbf{m} \leftarrow \mathtt{Invert}(\mathbf{c}, \mathbf{A}, \mathbf{T})$. Then, if $\mathbf{m} \in \mathbb{Z}_t^n$ it outputs $\mathbf{m}$, and otherwise it outputs $\perp$.

3.1 Correctness and Parameters

For the system to work correctly, we need to ensure that: (1) TrapGen can operate (i.e. $m \geq 6n \lg q$); (2) Lemma 8 holds; (3) Lemma 7 holds. To satisfy these requirements we set the parameters (q, p, m, n) as follows:

$$n = \lambda, \qquad q = \text{the prime nearest to } 2^{n^{\delta}}, \qquad m = \lceil 6n^{1+\delta} \rceil, \quad p = \lceil 120 n^{2+2\delta} \rceil,$$

where δ is constant between 0 and 1. Since $\mathbf{A}$ is uniformly random in $\mathbb{Z}_q^{n \times m}$ and $m \geq 6n^{1+\delta}$, with overwhelming probability this matrix will have rank n. According to the Lemma 7 and the Theorem 1 which we will give a proof in the next subsection. We obtain that the security of this scheme is based on the $\mathrm{LWE}_{q,d,m,\alpha}$, where $d \triangleq \frac{k \lg t - \omega(\lg \lambda)}{\lg q}$, and $1/\alpha = 2^{n^{\delta'}}$ $(0 < \delta' < \delta)$. Given the state of art algorithms, this problem is sub-exponentially hard. Furthermore, we can choose $k \lg t$ to be sub-linear. Therefore, our auxiliary inputs are sub-exponentially hard to invert.

The public key size, private key size, ciphertext size and ciphertext expansion factor in our scheme are $O(n^{2+2\delta})$, $O(n^{3+3\delta})$, $O(n^{1+\delta} \lg n)$, and $O(n^{\delta} \lg n / \lg t)$ respectively. To optimize the ciphertext expansion factor, we can choose $t = n$, which makes the ciphertext expansion factor to be $O(n^{\delta})$. In [11], these values are $n^2|\mathbb{G}|$, n^3, $n|\mathbb{G}|$ and $|\mathbb{G}|$ respectively,[5] where $|\mathbb{G}|$ denotes the length of elements in group $\mathbb{G}$ with order 2^n, It's easy to see that $|\mathbb{G}| \geq n$.

3.2 Security of The D-PKE Scheme

Theorem 1. *For any $k > (\lg q + \omega(\lg \lambda)) / \lg t$, $t = poly(\lambda) \leq q$. The D-PKE scheme is PRIV1-INDr-secure with respect to $2^{-k \lg t}$-hard-to-invert auxiliary inputs. If Lemma 7 holds, where $1/\beta \geq m \cdot p \cdot n^{\omega(1)}$, $q = n^{\omega(1)}$, and $p = poly(\lambda)$.*

Proof. For any distribution $\mathcal{M}$ over $\mathbb{Z}_t^n$, let $\mathcal{F} = \{f\}$ be $2^{-k \lg t}$-hard-to-invert with respect to distribution $\mathcal{M}$. To prove this theorem, we define a series of games, and give a reduction from the Lemma 7 with respect to distribution $\mathcal{M}$.

Game G_0 This game is the original PRIV1-INDr game with adversary $\mathcal{A}$. By $\mathbf{X}_i$, we denote the event $b = b'$ in Game G_i. By definition, $|\Pr[\mathbf{X}_0] - 1/2| = \mathbf{Adv}_{\text{D-PKE},\mathcal{A},\mathcal{F}}^{PRIV1\text{-}INDr}(\lambda)$.

Game G_1 This game is identical to game G_0, except that the challenger choose $\mathbf{A}$ uniformly at random in $\mathbb{Z}_q^{n \times m}$, and uses $\mathbf{A}$ as the public key given to $\mathcal{A}$. According to Lemma 4, it follows that $|\Pr[\mathbf{X}_1] - \Pr[\mathbf{X}_0]| \leq \mathrm{negl}(\lambda)$, for unbounded adversary $\mathcal{A}$.

Game G_2 This game is identical to game G_1, except the way to generate challenge ciphertext. The challenger samples $\mathbf{m} \leftarrow \mathcal{M}$, and samples $\mathbf{e} \leftarrow \mathcal{D}_{\mathbb{Z},\beta q}^m$. Let $\mathbf{b} = \mathbf{A}^t \mathbf{m} + \mathbf{e} \mod q$. The challenger

[5] One can encrypt large messages (other that bits) to reduce the ciphertext expansion factor, but in this case, it needs much more exponent arithmetics to decrypt.

sets $\mathbf{c}_0^* = \lfloor \mathbf{b} \rceil_p$, $\mathbf{c}_1^*$ as in game G_1, i.e. chosen at random in $\mathbb{Z}_p^m$. It outputs $(\mathbf{A}, \mathbf{c}_b^*, f(\mathbf{m}))$ to $\mathcal{A}$, but with one exception: we define a "bad event" $\mathtt{Bad}_2$ to be

$$\mathtt{Bad}_2 \triangleq \lfloor \mathbf{b} + [-B, B]^m \rceil_p \neq \{\lfloor \mathbf{b} \rceil_p\},$$

where $B = \beta q \sqrt{n}$. If $\mathtt{Bad}_2$ occurs on any of $\mathbf{b}$, the challenger immediately abort the game.

If $\mathtt{Bad}_2$ does not occur for the pair $(\mathbf{A}, \mathbf{b})$, then we have $\lfloor \mathbf{b} \rceil_p = \lfloor \mathbf{A}^t \mathbf{m} + \mathbf{e} \rceil_p = \lfloor \mathbf{A}^t \mathbf{m} \rceil_p$ with overwhelming probability over the choice of $\mathbf{e} \leftarrow \mathcal{D}_{\mathbb{Z}, \beta q}^m$, because $\|\mathbf{e}\|_\infty \leq \beta q \sqrt{n}$ with overwhelming probability according to Lemma 1. It follows that for any attacker $\mathcal{A}$,

$$|\Pr[\mathbf{X}_2] - \Pr[\mathbf{X}_1]| \leq \Pr[\mathtt{Bad}_2] + \mathrm{negl}(\lambda).$$

We do not directly bound the probability of $\mathtt{Bad}_2$ occurring in G_2, instead deferring it to the analysis of the next game, where we can show that it is indeed negligible.

Game G_3 In this game, the challenger chooses $\mathbf{b} \in \mathbb{Z}_q^m$ uniformly at random, and samples $\mathbf{m} \leftarrow \mathcal{M}$. It then sets $\mathbf{c}_0^* = \lfloor \mathbf{b} \rceil_p$, and chooses $\mathbf{c}_1^*$ uniformly at random in $\mathbb{Z}_p^m$. The challenger gives $(\mathbf{A}, \mathbf{c}_b^*, f(\mathbf{m}))$ to $\mathcal{A}$, subject to the same "bad event" $\mathtt{Bad}_3$ and abort condition as described in the game G_2 above. Under Lemma 7, and by the fact "bad event" can be tested efficiently given $\mathbf{b}$,[6] a straightforward reduction implies that $|\Pr[\mathbf{X}_3] - \Pr[\mathbf{X}_2]| \leq \mathrm{negl}(\lambda)$ for any efficient attacker $\mathcal{A}$. For the same reason, it also follows that

$$\big| \Pr[\mathtt{Bad}_3] - \Pr[\mathtt{Bad}_2] \big| \leq \mathrm{negl}(\lambda).$$

Now for each uniform $\mathbf{b}$, $\Pr[\mathtt{Bad}_3] \leq m(2B+1)p/q = \mathrm{negl}(\lambda)$, by assumption on q and β. It follows that

$$\Pr[\mathtt{Bad}_2] \leq \mathrm{negl}(\lambda) \quad \Rightarrow \quad |\Pr[\mathbf{X}_2] - \Pr[\mathbf{X}_1]| \leq \mathrm{negl}(\lambda).$$

Game G_4 This game is similar to game G_3, with $\mathbf{b}$ being chosen uniformly at random, $\mathbf{m}$ being sampled from $\mathcal{M}$, and $\mathtt{Bad}_4$ being defined similarly. However, in this game the challenger always returns $(\mathbf{A}, \mathbf{c}_b^*, f(\mathbf{m}))$ to $\mathcal{A}$, even when $\mathtt{Bad}_4$ occurs. By the analysis above, we have that for any adversary $\mathcal{A}$,

$$|\Pr[\mathbf{X}_4] - \Pr[\mathbf{X}_3]| \leq \Pr[\mathtt{Bad}_4] = \Pr[\mathtt{Bad}_3] \leq \mathrm{negl}(\lambda).$$

Since $f(\mathbf{m})$ is independent of $\mathbf{b}$ and the statistical distance between $U(\mathbb{Z}_q^{n\times m}, \mathbb{Z}_p^m)$ and $U(\mathbb{Z}_q^{n\times m}) \times \lfloor U(\mathbb{Z}_q^m) \rceil_p$ is at most $mp/q = \mathrm{negl}(\lambda)$ by assumption on q, so we have $|\Pr[\mathbf{X}_4] - 1/2| = \mathrm{negl}(\lambda)$ for any efficient adversary $\mathcal{A}$.

Finally, by the triangle inequality, we have $|\Pr[\mathbf{X}_0] - 1/2| \leq \mathrm{negl}(\lambda)$ for any efficient adversary $\mathcal{A}$, which completes the proof. □

The Multi-User Setting It's easy to extend the above theorem to multi-user setting where linear related messages $\mathbf{m}_1, ..., \mathbf{m}_k$ are encrypted under any polynomial number of public keys $\mathbf{A}_1, ..., \mathbf{A}_k$. Linear related messages mean that there exist invertible and efficiently computable matrices $\mathbf{V}_2, .., \mathbf{V}_k \subseteq \mathbb{Z}_q^{n\times n}$ and vectors $\mathbf{w}_2, ..., \mathbf{w}_k \in \mathbb{Z}_q^n$, such that $\mathbf{m}_i = \mathbf{V}_i \mathbf{m}_1 + \mathbf{w}_i$ $(2 \leq i \leq k)$. In this case, the joint distribution of ciphertexts is $(\lfloor \mathbf{A}_1^t \mathbf{m}_1 \rceil_p, ..., \lfloor \mathbf{A}_k^t \mathbf{m}_k \rceil_p)$. I.e., $(\lfloor \mathbf{A}_1^t \mathbf{m}_1 \rceil_p, \lfloor \mathbf{A}_2^t \mathbf{V}_2 \mathbf{m}_1 + \mathbf{A}_2^t \mathbf{w}_2 \rceil_p, ..., \lfloor \mathbf{A}_k^t \mathbf{V}_k \mathbf{m}_1 + \mathbf{A}_k^t \mathbf{w}_k \rceil_p)$. Since $\mathbf{V}_i$ is invertible and $\mathbf{A}_i$ is uniformly random for $2 \leq i \leq k$, then $\mathbf{A}_i^t \mathbf{V}_i$ is uniformly random. Because Lemma 7 holds for any $m = \mathit{poly}(n)$, $\mathbf{V}_i, \mathbf{w}_i$ are efficient computable, using the technique in the above proof, we can obtain that our D-PKE scheme is secure in the multi-user for linear related messages. We omit the proof here.

[6] Given $\mathbf{b} = (b_1, ..., b_m)$, for each b_i, one can compute $\lfloor b_i - B \rceil_p$ and $\lfloor b_i + B \rceil_p$ and tests these two values equal to $\lfloor b_i \rceil_p$ or not.

4 The D-IBE scheme

In this section, we describe our D-IBE scheme. Set the parameters p, q, n, m, σ as specified in Sec. 4.1. We treat an identity id as a non-zero sequence of ℓ bits, i.e, $id = (b_1, ..., b_\ell) \in \{0,1\}^\ell \backslash \{0^\ell\}$.

- **Setup.** Algorithm $\texttt{IBE.Setup}(\lambda)$ takes as input a security parameter λ. It uses the algorithm from Lemma 4 to generate a pair $(\mathbf{A}_0, \mathbf{T}) \leftarrow \texttt{TrapGen}(q, n, m)$. Select $\ell + 1$ uniformly random matrices $\mathbf{A}_1, ..., \mathbf{A}_\ell, \mathbf{B}$ in $\mathbb{Z}_q^{n \times m}$. It outputs $PP = (\mathbf{A}_0, \mathbf{A}_1, ..., \mathbf{A}_\ell, \mathbf{B})$, $MSK = \mathbf{T}$.

- **Key Generation.** Algorithm $\texttt{IBE.KGen}(PP, MSK, id)$ takes as input public parameters PP, a master secret key MSK, and an identity $id \in \{0,1\}^\ell$. It first computes $\mathbf{F}_{id} = [\mathbf{A}_0 | \sum_{i=1}^{\ell} b_i \mathbf{A}_i + \mathbf{B}]$, then it uses the algorithm in Lemma 5 to generate a basis of $\Lambda^\perp(\mathbf{F}_{id})$: $\mathbf{T}_{\mathbf{F}_{id}} \leftarrow \texttt{SampleLeft}(\mathbf{A}_0, \sum_{i=1}^{\ell} b_i \mathbf{A}_i + \mathbf{B}, \mathbf{T}, \sigma)$. It outputs $sk_{id} = \mathbf{T}_{\mathbf{F}_{id}}$.

- **Encryption.** Algorithm $\texttt{IBE.Enc}(id, \mathbf{m})$ takes as input public parameters PP, an identity $id \in \{0,1\}^\ell$, and a message $\mathbf{m} \leftarrow \mathbb{Z}_t^n$. It first computes $\mathbf{F}_{id} = [\mathbf{A}_0 | \sum_{i=1}^{\ell} b_i \mathbf{A}_i + \mathbf{B}]$, then let $\mathbf{c} = \lfloor \mathbf{F}_{id}^t \mathbf{m} \rceil_p$. It outputs $\mathbf{c}$.

- **Decryption.** Algorithm $\texttt{IBE.Dec}(PP, id, sk_{id}, \mathbf{c})$ takes as input public parameters PP, an identity id, a secret key sk_{id} and a ciphertext $\mathbf{c} \in \mathbb{Z}_p^{2m}$. It first computes $\mathbf{m} \leftarrow \texttt{Invert}(\mathbf{c}, \mathbf{F}_{id}, sk_{id})$. Then, if $\mathbf{m} \in \mathbb{Z}_t^n$ it outputs $\mathbf{m}$, and otherwise it outputs $\perp$.

4.1 Correctness of Parameters

To ensure the correctness condition, we require: (1) $\texttt{TrapGen}$ can operate (i.e. $m \geq 6n \lg q$); (2) Lemma 8 holds; (3) Lemma 7 holds; (4) σ is sufficiently large for $\texttt{SampleLeft}$ and $\texttt{SampleRight}$. To satisfy all these requirements, we set the parameters (q, p, m, n, σ) as follows:

$$n = \lambda, \;\; q = \text{ the prime nearest to } 2^{n^\delta}, \;\; m = \lceil 6n^{1+\delta} \rceil, \;\; \sigma = 6\ell n^{1.5+\delta}, \;\; p = \lceil 3\ell n^{3.5+3\delta} \rceil,$$

where δ is constant between 0 and 1. According to Lemma 7 and Theorem 2 which we will give a proof in the next subsection. We obtain an adaptively secure scheme whose security is based on the $\mathrm{LWE}_{q,d,m,\alpha}$, where $d \triangleq \frac{k \lg t - \omega(\lg \lambda)}{\lg q}$, and $1/\alpha = 2^{n^{\delta'}}$ $(0 < \delta' < \delta)$. Given the state of art algorithms, this problem is sub-exponentially hard. Furthermore, we can choose $k \lg t$ to be sub-linear. Therefore, our auxiliary inputs are sub-exponentially hard to invert.

The public key size, private key size, ciphertext size and ciphertext expansion factor in our scheme are $O(3(\ell+2)n^{2+2\delta})$, $O(n^{3+3\delta})$, $O(2n^{1+\delta} \lg \ell n)$, and $O(n^\delta \lg \ell n / \lg t)$ respectively. To optimize the ciphertext expansion factor, we can choose $t = \ell n$, which makes the ciphertext expansion factor to be $O(n^\delta)$.

Remark. We also give a more efficient selectively secure D-IBE, the security definition and the concrete construction are given in Appendix A.

4.2 Security of D-IBE.

Theorem 2. *For any $k > (\lg q + \omega(\lg \lambda))/\lg t$, $t = poly(\lambda)$, prime integer $q = n^{\omega(1)}$, and $p = poly(\lambda)$. Assume an adversary $\mathcal{A}$ on D-IBE's PRIV1-ID-INDr security with respect to $2^{-k \lg t}$-hard-to-invert auxiliary inputs, makes at most $Q(\lambda)$ secret key queries. Then for every polynomial $S(\lambda)$*

and $1/\beta \geq \ell m^2 \cdot p \cdot n^{\omega(1)}$ *we have*

$$\mathbf{Adv}_{\textit{D-IBE},\mathcal{A},\mathcal{F}}^{\textit{PRIV1-ID-INDr}}(\lambda) \leq \frac{2\mathbf{Adv}_{\textit{LWE}_{q,n,m,\beta,f}}(\lambda)}{\varDelta} + \frac{1}{S(\lambda)} + \textit{negl}(\lambda)$$

where $\varDelta = \frac{1}{8(\ell+1)Q}$*, and* f *is any* $2^{-k\lg t}$*-hard-to-invert function.*

According to Lemma 7 and because S is arbitrary, we obtain:

Corollary 1. *Let* $q = n^{\omega(1)}$ *be a prime integer,* $p = \textit{poly}(\lambda)$*,* $1/\beta \geq \ell m^2 \cdot p \cdot n^{\omega(1)}$*, and* $\alpha/\beta = \textit{negl}(\lambda)$*. Assuming* $\textit{LWE}_{q,d,m,\alpha}$ *assumption with* $d \triangleq \frac{k\lg t - \omega(\lg\lambda)}{\lg q}$*, then for any* $k > (\lg q + \omega(\lg\lambda))/\lg t$*,* $t = \textit{poly}(\lambda)$*, the D-IBE scheme is PRIV1-ID-INDr-secure with respect to* $2^{-k\lg t}$*-hard-to-invert auxiliary inputs.*

Proof. For any distribution $\mathcal{M}$ over $\mathbb{Z}_t^n$, let $\mathcal{F} = \{f\}$ be $2^{-k\lg t}$-hard-to-invert with respect to distribution $\mathcal{M}$. To prove this theorem, we define a series of games, and give a reduction from Lemma 7 with respect to distribution $\mathcal{M}$.

Game G_0 This game is the original PRIV1-ID-INDr game with adversary $\mathcal{A}$. We assume without loss of generality that $\mathcal{A}$ always makes exactly $Q = Q(\lambda)$ secret key queries. We denote these queries by id_j for $1 \leq j \leq Q$, and the challenge identity chosen by $\mathcal{A}$ as id^*. By $\mathbf{X}_i$, we denote the event $b = b'$ in Game G_i. By definition, $|\Pr[\mathbf{X}_0] - 1/2| = \mathbf{Adv}_{\text{D-IBE},\mathcal{A},\mathcal{F}}^{\textit{PRIV1-ID-INDr}}(\lambda)$. In the following, Let $\mathcal{ID}^Q = (id^*, id_1, ..., id_Q)$.

Game G_1 In this game, the challenger slightly changes the way to generate the matrices $\mathbf{A}_i, i \in [\ell]$ and $\mathbf{B}$. At the setup phase, the challenger first sets an integer $M = 4Q$, and chooses an integer k uniformly at random in between 0 and ℓ. It then chooses a random $\ell+1$-length vector, $\mathbf{x} = (x', x_1, ..., x_\ell)$, where x' is chosen uniformly at random in $\{1, ..., M\}$ and x_i for $i \in [\ell]$ are chosen uniformly at random in $\mathbb{Z}_M$. We define $F(id) = (q - kM) + x' + \sum_{i=1}^{\ell} b_i x_i$, note that $-kM + x' \neq 0$. And we define a binary function $K(id)$ as

$$K(id) = \begin{cases} 0 & \text{if } x' + \sum_{i=1}^{\ell} b_i x_i = 0 \mod M \\ 1 & \text{otherwise.} \end{cases}$$

Next it chooses matrices $\mathbf{B}'$ uniformly at random in $\mathbb{Z}_q^{n\times m}$, and chooses $\mathbf{R}_i \leftarrow \{-1,1\}^{m\times m}$ for $i \in [\ell]$. The challenger sets $\mathbf{B} = (q - kM + x')\mathbf{B}' \mod q$, and constructs $\mathbf{A}_i$ for $i \in [\ell]$ as $\mathbf{A}_i = \mathbf{A}_0\mathbf{R}_i + x_i\mathbf{B}'$. Since $\mathbf{B}'$ is uniform, and q is prime, then $\mathbf{B}$ is uniform (since $-kM + x' \mod q \neq 0$ for sufficiently large q). By Lemma 4, $\mathbf{A}_0$ is uniform with overwhelming probability, then according to Lemma 2, $\mathbf{A}_i$ is statistically close to uniform. Therefore, we have $|\Pr[\mathbf{X}_1] - \Pr[\mathbf{X}_0]| \leq \text{negl}(\lambda)$. Note that, in G_1,

$$\mathbf{F}_{id} = [\mathbf{A}_0|\mathbf{A}_0\sum_{i=1}^{\ell} b_i\mathbf{R}_i + (q - kM + x' + \sum_{i=1}^{\ell} b_i x_i)\mathbf{B}'] = [\mathbf{A}_0|\mathbf{A}_0\sum_{i=1}^{\ell} b_i\mathbf{R}_i + F(id)\mathbf{B}'] \mod q.$$

Furthermore, $F(id) = 0 \mod q$ implies $K(id) = 0$, since q is super-polynomial, and ℓ and M are polynomials, we can assume $q \gg \ell M$ for any reasonable values of q, ℓ and M.

Game G_2 In this game, after the adversary has terminated, the challenger throws an event $\texttt{Good}_2$ independently with probability $\varDelta = \frac{1}{8(\ell+1)Q}$. The challenger aborts the experiment (and outputs a uniformly random bit) if $\neg\texttt{Good}_2$ occurs. We get

$$\Pr[\mathbf{X}_2] - 1/2 = \Pr[\texttt{Good}_2](\Pr[\mathbf{X}_1] - 1/2) = \varDelta \cdot (\Pr[\mathbf{X}_1] - 1/2).$$

Game G_3 In this game, the challenger changes the abort policy. We define a function as

$$\tau(\mathbf{x}, \mathcal{ID}^Q) = \begin{cases} 0 & \text{if } (\wedge_{i=1}^{Q} K(id_i) = 1) \wedge x' + \sum_{i=1}^{\ell} b_i^* x_i = kM \\ 1 & \text{otherwise.} \end{cases}$$

Let E denote the event that $\tau(\mathbf{x}, \mathcal{ID}^Q)$ evaluates to 0 for a given choice of $\mathbf{x}$. According to the analysis in [23] (Claim 2), we know that $p_E = \Pr[E] \geq \varDelta = \frac{1}{8(\ell+1)Q}$. Ideally, we would like to replace event $\texttt{Good}_2$ from game G_2 with event E. Unfortunately, E might not be independent of $\mathcal{A}$'s view, so we use artificial abort techniques. That is, given the identities in all $\mathcal{ID}^Q$, we approximate p_E by sufficiently often sampling values of $\mathbf{x}$. Hoeffding's inequality yields that with $\lceil \lambda S/\varDelta \rceil$ samples, we can obtain an approximation $\hat{p_E} \geq \varDelta$ of p_E that satisfies $\Pr[|p_E - \hat{p_E}| \geq \varDelta/S] \leq 1/2^{\lambda}$. Now the challenger finally aborts if E does not occur. But even if E occurs (which might be with probability $p_E \geq \varDelta$), the challenger artificially enforces an abort with probability $1 - \varDelta/\hat{p_E}$. We call $\texttt{Good}_3$ be the event the challenger does not abort. We always have

$$\Pr[\texttt{Good}_3] = 1 - \Big((1 - p_E) + p_E(1 - \varDelta/\hat{p_E})\Big) = \varDelta \cdot p_E/\hat{p_E}.$$

Hence, except with probability $1/2^{\lambda}$,

$$|\Pr[\texttt{Good}_3] - \Pr[\texttt{Good}_2]| = |\varDelta - \varDelta \cdot p_E/\hat{p_E}| = \varDelta \cdot |(p_E - \hat{p_E})/\hat{p_E}| \leq \varDelta \cdot \varDelta/S\hat{p_E} \leq \varDelta/S.$$

Since the above inequality holds for arbitrary $\mathcal{ID}^Q$ except with probability $1/2^{\lambda}$, we obtain that the statistical distance between the output of game G_2 and G_3 is bounded by $\varDelta/S + 2^{-\lambda}$. Hence, $|\Pr[\mathbf{X}_3] - \Pr[\mathbf{X}_2]| \leq \varDelta/S + 2^{-\lambda}$.

Game G_4 In this game, the challenger makes the following conceptual change regarding secret key queries and challenge ciphertext. Namely, upon receiving a secret key query for $id \in \mathcal{ID}^Q \backslash id^*$, the challenger immediately aborts (with uniform output) if $K(id) = 0$. Upon receiving the challenge identity id^*, the challenger immediately aborts (with uniform output) if $x' + \sum_{i=1}^{\ell} b_i^* x_i \neq kM$. This change is purely conceptual: since $K(id) = 0$, for $id \in \mathcal{ID}^Q \backslash id^*$, or $x' + \sum_{i=1}^{\ell} b_i^* x_i \neq kM$, event E cannot occur, so the Game G_4 would eventually abort as well. We get $\Pr[\mathbf{X}_4] = \Pr[\mathbf{X}_3]$.

Game G_5 In this game, the challenger changes the ways to generate $\mathbf{A}_0, \mathbf{B}'$ and to answer secret key queries. By the change from game G_4, we may assume that $K(id) = 1$ for all $id \in \mathcal{ID}^Q \backslash id^*$ and $x' + \sum_{i=1}^{\ell} b_i^* x_i = kM$ for id^*. This implies that $F(id) \neq 0 \mod q$ for all $id \in \mathcal{ID}^Q \backslash id^*$, and $F(id^*) = 0 \mod q$. The challenger chooses $\mathbf{A}_0$ uniformly at random in $\mathbb{Z}_q^{n \times m}$ and use Lemma 4 to generate $\mathbf{B}'$ with a trapdoor $(\mathbf{B}', \mathbf{T}_{\mathbf{B}'}) \leftarrow \texttt{TrapGen}(q, n, m)$. From Lemma 4 we know that the distribution of $\mathbf{A}_0, \mathbf{B}'$ are statistically close. Upon receiving a secret query for id, the challenge use the algorithm $\mathbf{T}_{\mathbf{F}_{id}} \leftarrow \texttt{SampleRight}(\mathbf{A}_0, \sum_{i=1}^{\ell} b_i \mathbf{R}_i, \mathbf{B}', \mathbf{T}_{\mathbf{B}'}, \sigma)$, this could be done, since $F(id) \neq 0 \mod q$. This results in the same distribution of secret keys as in Game G_4 with sufficiently large σ, up to negligible statistical distance. Thus $|\Pr[\mathbf{X}_5] - \Pr[\mathbf{X}_4]| \leq \text{negl}(\lambda)$. Note that, in this case the matrix of the challenge ciphertext is as $\mathbf{F}_{id^*} = [\mathbf{A}_0 | \mathbf{A}_0 \mathbf{R}^*]$, where $\mathbf{R}^* = \sum_{i=1}^{\ell} b_i^* \mathbf{R}_i$.

Game G_6 In this game, the challenger changes the way to generate challenge ciphertext. The challenger samples $\mathbf{m} \leftarrow \mathcal{M}$, and sample error vector $\mathbf{e} \leftarrow \mathcal{D}_{\mathbb{Z}, \beta q}^m$. we denote $\mathbf{b} = \mathbf{A}_0^t \mathbf{m} + \mathbf{e} \mod q$. It sets $\hat{\mathbf{c}} = [\mathbf{b}^t | \mathbf{b}^t \mathbf{R}^*]^t$ and let $\mathbf{c}_0^* = \lfloor \hat{\mathbf{c}} \rceil_p$, $\mathbf{c}_1^*$ be as in the game G_5, i.e. chosen at random in $\mathbb{Z}_p^{2m}$. The challenger returns $(\mathbf{c}_b^*, f(\mathbf{m}))$ to $\mathcal{A}$, but with one exception: we define a "bad event" $\texttt{Bad}_6$ to be

$$\texttt{Bad}_6 \triangleq \lfloor \hat{\mathbf{c}} + [-B, B]^{2m} \rceil_p \neq \{\lfloor \hat{\mathbf{c}} \rceil_p\},$$

where $B = \ell\beta q\sqrt{nm}$. If $\mathtt{Bad}_6$ occurs on any of $\hat{\mathbf{c}}$, the challenger immediately abort the game.

Since $(\mathbf{R}^*)^t\mathbf{b} = (\mathbf{A}_0\mathbf{R}^*)^t\mathbf{m} + (\mathbf{R}^*)^t\mathbf{e}$, and $\mathbf{R}^* = \sum_{i=1}^{\ell} b_i^*\mathbf{R}_i$, where $\mathbf{R}_i \in \{-1,1\}^{m\times m}$, we have $\|(\mathbf{R}^*)^t\mathbf{e}\|_\infty \le \ell\beta q\sqrt{nm}$ with overwhelming probability, since $\mathbf{e} \leftarrow \mathcal{D}^m_{\mathbb{Z},\beta q}$ according to Lemma 1. If $\mathtt{Bad}_6$ does not occur for some $\hat{\mathbf{c}}$, then we have

$$\lfloor\hat{\mathbf{c}}\rceil_p = \begin{bmatrix} \lfloor\mathbf{A}_0^t\mathbf{m} + \mathbf{e}\rceil_p \\ \lfloor(\mathbf{A}_0\mathbf{R}^*)^t\mathbf{m} + (\mathbf{R}^*)^t\mathbf{e}\rceil_p \end{bmatrix} = \begin{bmatrix} \lfloor\mathbf{A}_0^t\mathbf{m}\rceil_p \\ \lfloor(\mathbf{A}_0\mathbf{R}^*)^t\mathbf{m}\rceil_p \end{bmatrix} = \lfloor\mathbf{F}_{id^*}^t\mathbf{m}\rceil_p.$$

It immediately follows that for any adversary $\mathcal{A}$

$$\Pr[\mathbf{X}_6] - \Pr[\mathbf{X}_5] \le \Pr[\mathtt{Bad}_6] + \mathrm{negl}(\lambda).$$

We do not directly bound the probability of $\mathtt{Bad}_6$ occurring in game G_6, instead deferring it to the analysis of the next game, where we can show that it is indeed negligible.

Game G_7 In this game the only difference is that challenger chooses $\mathbf{b} \in \mathbb{Z}_q^m$ uniformly at random, and samples $\mathbf{m} \leftarrow \mathcal{M}$. To generate the challenge ciphertext, it sets $\hat{\mathbf{c}} = [\mathbf{b}^t|\mathbf{b}^t\mathbf{R}^*]^t$, and let $\mathbf{c}_\mathbf{0}^* = \lfloor\hat{\mathbf{c}}\rceil_p$. It returns $(\mathbf{c}_b^*, f(\mathbf{m}))$ to $\mathcal{A}$, subject to the same “bad event” $\mathtt{Bad}_7$ and abort condition as described in the game G_6 above. Under Lemma 7 and by the fact “bad event” can be tested efficiently given $\hat{\mathbf{c}}$, this implies that $|\Pr[\mathbf{X}_7] - \Pr[\mathbf{X}_6]| \le \mathbf{Adv}_{\mathrm{LWE}_{q,n,m,\beta,f}}$ for any efficient attacker $\mathcal{A}$. For the same reason, it also follows that

$$\big|\Pr[\mathtt{Bad}_7] - \Pr[\mathtt{Bad}_6]\big| \le \mathbf{Adv}_{\mathrm{LWE}_{q,n,m,\beta,f}}.$$

Let us consider the pair $(\mathbf{b}^t, \mathbf{b}^t\mathbf{R}^*)$, where $\mathbf{b} \in \mathbb{Z}_q^m$ is uniformly random, $\mathbf{R}^* = \sum_{i=1}^{\ell} b_i^*\mathbf{R}_i$ and $\mathbf{R}_i$'s are pairwise independently chosen from $\{-1,1\}^m$ at random. Since $id^* \ne 0^\ell$, there exists j, such that $b_j^* = 1$. By Lemma 2 (when $n = 1$), we have that $(\mathbf{b}^t, \mathbf{b}^t\mathbf{R}_j)$ is statistically close to $U(\mathbb{Z}_q^{2m})$. Because $\mathbf{R}_i$'s are pairwise independent, we obtain that $(\mathbf{b}^t, \mathbf{b}^t\mathbf{R}^*)$ is statistically close to $U(\mathbb{Z}_q^{2m})$. This means that $\hat{\mathbf{c}}$ is statistically close to $U(\mathbb{Z}_q^{2m})$, therefore for each uniform $\hat{\mathbf{c}}$, $\Pr[\mathtt{Bad}_7] \le 2m(2B+1)p/q = \mathrm{negl}(\lambda)$, by assumption on q and β. It follows that

$$\begin{aligned} \Pr[\mathtt{Bad}_6] &\le \mathbf{Adv}_{\mathrm{LWE}_{q,n,m,\beta,f}} + \mathrm{negl}(\lambda) \\ &\Rightarrow |\Pr[\mathbf{X}_6] - \Pr[\mathbf{X}_5]| \le \mathbf{Adv}_{\mathrm{LWE}_{q,n,m,\beta,f}} + \mathrm{negl}(\lambda). \end{aligned}$$

Game G_8 This game is similar to game G_7, with $\mathbf{b} \in \mathbb{Z}_q^{2m}$ being chosen uniformly at random, $\mathbf{m}$ being sampled from $\mathcal{M}$, and $\mathtt{Bad}_8$ being defined similarly. However, in this game the challenger always returns $(\mathbf{c}_b^*, f(\mathbf{m}))$ to $\mathcal{A}$, even when $\mathtt{Bad}_8$ occurs. By the analysis above, we have that for any adversary $\mathcal{A}$,

$$|\Pr[\mathbf{X}_8] - \Pr[\mathbf{X}_7]| \le \Pr[\mathtt{Bad}_7] = \Pr[\mathtt{Bad}_6] \le \mathrm{negl}(\lambda).$$

According to the analysis in Game$_7$, we know $\hat{\mathbf{c}}$ is uniformly random, up to negligible statistical distance. Since $f(\mathbf{m})$ is independent of $\hat{\mathbf{c}}$ and the statistical distance between $U(\mathbb{Z}_p^{2m})$ and $\lfloor U(\mathbb{Z}_q^{2m})\rceil_p$ is at most $2mp/q = \mathrm{negl}(\lambda)$ by assumption on q, so we have $\Pr[\mathbf{X}_8] - 1/2 \le \mathrm{negl}(\lambda)$ for any efficient adversary $\mathcal{A}$.

Finally, by the triangle inequality, we obtain the result of Theorem 2. □

Acknowledgements Xiang Xie would like to thank Chris Peikert for many helpful discussions. The authors would like to thank the anonymous reviewers for their valuable comments. Xiang Xie and Rui Xue are supported by the Fund of the National Natural Science Foundation of China under Grant No. 61170280, the Strategic Priority Research Program of the Chinese Academy of

Sciences under Grant No. XDA06010701, and IIE's Cryptography Research Project. Rui Zhang is supported by the One Hundred Person Project of the Chinese Academy of Sciences, the Fund of the National Natural Science Foundation of China under Grant No. 61100225, and the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant No. XDA06010701.

References

1. M. Abdalla, M. Bellare, D. Catalano, E. Kiltz, T. Kohno, T. Lange, J. Malone-Lee, G. Neven, P. Paillier, and H. Shi. Searchable encryption revisited: Consistency properties, relation to anonymous ibe, and extensions. In *Advances in Cryptology–CRYPTO 2005*, pages 205–222. Springer, 2005.
2. S. Agrawal, D. Boneh, and X. Boyen. Efficient lattice (h)ibe in the standard model. *Advances in Cryptology–EUROCRYPT 2010*, pages 553–572, 2010.
3. S. Agrawal, D. Boneh, and X. Boyen. Lattice basis delegation in fixed dimension and shorter ciphertext hierarchical ibe. In *Proc. of Crypto'10*, volume 6223 of *LNCS*, pages 98–115, 2010.
4. J. Alwen and C. Peikert. Generating shorter bases for hard random lattices. *Theory of Computing Systems*, pages 1–19, 2011.
5. A. Banerjee, C. Peikert, and A. Rosen. Pseudorandom functions and lattices. *Advances in Cryptology–EUROCRYPT 2012 (to appear)*, 2012.
6. M. Bellare, A. Boldyreva, and A. O'Neill. Deterministic and efficiently searchable encryption. *Advances in Cryptology-CRYPTO 2007*, pages 535–552, 2007.
7. M. Bellare, M. Fischlin, A. O'Neill, and T. Ristenpart. Deterministic encryption: Definitional equivalences and constructions without random oracles. *Advances in Cryptology–CRYPTO 2008*, pages 360–378, 2008.
8. M. Bellare, E. Kiltz, C. Peikert, and B. Waters. Identity-based (lossy) trapdoor functions and applications. *Advances in Cryptology–EUROCRYPT 2012 (to appear)*, 2012.
9. A. Boldyreva, S. Fehr, and A. O'Neill. On notions of security for deterministic encryption, and efficient constructions without random oracles. *Advances in Cryptology–CRYPTO 2008*, pages 335–359, 2008.
10. D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano. Public key encryption with keyword search. In *Advances in Cryptology-EUROCRYPT 2004*, pages 506–522. Springer, 2004.
11. Z. Brakerski and G. Segev. Better security for deterministic public-key encryption: The auxiliary-input setting. *Advances in Cryptology–CRYPTO 2011*, pages 543–560, 2011.
12. D. Cash, D. Hofheinz, E. Kiltz, and C. Peikert. Bonsai trees, or how to delegate a lattice basis. *Advances in Cryptology–EUROCRYPT 2010*, pages 523–552, 2010.
13. Y. Dodis, S. Goldwasser, Y. Tauman Kalai, C. Peikert, and V. Vaikuntanathan. Public-key encryption schemes with auxiliary inputs. *Theory of Cryptography*, pages 361–381, 2010.
14. C. Gentry, C. Peikert, and V. Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 197–206. ACM, 2008.
15. S. Goldwasser, Y. Kalai, C. Peikert, and V. Vaikuntanathan. Robustness of the learning with errors assumption. *Innovations in Computer Science (ICS)*, 2010.
16. S. Goldwasser and S. Micali. Probabilistic encryption. *Journal of computer and system sciences*, 28(2):270–299, 1984.
17. D. Micciancio and C. Peikert. Trapdoors for lattices: Simpler, tighter, faster, smaller. *Advances in Cryptology–EUROCRYPT 2012 (to appear)*, 2012.
18. A. O'Neill. Deterministic public-key encryption revisited. Technical report, Cryptology ePrint Archive, Report 2010/533, 2010.
19. C. Peikert. Public-key cryptosystems from the worst-case shortest vector problem: extended abstract. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 333–342. ACM, 2009.
20. C. Peikert and B. Waters. Lossy trapdoor functions and their applications. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 187–196. ACM, 2008.
21. O. Regev. On lattices, learning with errors, random linear codes, and cryptography. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 84–93. ACM, 2005.
22. A. Shamir. Identity-based cryptosystems and signature schemes. In *Advances in cryptology*, pages 47–53. Springer, 1985.
23. B. Waters. Efficient identity-based encryption without random oracles. *Advances in Cryptology–EUROCRYPT 2005*, pages 114–127, 2005.

A The D-sIBE Scheme

Definition 3. *A deterministic identity-based public key encryption scheme* D-IBE=(`IBE.Setup`, `IBE.KGen`, `IBE.Enc`, `IBE.Dec`) *is* PRIV1-sID-INDr-secure *with respect to ϵ-hard-to-invert auxiliary inputs if for any probabilistic polynomial-time algorithm $\mathcal{A}$, for any efficiently sampleable distribution $\mathcal{M}$, and any efficiently computable $\mathcal{F} = \{f\}$ that is ϵ-hard-to-invert with respect to $\mathcal{M}$, such that the advantage of $\mathcal{A}$ in the following game is negligible.*

$$\mathbf{Adv}^{PRIV1\text{-}sID\text{-}INDr}_{\text{D-IBE},\mathcal{A},\mathcal{F}}(\lambda) = \Big| \Pr[id^* \leftarrow \mathcal{A}(\lambda); (PP, MSK) \leftarrow \texttt{IBE.Setup}(\lambda);$$
$$state \leftarrow \mathcal{A}^{\texttt{IBE.KGen}(\cdot)}(PP); b \leftarrow \{0,1\}; m \leftarrow \mathcal{M}; f \leftarrow \mathcal{F}; c_0^* = \texttt{IBE.Enc}(PP, id^*, m); c_1^* \leftarrow \mathcal{C};$$
$$b' \leftarrow \mathcal{A}^{\texttt{IBE.KGen}(\cdot)}(PP, c_b^*, f(m), state) : b = b'] - 1/2 \Big|.$$

Where $\mathcal{C}$ is the ciphertext space, and oracle `IBE.KGen`$(\cdot)$ *on input id generates a private key sk_{id} for the identity id with the restriction that $\mathcal{A}$ is not allowed to query id^*. The probability is taken over the choices of $m \leftarrow \mathcal{M}$, $(PP, MSK) \leftarrow$* `IBE.Setup`(λ)*, $sk_{id} \leftarrow$* `IBE.KGen`(PP, id, MSK)*, and over the internal coin tosses of $\mathcal{A}$.*

In this section, we propose a selectively secure deterministic IBE scheme in the auxiliary setting. Set the parameters p, q, n, m, σ as specified in Sec. A.1. We treat an identity id as an element in $\mathbb{Z}_q^n$.

- **Setup.** Algorithm `sIBE.Setup`(λ) takes as input a security parameter λ. It uses the algorithm from Lemma 4 to generate a pair $(\mathbf{A}_0, \mathbf{T}) \leftarrow$ `TrapGen`(q, n, m). Select two uniformly random matrices $\mathbf{A}_1, \mathbf{B}$ in $\mathbb{Z}_q^{n\times m}$, and a full-rank differences (FRD)[7] function H. It outputs $PP = (\mathbf{A}_0, \mathbf{A}_1, \mathbf{B}, H)$, $MSK = \mathbf{T}$.

- **Key Generation.** Algorithm `sIBE.KGen`(PP, MSK, id) takes as input public parameters PP, a master secret key MSK, and an identity $id \in \mathbb{Z}_q^n$. It first computes $\mathbf{F}_{id} = [\mathbf{A}_0|\mathbf{A}_1 + H(id)\mathbf{B}]$, then it uses the algorithm in Lemma 5 to generate a basis of $\Lambda^{\perp}(\mathbf{F}_{id})$: $\mathbf{T}_{\mathbf{F}_{id}} \leftarrow$ `SampleLeft`$(\mathbf{A}_0, \mathbf{A}_1 + H(id)\mathbf{B}, \mathbf{T}, \sigma)$. It outputs $sk_{id} = \mathbf{T}_{\mathbf{F}_{id}}$ as a secret key for id.

- **Encryption.** Algorithm `sIBE.Enc`$(PP, id, \mathbf{m})$ takes as input public parameters PP, an identity $id \in \mathbb{Z}_q^n$, and a message $\mathbf{m} \leftarrow \mathbb{Z}_t^n$. It first computes $\mathbf{F}_{id} = [\mathbf{A}_0|\mathbf{A}_1 + H(id)\mathbf{B}]$, then it computes $\mathbf{c} = \lfloor \mathbf{F}_{id}^t \mathbf{m} \rceil_p$. Finally it outputs $\mathbf{c}$.

- **Decryption.** Algorithm `sIBE.Dec`$(PP, id, sk_{id}, \mathbf{c})$ takes as input public parameter PP, an identity id, a secret key sk_{id} and a ciphertext $\mathbf{c} \in \mathbb{Z}_p^{2m}$. It first computes $\mathbf{m} \leftarrow$ `Invert`$(\mathbf{c}, \mathbf{F}_{id}, sk_{id})$. Then, if $\mathbf{m} \in \mathbb{Z}_t^n$ it outputs $\mathbf{m}$, and otherwise it outputs $\perp$.

A.1 Correctness and Parameters

For the system to work correctly, we need to ensure that: (1) `TrapGen` can operate (i.e. $m \geq 6n \lg q$); (2) Lemma 8 holds; (3) Lemma 7 holds; (4) σ is sufficiently large for `SampleLeft` and `SampleRight`. To satisfy these requirements we set the parameters (q, p, m, n, σ) as follows:

$$n = \lambda, \;\; q = \text{ the prime nearest to } 2^{n^\delta}, \;\; m = \lceil 6n^{1+\delta} \rceil, \;\; \sigma = 6n^{1.5+\delta}, \;\; p = \lceil 3n^{3.5+3\delta} \rceil,$$

[7] I.e., H maps $\mathbb{Z}_q^n$ to $\mathbb{Z}_q^{n\times n}$, and for $id \neq id'$, $H(id) - H(id')$ is full-rank.

where δ is constant between 0 and 1. According to Lemma 7 and Theorem 3 which we will give a proof in Appendix B. We obtain a selectively secure deterministic identity-based encryption scheme whose security is based on the $\mathrm{LWE}_{q,d,m,\alpha}$, where $d \triangleq \frac{k \lg t - \omega(\lg \lambda)}{\lg q}$, and $1/\alpha = 2^{n^{\delta'}}$ $(0 < \delta' < \delta)$. Given the state of art algorithms, this problem is sub-exponentially hard. Furthermore, we can choose $k \lg t$ to be sub-linear. Therefore, our auxiliary inputs are sub-exponentially hard to invert.

The public key size, private key size, ciphertext size and ciphertext expansion factor in our scheme are $O(3n^{2+2\delta})$, $O(n^{3+3\delta})$, $O(2n^{1+\delta} \lg n)$, and $O(n^{\delta} \lg n / \lg t)$ respectively. To optimize the ciphertext expansion factor, we can choose $t = n$, which makes the ciphertext expansion factor to be $O(n^{\delta})$.

B Security of D-sIBE

Theorem 3. *For any $k > (\lg q + \omega(\lg \lambda))/\lg t$, $t = ploy(\lambda) \leq q$. The D-sIBE scheme is PRIV1-sID-INDr-secure with respect to $2^{-k \lg t}$-hard-to-invert auxiliary inputs. If Lemma 7 holds, where $1/\beta \geq m^2 \cdot p \cdot n^{\omega(1)}$, $q = n^{\omega(1)}$, and $p = poly(\lambda)$.*

Proof. For any distribution $\mathcal{M}$ over $\mathbb{Z}_t^n$, let $\mathcal{F} = \{f\}$ be $2^{-k \lg t}$-hard-to-invert with respect to distribution $\mathcal{M}$. To prove this theorem, we define a series of games, and give a reduction from Lemma 7 with respect to distribution $\mathcal{M}$.

Game G_0 This game is the original PRIV1-sID-INDr game with adversary $\mathcal{A}$. By $\mathbf{X}_i$, we denote the event $b = b'$ in Game G_i. By definition, $|\Pr[\mathbf{X}_0] - 1/2| = \mathbf{Adv}^{PRIV1\text{-}sID\text{-}INDr}_{\text{D-sIBE},\mathcal{A},\mathcal{F}}(\lambda)$.

Game G_1 In this game, the challenger changes the ways to generate $\mathbf{A}_0, \mathbf{A}_1, \mathbf{B}$ and answer queries. The challenger chooses $\mathbf{A}_0$ uniformly from $\mathbb{Z}_q^{n \times m}$, and generate $(\mathbf{B}, \mathbf{T_B}) \leftarrow \mathtt{TrapGen}(q, n, m)$. Let $\mathbf{R} \leftarrow \{-1, 1\}^{m \times m}$, and let $\mathbf{A}_1 = \mathbf{A}_0\mathbf{R} - H(id^*)\mathbf{B}$. Since $\mathbf{A}_0$ is uniformly random, then by Lemma 2, $(\mathbf{A}_0, \mathbf{A}_0\mathbf{R})$ is statistically close to unform in $\mathbb{Z}_q^{n \times m} \times \mathbb{Z}_q^m$, and so is $(\mathbf{A}_0, \mathbf{A}_0\mathbf{R} - H(id^*)\mathbf{B})$. Because $\mathbf{B}$ is statistically close to uniform according to Lemma 4, the public parameters in game G_0 and G_1 are statistically close. When $\mathcal{A}$ queries on $id \neq id^*$, the challenger computes $\mathbf{F}_{id} = [\mathbf{A}_0 | \mathbf{A}_0\mathbf{R} + (H(id) - H(id^*))\mathbf{B}]$, where $H(id) - H(id^*)$ is full rank for $id \neq id^*$. The challenger uses `SampleRight` to generate $\mathbf{T}_{\mathbf{F}_{id}} \leftarrow \mathtt{SampleRight}(\mathbf{A}_0, \mathbf{R}, \mathbf{B}, \mathbf{T_B}, \sigma)$. For sufficiently large σ the distribution of $\mathbf{T}_{\mathbf{F}_{id}}$ is statistically close in game G_0 and G_1. It follows that $|\Pr[\mathbf{X}_1] - \Pr[\mathbf{X}_0]| \leq \text{negl}(\lambda)$ for any unbounded adversary $\mathcal{A}$.

Game G_2 This game is identical to game G_1, except the way to generate challenge ciphertext. The challenger samples $\mathbf{m} \leftarrow \mathcal{M}$, and samples $\mathbf{e} \leftarrow \mathcal{D}^m_{\mathbb{Z},\beta q}$. Let $\mathbf{b} = \mathbf{A}_0^t\mathbf{m} + \mathbf{e} \mod q$. It sets $\hat{\mathbf{c}} = [\mathbf{b}^t | \mathbf{b}^t\mathbf{R}]^t$. The challenger sets $\mathbf{c}_0^* = \lfloor \hat{\mathbf{c}} \rceil_p$, $\mathbf{c}_1^*$ as in the game G_1, i.e. chosen at random in $\mathbb{Z}_p^{2m}$. It outputs $(\mathbf{c}_b^*, f(\mathbf{m}))$ to $\mathcal{A}$, but with one exception: we define a "bad event" $\mathtt{Bad}_2$ to be

$$\mathtt{Bad}_2 \triangleq \lfloor \hat{\mathbf{c}} + [-B, B]^{2m} \rceil_p \neq \{\lfloor \hat{\mathbf{c}} \rceil_p\},$$

where $B = \beta q \sqrt{n} m$. If $\mathtt{Bad}_2$ occurs on any of $\hat{\mathbf{c}}$, the challenger immediately abort the game.

Since $\mathbf{R}^t\mathbf{b} = (\mathbf{A}_0\mathbf{R})^t\mathbf{m} + \mathbf{R}^t\mathbf{e}$, and $\mathbf{R} \in \{-1, 1\}^{m \times m}$, we have $\|\mathbf{R}^t\mathbf{e}\|_\infty \leq \beta q \sqrt{n} m$ with overwhelming probability, since $\mathbf{e} \leftarrow \mathcal{D}^m_{\mathbb{Z},\beta q}$ and according to Lemma 1. If $\mathtt{Bad}_2$ does not occur for some $\hat{\mathbf{c}}$, then we have

$$\lfloor \hat{\mathbf{c}} \rceil_p = \begin{bmatrix} \lfloor \mathbf{A}_0^t\mathbf{m} + \mathbf{e} \rceil_p \\ \lfloor (\mathbf{A}_0\mathbf{R})^t\mathbf{m} + \mathbf{R}^t\mathbf{e} \rceil_p \end{bmatrix} = \begin{bmatrix} \lfloor \mathbf{A}_0^t\mathbf{m} \rceil_p \\ \lfloor (\mathbf{A}_0\mathbf{R})^t\mathbf{m} \rceil_p \end{bmatrix} = \lfloor \mathbf{F}_{id^*}^t\mathbf{m} \rceil_p.$$

It immediately follows that for any adversary $\mathcal{A}$

$$|\Pr[\mathbf{X}_2] - \Pr[\mathbf{X}_1]| \leq \Pr[\mathtt{Bad}_2] + \mathrm{negl}(\lambda).$$

We do not directly bound the probability of $\mathtt{Bad}_2$ occurring in game G_2, instead deferring it to the analysis of the next game, where we can show that it is indeed negligible.

Game G_3 In this game the challenger chooses $\mathbf{b} \in \mathbb{Z}_q^m$ uniformly at random, and samples $\mathbf{m} \leftarrow \mathcal{M}$. To generate the challenge ciphertext, it sets $\hat{\mathbf{c}} = [\mathbf{b}^t | \mathbf{b}^t\mathbf{R}]^t$. The challenger sets $\mathbf{c}_0^* = \lfloor \hat{\mathbf{c}} \rceil_p$ and outputs $(\mathbf{c}_b^*, f(\mathbf{m}))$ to $\mathcal{A}$, subject to the same "bad event" $\mathtt{Bad}_3$ and abort condition as described in the game G_2 above. Under Lemma 7 and by the fact "bad event" can be tested efficiently given $\mathbf{b}$, this implies that $|\Pr[\mathbf{X}_3] - \Pr[\mathbf{X}_2]| \leq \mathrm{negl}(\lambda)$ for any efficient attacker $\mathcal{A}$. For the same reason, it also follows that $\left|\Pr[\mathtt{Bad}_2] - \Pr[\mathtt{Bad}_3]\right| \leq \mathrm{negl}(\lambda)$. Let us consider the pair $(\mathbf{b}^t, \mathbf{b}^t\mathbf{R})$, where $\mathbf{b} \in \mathbb{Z}_q^m$ is uniformly random, $\mathbf{R} \leftarrow \{-1, 1\}^{m \times m}$. From Lemma 2 (when $n = 1$), we have that $(\mathbf{b}^t, \mathbf{b}^t\mathbf{R})$ is statistically close to $U(\mathbb{Z}_q^{2m})$. This means that $\hat{\mathbf{c}}$ is statistically close to $U(\mathbb{Z}_q^{2m})$, therefore for each uniform $\hat{\mathbf{c}}$, $\Pr[\mathtt{Bad}_3] \leq 2m(2B+1)p/q = \mathrm{negl}(\lambda)$, by assumption on q and β. It follows that

$$\Pr[\mathtt{Bad}_2] \leq \mathrm{negl}(\lambda) \quad \Rightarrow \quad |\Pr[\mathbf{X}_2] - \Pr[\mathbf{X}_1]| \leq \mathrm{negl}(\lambda).$$

Game G_4 This game is similar to game G_3, with $\mathbf{b} \in \mathbb{Z}_q^m$ being chosen uniformly at random, $\mathbf{m}$ being sampled from $\mathcal{M}$, and $\mathtt{Bad}_4$ being defined similarly. However, in this game the challenger always returns $(\mathbf{c}_b^*, f(\mathbf{m}))$ to $\mathcal{A}$, even when $\mathtt{Bad}_4$ occurs. By the analysis above, we have that for any adversary $\mathcal{A}$,

$$|\Pr[\mathbf{X}_4] - \Pr[\mathbf{X}_3]| \leq \Pr[\mathtt{Bad}_4] = \Pr[\mathtt{Bad}_3] \leq \mathrm{negl}(\lambda).$$

According to the analysis in Game$_3$, we know $\hat{\mathbf{c}}$ is uniformly random, up to negligible statistical distance. Since $f(\mathbf{m})$ is independent of $\hat{\mathbf{c}}$ and the statistical distance between $U(\mathbb{Z}_p^{2m})$ and $\lfloor U(\mathbb{Z}_q^{2m}) \rceil_p$ is at most $2mp/q = \mathrm{negl}(\lambda)$ by assumption on q, so we have $|\Pr[\mathbf{X}_4] - 1/2| \leq \mathrm{negl}(\lambda)$ for any efficient adversary $\mathcal{A}$.

Finally, by the triangle inequality, we have $\mathbf{Adv}_{\text{D-sIBE},\mathcal{A},\mathcal{F}}^{PRIV1\text{-}sID\text{-}INDr}(\lambda) \leq \mathrm{negl}(\lambda)$ for any efficient adversary $\mathcal{A}$, which completes the proof. □