Direct CCA-Secure KEM and Deterministic PKE from Plain LWE

Direct CCA-Secure KEM and Deterministic PKE from Plain LWE

Xavier Boyen[1] *, Qinyi Li[2] **

[1] QUT, Brisbane, Australia
[2] Griffith University, Brisbane, Australia
qinyi.li@griffith.edu.au

Abstract. We present a particularly simple and efficient CCA-secure public-key encapsulation scheme without random oracles or costly sampling. The construction is direct in the sense that it eschews generic transformations via one-time signatures or MACs typically found in standard-model constructions. This gives us a compact, conceptually simpler, and computationally efficient operation, that in particular does not require any Gaussian sampling. Nevertheless, security is based on the hardness of the plain learning-with-errors (LWE) problem with polynomial modulus-to-noise ratio.
Of further interest, we also show how to obtain CCA-secure *deterministic* public-key encryption (for high-entropy messages), that is more compact and efficient than existing constructions.

1 Introduction

Public-key encryption (PKE) is a central cryptographic primitive to provide secure communication over insecure networks without prior secret-key agreement. In practice, due to its relative inefficiency, it is almost always used in conjunction with a secret-key cipher, where the former encrypts a *random* session key for the latter, which then encrypts the actual data. This flow is the motivation for "hybrid encryption" [8], which consists of a (public-)key encapsulation mechanism (KEM) and a data encapsulation mechanism (DEM). In terms of security, it is well known [8] that if both KEM and DEM are CCA-secure, then the hybrid encryption scheme is CCA-secure, which is the standard notion for security of PKE against active attacks. While DEMs are readily obtained from suitable symmetric-key modes of operation, in the case of KEMs substantial optimisations are to be gained by specialising them to work with random plaintexts only.

Constructing CCA-secure KEMs is easy in principle. Applying the Fujisaki-Okamoto transformations [10] to PKE/KEM schemes with weaker security guarantees (e.g., chosen-plaintext security) results in CCA-secure PKE/KEM schemes

* Research supported in part by ARC Discovery Project grant number DP140103885 and ARC Future Fellowship FT140101145 from the Australian Research Council.
** Corresponding author

in the random oracle model. While this approach often leads to practical constructions, one can only make heuristic security arguments for them. Moreover, when it comes to post-quantum security, these heuristic security arguments need to be made in quantum random-oracle models [6] which are not very well understood. For these reasons, designing an efficient and practical post-quantum KEM in the standard model (without random oracles) is already desirable and well motivated.

There are two somewhat generic ways to construct CCA-secure PKE/KEM from lattices in the standard model. The first one is via lossy trapdoor functions [19] (e.g., the constructions from [19,21,16]) and the second one is via the BCHK trasformation [5] from tag-based or identity-based encryption (IBE) (e.g. the constructions from [15]). Both of them require strongly unforgeable one-time signatures or message authentication codes (MACs) as building blocks. This introduces noticeable extra overheads, making the schemes less efficient and less compact.

In this paper, we primarily focus on constructing a KEM that is both conceptually very simple and computationally efficient, but without compromising its provable security. Specifically, we rely on a standard lattice problem (plain learning with errors, a.k.a. LWE [20,17]) in the standard model. [3]

1.1 Our Contributions

Our main contribution is a simple, compact, computationally efficient KEM scheme without random oracles. The construction makes use of identity-based/tag-based lattice trapdoor techniques [1,15]. The public key of our scheme includes two matrices $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ and $\mathbf{A}_1 \in \mathbb{Z}_q^{n \times w}$, where $w = n\lceil \log q \rceil$, and a target-collision-resistant compression or hash function $f : \mathbb{Z}_q^n \to \{0,1\}^\lambda$, where λ is the security parameter. The private key is a low norm matrix $\mathbf{R} \in \mathbb{Z}^{m \times w}$ such that $\mathbf{A}_1 = \mathbf{AR} \pmod q$. The ciphertext of our scheme contains two parts. The first part is $\mathbf{t} = f(\mathbf{s})$ where $\mathbf{s} \in \mathbb{Z}_q^n$ is the randomness of the encapsulation algorithm. The second part is a vector $\mathbf{c}^\top = \lfloor (p/q) \cdot \mathbf{s}^\top \cdot [\mathbf{A} | \mathbf{A}_1 + \varphi(\mathbf{t})\mathbf{G}] \rfloor$ where $\varphi : \mathbb{Z}_q^n \to \mathbb{Z}_q^{n \times n}$ is a full-rank difference encoding [1] (here $\mathbf{t}$ is encoded as a vector in $\mathbb{Z}_q^n$) and $\mathbf{G} \in \mathbb{Z}_q^{n \times w}$ is the gadget matrix [15]. The session key is obtained by applying a randomness extractor to $\mathbf{s}$. When $\mathbf{t}$ is non-zero (which happens with overwhelming probability), the lattice trapdoor $(\mathbf{R}, \mathbf{G})$ allows recovering $\mathbf{s}$ and, thus, reproducing the session key. The key idea of our construction is to make the identity/tag the hash value of the secret random vector $\mathbf{s}$ rather than a verification key or a commitment in the BCHK transformation. In terms of security, by using the LWE problem to (computationally) switch the rounding function $\lfloor (p/q) \cdot \mathbf{s}^\top \cdot [\mathbf{A} | \mathbf{A}_1 + \varphi(\mathbf{t})\mathbf{G}] \rfloor$ to the so-called "lossy mode" [2], the ran-

[3] We note that our approach here departs significantly from the recent NIST Post-Quantum KEM competition, wherein most submitters chose to embrace random oracles and stronger hardness assumptions (e.g., many variants of ring-LWE), to address its rather idiosyncratic rules and success criteria.

dom vector $\mathbf{s}$ retains sufficient min-entropy (even conditioned on $\mathbf{c}$ and $\mathbf{t}$) that the session key would be random.

Our construction can be seen as a "direct" CCA-secure PKE/KEM construction from identity-based/tag-based encryption in the sense that it does not employ generic transformations. Such kind of direct constructions from pairing-based IBE are known, e.g., [7,13,14]. From a high level idea, our KEM construction also has similarities to the CCA-secure PKE scheme from a lossy trapdoor function (LTF) and all-but-one lossy trapdoor function (ABO-LTF) from [19]. In [19], the encryption is roughly done by evaluating a LTF and an ABO-LTF (both are invertible) on the randomness. The well-formness of the ciphertext is guaranteed by signing theses two evaluations with a one-time signature scheme (the verification key also serves as the tag for the ABO-LTF). Our construction "shrinks" this further by using only one (ABO) LTF plus a compression hash function. For our KEM construction, the hash function, which is much lighter than an LTF, is already lossy and enough to ensure that the ciphertext is well-formed. One should also note that our construction is for CCA-secure KEM which is a more specialised primitive than CCA-secure PKE studied in certain earlier constructions.

Our KEM construction is of good computational efficiency. First, the encryption process essentially involves a vector-matrix multiplication, a rounding operation and a target-collision-resistant hash function. In particular, discrete Gaussian sampling is avoided. Second, the decryption can be done efficiently in a parallel fashion by using the so-called "gadget" trapdoor inversion first proposed in [15].

In terms of space efficiency, since our KEM scheme is based on a relatively stronger LWE assumption (but still with polynomial modulus-to-noise ratio), compared to the most efficient existing CCA-secure lattice PKE/KEM constructions in the standard model, e.g., [15], our construction would need relatively larger matrix dimensions (to provide sufficient hardness for the LWE problem). However, since our KEM ciphertext only consists of a single vector over a small field and a small hash value (whose bit-size is the security parameter, e.g., 128), and since our KEM private key is a low-norm matrix with very small entries (-1 and 1), the impact of requiring larger dimensions is rather limited.

As a by-product of our KEM scheme and its structure, we also give a CCA-secure deterministic lattice PKE system. Deterministic PKE has useful direct and indirect applications such as efficient searchable encryption and de-duplication of encrypted databases. Our construction is efficient and compact than what one would get through generic transformations (e.g., [4]). One drawback of our deterministic PKE is that it requires an LWE hardness assumption here with super-polynomial modulus-to-noise ratio, which is stronger than what we need in the (randomised) KEM scheme.

2 Preliminaries

Notation. We denote the security parameter by λ. We use bold lowercase letters (e.g. $\mathbf{a}$) to denote vectors and bold capital letters (e.g. $\mathbf{A}$) to denote matrices. For a positive integer $q \geq 2$, let $\mathbb{Z}_q$ be the ring of integers modulo q. We denote the group of $n \times m$ matrices in $\mathbb{Z}_q$ by $\mathbb{Z}_q^{n\times m}$. Vectors are treated as column vectors. The transpose of a vector $\mathbf{a}$ is denoted by $\mathbf{a}^\top$. For $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ and $\mathbf{B} \in \mathbb{Z}_q^{n\times m'}$, let $[\mathbf{A}|\mathbf{B}] \in \mathbb{Z}_q^{n\times(m+m')}$ be the concatenation of $\mathbf{A}$ and $\mathbf{B}$. We denote by $x \leftarrow X$ the process of sampling x according to the distribution X. We denote $s \leftarrow_\$ S$ the process that of sampling element x uniformly from the set S.

For $x \in \mathbb{Z}_p$, define $\mathsf{Transform}_q(x) = \lceil (q/p)\cdot x \rceil$. For $x \in \mathbb{Z}_q$, define the rounding function $\lfloor x \rceil_p = \lfloor (p/q)\cdot x \rceil$. The functions $\mathsf{Transform}_q(\cdot)$ and $\lfloor \cdot \rceil_p$ naturally extend to vectors by applying them component-wise.

For a security parameter λ, a function $\mathsf{negl}(\lambda)$ is negligible in λ if it is smaller than all polynomial fractions for a sufficiently large λ.

Definition 1 (Bounded Distribution, [2]). *For a distribution χ over the reals, and a bound β, we say that χ is β-bounded if the average absolute value of $x \leftarrow \chi$ is less than β, i.e., if $\mathbb{E}[|x|] \leq \beta$.*

Lemma 1. *Let χ be a B-bounded distribution over $\mathbb{Z}$. Let $q \geq p \cdot (2B+1) \cdot n^{\omega(1)}$ be a prime. For $e \leftarrow \chi$, $u \leftarrow_\$ \mathbb{Z}_q$, we have $\lfloor u + e \rceil_p \neq \lfloor u \rceil_p$ with probability $\leq (2B+1) \cdot p/q$ which is negligible in n.*

We recall the notion of full-rank-difference encodings (FRD). Agrawal et al. [1] gave an explicit construction of FRD, which we adapt in our construction.

Definition 2. *Let $n \geq 1$ be an integer and q be a prime. We say that a function $\varphi : \mathbb{Z}_q^n \to \mathbb{Z}_q^{n\times n}$ is an encoding with full-rank differences (FRD) if:*

1. *φ is computable in polynomial time;*
2. *for all distinct $\mathbf{u}, \mathbf{v} \in \mathbb{Z}_q^n$, $\varphi(\mathbf{u}) - \varphi(\mathbf{v}) \in \mathbb{Z}_q^{n\times n}$ is full rank (or invertible).*

Definition 3. *Let λ be a security parameter, $n = n(\lambda)$, $\ell = \ell(\lambda)$ and S be a distribution over D. A set of functions $\mathcal{F} = \{f : D \to R\}$ is a family of compression hash functions if (1) There exists a p.p.t algorithm that takes as input a security parameter 1^λ and uniformly samples a function f from $\mathcal{F}$; (2) Given f, $x \in D$, the computation of $f(x)$ can be done in p.p.t; (3) $\log|R| < \log|D|$. We say $\mathcal{F}$ is second pre-image resistant if for all p.p.t algorithm $\mathcal{A}$, the advantage*

$$\mathsf{Adv}^{\mathsf{tcr}}_{\mathcal{F},\mathcal{A}}(\lambda) = \left[\begin{array}{c} x \neq x^* \\ \text{and } f(x^*) = f(x) \end{array} : \begin{array}{c} f \leftarrow_\$ \mathcal{F} \ ; \ x^* \leftarrow S \\ x \leftarrow \mathcal{A}(1^\lambda, f, x^*) \end{array} \right] \leq \mathsf{negl}(\lambda)$$

We say $\mathcal{F}$ is ϵ-hard-to-invert w.r.t S if for all p.p.t algorithm $\mathcal{A}$,

$$\Pr[\mathcal{A}(f(x), f) = x) \ : \ f \leftarrow_\$ \mathcal{F}, x \leftarrow S] \leq \epsilon.$$

A collection of compression hash functions is collision-resistant if it is second pre-image resistant and $\mathsf{negl}(\lambda)$-hard-to-invert.

2.1 Public-Key Encapsulation

A public-key encapsulation (KEM) scheme $\Pi = (\mathsf{KeyGen}, \mathsf{Encap}, \mathsf{Decap})$ with key space $\mathcal{K}_\lambda$ consists of three polynomial-time algorithms. The key generation algorithm $\mathsf{KeyGen}(1^\lambda)$ generates a public key Pk and private key Sk. The randomised key encapsulation algorithm $\mathsf{Encap}(\mathsf{Pk})$ generates a session key $K \in \mathcal{K}_\lambda$ and a ciphertext Ct. The decapsulation algorithm $\mathsf{Decap}(\mathsf{Pk}, \mathsf{Sk}, \mathsf{Ct})$ returns the session key K or the error symbol $\perp$. The correctness of a KEM scheme requires that for all $\lambda \in \mathbb{N}$, and all $(K, \mathsf{Ct}) \leftarrow \mathsf{Encap}(\mathsf{Pk})$,

$$\Pr[\mathsf{Decap}(\mathsf{Pk}, \mathsf{Sk}, \mathsf{Ct}) = K] \geq 1 - \mathsf{negl}(\lambda)$$

where the probability is taken over the choice of $(\mathsf{Pk}, \mathsf{Sk}) \leftarrow \mathsf{KeyGen}(1^\lambda)$ and the random coins of Encap and Decap.

We recall the chosen-ciphertext security of KEM. The IND-CCA security of a KEM scheme Π with session key space $\mathcal{K}_\lambda$ is defined by the following security game. The challenger $\mathcal{C}$ runs $(\mathsf{Pk}, \mathsf{Sk}) \leftarrow \mathsf{KeyGen}(1^\lambda)$, chooses a random coin $\mu \leftarrow_\$ \{0,1\}$, samples $K_0^* \leftarrow_\$ \mathcal{K}_\lambda$, and computes $(K_1^*, \mathsf{Ct}^*) \leftarrow \mathsf{Encap}(\mathsf{Pk})$. Then $\mathcal{C}$ passes $(\mathsf{Pk}, K_\mu^*, \mathsf{Ct}^*)$ to the adversary. The adversary launches adaptive chosen-ciphertext attacks: It repeatedly chooses any $\mathsf{Ct} \neq \mathsf{Ct}^*$ and sends it over to $\mathcal{C}$, to which $\mathcal{C}$ returns $\mathsf{Decap}(\mathsf{Pk}, \mathsf{Sk}, \mathsf{Ct})$. Finally, $\mathcal{A}$ outputs μ' and wins if $\mu' = \mu$. We define $\mathcal{A}$'s advantage in the above security game as

$$\mathsf{Adv}^{\mathsf{ind\text{-}cca}}_{\mathcal{A},\Pi}(\lambda) = |\Pr[\mu' = \mu] - 1/2|.$$

We say Π is IND-CCA-secure if $\mathsf{Adv}^{\mathsf{ind\text{-}cca}}_{\mathcal{A},\Pi}(\lambda)$ is negligible in λ.

2.2 Randomness Extraction

The statistical distance between two random variables X and Y over a finite set S is $\mathsf{SD}(X,Y) = \frac{1}{2}\sum_{s\in S} |\Pr[X = s] - \Pr[Y = s]|$. For any $\epsilon > 0$, we say X and Y are ϵ-close if $\mathsf{SD}(X,Y) \leq \epsilon$. The min-entropy of a random variable X is $H_\infty(X) = -\log(\max_{s\in S} \Pr[X = s])$. The *average-case* conditional min-entropy of X given Y is $\tilde{\mathrm{H}}_\infty(X|Y) = -\log\left(\mathbb{E}_{y\leftarrow Y}\left[\max_x \Pr[X = x|Y = y]\right]\right)$. A distribution (or a random variable) X is called k-source if $\mathrm{H}_\infty(X) \geq k$.

Lemma 2 ([9], Lemma 2.2). *Let X, Y and Z be random variables where Z has at most 2^λ positive-probability values. Then $\tilde{\mathrm{H}}_\infty(X|Y,Z) \geq \tilde{\mathrm{H}}_\infty(X|Y) - \lambda$, and in particular $\tilde{\mathrm{H}}_\infty(X|Z) \geq \mathrm{H}_\infty(X) - \lambda$.*

Definition 4. *A collection of functions $\mathcal{H} = \{h : D \to R\}$ is universal if for any $x_1, x_2 \in D$ such that $x_1 \neq x_2$ it holds that $\Pr_{H\leftarrow\mathcal{H}}[H(x_1) = H(x_2)] = 1/|R|$.*

Lemma 3. *Let X, Y be random variables such that $X \in \{0,1\}^n$, and $\tilde{\mathrm{H}}_\infty(X|Y) \geq k$. Let $\mathcal{H}$ be a collection of universal hash functions from $\{0,1\}^n$ to $\{0,1\}^\ell$ where $\ell \leq k - 2\log(1/\epsilon)$. It holds that for $h \leftarrow_\$ \mathcal{H}$, and $r \leftarrow_\$ \{0,1\}^\ell$,*

$$\mathit{SD}\left((h, h(X), Y), (h, r, Y)\right) \leq \epsilon$$

Lemma 4 ([1], Lemma 4). *Suppose that* $m > (n+1)\log q + \omega(\log n)$ *and that* $q > 2$ *is prime. Let* $\mathbf{R}$ *be an* $m \times k$ *matrix chosen uniformly in* $\{1, -1\}^{m \times k} \mod q$ *where* $k = k(n)$ *is polynomial in* n*. Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be matrices chosen uniformly in* $\mathbb{Z}_q^{n \times m}$ *and* $\mathbb{Z}_q^{n \times k}$ *respectively. Then the distribution* $(\mathbf{A}, \mathbf{AR})$ *is statistically close to the distribution* $(\mathbf{A}, \mathbf{B})$*.*

2.3 Computational Assumptions

We recall the LWE problem that was introduced by Regev [20].

Definition 5. *Let* λ *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* $\mathsf{LWE}_{n,m,q,\chi}$ *problem asks for distinguishing the following two distributions:*

$$\mathsf{Real} = (\mathbf{A}, \mathbf{s}^\top \mathbf{A} + \mathbf{e}^\top) \quad \text{and} \quad \mathsf{Rand} = (\mathbf{A}, \mathbf{c}^\top)$$

where $\mathbf{A} \leftarrow_\$ \mathbb{Z}_q^{n \times m}$*,* $\mathbf{s} \leftarrow_\$ \mathbb{Z}_q^n$*,* $\mathbf{e} \leftarrow \chi^m$*, and* $\mathbf{c} \leftarrow_\$ \mathbb{Z}_q^n$*. We define the advantage that an adversary* $\mathcal{A}$ *has in solving the LWE problem by*

$$\mathsf{Adv}_{\mathcal{A}}^{\mathsf{LWE}_{n,m,q,\chi}}(\lambda) = |\Pr[\mathcal{A}(1^\lambda, \mathsf{Real}) = 1] - \Pr[\mathcal{A}(1^\lambda, \mathsf{Rand})]|.$$

We say the LWE assumption holds if for every p.p.t. algorithm $\mathcal{A}$*,* $\mathsf{Adv}_{\mathcal{A}}^{\mathsf{LWE}_{n,m,q,\chi}}(\lambda)$ *is negligible in* λ*.*

Usually, the distribution χ is the discrete Gaussian distribution $D_{\mathbb{Z},\alpha q}$ where the parameter $\alpha \in (0,1)$ and $\alpha q \geq \sqrt{n}$. We refer to [11] for details on discrete Gaussian distributions and [17] for the recent result on the hardness of LWE.

In our construction, we consider the amortised LWE problem that asks to distinguish between distributions $(\mathbf{B}, \mathbf{CB} + \mathbf{F})$ and $(\mathbf{B}, \mathbf{A})$ where $\mathbf{B} \leftarrow_\$ \mathbb{Z}_q^{\ell \times m}$, $\mathbf{C} \leftarrow_\$ \mathbb{Z}_q^{n \times \ell}$, $\mathbf{F} \leftarrow \chi^{n \times m}$ and $\mathbf{A} \leftarrow \mathbb{Z}_q^{n \times m}$. It was shown, e.g., in [18] (Lemma 7.3), that a p.p.t. algorithm that distinguishes the two distributions of the amortised LWE problem with probability ϵ can be efficiently turned into a p.p.t. algorithm that breaks the $\mathsf{LWE}_{\ell,m,q,\chi}$ problem (per Definition 5) with advantage ϵ/n.

We recall the following Lemma, first proven by Goldwasser et al. [12], and used by Xie et al. [22]. It says that, for certain parameters, the LWE problem remains hard even if the secret is chosen from an arbitrary distribution with sufficient min-entropy in the presence of hard-to-invert auxiliary input.

Lemma 5. *Let* $k \geq \log q$ *and* $\mathcal{F} = \{f : \{0,1\}^n \to \{0,1\}^*\}$ *be a family of one-way functions that are* 2^{-k} *hard to invert with respect to distribution* S *over* $\{0,1\}^n$*. For any super-polynomial* $q = q(\lambda)$ *and any* $m = \mathsf{poly}(n)$*, any* $\beta, \gamma \in (0,1)$ *such that* $\gamma/\beta = \mathsf{negl}(n)$*, the distributions*

$$\big(\mathbf{A}, \mathbf{s}^\top \mathbf{A} + \mathbf{e}^\top, f(\mathbf{s})\big) \quad \text{and} \quad \big(\mathbf{A}, \mathbf{c}^\top, f(\mathbf{s})\big)$$

are computationally indistinguishable where $\mathbf{A} \leftarrow_\$ \mathbb{Z}_q^{n \times m}$*,* $\mathbf{s} \leftarrow S$*,* $\mathbf{c} \leftarrow_\$ \mathbb{Z}_q^m$*,* $\mathbf{e} \leftarrow D_{\mathbb{Z},\beta q}^m$*, assuming the* $\mathsf{LWE}_{\ell.m,q,D_{\mathbb{Z},\gamma q}}$ *assumption holds where* $\ell \geq \frac{k - \omega(\log n)}{\log q}$*.*

2.4 Lattice Trapdoors

Let $n \geq 1$, $q \geq 2$ and $p \leq q$. Set $k = \lceil \log q \rceil$ and $w = nk$, and define the n-by-w gadget matrix $\mathbf{G} = \mathbf{I}_n \otimes [1, 2, 4, ..., 2^{k-1}] \in \mathbb{Z}_q^{n \times w}$. We recall the following lemma that applies the gadget trapdoor [15] to invert the LWE and LWR functions. The lemma stems from Lemma 7.2 of [2] (the algorithm BigInvert). Here we use the fact that the gadget matrix $\mathbf{G}$ has a (publicly known) trapdoor matrix $\mathbf{T} \in \mathbb{Z}^{w \times w}$ s.t. $\mathbf{GT} = \mathbf{0} \bmod q$ and $\|\mathbf{T}\| \leq \sqrt{5}$. (See [15], Proposition 4.2 for details).

Lemma 6 ([3] Lemma 7.2). *Let $n \geq 1$, $q \geq 2$, $w = n\lceil \log q \rceil$ and $m = \bar{m} + w$. Set , $\bar{m} > (n+1)\log q + \omega(\log n)$. Let $\mathbf{F} = [\mathbf{A}|\mathbf{AR} + \mathbf{HG}]$ where $\mathbf{A} \in \mathbb{Z}_q^{n \times \bar{m}}$, $\mathbf{R} \leftarrow_{\$} \{-1, 1\}^{\bar{m} \times w}$ and $\mathbf{H} \in \mathbb{Z}_q^{n \times n}$ be an invertible matrix. We have for $\mathbf{c}^\top = \lfloor \mathbf{s}^\top \mathbf{F} \rceil_p$ where $\mathbf{s} \in \mathbb{Z}_q^n$, $p \geq O(\bar{m}\sqrt{n \log q})$, there is a p.p.t algorithm $\mathsf{Invert}(\mathsf{Transform}_q(\mathbf{c}), \mathbf{F}, \mathbf{H}, \mathbf{R})$ that outputs $\mathbf{s}$.*

The following lemma is derived from Lemma 3.3 and Theorem 7.3 of [3].

Lemma 7. *Let λ be the security parameter. Let n, m, ℓ, p, γ be positive integers, χ be a β-bounded distribution, $w = n\lceil \log q \rceil$, and $q \geq \bar{m}\beta\gamma n(\bar{m}+w)p$ be a prime. Then it holds that for $\mathbf{s} \leftarrow_{\$} \mathbb{Z}_q^n$, $\mathbf{A} = \mathbf{CB} + \mathbf{F} \in \mathbb{Z}_q^{n \times \bar{m}}$, $\mathbf{R} \leftarrow_{\$} \{-1, 1\}^{\bar{m} \times w}$*

$$\tilde{\mathrm{H}}_\infty(\mathbf{s} | \left\lfloor \mathbf{s}^\top [\mathbf{A}|\mathbf{AR}] \right\rceil_p) \geq n \log(2\gamma) - (\ell + \lambda) \log q$$

where $\mathbf{B} \leftarrow_{\$} \mathbb{Z}_q^{\ell \times \bar{m}}$, $\mathbf{C} \leftarrow_{\$} \mathbb{Z}_q^{n \times \ell}$ and $\mathbf{F} \leftarrow \chi^{n \times \bar{m}}$.

3 The KEM Scheme

Let λ be the security parameter. The scheme uses a full-rank difference encoding function $\varphi : \mathbb{Z}_q^n \to \mathbb{Z}_q^{n \times n}$ which can be instatiated by the construction given by Agrawal et al. [1]. The scheme also employs a family of hash functions $\mathcal{F} = \{f : \mathbb{Z}_q^n \to \{0,1\}^\lambda\}$ that is second pre-image resistant, and a family of universal hash functions $\mathcal{H} = \{h : \mathbb{Z}_q^n \to \{0,1\}^\lambda\}$ for which efficient constructions are known. Let χ be a β-bounded distribution over $\mathbb{Z}_q$. Given the lattice dimension $\ell \geq \lambda$ for LWE problem, we set the parameters for our KEM scheme as follows.

- Let $\delta > 0$ be a constant. Set the matrix dimension n large enough such that $\frac{n-4\lambda}{n^\delta} \geq \ell$ for Lemma 7 (ensuing that $\mathbf{s}$ sufficient leftover min-entropy).
- Set the matrix dimension $\bar{m} = n^{1+\delta}$ to ensure that Lemma 4 applies. Here we assume $n^\delta = 2 \log q$.
- The rounding parameter $p = 3\bar{m}^{1.5}$ for Lemma 6.
- The parameter $\gamma = 1$ for Lemma 7
- Set $\beta = \sqrt{\ell}$ as required by the hardness of LWE problem.
- The LWE modulus $q = 12\bar{m}^5$ that satisfies Lemma 7.

KeyGen(1^λ)**:** On input the security parameter λ, the key generation algorithm does:

1. Choose $\mathbf{A} \leftarrow_{\$} \mathbb{Z}_q^{n \times \bar{m}}$, $\mathbf{R} \leftarrow_{\$} \{-1,1\}^{\bar{m} \times w}$; Set $\mathbf{A}_1 = \mathbf{AR} \bmod q$.
2. Randomly sample a hash function $f \leftarrow_{\$} \mathcal{F}$ and a universal hash function $h \leftarrow_{\$} \mathcal{H}$.
3. Set $\mathsf{Pk} = (\mathbf{A}, \mathbf{A}_1, f, h)$ and $\mathsf{Sk} = \mathbf{R}$.

Encap(Pk): On input the public key Pk, the encapsulation algorithm does:

1. Select $\mathbf{s} \leftarrow_{\$} \mathbb{Z}_q^n$ and compute $\mathbf{t} \leftarrow f(\mathbf{s})$.
2. Encode $\mathbf{t}$ as a vector in $\mathbb{Z}_q^n$ and compute $\mathbf{c}^\top = \left\lfloor \mathbf{s}^\top \cdot [\mathbf{A} | \mathbf{A}_1 + \varphi(\mathbf{t})\mathbf{G}] \right\rceil_p$
3. Set $K \leftarrow h(\mathbf{s})$ and $\mathsf{Ct} = (\mathbf{c}, \mathbf{t})$.

Decap($\mathsf{Pk}, \mathsf{Sk}, \mathsf{Ct}$): On input the private key Sk and a ciphertext $\mathsf{Ct} = (\mathbf{c}, \mathbf{t})$, the decapsulation algorithm does:

1. Runs $\mathsf{Invert}(\mathsf{Transform}_q(\mathbf{c}), [\mathbf{A} | \mathbf{A}_1 + \varphi(\mathbf{t})\mathbf{G}], \mathbf{R})$ to get $\mathbf{s}' \in \mathbb{Z}_q^n$.
2. Compute $\mathbf{t}' = f(\mathbf{s}')$ and return $\perp$ if $\mathbf{t}' \neq \mathbf{t}$.
3. Return $K \leftarrow h(\mathbf{s}')$.

The decryption correctness can be checked by the correctness of Invert as stated in Lemma 6.

Theorem 1. *If the family of hash functions $\mathcal{F}$ is second pre-image resistant and the $\mathsf{LWE}_{\ell,\bar{m},q,\chi}$ assumption holds, then the KEM scheme is IND-CCA-secure. More specifically, let λ be the security parameter. Given a p.p.t adversary $\mathcal{A}$ that breaks the KEM scheme Π with advantage $\mathsf{Adv}^{\mathsf{ind\text{-}cca}}_{\Pi,\mathcal{A}}(\lambda)$, there exist a p.p.t algorithm $\mathcal{B}_1$ that breaks the second pre-image resistance of $\mathcal{F}$ with advantage $\mathsf{Adv}^{\mathsf{tcr}}_{\mathcal{F},\mathcal{B}_1}(\lambda)$ and a p.p.t algorithm $\mathcal{B}_2$ that breaks $\mathsf{LWE}_{\ell,\bar{m},q,\chi}$ with advantage $\mathsf{Adv}^{\mathsf{LWE}_{\ell,\bar{m},q,\chi}}_{\mathcal{B}_2}(\lambda)$, such that $\mathsf{Adv}^{\mathsf{ind\text{-}cca}}_{\Pi,\mathcal{A}}(\lambda) \leq \mathsf{Adv}^{\mathsf{tcr}}_{\mathcal{F},\mathcal{B}_1}(\lambda) + \mathsf{Adv}^{\mathsf{LWE}_{\ell,\bar{m},q,\chi}}_{\mathcal{B}_2}(\lambda) + \mathsf{negl}(\lambda)$ where $\mathsf{negl}(\lambda)$ is negligible in λ.*

Proof. We proceed with the proof as a sequence of games. For $i = \{0,1,2,3,4\}$, we denote the i-th game by $Game_i$. We denote by $Game_i \Rightarrow 1$ the event that the adversary wins the security game, i.e., it outputs μ' such that $\mu' = \mu$.

The first game $Game_0$ is the same as the IND-CCA security game. That is, the adversary $\mathcal{A}$ receives a public key $\mathsf{Pk} = (\mathbf{A}, \mathbf{A}_1, f, h)$ and a challenge ciphertext $\mathsf{Ct}^* = (\mathbf{c}^*, \mathbf{t}^*)$, where

$$\mathbf{t}^* = f(\mathbf{s}^*) \quad ; \quad \mathbf{c}^{*\top} = \left\lfloor \mathbf{s}^{*\top} \cdot [\mathbf{A} | \mathbf{A}_1 + \varphi(\mathbf{t}^*)\mathbf{G}] \right\rceil_p$$

for some $\mathbf{s}^* \leftarrow_{\$} \mathbb{Z}_q^n$, and a session key K^*_μ, which is either a random value from $\{0,1\}^\lambda$ or $h(\mathbf{s}^*)$, from the challenger $\mathcal{B}$. Then $\mathcal{A}$ adaptively issues decryption queries $\mathsf{Ct} = (\mathbf{c}, \mathbf{t}) \neq \mathsf{Ct}^*$ and $\mathcal{B}$ runs the real decryption algorithm to return the answers. Finally, $\mathcal{A}$ outputs a bit value μ' indicating that Ct^* encapsulates a real session key or a random session key. According to the definition, we have

$$\Pr[Game_0 \Rightarrow 1] = \Pr[\mu' = \mu] = \mathsf{Adv}^{\mathsf{ind\text{-}cca}}_{\Pi,\mathcal{A}}(\lambda) + 1/2$$

In $Game_1$, we make a change in the way of answering decryption queries: $\perp$ is returned if for the given decryption query $\mathsf{Ct} = (\mathbf{c}, \mathbf{t})$, $\mathbf{t} = \mathbf{t}^*$; otherwise, Ct is processed with the real decapsulation algorithm as in $Game_0$. We argue that

unless the adversary breaks the second pre-image resistant property of the hash function f, this change is not noticeable.

First of all, we must have $\mathbf{c} \neq \mathbf{c}^*$ (otherwise the decryption query is invalid as it is the challenge ciphertext itself). To make the decryption oracle not to output $\perp$, there must be a unique $\mathbf{s} \neq \mathbf{s}^*$ such that $\mathbf{c}^\top = \lfloor \mathbf{s}^\top \cdot [\mathbf{A}|\mathbf{A}_1 + \varphi(\mathbf{t}^*)\mathbf{G}] \rceil_p$ (and such an $\mathbf{s}$ can be found by the algorithm Invert since the private key $\mathbf{R}$ is known). Therefore we must have $f(\mathbf{s}) = f(\mathbf{s}^*) = \mathbf{t}^*$ which makes $\mathbf{s}$ a valid second pre-image for $\mathbf{t}^*$. So, we have

$$|\Pr[Game_1 \Rightarrow 1] - \Pr[Game_0 \Rightarrow 1]| \leq \mathsf{Adv}^{\mathsf{tcr}}_{\mathcal{F},\mathcal{B}_1}(\lambda)$$

for some proper adversary $\mathcal{B}_1$.

In $Game_2$, we make the following changes on generating the matrix $\mathbf{A}_1$ from the public key Pk. Firstly, we pick $\mathbf{s}^* \leftarrow_\$ \mathbb{Z}_q^n$ and set $\mathbf{t}^* = f(\mathbf{s}^*)$. Then we sample $\mathbf{R} \leftarrow_\$ \{-1, 1\}^{\bar{m} \times w}$ and set $\mathbf{A}_1 \leftarrow \mathbf{AR} - \varphi(\mathbf{t}^*)\mathbf{G} \bmod q$. $\mathbf{s}^*$ is also used to construct the challenge ciphertext:

$$\mathbf{t}^* \leftarrow f(\mathbf{s}^*) \quad ; \quad \mathbf{c}^{*\top} \leftarrow \left\lfloor \mathbf{s}^{*\top} \cdot [\mathbf{A}|\mathbf{A}_1 + \varphi(\mathbf{t}^*)\mathbf{G}] \right\rceil_p$$

The decryption oracle is implemented as in $Game_1$.

We argue that the adversary's views in $Game_2$ and $Game_1$ are statistically close. First, by Lemma 4, the distributions of $\mathbf{A}_1$ in these two games are statistically close. This means that Pk generated in these two games are statistically indistinguishable for $\mathcal{A}$. Then we note that the decryption queries will be answered properly. This is because by the standard technique of Agrawal et al. [1], knowledge of the binary matrix $\mathbf{R}$ lets us transform the trapdoor for $\mathbf{G}$ into a trapdoor for the whole matrix, as long as $\mathbf{H}$ is invertible. The simulator can thus answer in the same way as the previous games, except for the ciphertexts $\mathsf{Ct} = (\mathbf{c}, \mathbf{t}^*)$, which however, are already excluded:

$$\begin{aligned}[\mathbf{A}|\mathbf{A}_1 + \varphi(\mathbf{t})\mathbf{G}] &= [\mathbf{A}|\mathbf{AR} + (\varphi(\mathbf{t}) - \varphi(\mathbf{t}^*))\mathbf{G}] \\ &= [\mathbf{A}|\mathbf{AR} + \mathbf{HG}]\end{aligned}$$

where, by the property of FRD, $\mathbf{H} \in \mathbb{Z}_q^{n \times n}$ is invertible. So we have

$$|\Pr[Game_2 \Rightarrow 1] - \Pr[Game_1 \Rightarrow 1]| \leq \mathsf{negl}_1(\lambda)$$

for some negligible statistical error $\mathsf{negl}_1(\lambda)$.

In $Game_3$ we change the way that the matrix $\mathbf{A}$ is constructed. In particular, we obtain $\mathbf{A} \leftarrow \mathbf{CB} + \mathbf{F}$ where $\mathbf{B} \leftarrow_\$ \mathbb{Z}_q^{\ell \times \bar{m}}$, $\mathbf{C} \leftarrow_\$ \mathbb{Z}_q^{n \times \ell}$ and $\mathbf{F} \leftarrow \chi^{n \times \bar{m}}$. By the LWE assumption (amortised version) we immediately have

$$\Pr[Game_3 \Rightarrow 1] - \Pr[Game_2 \Rightarrow 1] \leq \mathsf{Adv}^{\mathsf{LWE}_{\ell,\bar{m},q,\chi}}_{\mathcal{B}_2}(\lambda)$$

for some proper adversary $\mathcal{B}_2$.

In $Game_4$, we change the way of generating the challenge session key. In particular, K_1^* is chosen randomly from $\{0,1\}^\lambda$ (recall that K_0^* is chosen uniformly

random from $\{0,1\}^\lambda$ in all previous games). We argue that $Game_3$ and $Game_4$ are statistically indistinguishable. First of all, $\mathbf{t}^*$ in the challenge ciphertext has at most 2^λ values. Second, by the construction of the matrix $\mathbf{A}$ and $\mathbf{c}^*$, we have

$$\begin{aligned}\mathbf{c}^{*\top} &= \left\lfloor \mathbf{s}^{*\top} \cdot [\mathbf{A}|\mathbf{A}_1 + \varphi(\mathbf{t}^*)\mathbf{G}] \right\rceil_p \\ &= \left\lfloor \mathbf{s}^{*\top} \cdot [\mathbf{A}|\mathbf{AR} + (\varphi(\mathbf{t}^*) - \varphi(\mathbf{t}^*))\mathbf{G}] \right\rceil_p \\ &= \left\lfloor \mathbf{s}^{*\top} \cdot [\mathbf{A}|\mathbf{AR}] \right\rceil_p\end{aligned}$$

By Lemma 2,

$$\begin{aligned}\tilde{\mathrm{H}}_\infty(\mathbf{s}^*|\mathbf{c}^*, \mathbf{t}^*) &\geq \tilde{\mathrm{H}}_\infty(\mathbf{s}^*|\mathbf{c}^*) - \lambda \\ &\geq n\log(2\gamma) - (\lambda+\ell)\log q - \lambda \\ &\geq n - 2\ell\log q - \lambda \\ &= n - \ell \cdot n^\delta - \lambda \\ &\geq 4\lambda - \lambda \\ &= 3\lambda\end{aligned}$$

Let $\epsilon = 2^{-\lambda}$. So, we have $\tilde{\mathrm{H}}_\infty(\mathbf{s}^*|\mathbf{c}^*, \mathbf{t}^*) \geq 2\log(1/\epsilon) + \lambda$. Applying Lemma 3 results in $\mathsf{SD}((\mathbf{c}^*, \mathbf{t}^*, h, h(\mathbf{s}^*), (\mathbf{c}^*, \mathbf{t}^*, h, K_1^*)) \leq \epsilon = 2^{-\lambda}$ where $K_1^* \leftarrow \{0,1\}^\lambda$. We therefore obtain

$$|\Pr[Game_4 \Rightarrow 1] - \Pr[Game_3 \Rightarrow 1]| \leq 2^{-\lambda}$$

Additionally, In $Game_4$, K_0^* and K_1^* are all random strings chosen from $\{0,1\}^\lambda$. So the adversary $\mathcal{A}$ has exactly probability $1/2$ of correctly guessing μ, i.e.,

$$\Pr[Game_4 \Rightarrow 1] = 1/2$$

Combining the above steps gives us

$$\mathsf{Adv}^{\mathsf{ind\text{-}cca}}_{\Pi,\mathcal{A}}(\lambda) \leq \mathsf{Adv}^{\mathsf{tcr}}_{\mathcal{F},\mathcal{B}_1}(\lambda) + \mathsf{Adv}^{\mathsf{LWE}_{\ell,\bar{m},q,\chi}}_{\mathcal{B}_2}(\lambda) + \mathsf{negl}(\lambda)$$

where $\mathsf{negl}(\lambda) = \mathsf{negl}_1(\lambda) + 2^{-\lambda}$ is negligible. This completes the proof. □

4 CCA-Secure Deterministic Public-Key Encryption

In this section, we show a construction of CCA-secure deterministic public-key encryption (D-PKE) in the standard model. Deterministic public-key encryption only makes sense for high-min-entropy plaintexts, to preclude the obvious guess-and-encrypt attack, but for such messages it has practical applications ranging from encrypted keyword search to encrypted cloud storage with deduplication.

Our CCA-secure D-PKE has a similar structure as our KEM. We consider the so-called PRIV-CCA security notion for single hard-to-guess message as in

[4]. [4] The security of our construction is again based on the hardness of the LWE problem with high-min-entropy secret in the presence of hard-to-invert auxiliary inputs, which is as hard as the standard form of LWE with certain parameters [12]. Our construction is more efficient than the generic constructions by Boldyreva et al. [4] which requires double encryption (e.g., using two lossy trapdoor functions when instantiated with lattice-based primitives).

A D-PKE scheme consists of three algorithms. On input a security parameter 1^λ, the randomised key generation algorithm $\mathsf{KeyGen}(1^\lambda)$ outputs a pair of public and private keys $(\mathsf{Pk}, \mathsf{Sk})$. The *deterministic* encryption algorithm $\mathsf{Enc}(\mathsf{Pk}, \mathsf{m})$ returns a ciphertext Ct. The decryption algorithm $\mathsf{Dec}(\mathsf{Pk}, \mathsf{Sk}, \mathsf{Ct})$ returns the message m or $\perp$. The correctness is required that for all m, $(\mathsf{Pk}, \mathsf{Sk}) \leftarrow \mathsf{KeyGen}(1^\lambda)$,

$$\Pr[\mathsf{Dec}(\mathsf{Pk}, \mathsf{Sk}, \mathsf{Enc}(\mathsf{Pk}, \mathsf{m})) = \mathsf{m}] \geq 1 - \mathsf{negl}(\lambda).$$

We recall the indistinguishability-based security definition of D-PKE for single high-min-entropy messages. Here we consider a stronger version where we require ciphertext pseudorandomness, i.e., that ciphertexts be computationally indistinguishable from random stings. The security game with a D-PKE scheme Π is defined as follows. The adversary $\mathcal{A}$ outputs a distribution M over the message space. where $\mathrm{H}_\infty(M) \geq k$ (i.e., M is a k-source). The challenger $\mathcal{B}$ runs $(\mathsf{Pk}, \mathsf{Sk}) \leftarrow \mathsf{KeyGen}(1^\lambda)$. It flips a coin $\mu \leftarrow_\$ \{0,1\}$. If $\mu = 0$ it computes $\mathsf{Ct}^* \leftarrow \mathsf{Enc}(\mathsf{Pk}, \mathsf{m}^*)$ where $\mathsf{m}^* \leftarrow M$. Otherwise it chooses Ct^* uniformly at random from the ciphertext space. $\mathcal{B}$ returns $(\mathsf{Pk}, \mathsf{Ct}^*)$ to $\mathcal{A}$. $\mathcal{A}$ then launches adaptive decryption queries $\mathsf{Ct} \neq \mathsf{Ct}^*$ to which $\mathcal{B}$ returns $\mathsf{Dec}(\mathsf{Pk}, \mathsf{Sk}, \mathsf{Ct})$. Finally, $\mathcal{A}$ outputs μ' and wins if $\mu' = \mu$. We define $\mathcal{A}$'s advantage in the security game as

$$\mathsf{Adv}^{\mathsf{priv1\text{-}cca}}_{\Pi,\mathcal{A}}(\lambda) = |\Pr[\mu' = \mu] - 1/2|.$$

We say a D-PKE scheme Π is PRIV-CCA-secure w.r.t. a k-source single message if for every p.p.t. adversary $\mathcal{A}$, the advantage is negligible in λ.

Construction. Our construction uses a full-rank difference encoding function $\varphi : \mathbb{Z}_q^n \to \mathbb{Z}_q^{n \times n}$ as in our construction of KEM. The construction also uses a family of second pre-image resistant functions $\mathcal{F} = \{f : \{0,1\}^n \to \{0,1\}^{2\lambda}\}$ that is universal and 2^{-k}-hard-to-invert with respect to a k-source M over $\{0,1\}^n$. Such a family of functions can be built from the standard Short Integer Solution (SIS) problem.

The security of the construction is based on the hardness of $\mathsf{LWE}_{\ell,q,\bar{m},D_{\mathbb{Z},\gamma q}}$ where we need, for Lemma 5, $\ell \geq \frac{k-\omega(\log n)}{\log q}$, $\gamma \in (0,1)$ such that $\gamma/\beta = \mathsf{negl}(n)$. We set the parameters for decryption correctness and security as follows.

- Set the LWE modulus $q = n^{\omega(1)}$ and parameter $\beta = \sqrt{\ell}/q$ for the LWE hardness results of, e.g. [20,17].
- Set the dimension $\bar{m} = n^{1+\delta}$ where $n^\delta = O(\log q)$ for Lemma 4.

[4] It was shown in [4] that such a security notion is equevalent to the PRIV-CCA security notion for multiple messages that form a block source. See [4] for details.

– The rounding parameter $p = 3\bar{m}^{1.5}$, to ensure that Lemma 6 applies.
– Finally, $\beta = 1/(2p\sqrt{n\bar{m}}) \cdot n^{-\omega(1)}$ for applying Lemma 1.

KeyGen(1^λ)**:** On input the security parameter λ, the algorithm does:
1. Choose $\mathbf{A} \leftarrow_\$ \mathbb{Z}_q^{n\times\bar{m}}$, $\mathbf{R} \leftarrow_\$ \{-1,1\}^{\bar{m}\times w}$; Set $\mathbf{A}_1 = \mathbf{AR} \bmod q$.
2. Sample a *universal*, second pre-image resistant, 2^k-hard-to-invert hash function $f \leftarrow_\$ \mathcal{F}$.
3. Set $\mathsf{Pk} = (\mathbf{A}, \mathbf{A}_1, f)$ and $\mathsf{Sk} = \mathbf{R}$.

Enc$(\mathsf{Pk}, \mathsf{m})$**:** On input the public key Pk and message $\mathsf{m} \in \{0,1\}^n$ which comes from some k-source, the algorithm does:
1. Compute $\mathbf{t} \leftarrow f(\mathsf{m})$ and encode $\mathbf{t}$ as a vector in $\mathbb{Z}_q^n$.
2. Compute $\mathbf{c}^\top = \left\lfloor \mathsf{m}^\top \cdot [\mathbf{A}|\mathbf{A}_1 + \varphi(\mathbf{t})\mathbf{G}] \right\rceil_p$
3. Set $\mathsf{Ct} = (\mathbf{c}, \mathbf{t})$.

Dec$(\mathsf{Sk}, \mathsf{Ct})$**:** On input the private key Sk and a ciphertext $\mathsf{Ct} = (\mathbf{c}, \mathbf{t})$, the decryption algorithm does:
1. Runs $\mathsf{Invert}(\mathsf{Transform}_q(\mathbf{c}), [\mathbf{A}|\mathbf{A}_1 + \varphi(\mathbf{t})\mathbf{G}], \mathbf{R})$ to get $\mathsf{m}' \in \{0,1\}^n$.
2. Compute $\mathbf{t}' = f(\mathsf{m}')$. Return m' if $\mathbf{t}' = \mathbf{t}$ or return $\perp$ otherwise.

Security Proof. Now we give the security proof.

Theorem 2. *Let $k \geq 2\log(1/n^{-\omega(1)}) + \lambda$. If the family of functions $\mathcal{F}$ is universal, 2^{-k}-hard-to-invert, second pre-image resistant, and Lemma 5 holds, the above construction of D-PKE scheme is PRIV-CCA-secure for k-source single message.*

Proof. We proceed the proof by a sequence of games. For $i = \{0,1,2,3,4\}$, we denote the i-th game by $Game_i$. We denote by $Game_i \Rightarrow 1$ the event that the adversary wins the security game, i.e., it outputs μ' such that $\mu' = \mu$.

The first game $Game_0$ is the original PRIV-CCA security game. That is, the adversary $\mathcal{A}$ generates a k-source distribution M. The challenger samples a challenge message $\mathsf{m}^* \leftarrow M$ and a fair coin $\mu \leftarrow_\$ \{0,1\}$. It then returns the public key $(\mathbf{A}, \mathbf{A}_1, f)$ and the challnege ciphertext Ct_μ^* to $\mathcal{A}$, where $\mathsf{Ct}_0^* \leftarrow \mathsf{Enc}(\mathsf{Pk}, \mathsf{m}^*)$ and Ct_1^* is uniformly chosen from the ciphertext space. $\mathcal{A}$ then launches adaptive chosen-ciphertext queries Ct subject to the condition that $\mathsf{Ct} \neq \mathsf{Ct}^*$. Finally, $\mathcal{A}$ outputs μ' and it wins if $\mu' = \mu$. By definition we have

$$|\Pr[Game_0 \Rightarrow 1] - 1/2| = \mathsf{Adv}_{\Pi,\mathcal{A}}^{\mathsf{priv1\text{-}cca}}(\lambda).$$

In the second game $Game_1$, we slightly change the way of answering the decryption query: Let the challenge ciphertext $\mathsf{Ct}_\mu^* = (\mathbf{c}^*, \mathbf{t}^*)$. A decryption query $\mathsf{Ct} = (\mathbf{c}, \mathbf{t})$ is rejected if $\mathbf{t} = \mathbf{t}^*$. First, we must have $\mathsf{Ct} \neq \mathsf{Ct}_\mu^*$ by security definition. Second, if $\mathbf{c} \neq \mathbf{c}^*$, there is a $\mathsf{m}' \in \{0,1\}^n$ such that $\mathbf{c}^\top = \left\lfloor \mathsf{m}'^\top [\mathbf{A}|\mathbf{A}_1 + \varphi(\mathbf{t}^*)] \right\rceil_p$. (In the case that Ct_0^* was returned, we must have $\mathsf{m}' \neq \mathsf{m}^*$.) Therefore, m' is a valid second pre-image of $\mathbf{t}^*$ on f, and m' can be recovered efficiently through the decryption procedure. So a p.p.t distinguisher

between $Game_0$ and $Game_1$ leads to a second-pre-image inversion algorithm for $\mathcal{F}$ and we have

$$|\Pr[Game_1 \Rightarrow 1] - \Pr[Game_0 \Rightarrow 1]| \leq \mathsf{negl}_1(\lambda).$$

In $Game_2$ we set $\mathbf{A}_1 \leftarrow \mathbf{AR} - \varphi(\mathbf{t}^*)\mathbf{G}$. By making this change we have challenge ciphertext $\mathsf{Ct}_0^* = (\mathbf{t}^* = f(\mathsf{m}^*),\ \mathbf{c}^{*T} = \lfloor \mathsf{m}^{*\top}[\mathbf{A}|\mathbf{AR}] \rfloor_p)$. By Lemma 4, $\mathbf{A}_1$ is distributed properly except for a negligible statistical error $\mathsf{negl}_1\lambda$. So we have

$$|\Pr[Game_2 \Rightarrow 1] - \Pr[Game_1 \Rightarrow 1]| \leq \mathsf{negl}_2(\lambda).$$

In $Game_3$, we make changes on computing the challenge ciphertext Ct_0^*. Specifically, given the challenge message $\mathsf{m}^* \leftarrow M$, we sample $\mathbf{e} \leftarrow D_{\mathbb{Z},\beta q}^{\bar{m}}$ and compute

$$\begin{aligned}\mathbf{c}^{*\top} &= \left\lfloor \mathsf{m}^{*\top}[\mathbf{A}|\mathbf{A}_1 + \varphi(\mathbf{t}^*)\mathbf{G}] + [\mathbf{e}^\top|\mathbf{e}^\top\mathbf{R}] \right\rfloor_p \\ &= \left\lfloor \mathsf{m}^{*\top}[\mathbf{A}|\mathbf{AR}] + [\mathbf{e}^\top|\mathbf{e}^\top\mathbf{R}] \right\rfloor_p\end{aligned}$$

where $\mathbf{R}$ is chosen as in the key generation phase. Since m^* is a sample from the distribution M which is chosen independent of $\mathbf{A}$ and $\mathbf{AR}$, so $\mathsf{m}^{*\top}[\mathbf{A}|\mathbf{AR}]$ is a random sample from the uniform distribution over $\mathbb{Z}_q^{\bar{m}+w}$ (Recall $\mathbf{A}$ is randomly chosen and $\mathbf{AR}$ statistically close to uniform as per Lemma 4). By Lemma 1 and the fact that $\|\mathbf{e}^\top\mathbf{R}\|_\infty \leq \beta q\sqrt{n\bar{m}}$, with all but negligible probability $\mathsf{negl}_3(\lambda)$,

$$\mathbf{c}^{*\top} = \left\lfloor \mathsf{m}^{*\top}[\mathbf{A}|\mathbf{AR}] \right\rfloor_p$$

as produced in $Game_2$. This shows that

$$|\Pr[Game_3 \Rightarrow 1] - \Pr[Game_2 \Rightarrow 1]| \leq \mathsf{negl}_3(\lambda).$$

In $Game_4$, we set $\mathsf{Ct}_0^* = (\mathbf{t}^* = f(\mathsf{m}^*),\ \mathbf{c}^* = \lfloor [\mathbf{b}^\top|\mathbf{b}^\top\mathbf{R}] \rfloor_p$ where $\mathbf{b} \leftarrow_\$ \mathbb{Z}_q^{\bar{m}}$ and $\mathsf{m}^* \leftarrow M$. by Lemma 5, the distributions $(\mathbf{A}, \mathbf{b}^\top, f(\mathsf{m}^*))$ and $(\mathbf{A}, \mathsf{m}^{*\top}\mathbf{A} + \mathbf{e}^\top, f(\mathsf{m}^*))$ are computationally indistinguishable under the $\mathsf{LWE}_{\ell,\bar{m},q,D_{\mathbb{Z},\gamma q}}$ assumption, where m^* is from an arbitrary k-source distribution over $\mathbb{Z}_q^n$, $\mathbf{e} \leftarrow D_{\mathbb{Z},\beta q}^{\bar{m}}$ and 2^{-k}-hard-to-invert function f, and $\mathbf{b} \leftarrow \mathbb{Z}_q^{\bar{m}}$. So the challenge ciphertext Ct_0^* in $Game_4$ is indistinguishable from

$$\begin{aligned}&\left(f(\mathsf{m}^*),\ \lfloor [\mathsf{m}^{*\top}\mathbf{A} + \mathbf{e}^\top|(\mathsf{m}^{*\top}\mathbf{A} + \mathbf{e}^\top)\mathbf{R}] \rfloor_p\right) \\ =&\left(f(\mathsf{m}^*),\ \lfloor \mathsf{m}^{*\top}[\mathbf{A}|\mathbf{AR}] + [\mathbf{e}^\top|\mathbf{e}^\top\mathbf{R}] \rfloor_p\right) \\ =&\left(f(\mathsf{m}^*),\ \lfloor \mathsf{m}^{*\top}[\mathbf{A}|\mathbf{A}_1 + \varphi(\mathbf{t}^*)\mathbf{G}] + [\mathbf{e}^\top|\mathbf{e}^\top\mathbf{R}] \rfloor_p\right)\end{aligned}$$

which is the challenge ciphertext Ct_0^* produced in $Game_3$. We have

$$\Pr[Game_4 \Rightarrow 1] - \Pr[Game_3 \Rightarrow 1] \leq \mathsf{Adv}_{\mathcal{B}}^{\mathsf{LWE}_{\ell,\bar{m},q,D_{\mathbb{Z},\gamma q}}}(\lambda) + \mathsf{negl}_3(\lambda)$$

for some LWE adversary $\mathcal{B}$.

Furthermore, since the challenge message m^*, a k-source sample, is independent of $\mathbf{c}^*$, $\mathbf{t}^* = f(\mathsf{m}^*)$ is distributed uniformly over $\{0,1\}^{2\lambda}$ except for the negligible probability $\lambda^{-\omega(1)}$ (by the fact that $k \geq 2\log(1/\lambda^{-\omega(1)}) + \lambda$, the universality of f, and Lemma 3). Since $\mathbf{b}$ is chosen uniformly at random from $\mathbb{Z}_q^{\bar{m}}$, by Lemma 4, $\mathbf{c}^* = \lfloor [\mathbf{b}^\top | \mathbf{b}^\top \mathbf{R}] \rceil_p$ is statistically close to the uniform distribution over $\mathbb{Z}_q^{\bar{m}+w}$ with up to a negligible distance $p/q = \mathsf{negl}_4(\lambda)$. This shows that Ct_0^* in $Game_5$ is statistically close to a random ciphertext, e.g., Ct_1^*. We have

$$|\Pr[Game_4 \Rightarrow 1] - 1/2| \leq \lambda^{-\omega(1)} + \mathsf{negl}_4(\lambda).$$

To sum up, we have

$$\mathsf{Adv}^{\mathsf{priv1\text{-}cca}}_{\Pi,\mathcal{A}}(\lambda) \leq \mathsf{Adv}^{\mathsf{LWE}_{\ell,\bar{m},q,D_{\mathbb{Z},\gamma q}}}_{\mathcal{B}}(\lambda) + \mathsf{negl}(\lambda)$$

where $\mathsf{negl}(\lambda)$ accounts for the sum of all negligible terms appeared in the proof. This completes the proof. □

References

1. Shweta Agrawal, Dan Boneh, and Xavier Boyen. Efficient lattice (H)IBE in the standard model. In Henri Gilbert, editor, *Advances in Cryptology - EUROCRYPT 2010*, volume 6110 of *Lecture Notes in Computer Science*, pages 553–572. Springer Berlin Heidelberg, 2010.
2. Joël Alwen, Stephan Krenn, Krzysztof Pietrzak, and Daniel Wichs. Learning with rounding, revisited. In *Advances in Cryptology–CRYPTO 2013*, pages 57–74. Springer, 2013.
3. Joël Alwen, Stephan Krenn, Krzysztof Pietrzak, and Daniel Wichs. Learning with rounding, revisited: New reduction, properties and applications. Cryptology ePrint Archive, Report 2013/098, 2013. `https://eprint.iacr.org/2013/098`.
4. Alexandra Boldyreva, Serge Fehr, and Adam O'Neill. On notions of security for deterministic encryption, and efficient constructions without random oracles. In *CRYPTO 2008*, pages 335–359. Springer, 2008.
5. Dan Boneh, Ran Canetti, Shai Halevi, and Jonathan Katz. Chosen-ciphertext security from identity-based encryption. *SIAM J. Comput.*, 36(5):1301–1328, December 2006.
6. Dan Boneh, Özgür Dagdelen, Marc Fischlin, Anja Lehmann, Christian Schaffner, and Mark Zhandry. Random oracles in a quantum world. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 41–69. Springer, 2011.
7. Xavier Boyen, Qixiang Mei, and Brent Waters. Direct chosen ciphertext security from identity-based techniques. In *Proceedings of the 12th ACM conference on Computer and communications security*, pages 320–329. ACM, 2005.
8. Ronald Cramer and Victor Shoup. Design and analysis of practical public-key encryption schemes secure against adaptive chosen ciphertext attack. *SIAM Journal on Computing*, 33(1):167–226, 2003.
9. Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM journal on computing*, 38(1):97–139, 2008.

10. Eiichiro Fujisaki and Tatsuaki Okamoto. Secure integration of asymmetric and symmetric encryption schemes. *Journal of cryptology*, 26(1):80–101, 2013.
11. Craig Gentry, Chris Peikert, and Vinod Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, STOC '08, pages 197–206, New York, NY, USA, 2008. ACM.
12. Shafi Goldwasser, Yael Kalai, Chris Peikert, and Vinod Vaikuntanathan. Robustness of the learning with errors assumption. *Innovations in Computer Science*, pages 230–240, 2010.
13. Eike Kiltz and David Galindo. Direct chosen-ciphertext secure identity-based key encapsulation without random oracles. *Theoretical Computer Science*, 410(47-49):5093–5111, 2009.
14. Junzuo Lai, Robert H Deng, Shengli Liu, and Weidong Kou. Efficient cca-secure pke from identity-based techniques. In *Cryptographers Track at the RSA Conference*, pages 132–147. Springer, 2010.
15. Daniele Micciancio and Chris Peikert. Trapdoors for lattices: Simpler, tighter, faster, smaller. In David Pointcheval and Thomas Johansson, editors, *Advances in Cryptology – EUROCRYPT 2012*, volume 7237 of *Lecture Notes in Computer Science*, pages 700–718. Springer Berlin Heidelberg, 2012.
16. Petros Mol and Scott Yilek. Chosen-ciphertext security from slightly lossy trapdoor functions. In *International Workshop on Public Key Cryptography*, pages 296–311. Springer, 2010.
17. Chris Peikert, Oded Regev, and Noah Stephens-Davidowitz. Pseudorandomness of ring-LWE for any ring and modulus. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 461–473. ACM, 2017.
18. Chris Peikert, Vinod Vaikuntanathan, and Brent Waters. A framework for efficient and composable oblivious transfer. In *Annual international cryptology conference*, pages 554–571. Springer, 2008.
19. Chris Peikert and Brent Waters. Lossy trapdoor functions and their applications. *SIAM Journal on Computing*, 40(6):1803–1844, 2011.
20. Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, STOC '05, pages 84–93, New York, NY, USA, 2005. ACM.
21. Alon Rosen and Gil Segev. Chosen-ciphertext security via correlated products. In *Theory of Cryptography Conference*, pages 419–436. Springer, 2009.
22. Xiang Xie, Rui Xue, and Rui Zhang. Deterministic public key encryption and identity-based encryption from lattices in the auxiliary-input setting. In *International Conference on Security and Cryptography for Networks*, pages 1–18. Springer, 2012.