数理解析研究所講究録
第 1554 巻 2007 年 40-47

A Lattice-Based Cryptosystem and Proof of Knowledge on Its Secret Key

草川 恵太* Keita Xagawa　　河内 亮周* Akinori Kawachi　　田中 圭介* Keisuke Tanaka

Abstract— We propose a lattice-based cryptosystem by modifying the Regev'05 cryptosystem (STOC 2005), and design a proof of secret-key knowledge. Lattice-based public-key identification schemes have already been proposed. However, it is not known that their public keys can be used for the public keys of encryption schemes. Our modification admits the proof of knowledge on its secret key, although we need a stronger assumption than that required by the original cryptosystem.

Keywords: lattice-based cryptosystems, proof of knowledge, secret keys.

1 Introduction

Lattice-Based Cryptosystems. Since Ajtai's seminal results on the average-case/worst-case connection of lattice problems [1], the lattice-based cryptosystems have been studied. Ajtai and Dwork proposed a public-key cryptosystem [4] based on the worst-case hardness of unique shortest vector problem (uSVP). After their results, Regev proposed a cryptosystem [21] based on the worst-case hardness of uSVP. In 2005, Regev introduced a cryptosystem R05 [22] based on the approximation version of SVP and Ajtai introduced another cryptosystem [3]. In the Regev'05 cryptosystem and the Ajtai05 cryptosystem, the size of the public key is $\tilde{O}(n^2)$ in the bare model and $\tilde{O}(n)$ or in the common reference string model. Their cryptosystems are more suitable for practical use than the Ajtai-Dwork cryptosystem.

However, there were no applications of lattice-based cryptosystems, except Micciancio and Vadhan [18] and Goldwasser and Kharchenko [11]. The former is a zero-knowledge proof for a gap version of closest vector problem (GapCVP_γ), which we refer as the MV protocol. The latter is a proof of plaintext knowledge for the Ajtai-Dwork cryptosystem. Thus, we consider another application for lattice-based cryptosystems, a proof of knowledge on its secret key.

Summary. We propose a modified Regev'05 cryptosystem and introduce a proof of knowledge on its secret key in the common reference string (CRS) model. We consider the relation between the private key and the public key as that between the message and the codeword with the error in coding theory. To construct a proof of knowledge, we modify generation of the error. This modification admits a prover to prove the knowledge of the error and the message based on Stern [24]. Thus, we obtain a proof of knowledge on a secret key of our cryptosystem.

Related Results. There already exist public-key identification schemes based on lattice and coding problems. In 1989, Shamir showed an identification scheme based on permuted kernel problem [23]. Stern proposed public-key identification based on syndrome decoding problem in 1996 [24]. Micciancio and Vadhan introduced a zero-knowledge proof with efficient prover for GapCVP_γ and discussed public-key identification schemes [18]. Recently, Hayashi and Tada showed public-key identification schemes based on binary non-negative exact length vector problem (or integer subset sum problem) [14]. Unfortunately, it is unknown whether their public keys can be used as a public key of cryptosystems or not. We stress that in our identification schemes, information for identification is indeed a public key of cryptosystems.

Why can we not apply the MV protocol to R05? Before description of our idea, we briefly review the key generation of R05 and explain why the same approach with the Micciancio-Vadhan protocol [18] fails for our goal. (We abbreviate it to "the MV protocol").

In R05, the secret key is $\mathbf{s} \in \mathbb{Z}_q^n$ and the public key is $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_m] \in \mathbb{Z}_q^{n \times m}$ and $\mathbf{b} = {}^t\mathbf{A}\mathbf{s} + \mathbf{e}$, where $\mathbf{e} \in \mathbb{Z}_q^m$ and each coordinate of $\mathbf{e}$ is close to 0. From a coding-theoretical view, we can regard ${}^t\mathbf{A}$ as a generator matrix, $\mathbf{s}$ as a message, and $\mathbf{e}$ as an error. Remark that the length of $\mathbf{e}$ is short. Hence, one would think we can apply the MV protocol to proofs of knowledge for a secret key $\mathbf{s}$. However, we cannot apply it in a naive way. We explain more details.

We first review the intuition which is used in the MV protocol. (See [18] for more details.) Let $(\mathbf{B}, \mathbf{y}, t)$

* Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology. W8-55, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan. {xagawa5, kawachi, keisuke}@is.titech.ac.jp. Supported in part by NTT Information Sharing Platform Laboratories and Grant-in-Aid for Scientific Research, Ministry of Education, Culture, Sports, Science, and Technology, 16092206.

be an instance of GapCVP_γ.[1] Let $B_m(\mathbf{c}, r)$ be an m-dimensional hyperball whose center is $\mathbf{c}$ and radius is r. In their protocol, the prover chooses a random bit c and a random vector $\mathbf{r}$ from $B_m(\mathbf{0}, \gamma t/2)$. The prover computes $\mathbf{m} = c\mathbf{y} + \mathbf{r} \bmod \mathbf{B}$ and sends $\mathbf{m}$ to the verifier. The verifier sends a challenge bit δ to the prover. Note that if $(\mathbf{B}, \mathbf{y}, t)$ is a YES instance then the ratio between the volume of $(B_m(\mathbf{0}, \gamma t/2) \bmod \mathbf{B}) \cap (B_m(\mathbf{y}, \gamma t/2) \bmod \mathbf{B})$ and that of $B(\mathbf{0}, \gamma t/2)$ is at least $1/\text{poly}(n)$. If $\mathbf{m} \in (B_m(\mathbf{0}, \gamma t/2) \bmod \mathbf{B}) \cap (B_m(\mathbf{y}, \gamma t/2) \bmod \mathbf{B})$ the prover can flip a bit c. The prover sends the proof that $\mathbf{m}$ is chosen from $B_m(c\mathbf{y}, \gamma t/2)$. Note that if $(\mathbf{B}, \mathbf{y}, t)$ is a NO instance then $(B_m(\mathbf{0}, \gamma t/2) \bmod \mathbf{B}) \cap (B_m(\mathbf{y}, \gamma t/2) \bmod \mathbf{B}) = \emptyset$. Therefore the prover can not flip a bit c after a reception of the challenge bit.

Next, we consider applying their protocol to the Regev'05 cryptosystem, i.e., a proof of knowledge that, on input $(\mathbf{A}, \mathbf{b})$, the prover knows $\mathbf{s}$ such that $\mathbf{b} = {}^t\mathbf{A}\mathbf{s} + \mathbf{e}$, where $\mathbf{e} \in B_m(\mathbf{0}, t)$.[2] Note that a linear code is $\mathbb{Z}_q$-module in $\mathbb{Z}_q^m$ and a lattice is $\mathbb{Z}$-module in $\mathbb{R}^m$. Therefore, instead of reducing modulo $\mathbf{B}$, we multiply a parity-check matrix $\mathbf{H}$ of ${}^t\mathbf{A}$ to the vector in $\mathbb{Z}_q^m$. Suppose that $B_m(\mathbf{0}, \gamma t/2)$ and $B_m(\mathbf{b}, \gamma t/2)$ do not intersect. Unfortunately, we cannot ensure that $\mathbf{H}B_m(\mathbf{0}, \gamma t/2)$ and $\mathbf{H}B_m(\mathbf{b}, \gamma t/2)$ do not intersect because the dimension of $\mathbf{H}\mathbb{Z}_q^m$ is $m - n < m$. On such NO instance $(\mathbf{A}, \mathbf{b})$, the prover can cheat the verifier on which hyperball he chose $\mathbf{m}$ from. Hence the soundness of the protocol fails. Thus, we cannot apply their protocol to the Regev'05 cryptosystem in a straightforward way.

Main Idea. As seen in the above paragraphs, we cannot apply the protocol [18] to the Regev'05 cryptosystem straightforwardly. Let us reconsider multiplying a parity-check matrix $\mathbf{H}$. Let $\mathbf{s} \in \mathbb{Z}_q^n$ be a private key and let $(\mathbf{A}, \mathbf{b})$ be a public key, where $\mathbf{b} = {}^t\mathbf{A}\mathbf{s} + \mathbf{e}$. Multiplying a parity-check matrix $\mathbf{H}$ to the equation $\mathbf{b} = {}^t\mathbf{A}\mathbf{s} + \mathbf{e}$, we obtain that $\mathbf{H}\mathbf{b} = \mathbf{H}\mathbf{e}$. The prover should prove the knowledge of $\mathbf{e}$ that satisfies the equation and each coordinate of $\mathbf{e}$ is in certain range. The difficulty to construct the protocol is to combine protocols that prove sufficiency of the equation and lying in the range.

Then, we modify a public key as follows: The secret key is $\mathbf{s} \in \mathbb{Z}_q^n$ and $\mathbf{s}' \in \{0,1\}^{m_1}$, whose Hamming weight is m_2. The public key is $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ and $\mathbf{E} \in \mathbb{Z}_q^{m \times m_1}$ and $\mathbf{b} = {}^t\mathbf{A}\mathbf{s} + \mathbf{E}\mathbf{s}'$. In this case, by multiplying a parity-check matrix $\mathbf{H}$, we have that $\mathbf{H}\mathbf{b} = \mathbf{H}\mathbf{E}\mathbf{s}'$. Translating a matrix $\mathbf{HE}$ as a parity-check matrix, we have an instance $(\mathbf{HE}, \mathbf{Hb}, m_2)$ and a witness $\mathbf{s}'$ of Syndrome Decoding Problem (SDP).[3] Since Stern proposed a proof of knowledge for SDP in 1996 [24], we adopt it to prove knowledge of secret key $\mathbf{s}'$.

The proof of knowledge for SDP needs a statistically-hiding and computationally-binding commitment scheme. Fortunately, if $\mathbf{A}$ is chosen randomly then the function $f_\mathbf{A} : \{0,1\}^m \to \mathbb{Z}_q^n : \mathbf{m} \mapsto \mathbf{Am}$ is a collision-resistant function based on the approximation version of SVP [2, 10, 7, 15, 17]. Thus we employ that function to develop a statistically-hiding and computationally-binding string commitment scheme. Our construction of a string commitment is more straightforward than Damgård, Pedersen, and Pfizmann [8, 9] and Halevi and Micali [13], which used the universal hash functions.

We also show the security of the modified R05, mR05. Unfortunately, we need a stronger assumption than the original one. The stronger assumption is the worst-case hardness of certain learning problem, which is based on well-known problem Learning With Error (LWE).

Organization. The rest of this chapter is organized as follows. We briefly note basic notions and notations in Section 2. We describe the Regev'05 cryptosystem and our modified cryptosystem in Section 3. Next, we give our main results, a proof of knowledge on a secret key, in Section 4. Finally, we conclude in Section 5.

2 Preliminaries

Let $w_H(\mathbf{x})$ denote Hamming weight of $\mathbf{x}$, i.e., the number of nonzero elements in $\mathbf{x}$. For an element $x \in \mathbb{Z}_q$ we define $|x|_q$ as the integer x if $x \in \{0, 1, \dots, \lfloor q/2 \rfloor\}$ and as the integer $q - x$ otherwise. In other words, $|x|_q$ represents the distance of x from 0 in $\mathbb{Z}_q$.

Gaussian and other distributions. The normal distribution with mean 0 and variance σ^2 is the distribution on $\mathbb{R}$ given by the density function $(1/\sqrt{2\pi}\sigma)\exp(-(x/\sigma)^2/2)$. For any distribution ϕ, we consider the distribution $\phi^{(n)}$ obtained as follows: (1) take n samples $x_1, \dots, x_n$ from ϕ independently and (2) output ${}^t(x_1, \dots, x_n)$. For a n-dimensional vector $\mathbf{x}$ and any $s > 0$, let $\rho_s^{(n)}(\mathbf{x}) = \exp(-\pi\|\mathbf{x}/s\|^2)$ be a Gaussian function scaled by a factor of s. Also, $\nu_s^{(n)} := \rho_s^{(n)}/s^n$ is an n-dimensional probability density function. For $\alpha \in \mathbb{R}^+$ the distribution Ψ_α is the distribution on $[0, 1)$ obtained by sampling from a normal variable with mean 0 and variance $\alpha^2/(2\pi)$ and reducing the result modulo 1: $\Psi_\alpha(r) := \sum_{k\in\mathbb{Z}}(1/\alpha)\exp(-\pi((r-k)/\alpha)^2)$. For an arbitrary probability distribution with density function $\phi : \mathbb{T} \to \mathbb{R}^+$ and some integer $q > 0$, we define its discretization $\bar{\phi} : \mathbb{Z}_q \to \mathbb{R}^+$ as the discrete

---

[1] $(\mathbf{B}, \mathbf{y}, t)$ is a YES instance if there exists $\mathbf{w} \in \mathbb{Z}^n$ such that $\|\mathbf{Bw} - \mathbf{y}\| \le t$. It is a NO instance if for any vector $\mathbf{w} \in \mathbb{Z}^n$, $\|\mathbf{Bw} - \mathbf{y}\| \ge \gamma t$. Although they consider only full-rank lattices in [18], we consider not only full-rank lattices. That is, an instance of GapCVP_γ consists of $\mathbf{B}$, which is a basis of a lattice whose rank is n, $\mathbf{y} \in \mathbb{R}^m$, $\gamma \ge 1$.

[2] We abuse the notation $B_m(\cdot, \cdot)$.

[3] Syndrome Decoding Problem: Given input $(\mathbf{H}, \mathbf{y}, m)$, where $\mathbf{H} \in \mathbb{Z}_2^{(n-k)\times n}$, $\mathbf{y} \in \mathbb{Z}_2^{n-k}$, $m \ge 0$, find $\mathbf{x} \in \mathbb{Z}_2^n$ such that $\mathbf{Hx} = \mathbf{y}$ and Hamming weight of $\mathbf{x}$ is exactly m.

probability distribution obtained by sampling from ϕ, multiplying by q, and rounding to the closest integer modulo q. More formally, $\bar{\phi}(i) := \int_{(i-1/2)q}^{(i+1/2)q} \phi(x)dx$.

For integers $m_1 \geq m_2 \geq 0$, we define $\mathrm{Set}_{m_2} := \{\mathbf{s}' \in \{0,1\}^{m_1} \mid w_H(\mathbf{s}') = m_2\}$. For any $\mathbf{s} \in \mathbf{Z}_q^m$, we define $A_{\mathbf{s}}$ obtained as follows: (1) Choose a random vector $\mathbf{a} \in \mathbf{Z}_q^m$. (2) Choose a random element $e \in \mathbf{Z}_q$ according to $\bar{\Psi}_\alpha$. (3) Outputs $(\mathbf{a}, \langle \mathbf{a}, \mathbf{s}\rangle + e)$. For any $\mathbf{s} \in \mathbf{Z}_q^m$ and any $\mathbf{s}' \in \mathrm{Set}_{m_2}$, we define $A_{\mathbf{s},\mathbf{s}'}$ as the distribution on $\mathbf{Z}_q^n \times \mathbf{Z}_q^{m_1} \times \mathbf{Z}_q$ obtained as follows: (1) Choose a random vector $\mathbf{a} \in \mathbf{Z}_q^m$. (2) Choose a random vector $\mathbf{e} \in \mathbf{Z}_q^{m_1}$ according to $\bar{\Psi}_{\alpha/m_2}^{(m_1)}$. (3) Set $b := \langle \mathbf{a}, \mathbf{s}\rangle + \langle \mathbf{e}, \mathbf{s}'\rangle$ and output $(\mathbf{a}, \mathbf{e}, b)$. We also define U' as the distribution on $\mathbf{Z}_q^n \times \mathbf{Z}_q^{m_1} \times \mathbf{Z}_q$ obtained as follows: (1) Choose a random vector $\mathbf{a} \in \mathbf{Z}_q^m$. (2) Choose a random vector $\mathbf{e} \in \mathbf{Z}_q^{m_1}$ according to $\bar{\Psi}_{\alpha/m_2}^{(m_1)}$. (3) Choose a random elements $u \in \mathbf{Z}_q$ and output $(\mathbf{a}, \mathbf{e}, u)$.

We consider the following learning problems.

Definition 2.1 (Learning With Errors, $\mathrm{LWE}_{q,\bar{\Psi}_\alpha}$). Given samples from $A_{\mathbf{s}}$, find $\mathbf{s}$.

Definition 2.2 (Learning With Known Errors, $\mathrm{LWKE}_{q,\bar{\Psi}_\alpha}$). Given parameters, m_1 and m_2, and samples from $A_{\mathbf{s},\mathbf{s}'}$, find $\mathbf{s}$.

We note that if there exists an adversary $\mathcal{A}$ that solves $\mathrm{LWE}_{q,\bar{\Psi}_\alpha}$ with non-negligible probability then there exists an adversary $\mathcal{A}'$ that solves $\mathrm{LWKE}_{q,\bar{\Psi}_\alpha}$ with non-negligible probability. If $\mathcal{A}$ needs $k = \mathrm{poly}(n)$ samples, then $\mathcal{A}'$ takes k samples $(\mathbf{a}_i, \mathbf{e}_i, b_i)$ from $A_{\mathbf{s},\mathbf{s}'}$. $\mathcal{A}'$ inputs $\{(\mathbf{a}_i, b_i)\}_{i=1,\dots,k}$ to $\mathcal{A}$ and obtains an output $\mathbf{s}$. $\mathcal{A}'$ outputs $\mathbf{s}$. Using the reproducibility of Gaussian distributions, we show that the sum of m_2 samples according to $\bar{\Psi}_{\alpha/m_2}$ is, in fact, distributed according to $\bar{\Psi}_\alpha$, and hence $\{(\mathbf{a}_i, b_i)\}_{i=1,\dots,k}$ which $\mathcal{A}'$ computes is indeed samples from $A_{\mathbf{s}}$.

Given two probability density functions ϕ_1, ϕ_2 on $\mathbf{R}^n$, we define the statistical distance between them as $\Delta(\phi_1, \phi_2) := \frac{1}{2}\int_{\mathbf{R}^n} |\phi_1(\mathbf{x}) - \phi_2(\mathbf{x})| d\mathbf{x}$. A similar definition holds for discrete random variables. We sometimes abuse such notation, and use the same notation for two arbitrary functions. Note that the acceptance probability of any algorithm on inputs from X differs from its acceptance probability on inputs from Y by at most $\Delta(X, Y)$.

We say that an algorithm $\mathcal{D}$ with oracle access is a distinguisher between two distributions if its acceptance probability when the oracle outputs samples of the first distribution and when the oracle outputs samples of the second distribution differ by a non-negligible amount.

Lattices. An n-dimensional lattice in $\mathbf{R}^n$ is the set $L(\mathbf{b}_1, \dots, \mathbf{b}_n) = \{\sum_{i=1}^n \alpha_i \mathbf{b}_i \mid \alpha_i \in \mathbf{Z}\}$ of all integral combinations of n linearly independent vectors $\mathbf{b}_1, \dots, \mathbf{b}_n$. The sequence of vectors $\mathbf{b}_1, \dots, \mathbf{b}_n$ is called a *basis* of the lattice L. For more details on lattices, see the textbook by Micciancio and Goldwasser [16].

We give well-known lattice problems, Shortest Vector Problem (SVP) and Shortest Independent Vector Problem (SIVP) and their approximation version.

Definition 2.3 (Shortest Vector Problem, SVP). Given a basis $\mathbf{B}$ of a lattice L, find a non-zero vector $\mathbf{v} \in L$ such that for any non-zero vector $\mathbf{x} \in L$, $\|\mathbf{v}\| \leq \|\mathbf{x}\|$.

Definition 2.4 (SVP_γ). Given a basis $\mathbf{B}$ of a lattice L, find a non-zero vector $\mathbf{v} \in L$ such that for any non-zero vector $\mathbf{x} \in L$, $\|\mathbf{v}\| \leq \gamma\|\mathbf{x}\|$.

Definition 2.5 (Shortest Independent Vector Problem, SIVP). Given a basis $\mathbf{B}$ of a lattice L, find a sequence of n linearly independent vectors $\mathbf{v}_1, \dots, \mathbf{v}_n \in L$ such that for any sequence of n linearly independent vectors $\mathbf{x}_1, \dots, \mathbf{x}_n \in L$, $\max_i \|\mathbf{v}_i\| \leq \max_i \|\mathbf{x}_i\|$.

Definition 2.6 (SIVP_γ). Given a basis $\mathbf{B}$ of a lattice L, find a sequence of n linearly independent vectors $\mathbf{v}_1, \dots, \mathbf{v}_n \in L$ such that for any sequence of n linearly independent vectors $\mathbf{x}_1, \dots, \mathbf{x}_n \in L$, $\max_i \|\mathbf{v}_i\| \leq \gamma \max_i \|\mathbf{x}_i\|$.

Codes. Let $\mathbf{F}_q$ denote a field with q elements, where q is a prime power. A q-ary linear code C is a linear subspace of $\mathbf{F}_q^n$. If C has dimension k then C is called an $[n,k]_q$ code. A generator matrix $\mathbf{G}$ for a linear code C is a n by k matrix for which the columns are a basis of C. Note that $C := \{\mathbf{Gm} \mid \mathbf{m} \in \mathbf{F}_q^k\}$. We say that $\mathbf{G}$ is in standard form if $\mathbf{G} = \binom{\mathbf{I}_k}{\mathbf{P}}$. For an $[n,k]_q$ code C, we define the dual code $C^\perp$ by $C^\perp := \{\mathbf{y} \in \mathbf{F}_q^n \mid \text{for any } \mathbf{x} \in C, \langle \mathbf{x}, \mathbf{y}\rangle = 0\}$. If $\mathbf{G} = \binom{\mathbf{I}_k}{\mathbf{P}}$ is a generator matrix in standard form of the code C, then $\mathbf{H} = \binom{-{}^t\mathbf{P}}{\mathbf{I}_{n-k}}$ is a generator matrix of the code $C^\perp$. This follows from the fact that $\mathbf{H}$ has the right size and rank and that ${}^t\mathbf{H}\mathbf{G} = \mathbf{0}$, which implies every codeword $\mathbf{Gm}$ has inner product 0 with every column of $\mathbf{H}$. In other words, $\mathbf{x} \in C$ if and only if ${}^t\mathbf{H}\mathbf{x} = \mathbf{0}$. Thus, we call $\mathbf{H}$ a parity-check matrix. We note that, given any generator matrix $\mathbf{G}$ of the code C, we can efficiently compute C's generator matrix $\mathbf{G}'$ in standard form and C's parity-check matrix $\mathbf{H}$.

If C is a linear code with a parity-check matrix $\mathbf{H}$ then for every $\mathbf{x} \in \mathbf{F}_q^n$ we call ${}^t\mathbf{H}\mathbf{x}$ the syndrome of $\mathbf{x}$.

It is well known that the question of finding the nearest codeword to a vector (Nearest Codeword Problem, NCP) is NP-hard even in approximation version [5]. It is also difficult to find a word of a given weight from its syndrome [6].

Definition 2.7 (Symdrome Decoding Problem, SDP). Given a parity-check matrix $\mathbf{H} \in \mathbf{Z}_2^{n \times m}$, a binary nonzero vector $\mathbf{y} \in \mathbf{Z}_2^m$, and a positive integer w, find a binary vector $\mathbf{x} \in \mathbf{Z}_2^n$ with no more than w 1's such that ${}^t\mathbf{H}\mathbf{x} = \mathbf{y}$.

Zero Knowledge and Proof of Knowledge. In this section, we recall definitions and notations of zero knowledge and proof of knowledge.

Definition 2.8 (Auxiliary-Input Zero Knowledge). An interactive proof system (P, V) for a language L is (perfect/statistical/computational) *auxiliary-input zero knowledge* if for every probabilistic polynomial-time machine V^* and polynomial $p(\cdot)$, there exists a probabilistic polynomial-time machine S such that the ensembles $\{(P, V^*(z))(x)\}$ and $\{S(x, z)\}$ are (perfectly/statistically/computationally) indistinguishable on the set $\{(x, z) : x \in L, |z| = p(|x|)\}$.

For a relation $R \subseteq \{0,1\}^* \times \{0,1\}^*$ and $x \in \{0,1\}^*$, we define a set of witness $R(x) := \{y \mid (x, y) \in R\}$.

Definition 2.9 (Proof of Knowlegde). Let $\eta \in (0, 1)$. An interactive protocol (P, V) with a prover P and a verifier V is a *proof of knowledge system with knowledge error κ for a relation R* if the following holds:

Completeness: For every common input x for which there exists y such that $(x, y) \in R$ the verifier V always accepts interacting with the prover P.

Validity with error η: There exists a polynomial-time interacting oracle Turing machine K and a constant $c > 0$ such that for every $x \in \{0,1\}^*$ such that $R(x) \neq \emptyset$ and for every prover P^* the following holds: $K^{P^*}(x) \in R(x) \cup \{\perp\}$ and $\Pr[K^{P^*}(x) \in R(x)] \geq (p - \kappa)^c$, where $p > \kappa$ is the probability that V accepts while interacting with P^* on common input x.

2.1 String Commitments

We explain the notation for commitment schemes in the common reference string (CRS) model. Assume that there exists a trusted third party (TTP). Let $\mathrm{Com}_{(\cdot)}(\cdot;\cdot)$ be an indexed function which maps a pair of a message string and a random string to a commitment string. First, TTP on input 1^n outputs a random string a, which is the CRS and the index of the commitment function. To commit to a string s, the sender chooses a random string r, computes $c = \mathrm{Com}_a(s; r)$, and sends c to the receiver. To reveal commitment c, the sender sends s and r to the receiver. The receiver accepts if $c = \mathrm{Com}_a(s; r)$ or rejects otherwise.

Definition 2.10. We say a string commitment scheme $\mathrm{Com}_{(\cdot)}(\cdot;\cdot)$ is statistically hiding and computationally binding if it has the following properties:

Statistical Hiding: For any two strings s and s', the statistical distance between $(a, \mathrm{Com}_a(s; r))$ and $(a, \mathrm{Com}_a(s'; r'))$ is negligible, where a, r, r' are random and independent.

Computational Binding: For any probabilistic polynomial-time machine $\mathcal{A}$, if a is randomly chosen by TTP, then the probability that, given an input a, $\mathcal{A}$ outputs (s, r) and (s', r') such that $\mathrm{Com}_a(s; r) = \mathrm{Com}_a(s'; r')$ is negligible.

2.2 Subset-Sum Hash Functions and a String Commitment Scheme

As explained in Section 1 we need a string commitment scheme to construct a proof of knowledge of a secret key. We first argue the family of subset-sum hash functions and the string commitment scheme.

Let n be a security parameter (or a dimension of underlying lattice problems). For a prime $q = q(n) = n^{O(1)}$ and an integer $m = m(n) > n \log q(n)$, we define a family of hash functions, $\mathcal{H}_{q,m} = \{f_{\mathbf{A}} : \{0,1\}^{m(n)} \to \mathbb{Z}_{q(n)}^n \mid \mathbf{A} \in \mathbb{Z}_{q(n)}^{n \times m(n)}\}$, where $f_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x} \bmod q(n)$.

Originally, Ajtai [1] showed $\mathcal{H}_{q,m}$ is a family of one-way functions under the assumption that SVP with some polynomial approximation factor is hard in the worst case for suitably chosen $q(n)$ and $m(n)$. It is known that $\mathcal{H}_{q,m}$ is indeed a family of collision-resistant hash functions for suitably chosen q and m by Goldreich, Goldwasser, and Halevi [10], Cai and Nerurkar [7] and Micciancio [15]. Recently, Micciancio and Regev showed $\mathcal{H}_{q,m}$ is a family of collision-resistant hash functions under the assumption $\mathrm{SVP}_{\tilde{O}(n)}$ is hard in the worst case [17].

We construct a statistically-hiding and computationally-binding string commitment scheme based on the above hash functions. It is well known that if there exists a collision-resistant hash function then there exists a statistically hiding and computationally binding string commitment scheme [8, 9, 13]. Their construction used universal hash functions for the statistically hiding property. However, our construction do not use it, because if m is sufficiently large and a plaintext $\mathbf{s}$ is randomized, $\mathbf{A}\mathbf{s}$ is distributed statistically close to the uniform distribution. To prove the statistically-hiding property, we use Claim 2.12 below in [22].

We describe how to achieve a string commitment scheme in the CRS model. We first split the domain $\{0,1\}^m$ into two domain $\{0,1\}^{m/2} \times \{0,1\}^{m/2}$. The first domain is used for randomization. The second domain is for message. We define $\mathrm{Com}_{\mathbf{A}}(s; r) := \mathbf{A}\mathbf{x}$, where $\mathbf{x} = {}^t(r_0, \ldots, r_{m/2}, s_1, \ldots, s_{m/2})$, $r = r_1 \ldots r_{m/2}$, and $s = s_1 \ldots s_{m/2}$.

Lemma 2.11. *For a prime $q = q(n) = n^{O(1)}$ and an integer $m = m(n) > 10n \log q$, if $\mathcal{H}_{q,m}$ is collision resistant and a trusted third party gives a random matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$, then $\mathrm{Com}_{\mathbf{A}}$ is a statistically hiding and computationally binding string commitment scheme in the CRS model.*

Proof. The computationally-binding property immediately follows from the collision-resistant property.

Next, we consider the statistically-hiding property. Using Claim 2.12 below, we have that with probability exponentially close to 1 the statistical distance between the distribution of $(\mathbf{A}, \mathrm{Com}_{\mathbf{A}}(0^{m/2}; r))$ and that of $(\mathbf{A}, \mathbf{u})$ is negligible in n, where r and $\mathbf{u}$ are random variables according to the uniform distribution on $\{0,1\}^{m/2}$ and $\mathbf{Z}_q^n$, respectively. Hence, for any two messages $m_1, m_2 \in \{0,1\}^{m/2}$, the statistical distance between the distribution of $(\mathbf{A}, \mathrm{Com}_{\mathbf{A}}(m_1; r_1))$ and that of $(\mathbf{A}, \mathrm{Com}_{\mathbf{A}}(m_2; r_2))$ is negligible in n with probability exponentially close to 1, where r_1 and r_2 are random variables according to the uniform distribution on $\{0,1\}^{m/2}$. This completes the proof. □

Claim 2.12 (Claim 5.3, [22]). *Let G be a finite Abelian group and let $l = c \log |G|$. For $c \geq 5$, when choosing l elements $g_1, \ldots, g_l$ uniformly from G the probability that the statistical distance between the uniform distribution on G and the distribution given by the sums of random subsets of $g_1, \ldots, g_l$ is more than $2/|G|$ is at most $1/|G|$.*

3 The Regev'05 Cryptosystem and Modified Regev'05 Cryptosystem

3.1 The Regev'05 Cryptosystem

Regev proposed a lattice-based cryptosystem in 2005 [22]. We briefly review the Regev'05 cryptosystem, R05.

Cryptosystem 3.1 (R05, [22]). Let n be a security parameter (or a dimension of the underlying lattice problem). Let q be a prime and α be a parameter to define the variance of Gaussian distribution such that $\alpha q > 2\sqrt{n}$. Let m be an integer at least $5(n+1)\log q$.

Private Key: Choose $\mathbf{s} \in \mathbf{Z}_q^n$ randomly.

Public Key: Choose $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbf{Z}_q^n$ randomly. Choose $e_1, \ldots, e_m$ according to the distribution $\bar{\Psi}_\alpha$. Compute $b_i = \langle \mathbf{a}_i, \mathbf{s} \rangle + e_i \bmod q$. The public key is $\{(\mathbf{a}_i, b_i)\}_{i=1,\ldots,m}$.

Encryption: A plaintext is $\sigma \in \{0,1\}$. Choose $S \subseteq_R \{1, \ldots, m\}$ randomly. The ciphertext is $(\sum_{i \in S} \mathbf{a}_i, \sigma \lfloor q/2 \rfloor + \sum_{i \in S} b_i)$.

Decryption: Let $(\mathbf{a}, b) \in \mathbf{Z}_q^n \times \mathbf{Z}_q$ be a received ciphertext. If $|b - \langle \mathbf{a}, \mathbf{s} \rangle|_q \leq q/4$ then decrypt to 0. Otherwise decrypt to 1.

The size of a public key and a private key are $O(mn \log q) = O(n^2 \log^2 q)$ and $O(n \log q) = O(n \log n)$ respectively. If $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is the CRS, this is the idea from Ajtai [3], the size of a public key is $O(m \log q) = O(n \log^2 q)$. We summarize the security and decryption errors of R05.

Theorem 3.2 (Thereom 3.1, Lemma 4.4, and Lemma 5.4, [22]). *Let $\alpha = \alpha(n)$ be a real number on $(0,1)$ and $q = q(n)$ a prime such that $\alpha q > 2\sqrt{n}$. For $m \geq 5(n+1)\log q$, if there exists a polynomial time algorithm that distinguishes between encryptions of 0 and 1 then there exists a distinguisher that distinguishes between $A_{\mathbf{s}}$ and $U(\mathbf{Z}_q^n \times \mathbf{Z}_q)$ for a non-negligible fraction of all possible $\mathbf{s}$.*

Next, assume there exists a distinguisher that distinguishes $A_{\mathbf{s}}$ from $U(\mathbf{Z}_q^n \times \mathbf{Z}_q)$ for a non-negligible fraction of all possible $\mathbf{s}$. Then, there exists an efficient algorithm that solves $\mathrm{LWE}_{q,\bar{\Psi}_\alpha}$.

Finally, assume there exists an efficient (possibly quantum) algorithm that solves $\mathrm{LWE}_{q,\bar{\Psi}_\alpha}$. Then there exists an efficient quantum algorithm for solving the worst-case of $\mathrm{SVP}_{\tilde{O}(n/\alpha)}$ and $\mathrm{SIVP}_{\tilde{O}(n/\alpha)}$.

Lemma 3.3 (Lemma 5.1, [22] (Correctness)). *The decryption error probability is at most $2^{-\Omega(1/(m\alpha^2))} + 2^{-\Omega(n)}$.*

Remark 3.4. The reduction in Theorem 3.2 is quantum. Therefore, the security of R05 depends on the worst-case hardness of $\mathrm{LWE}_{q,\bar{\Psi}_\alpha}$ in the classical sense.

3.2 Modified Regev'05 Cryptosystem

We modify the Regev'05 cryptosystem to obtain a new cryptosystem mR05.

Cryptosystem 3.5 (mR05). Let n be a security parameter (or a dimension of the underlying lattice problem). Let q be a prime and α be a parameter to define the variance of Gaussian distribution such that $\alpha q > 2\sqrt{n}$. Let t_α be a threshold such that $\Pr_{e \sim \bar{\Psi}_{\alpha/m_2}}[|e|_q \geq t_\alpha]$ is negligible in n (i.e., $t_\alpha = \omega(\alpha q \log n / m_2)$.) Let m be an integer at least $10(n+1)\log q$. Let m_1 and m_2 be integers such that $m_1, m_2 = \mathrm{poly}(n)$ and $\binom{m_1}{m_2}$ is exponential in n. Let $\mathrm{Set}_{m_1,m_2} := \{\mathbf{s}' \in \{0,1\}^{m_1} \mid w_H(\mathbf{s}') = m_2\}$. We need $4mm_2t_\alpha < q$ to ensure the correctness of the cryptosystem.

Private Key: Choose $\mathbf{s} \in \mathbf{Z}_q^n$ randomly. Choose $\mathbf{s}' \in \mathrm{Set}_{m_1,m_2}$ randomly.

Public Key: Choose $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbf{Z}_q^n$ randomly and $\mathbf{e}_1, \ldots, \mathbf{e}_{m_1}$ according to the distribution $\bar{\Psi}_{\alpha/m_2}^{(m)}$. Let $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_m]$ and $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_{m_1}]$. Check for any i, $\mathbf{e}_i$'s coordinates are at most t_α in the sense of $|\cdot|_q$. Compute $\mathbf{e} := \mathbf{E}\mathbf{s}'$. Let $\mathbf{b} := {}^t\mathbf{A}\mathbf{s} + \mathbf{e} \in \mathbf{Z}_q^m$. The public key is $(\mathbf{A}, \mathbf{E}, \mathbf{b})$. The secret key is $\mathbf{s}, \mathbf{s}'$.

Encryption: A plaintext is $\sigma \in \{0,1\}$. Choose $S \subseteq_R \{1, \ldots, m\}$ randomly. The ciphertext is $(\sum_{i \in S} \mathbf{a}_i, \sigma \lfloor q/2 \rfloor + \sum_{i \in S} b_i)$.

Decryption: Let $(\mathbf{a}, b) \in \mathbf{Z}_q^n \times \mathbf{Z}_q$ be a received ciphertext. If $|b - \langle \mathbf{a}, \mathbf{s} \rangle|_q \leq q/4$ then decrypt to 0. Otherwise decrypt to 1.

For example, we set $q = \Theta(n^3)$, $m = 10(n+1)\log q$, $\alpha = 1/m^2$, $t_\alpha = n/\log n$, $m_1 = m$, and $m_2 = \sqrt{m}$. Note that, with such parameters, we have that $4mm_2t_\alpha < q$.

The size of a public key and a private key are $O(mn\log q + m_1 n \log q) = O(n^2 \log^2 q)$ and $O(n\log q + m_1 \log q) = O(n \log^2 n)$ respectively. If $\mathbf{A}$ and $\mathbf{E}$ are the CRSs the size of a public key is $O(m\log q) = O(n\log^2 q)$. Note that, from a coding-theoretical view, ${}^t\mathbf{A}$ is a generator matrix and we can compute a parity check matrix $\mathbf{H}$ such that, for any $\mathbf{s} \in \mathbb{Z}_q^n$, $\mathbf{H}^t\mathbf{A}\mathbf{s} = \mathbf{0} \in \mathbb{Z}_q^{m-n}$.

First, we see the correctness of mR05.

Lemma 3.6 (Correctness). *There exist no decryption errors in* mR05.

Proof. Suppose that $(\mathbf{a}, b)$ is a valid ciphertexts of 0, i.e., $(\mathbf{a}, \mathbf{b}) = (\sum_{i=1}^m r_i\mathbf{a}_i, \sum_{i=1}^m r_i b_i)$ for some $r \in \{0,1\}^m$. We have

$$|b - \langle \mathbf{a}, \mathbf{s}\rangle|_q = \left|\sum_{i=1}^m r_i b_i - \langle \sum_{i=1}^m r_i \mathbf{a}_i, \mathbf{s}\rangle\right|_q$$
$$= \left|\sum_{i=1}^m r_i e_i\right|_q \le \left|\sum_{i=1}^m e_i\right|_q \le m|e_i|_q \le mm_2 t_\alpha,$$

where e_i is i-th coordinate of $\mathbf{e} = \mathbf{E}\mathbf{s}'$. Since we set $4mm_2t_\alpha < q$, we obtain $|b - \langle \mathbf{a}, \mathbf{s}\rangle|_q < q/4$. Next we consider the case $(\mathbf{a}, b)$ is a valid ciphertexts of 1, i.e., $(\mathbf{a}, \mathbf{b}) = (\sum_{i=1}^m r_i\mathbf{a}_i, \lfloor q/2 \rfloor + \sum_{i=1}^m r_i b_i)$ for some $r \in \{0,1\}^m$. Similarly to the case of 0, we have

$$|b - \langle \mathbf{a}, \mathbf{s}\rangle|_q \ge \lfloor q/2 \rfloor - mm_2 t_\alpha \ge q/4$$

and we can decrypt correctly. □

Combining Lemma 3.8, 3.9, and 3.10 below, we obtain the following theorem on security of mR05.

Theorem 3.7 (Security). *For $m \ge 10(n+1)\log q$, if there exists a polynomial-time algorithm $\mathcal{D}$ that distinguishes between encryptions of* 0 *and* 1 *with its public key, then there exists a polynomial-time algorithm $\mathcal{A}$ that solves* $\mathrm{LWKE}_{q,\bar{\Psi}_\alpha}$ *in the worst case.*

Lemma 3.8. *For $m \ge 5(n+1)\log q$, if there exists a polynomial time algorithm $\mathcal{D}$ that distinguishes between encryptions of* 0 *and* 1 *with its public key, then there exists a distinguisher $\mathcal{D}'$ that distinguishes between $A_{\mathbf{s},\mathbf{s}'}$ and U' for a non-negligible fraction of all possible* $\mathbf{s}$ *and* $\mathbf{s}'$.

We omit the proof, because the proof is quite similar to the security proof in [22].

Lemma 3.9 (Average-case to Worst-case). *Assume there exists a distinguisher $\mathcal{D}$ that distinguishes $A_{\mathbf{s},\mathbf{s}'}$ from U' for a non-negligible fraction of all possible* $\mathbf{s}$ *and* $\mathbf{s}'$. *Then there exists an algorithm $\mathcal{D}'$ that for all* $\mathbf{s}$ *and* $\mathbf{s}'$ *accepts with probability exponentially close to* 1 *on inputs from $A_{\mathbf{s},\mathbf{s}'}$ and rejects with probability exponentially close to* 1 *on inputs from U'.*

Proof. As similar to Regev's proof [22], we prove the lemma based on the following transformation. For any $\mathbf{t} \in \mathbb{Z}_q^n$ and any permutation $\pi \in S_{m_1}$ consider the function $f_{\mathbf{t},\pi} : \mathbb{Z}_q^n \times \mathbb{Z}_q^{m_1} \times \mathbb{Z}_q \to \mathbb{Z}_q^n \times \mathbb{Z}_q^{m_1} \times \mathbb{Z}_q$ defined by

$$f_{\mathbf{t},\pi}(\mathbf{a}, \mathbf{e}, b) = (\mathbf{a}, \pi(\mathbf{e}), b + \langle \mathbf{a}, \mathbf{t}\rangle).$$

This function transforms the distribution $A_{\mathbf{s},\mathbf{s}'}$ into $A_{\mathbf{s}+\mathbf{t},\pi(\mathbf{s}')}$. Moreover, it transforms the distribution U' into itself.

Next we consider a random statistical test. Assume that for n^{-c_1} fraction of all possible $(\mathbf{s}, \mathbf{s}')$, the acceptance probability of W on inputs from $A_{\mathbf{s},\mathbf{s}'}$ and on inputs from U' differ by at least n^{-c_2}. We construct the distinguisher $\mathcal{D}'$ as follows. Let R denote the unknown input distribution. (0) Repeat the following n^{c_1+1} times. (1) Choose a vector $\mathbf{t} \in \mathbb{Z}_q^n$ and a permutation $\pi \in S_{m_1}$ uniformly at random. (2) Estimate p_R, the acceptance probability of $\mathcal{D}$ on $f_{\mathbf{t},\pi}(R)$, by calling $\mathcal{D}$ $T = n^{2c_2+1}$ times. Let x_R be the number of 1 in the outputs of $\mathcal{D}$. (3) Estimate p_U, the acceptance probability of $\mathcal{D}$ on U', by calling $\mathcal{D}$ T times. Let x_U be the number of 1 in the outputs of $\mathcal{D}$. (4) If $|x_U - x_R|/T \ge n^{-c_2}/2$ then stop and accept. Otherwise continue. (5) If the procedure ends without accepting, stop and reject.

When R is U', the probability that $|p_U - x_U/T| \ge n^{-c_2}/8$ is exponentially small by the Hoeffding bound. Since $f_{\mathbf{t},\pi}(U') = U'$, the probability that $|p_U - x_R/T| \ge n^{-c_2}/8$ is exponentially small. Therefore, the acceptance probability of $\mathcal{D}'$ is exponentially close to 0.

When R is $A_{\mathbf{s},\mathbf{s}'}$ for some $\mathbf{s}, \mathbf{s}'$. In each of the iterations, we are considering the distribution $f_{\mathbf{t},\pi}(A_{\mathbf{s},\mathbf{s}'}) = A_{\mathbf{s}+\mathbf{t},\pi(\mathbf{s}')}$ for some uniformly chosen $\mathbf{t}$ and π. Hence, with probability exponentially close to 1, in one of the n^{c_1+1} iterations, $(\mathbf{s}+\mathbf{t}, \pi(\mathbf{s}'))$ is such that the acceptance probability of $\mathcal{D}$ on inputs from $A_{\mathbf{s}+\mathbf{t},\pi(\mathbf{s}')}$ and on inputs from U' differ by at least n^{-c_2}. In this case, from the Hoeffding bound, the probability that $|p_U - x_U/T| \ge n^{-c_2}/8$ and $|p_R - x_R/T| \ge n^{-c_2}/8$ is exponentially small. Hence, $\mathcal{D}'$ accepts with probability exponentially close to 1. □

Lemma 3.10 (Decision to Search). *Let $n \ge 1$ be some integer and $q \ge 2$ be a prime. Assume there exists an algorithm $\mathcal{D}$ that for all $\mathbf{s}, \mathbf{s}'$ accepts with probability exponentially close to* 1 *on inputs from $A_{\mathbf{s},\mathbf{s}'}$ and rejects with probability exponentially close to* 1 *on inputs from U'. Then, there exists an algorithm $\mathcal{D}'$ that, given samples from $A_{\mathbf{s},\mathbf{s}'}$ for some* $\mathbf{s}$, *outputs* $\mathbf{s}$ *with probability exponentially close to* 1.

Proof. We only show how $\mathcal{D}'$ find the first coordinate of $\mathbf{s}$ $s_1 \in \mathbb{Z}_q$. For any $k \in \mathbb{Z}_q$, consider the following transformation. Given a tuple $(\mathbf{a}, \mathbf{e}, b)$ we output the tuple $(\mathbf{a} + {}^t(l, 0, \ldots, 0), \mathbf{e}, b + lk)$ where $l \in \mathbb{Z}_q$ is chosen uniformly at random. This random transformation takes U' into itself. Moreover, if $k = s_1$ then this transformation also takes $A_{\mathbf{s},\mathbf{s}'}$ into itself. Finally, if $k \ne s_1$

then it transforms $A_{s,s'}$ to U'. Therefore, using $\mathcal{D}$, we can test whether $k = s_1$ or not. Since there are only $q < \mathrm{poly}(n)$ possibilities for s_1, we can try all of them. □

Remark 3.11. The hardness of the worst case of $\mathrm{LWKE}_{q,\bar{\Psi}_\alpha}$ implies the hardness of the worst case of $\mathrm{LWE}_{q,\bar{\Psi}_\alpha}$. Note that it is unknown if the converse statement holds. We also note that, from Theorem 3.2, there exists a quantum reduction from $\mathrm{LWE}_{q,\bar{\Psi}_\alpha}$ to $\mathrm{SVP}_{\tilde{O}(n/\alpha)}$ and $\mathrm{SIVP}_{\tilde{O}(n/\alpha)}$.

4 Main Protocol

Recall that we can consider ${}^t\mathbf{A}$ as a generator matrix from a coding-theoretical view and a parity-check matrix $\mathbf{H}$ is easily computed. Informally, if Alice wants to prove that she has a secret key corresponding to a public key $\mathbf{b}$, it is sufficient that she proves that she has an error key $\mathbf{s}'$ such that $\mathbf{HEs}' = \mathbf{Hb}$.

Definition 4.1 (Relation R_{mR05}). Let $(\mathbf{A}, \mathbf{E}, \mathbf{b})$ be a public key of mR05, $\mathbf{H}$ a parity-check matrix of $\mathbf{A}$, $\mathbf{s}$ a vector in $\mathbb{Z}_q^n$, and $\mathbf{s}'$ a vector in $\mathbb{Z}_q^{m_1}$. We say that input $(\mathbf{A}, \mathbf{H}, \mathbf{E}, \mathbf{b})$ and witness $(\mathbf{s}, \mathbf{s}')$ are in R_{mR05} if $\mathbf{s}' \in \mathrm{Set}_{m_1,m_2}$, $\mathbf{As} + \mathbf{Es}' = \mathbf{b}$, and $\mathbf{HEs}' = \mathbf{Hb}$.

Next, we describe the protocol for a proof of knowledge for a secret key, which is mainly based on a proof of knowledge for SDP by Stern [24].

Protocol 4.2 (Protocol PSK). Let P and V be a prover and a verifier respectively. The CRS is $\mathbf{A}, \mathbf{E}$. The common input is $\mathbf{b}$. The auxiliary inputs to the prover are $\mathbf{s}$ and $\mathbf{s}'$ such that $\mathbf{b} = {}^t\mathbf{As} + \mathbf{Es}'$. Let $\mathrm{Com}(\cdot;\cdot) = \mathrm{Com}_{\mathbf{A}}(\cdot;\cdot)$.

Step P1 Choose a random permutation π for $\{1, \ldots, m_1\}$ and a random vector $\mathbf{y} \in \mathbb{Z}_q^{m_1}$. Compute $c_1 = \mathrm{Com}(\pi, \mathbf{HEy}; r_1)$, $c_2 = \mathrm{Com}(\pi(\mathbf{y}); r_2)$ and $c_3 = \mathrm{Com}(\pi(\mathbf{y} + \mathbf{s}'); r_3)$. Send c_1, c_2, c_3 to V.

Step V1 V sends a random challenge bit $\delta \in_R \{1, 2, 3\}$ to P.

Step P2 If $\delta = 1$, P opens c_1 and c_2 (i.e., sends $\pi, \mathbf{y}, r_1$, and r_2 to V). If $\delta = 2$, P opens c_1 and c_3 (i.e., sends $\pi, \mathbf{y} + \mathbf{s}', r_1$ and r_3 to V). If $\delta = 3$, P opens c_2 and c_3 (i.e., sends $\pi(\mathbf{s}'), \pi(\mathbf{y}), r_2$, and r_3 to V).

Step V2 If $\delta = 1$, received $\tilde{\pi}$, $\tilde{\mathbf{y}}$, $\tilde{r}_1$, and $\tilde{r}_2$, check the commitments c_1 and c_2 were correct (i.e., $c_1 = \mathrm{Com}(\tilde{\pi}, \mathbf{HE}\tilde{\mathbf{y}}; \tilde{r}_1)$ and $c_2 = \mathrm{Com}(\tilde{\pi}(\tilde{\mathbf{y}}); \tilde{r}_2)$). If $\delta = 2$, received $\tilde{\pi}$, $\tilde{\mathbf{x}}$, $\tilde{r}_1$, and $\tilde{r}_3$, check that the commitments c_1 and c_3 were correct (i.e., $c_1 = \mathrm{Com}(\tilde{\pi}, \mathbf{HE}\tilde{\mathbf{x}} - \mathbf{Hb}; \tilde{r}_1)$ and $c_3 = \mathrm{Com}(\tilde{\pi}(\tilde{\mathbf{x}}); \tilde{r}_3)$). If $\delta = 3$, received $\tilde{\mathbf{x}}_1$, $\tilde{\mathbf{x}}_2$, $\tilde{r}_2$, and $\tilde{r}_3$, check that the commitments c_2 and c_3 were correct (i.e., $c_2 = \mathrm{Com}(\tilde{\mathbf{x}}_1; \tilde{r}_2)$ and $c_3 = \mathrm{Com}(\tilde{\mathbf{x}}_1 + \tilde{\mathbf{x}}_2; \tilde{r}_3)$) and that $w_H(\tilde{\mathbf{x}}_2) = m_2$.

Theorem 4.3. *Interactive protocol (P, V) is a proof of knowledge system with knowledge error $2/3$ for R_{mR05}. Moreover, the protocol (P, V) is a statistical zero-knowledge argument for R_{mR05} in CRS model under the assumption that the worst case of* $\mathrm{LWKE}_{q,\bar{\Psi}_\alpha}$ *and* $\mathrm{SVP}_{\tilde{O}(n)}$ *is hard.*

Proof of completeness. We omit the proof since it is evident. □

We use Lemma 4.4 below in [24] in the proof of knowledge error.

Lemma 4.4 (Theorem 1 and Lemma 1, [24]). *Assume that some probabilistic polynomial-time adversary P^* is accepted with probability at least $(2/3)^r + \epsilon$, $\epsilon > 0$, after playing the identification protocol r times. Then there exists a polynomial-time probabilistic machine K such that outputs the witness $\mathbf{s}'$ from the common input or else finds collisions for the hash function with probability larger than $\epsilon^3/10$.*

The idea of Lemma 4.4 is follows: Assume that P^* can output response to all V's challenges correctly. Let P's response to V's challenge 1 be $\tilde{\pi}_1$, $\tilde{\mathbf{y}}$, $\tilde{r}_{1,1}$, and $\tilde{r}_{1,2}$. Let P's response to V's challenge 2 be $\tilde{\pi}_2$, $\tilde{\mathbf{x}}$, $\tilde{r}_{2,1}$, and $\tilde{r}_{2,3}$. Finally, let P's response to V's challenge 3 be $\tilde{\mathbf{x}}_1$, $\tilde{\mathbf{x}}_2$, $\tilde{r}_{3,2}$ and $\tilde{r}_{3,3}$. Since all response are correct, we obtain that

$$
\begin{aligned}
c_1 &= \mathrm{Com}(\tilde{\pi}_1, \mathbf{HE}\tilde{\mathbf{y}}; \tilde{r}_{1,1}) &&= \mathrm{Com}(\tilde{\pi}_2, \mathbf{HE}\tilde{\mathbf{x}} - \mathbf{Hb}; \tilde{r}_{2,1})\\
c_2 &= \mathrm{Com}(\tilde{\pi}_1(\tilde{\mathbf{y}}); \tilde{r}_{1,2}) &&= \mathrm{Com}(\tilde{\mathbf{x}}_1; \tilde{r}_{3,2})\\
c_3 &= \mathrm{Com}(\tilde{\pi}_2(\tilde{\mathbf{x}}); \tilde{r}_{2,3}) &&= \mathrm{Com}(\tilde{\mathbf{x}}_1 + \tilde{\mathbf{x}}_2; \tilde{r}_{3,3})
\end{aligned}
$$

If there exists a distinct pair in the inputs of commitment, we find a collision. Then, we assume there exists no distinct pair in P^*'s responses. Since P^* is accepted, $w_H(\tilde{\mathbf{x}}_2) = m_2$. From c_1's equation, $\tilde{\pi}_1 = \tilde{\pi}_2$. Combining $\tilde{\pi}_1 = \tilde{\pi}_2$ and c_3's equations, we obtain $\tilde{\mathbf{x}} = \tilde{\pi}_2^{-1}(\tilde{\mathbf{x}}_1) + \tilde{\pi}_2^{-1}(\tilde{\mathbf{x}}_2)$. From c_2's equation, we have that $\tilde{\mathbf{y}} = \tilde{\pi}_2^{-1}(\tilde{\mathbf{x}}_1)$. Therefore, combining the above argument and c_1's equation, we obtain $\mathbf{Hb} = \mathbf{HE}(\tilde{\mathbf{x}} - \tilde{\mathbf{y}}) = \mathbf{HE}\tilde{\pi}_2^{-1}(\tilde{\mathbf{x}}_2)$ and a witness $\tilde{\pi}_2^{-1}(\tilde{\mathbf{x}}_2)$. Thus, we obtain a collision or a witness using P^*.

Proof of knowledge error with $2/3$. Assume that some probabilistic polynomial-time adversary P^* in Lemma 4.4. Using Lemma 4.4, we obtain K in the above. In Stern's proof, he consider binary linear codes. Although we play the protocol in q-ary linear codes, we can apply Stern's proof to q-ary codes. Note that, under the assumption that the worst case of $\mathrm{SVP}_{\tilde{O}(n)}$ is hard, finding collision is hard [17]. Therefore if assume that $\mathrm{SVP}_{\tilde{O}(n)}$ is hard in the worst case, we obtain a knowledge extractor K. □

Proof of zero knowledge. Since Com is statistically hiding, the simulator's output the transcript when the simulator did not output $\perp$ can be statistically close to the real transcript. We omit the detail of the simulator due to lack of space. □

5 Concluding Remarks

In this chapter, we have proposed a modified Regev'05 cryptosystem (mR05) and introduced a proof of knowledge on its secret key.

At the end, we list up a few open problems: (1) A proof of knowledge on a secret key of the original Regev'05 cryptosystem (R05); mR05 needs stronger assumption than one which R05 needs. (2) Relation between LWE and LWKE; we have failed to show a reduction from LWE to LWKE. (3) Zero knowledge on coding problems; As seen in Section 1, the MV protocol can not apply to coding problems. Thus, we need a direct protocol for coding problems.

References

[1] M. Ajtai. Generating hard instances of lattice problems. *ECCC*, TR96-007, 1996.

[2] M. Ajtai. Generating hard instances of lattice problems (extended abstract). In *STOC '96*, pages 99–108, 1996. See also ECCC TR96-007.

[3] M. Ajtai. Representing hard lattices with $O(n \log n)$ bits. In *STOC 2005*, pages 94–103, 2005.

[4] M. Ajtai and C. Dwork. A public-key cryptosystem with worst-case/average-case equivalence. In *STOC '97*, pages 284–293, 1997.

[5] S. Arora, L. Babai, and Z. Stern, Jacques Sweedyk. The hardness of approximate optima in lattices, codes, and systems of linear equations. *Journal of Computer and System Sciences*, 54(2):317–331, 1997.

[6] E. R. Berlekamp, R. J. McEliece, and H. C. A. van Tilborg. On the inherent intractability of certain coding problems. *IEEE Transactions on Information Theory*, 24(3):384–386, MAY 1978.

[7] J.-Y. Cai and A. Nerurkar. An improved worst-case to average-case connection for lattice problems. In *FOCS '97*, pages 468–477, 1997.

[8] I. B. Damgård, T. P. Pedersen, and B. Pfizmann. On the existence of statistically hiding bit commitment schemes and fail-stop signatures. *Journal of Cryptology*, 10(3):163–194, 1997.

[9] I. B. Damgård, T. P. Pedersen, and B. Pfizmann. Statistical secrecy and multibit commitments. *IEEE Transactions on Information Theory*, 44(3):1143–1151, MAY 1998.

[10] O. Goldreich, S. Goldwasser, and S. Halevi. Collision-free hashing from lattice problems. *ECCC*, TR96-042, 1996.

[11] S. Goldwasser and D. Kharchenko. Proof of plaintext knowledge for the Ajtai-Dwork cryptosystem. In *TCC 2005*, pages 529–555, 2005.

[12] S. Goldwasser and Y. Tauman Kalai. On the (in)security of the Fiat-Shamir paradigm. *ECCC*, TR03-015, 2003.

[13] S. Halevi and S. Micali. Practical and provably-secure commitment scheme from collision-free hashing. In *CRYPTO '96*, pages 201–215, 1996.

[14] S. Hayashi and M. Tada. A lattice-based public-key identification scheme. In *ISITA 2006*, 2006.

[15] D. Micciancio. Almost perfect lattices, the covering radius problem, and applications to Ajtai's connection factor. *SIAM Journal on Computing*, 34(1):118–169, 2004.

[16] D. Micciancio and S. Goldwasser. *Complexity of Lattice Problems: a cryptographic perspective*. Kluwer Academic Publishers, 2002.

[17] D. Micciancio and O. Regev. Worst-case to average-case reductions based on Gaussian measures. In *FOCS 2004*, pages 372–381, 2004.

[18] D. Micciancio and S. Vadhan. Statistical zero-knowledge proofs with efficient provers: Lattice problems and more. In *CRYPTO 2003*, pages 282–298, 2003.

[19] K. Ohta and T. Okamoto. On concrete security treatment of signatures derived from identification. In *CRYPTO '98*, pages 354–369, 1998.

[20] D. Pointcheval and J. Stern. Security proofs for signature schemes. In *EUROCRYPT '96*, pages 387–398, 1996.

[21] O. Regev. New lattice-based cryptographic constructions. *Journal of the ACM*, 51(6):899–942, 2004.

[22] O. Regev. On lattices, learning with errors, random linear codes, and cryptography. In *STOC 2005*, pages 84–93, 2005.

[23] A. Shamir. An efficient identification scheme based on permuted kernels (extended abstract). In *CRYPTO '89*, pages 606–609, 1989.

[24] J. Stern. A new paradigm for public key identification. *IEEE Transactions on Information Theory*, 42(6):749–765, November 1996.