Advance Access publication on 11 June 2018
doi:10.1093/comjnl/bxy060

Klepto for Ring-LWE Encryption

Dianyan Xiao[1] and Yang Yu[2*]

[1]Institute for Advanced Study, Tsinghua University, Beijing 100084, China
[2]Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
[*]Corresponding author: y-y13@mails.tsinghua.edu.cn

Due to its great efficiency and quantum resistance, public key cryptography based on Ring-LWE problem has drawn much attention in recent years. A batch of cryptanalysis works provided ever-improved security estimations for various Ring-LWE schemes, but few works discussed the security of Ring-LWE cryptography from kleptographic aspect. In this paper, we show how to embed a backdoor into a classic Ring-LWE encryption scheme so that partial bits of the plaintext are leaked to the owner of the backdoor. By theoretical analysis and experimental observations, we argue that the klepto Ring-LWE encryption scheme with such backdoor is feasible and practical.

Keywords: lattice-based cryptography; kleptography; public key encryption; ring learning with errors

Received 11 January 2018; revised 9 May 2018; editorial decision 11 May 2018
Handling editor: Albert Levi

1. INTRODUCTION

Kleptography, introduced by Young and Yung [1–4], is the study of exploiting cryptographic backdoors to steal information securely and subliminally. A typical klepto scheme is a black-box implementation whose output should be indistinguishable from that of the legitimate cryptosystem for anyone but the owner of the backdoor key. A klepto scheme can be designed to leak the message, the private key or the state of the pseudorandom number generator, and the attacker can decrypt the leaked information using his backdoor key. After the dramatic revelations of Edward Snowden, the cryptographic research community realized that kleptographic attack indeed had been deployed and likely used for worldwide surveillance, which rekindled the interest in kleptography.

With the threat that quantum computers pose to most of the current cryptosystems, post-quantum cryptography has been gaining much attention in recent years. Due to the great performance and strong security guarantee, lattice-based cryptography is considered as a desirable quantum-safe alternative to classical schemes based on integer factorization or discrete logarithms. NTRU and LWE are two most widely used families of lattice-based cryptography. NTRU [5] is one of the earliest lattice-based schemes and has been standardized by IEEE. Through more than 20 years' study, NTRU is believed very efficient and secure. However, the security of classical NTRU relies on heuristic arguments and provably secure NTRU variants [6–8] are impractical. LWE (*Learning With Errors*) was introduced by Regev in [9]. In terms of compactness and efficiency, Ring-LWE [10], an algebraic variant of LWE, enjoys better popularity than usual LWE in practical applications. Moreover, Ring-LWE has been proved to be as hard as certain worst-case problems over ideal lattices, and this provides a firm theoretical grounding for the security of Ring-LWE schemes. Therefore, Ring-LWE schemes seem to reach an ideal balance between efficiency and security.

With the upcoming post-quantum cryptography standardization by the NIST, it is pressing to provide a comprehensive cryptanalysis for lattice-based cryptosystems. From the mathematical and algorithmic aspects, people have developed various attacks against lattice-based cryptosystems, such as lattice-reduction attacks [11, 12] and combinatorial attacks [13, 14]. All these cryptanalysis results seem to form a somewhat systematical methodology for estimating the security of lattice-based cryptography. However, from the kleptographic aspect, there are only few results related to lattice schemes. In [15], the authors discussed a class of possible backdoors for NewHope, a Ring-LWE key exchange. As claimed in [15], their backdoors apply to fixed public parameter and can be prevented by the 'nothing-up-my-sleeve' process which is to choose the public parameter as the hash of a common universal string. In a very recent paper [16], Kwant, Lange and Thissen targeted NTRU scheme [5] and proposed a klepto scheme with an ECC-based backdoor. Also, they discussed the impact of the NTRU backdoor and countermeasures against the klepto scheme.

In this paper, we show how to modify a classic Ring-LWE encryption scheme into a klepto scheme. The backdoor we set is also based on Ring-LWE itself which makes the whole scheme accord with post-quantum setting. Our technical idea is to 'encode' a polynomial of low degree but large coefficients into a new polynomial of high degree but small coefficients. Exploiting this idea, we are able to infect the Ring-LWE ciphertext using a polynomial of small coefficients so that the infected ciphertext can be translated into a backdoor ciphertext, and at the same time the decryption would not be affected too much by this modification. By studying the backdoor theoretically and experimentally, we claim that such a klepto scheme is practical.

The rest of this paper is organized as follows. After some preliminaries in Section 2, we introduce our Ring-LWE encryption backdoor in Section 3. In Section 4, we analysis the impact and quality of our backdoor theoretically. In Section 5, we provide experimental results to check the quality of the backdoor. Finally, we conclude in Section 6.

2. PRELIMINARIES

2.1. Notations

For an integer $q \geq 2$, we identify $\mathbb{Z}_q$ with $[0, q) \cap \mathbb{Z}$. The ring $\mathcal{R}$ that we will work with is a power-of-2 cyclotomic ring, i.e. $\mathcal{R} = \mathbb{Z}[X]/(X^n + 1)$ where n is a power of 2. Let $\mathcal{R}_q = \mathbb{Z}_q[X]/(X^n + 1)$. For any $t = \sum_{i=0}^{n-1} t_i X^i \in \mathcal{R}$, we call $(t_0, \ldots, t_{n-1}) \in \mathbb{Z}^n$ the coefficient vector of t. We denote by $\|t\|$ (resp. $\|t\|_\infty$) the Euclidean (resp. ℓ_∞) norm of the coefficient vector of t.

A function $f(n)$ is *negligible*, if $f(n) = o(n^{-c})$ for any constant c. Generally, we denote by $\mathrm{negl}(n)$ as a negligible function with respect to n. We say that a probability is *overwhelming* if it is $1 - \mathrm{negl}(n)$. The notations $\log(\cdot)$ and $\ln(\cdot)$ represent the base 2 and natural logarithms, respectively.

We write $z \hookleftarrow D$ when the random variable z is sampled from the distribution D, and denote by $D(x)$ the probability of $z = x$. For a finite domain E, let $U(E)$ be the uniform distribution over E. For two distributions D_1, D_2 over a same discrete domain E, their statistical distance is $\Delta(D_1; D_2) = \frac{1}{2}\sum_{x\in E}|D_1(x) - D_2(x)|$. We say D_1 and D_2 are *statistically close* with respect to n if $\Delta(D_1; D_2)$ is negligible.

2.2. Kleptography

The core of a klepto scheme is a SETUP (*Secretly Embedded Trapdoor with Universal Protection*) that was introduced by Young and Yung in [1].

DEFINITION 2.1 (Adapted from Definition 1 in [1]). *Let C be a publicly known cryptosystem. A SETUP mechanism is an algorithmic modification made to C to get C' such that*

- *The input of C' agrees with the public specifications of the input of C.*
- *C' computes using the attacker's public encryption function E (and possibly other functions as well), contained within C'.*
- *The attacker's private decryption function D is not contained within C' and is known only by the attacker.*
- *The output of C' agrees with the public specifications of the output of C. At the same time, it contains published bits (of the plaintext or user's secret key) which are easily derivable by the attacker but are otherwise hidden.*
- *Furthermore, the output of C and C' are polynomially indistinguishable to everyone (including those who have access to the code of C') except the attacker.*

As explained in [16], the SETUP still works well in practice even if we use a relaxed Condition 5 in which the output of C and C' are only required to be fairly close rather than polynomially indistinguishable, because the end user does not know the code of C' and often does not know the distribution of the output of C exactly. We will follow the relaxed SETUP setting later.

2.3. Gaussian measures

We denote by D_σ the discrete Gaussian distribution over $\mathbb{Z}$ with deviation σ.[1] Let $\rho_\sigma(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right)$, then the probability of $x \in \mathbb{Z}$ under D_σ is $D_\sigma(x) = \rho_\sigma(x)/\rho_\sigma(\mathbb{Z})$ where $\rho_\sigma(\mathbb{Z}) = \sum_{x\in\mathbb{Z}}\rho_\sigma(x)$. For $\mathcal{R} = \mathbb{Z}[X]/(X^n + 1)$, let $D_{\sigma,\mathcal{R}}$ be the distribution of $v \in \mathcal{R}$ where each coefficient of v is sampled from D_σ independently.

For $\epsilon > 0$, let $\eta_\epsilon(\mathcal{R}) = \min\{s > 0 \mid \sum_{v\in\mathcal{R}}\rho_{1/(\sqrt{2\pi}s)}(\|v\|) \leq 1 + \epsilon\}$ that actually equals the so-called smoothing parameter of $\mathbb{Z}^n$. Now we recall some basic properties of Gaussian.

LEMMA 2.1 (Adapted from Lemma 3.3 in [17]). *Let $\mathcal{R} = \mathbb{Z}[X]/(X^n + 1)$ and $\epsilon \in (0, 1)$. Then $\eta_\epsilon(\mathcal{R}) \leq \sqrt{\ln(2n(1 + 1/\epsilon))/\pi}$.*

LEMMA 2.2 (Lemma 1.5 in [18]). *Let $\mathcal{R} = \mathbb{Z}[X]/(X^n + 1)$. Let $c > 1$ and $C = c \cdot \exp\left(\frac{1 - c^2}{2}\right) < 1$. Then $\Pr_{y\hookleftarrow D_{\sigma,\mathcal{R}}}(\|y\| \geq c\sigma\sqrt{n}) \leq C^n$.*

LEMMA 2.3 (Adapted from Lemma 10 in [19]). *Let $\mathcal{R} = \mathbb{Z}[X]/(X^n + 1)$ and $\epsilon \in \left(0, \frac{1}{2m + 1}\right)$. For $z_1, \ldots, z_m \in \mathcal{R}$ and $\sigma \geq \eta_\epsilon(\mathcal{R})/\sqrt{2\pi}$, we have*

[1] Discrete Gaussian is sometimes defined by its width $s = \sqrt{2\pi}\sigma$.

$$\Pr_{y_i \hookleftarrow D_{\sigma,\mathcal{R}}}\left(\left\|\sum_{i=1}^{m} y_i z_i\right\|_\infty \geq \sqrt{2\pi}\sigma t \cdot \sqrt{\sum_{i=1}^{m}\|z_i\|^2}\right) \leq \frac{1+\epsilon}{1-\epsilon}\cdot 2tn\sqrt{2\pi e}\cdot \mathrm{e}^{-\pi t^2}.$$

In particular, the above probability is negligible for $t = \omega(\sqrt{\log n})$.

2.4. Ring-LWE

For $s \in \mathcal{R}_q$ and ψ a distribution over $\mathcal{R}$, the Ring-LWE distribution $A_{s,\psi}$ is the distribution over $\mathcal{R}_q \times \mathcal{R}_q$ obtained by sampling the pair $(a, as + e \bmod q)$ where $a \hookleftarrow U(\mathcal{R}_q)$ and $e \hookleftarrow \psi$. The *search* Ring-LWE problem is to find s given arbitrarily many independent samples from $A_{s,\psi}$. The *decision* version is defined as follows: given some samples from $A_{s,\psi}$ where $s \hookleftarrow \psi$ and the same number of samples from $U(\mathcal{R}_q \times \mathcal{R}_q)$, distinguish them with an advantage $1/\mathrm{poly}(n)$. As shown in [10], for certain parameters and error distribution ψ, both search and decision Ring-LWE problems are as hard as the worst-case approximate shortest vector problem with polynomial factor over ideal lattices. Currently, it is believed that Ring-LWE with proper parameters is against subexponential quantum attacks. We refer to [10, 20] for more details of Ring-LWE.

The Ring-LWE encryption scheme that we will discuss later was first described in [10]. The issues of parameter selection and implementation details were well-studied in [21]. We denote by $RLWE_{n,q,\sigma}$ the Ring-LWE encryption scheme specified by the tuple (n, q, σ) where n is a power of 2, q is the modulus,[2] and σ is the deviation of the discrete Gaussian used as Ring-LWE error distribution. We may omit the subscripts when they are clear from the context. The ring $\mathcal{R} = \mathbb{Z}[X]/(X^n + 1)$ and the plaintext space is $\mathcal{R}_2$. We list below three main algorithms, i.e. key generation, encryption and decryption:

- *RLWE-KeyGen*: Choose $a \hookleftarrow U(\mathcal{R}_q)$ and $s, e \hookleftarrow D_{\sigma,\mathcal{R}}$. Let $b = as + e \in \mathcal{R}_q$. The public key is the pair $(a, b) \in \mathcal{R}_q \times \mathcal{R}_q$, and the secret key is $s \in \mathcal{R}$.
- *RLWE-Enc* (m): Choose $r, e_1, e_2 \hookleftarrow D_{\sigma,\mathcal{R}}$. The ciphertext is
$$(u, v) = \left(ar + e_1, br + e_2 + \left\lfloor \frac{q}{2} \right\rceil m\right) \in \mathcal{R}_q \times \mathcal{R}_q.$$
- *RLWE-Dec* $((u, v))$: Calculate $\overline{m} = v - us \bmod q$. The coefficients of m can be recovered by rounding the coefficients of $\overline{m}$ back to either 0 or $\left\lfloor \frac{q}{2} \right\rceil$, whichever is closest modulo q.

For appropriate parameters, it can be shown (see [20]) that $\|re + e_2 - e_1 s\|_\infty$ would be less than $\frac{q}{4}$ with overwhelming probability. Notice that

$$v - us = re + e_2 - e_1 s + \left\lfloor \frac{q}{2} \right\rceil m \bmod q,$$

thus we know that decryption is correct with overwhelming probability.

Assuming the hardness of Ring-LWE, it follows that the Ring-LWE encryption scheme is semantically secure and the ciphertext is pseudorandom, which was explained in [20].

3. THE BACKDOOR OF RING-LWE ENCRYPTION

In this section, we will propose a modified Ring-LWE encryption scheme with a backdoor using a SETUP that is based on a smaller Ring-LWE scheme. By setting different parameters, the klepto Ring-LWE scheme leaks a different proportion of the message to the third party owning the backdoor key.

The klepto scheme is specified by two sets of parameters: the public one is (n, q, σ) determining the Ring-LWE encryption scheme and the secret one is (n', q', σ', τ) determining the Ring-LWE-based backdoor with parameter τ for adjusting the proportion of leaked message. To make the backdoor workable and compact, we set $n' < n$ and $\tau \leq n/n'$.

Let $\mathcal{R} = \mathbb{Z}[X]/(X^n + 1)$ and $\mathcal{R}' = \mathbb{Z}[X]/(X^{n'} + 1)$ where n is a power of 2 and n' is a factor of n. We identify a polynomial and its coefficient vector. Let $k = n/n'$, $p = \left\lceil q'^{\frac{1}{k}} \right\rceil$ and $d = 2^\tau$. Now we are to define two maps that will be used in encryption and decryption algorithms. The first one is that

$$\theta_{\mathrm{ext}} : \mathcal{R}'_{q'} \to \mathcal{R}_p$$
$$(v'_0, v'_1, \ldots, v'_{n'-1}) \mapsto (v_0, v_1, \ldots, v_{n-1}),$$

where $(v_{ik} v_{ik+1} \cdots v_{ik+k-1})$ is the p-adic expansion of v'_i for $i = 0, 1, \ldots, n' - 1$. The second one is that

$$\theta_{\mathrm{com}} : \mathcal{R}_2 \to \mathcal{R}'_d$$
$$(v_0, v_1, \ldots, v_{n-1}) \mapsto (v'_0, v'_1, \ldots, v'_{n'-1}),$$

where $(v_{i\tau} v_{i\tau+1} \cdots v_{i\tau+\tau-1})$ is the 2-adic expansion of v'_i for $i = 0, 1, \ldots, n' - 1$. It can be observed that θ_{ext} extends a polynomial in $\mathcal{R}'_{q'}$ to a polynomial in $\mathcal{R}$ of bounded infinity norm and θ_{com} compresses the first $\tau n'$ bits of a binary polynomial in $\mathcal{R}$ into a polynomial in $\mathcal{R}_d$. Furthermore, it is worth noting that θ_{ext} is injective and easy-to-invert, and so is θ_{com} restricted to $\{0, 1\}^{\tau n'} \times 0^{n-\tau n'}$.

[2] In practice, the modulus q is usually chosen to be a number of special property, such as a prime congruent to 1 modulo $2n$, which leads to a faster implementation.

We also introduce a slightly modified Ring-LWE encryption, denoted by $RLWE'_{n,q,\sigma,\tau}$, whose plaintext space is $\mathcal{R}_d$. The key generation algorithm is exactly the same as that of $RLWE_{n,q,\sigma}$. The encryption and decryption algorithms are listed as follows:

- *RLWE'-Enc* (m): Choose $r, e_1, e_2 \hookleftarrow D_{\sigma,\mathcal{R}}$. The ciphertext is
$$(u, v) = \left(ar + e_1, br + e_2 + \left\lfloor \frac{q}{2^\tau} \right\rfloor m\right) \in \mathcal{R}_q \times \mathcal{R}_q.$$
- *RLWE'-Dec*$((u, v))$: Calculate $\overline{m} = v - us \bmod q$. The coefficients of m can be recovered by rounding the coefficients of $\overline{m}$ back to a certain multiple of $\left\lfloor \frac{q}{2^\tau} \right\rfloor$, whichever is closest modulo q.

Intuitively, the modified Ring-LWE scheme works well under proper parameters like the original Ring-LWE scheme. Further discussion will be shown in next section. We now describe three main algorithms of the klepto Ring-LWE scheme.

The public key generation algorithm is totally the same as that in Ring-LWE encryption. The kleptographic attacker generated his backdoor Ring-LWE key pair $((a', b'), s')$ and picked $\lfloor \tau n' \rfloor$ bits to locate the leaked message in secret and in advance. For simplicity, we will only discuss the case where τ is a positive integer and the attacker targeted the first $\tau n'$ bits of the message.[3]

The encryption algorithm is changed as follows:

- *Klepto-Enc*(m).
  (1) Run $RLWE_{n,q,\sigma}$-*Enc* (m) and obtain the ciphertext $(u, v) \in \mathcal{R}_q \times \mathcal{R}_q$.
  (2) Run $RLWE'_{n',q',\sigma',\tau}$-*Enc* $(\theta_{\text{com}}(m))$ and obtain the ciphertext $(u', v') \in \mathcal{R}'_{q'} \times \mathcal{R}'_{q'}$.
  (3) Calculate $u'' = \theta_{\text{ext}}(u') \in \mathcal{R}_p$ and $v'' = \theta_{\text{ext}}(v') \in \mathcal{R}_p$.
  (4) Calculate $\Delta_u \in \mathcal{R}$ with all coefficients in $[-p/2, p/2)$ such that $u'' = \bar{u} \bmod p$ where $\bar{u} = u + \Delta_u \bmod q$; if such Δ_u does not exist, back to 1. Calculate $\Delta_v \in \mathcal{R}$ with all coefficients in $[-p/2, p/2)$ such that $v'' = \bar{v} \bmod p$ where $\bar{v} = v + \Delta_v \bmod q$; if such Δ_v does not exist, back to 1.
  (5) The ciphertext is $(\bar{u}, \bar{v}) \in \mathcal{R}_q \times \mathcal{R}_q$.

There are two different decryption algorithms for the legitimate receiver and the attacker. The legitimate receiver uses his secret key s and follows the original Ring-LWE decryption. The middle term he calculated is that

$$\begin{aligned}\bar{v} - \bar{u}s &= v + \Delta_v - us - \Delta_u s \\ &= (re + e_2 - e_1 s) + \Delta_v - \Delta_u s + \left\lfloor \frac{q}{2} \right\rfloor m \bmod q.\end{aligned}$$

When the magnitude of $p\left(=\left\lceil q'^{\frac{1}{k}} \right\rceil\right)$ is significantly less than q, it would still hold that $\|\Delta_v - \Delta_u s + re + e_2 - e_1 s\|_\infty < \frac{q}{4}$ with a high probability in practice and thus the receiver recovers the message correctly in this case.

For the attacker, the decryption algorithm is shown as follows:

- *Klepto-Dec*$((\bar{u}, \bar{v}))$.
  (1) Calculate $(u'', v'') = (\bar{u} \bmod p, \bar{v} \bmod p)$ and then calculate $u', v' \in \mathcal{R}'_{q'}$ such that $u'' = \theta_{\text{ext}}(u')$, $v'' = \theta_{\text{ext}}(v')$.
  (2) Run $RLWE'_{n',q',\sigma',\tau}$-*Dec*$((u', v'))$ and obtain a middle term $m' \in \mathcal{R}'_d$.
  (3) Calculate $\overline{m} \in \{0, 1\}^{\tau n'} \times 0^{n - \tau n'}$ such that $m' = \theta_{\text{com}}(\overline{m})$. The leaked message is $\overline{m}$.

The backdoor discussed in [15] targets a Ring-LWE key exchange, while ours targets a Ring-LWE encryption scheme. More importantly, the backdoor in [15] modified the public parameter a as an NTRU-like public key f/g where f, g are small polynomials, while our backdoor is embedded in the implementation of encryption and never changes the public key. Indeed, the public key could be generated elsewhere and chosen to be the hash of a universal string, in which the backdoor in [15] does not work. Compared with the NTRU klepto scheme [16], our klepto scheme is built on a totally different cryptosystem and our backdoor follows post-quantum setting rather than ECC setting. Furthermore, the parameter selection and analysis of our backdoor are quite different from that in [16], which is shown in the next section.

4. ANALYSIS OF THE BACKDOOR

In this section, we are to report on the impact and quality of the backdoor in Ring-LWE encryption. More specifically, we will discuss how the backdoor parameters affect the decryption for the attacker and legitimate receiver, how much infected ciphertexts behave like uninfected ones and how middle terms in decryption behave different with respect to infected and uninfected ciphertexts.

4.1. Decryption failures for the attacker

There are two kinds of operations in backdoor decryption: inversions of θ_{ext} and θ_{com} and a modified Ring-LWE decryption. The decryption failure of backdoor decryption is

[3] For different targeted bits, we only need to modify the map θ_{com}.

equivalent to the decryption failure of the modified Ring-LWE scheme. By a routine computation, we have that the middle term in the modified Ring-LWE decryption is

$$v' - u's' = r'e' + e_2' - e_1's' + \left\lfloor \frac{q'}{2^\tau} \right\rfloor m' \bmod q'.$$

Thus, a successful decryption relies on the fact that $\|r'e' + e_2' - e_1's'\|_\infty < \frac{1}{2}\left\lfloor \frac{q'}{2^\tau} \right\rfloor$. It then leads to that the probability of decryption failure for the attacker is

$$P_{\text{attacker}} = \Pr\left(\|r'e' + e_2' - e_1's'\|_\infty \geq \frac{1}{2}\left\lfloor \frac{q'}{2^\tau} \right\rfloor \right)$$

where $r', e', e_1', e_2', s' \hookleftarrow D_{\sigma',\mathcal{R}'}$.

We choose $\sigma' \geq \sqrt{\ln(2n'(1 + 1/\epsilon))}/(\sqrt{2}\pi)$ for a small $\epsilon \in \left(0, \frac{1}{7}\right)$. For $t' = \omega\left(\sqrt{\log n'}\right)$, Lemmas 2.1 and 2.2 show that

$$\|s'\|, \|r'\| \geq c\sqrt{n'}\sigma'$$

with negligible probability. Combining Lemma 2.3, we have that

$$\|r'e' + e_2' - e_1's'\|_\infty \geq \sqrt{2\pi}\sigma't'\sqrt{2n'c^2\sigma'^2 + 1}.$$

with negligible probability. Consequently, when

$$q' > 2^{\tau+1}\sqrt{2\pi}\sigma't'\sqrt{2n'c^2\sigma'^2 + 1} + 2^{\tau-1}, \tag{1}$$

the probability P_{attacker} is negligible, which implies that the attacker recovers correctly all targeted bits of message with overwhelming probability.

4.2. Decryption failures for the legitimate receiver

In the klepto Ring-LWE encryption scheme, the ciphertext is infected by two extra terms Δ_u and Δ_v, and the middle term in decryption becomes

$$\bar{v} - \bar{u}s = (re + e_2 - e_1 s) + \Delta_v - \Delta_u s + \left\lfloor \frac{q}{2} \right\rfloor m \bmod q.$$

According to the definitions, we know that $\|\Delta_u\|_\infty$, $\|\Delta_v\|_\infty \leq \frac{p}{2}$ and then $\|\Delta_u\| \leq \frac{\sqrt{n}p}{2}$. Thus, the probability of decryption failure for legitimate receivers is bounded by

$$P_{\text{klepto}} \leq \Pr\left(\|re + e_2 - e_1 s - \Delta_u s\|_\infty \geq \frac{1}{2}\left\lfloor \frac{q}{2} \right\rfloor - \frac{p}{2} \right)$$

where $r, e, e_1, e_2, s \hookleftarrow D_{\sigma,\mathcal{R}}$.

For $t = \omega(\sqrt{\log n})$, $\epsilon \in \left(0, \frac{1}{9}\right)$ and $\sigma \geq \sqrt{\ln(2n(1 + 1/\epsilon))}/(\sqrt{2}\pi)$, by similar arguments shown in last subsection, we have that when

$$q > 4\sqrt{2\pi}\sigma t\sqrt{2nc^2\sigma^2 + 1 + \frac{np^2}{4}} + 2p + 1, \tag{2}$$

the probability P_{klepto} is negligible, which implies that legitimate receivers can hardly detect the klepto scheme only from decryption failures.

Remark 1. Equations (1) and (2) provide quantitative parameter relations to ensure a negligible probability of decryption failures, which sets a theoretical grounding for backdoor parameter selection. However, in practice, people may choose tighter parameters to achieve better efficiency and an acceptable decryption failure rate.

4.3. Distinctions between infected and uninfected ciphertexts

Next we are to report on a few distinctions between infected and uninfected ciphertexts. We follow the notations in Section 3. Let $q = lp + w$ where $l \in \mathbb{Z}$ and $w \in \mathbb{Z}_p$. We start with the following heuristics profiling the distributions of ciphertexts:

(1) We model the distributions of (u, v) and (u', v') as $U(\mathcal{R}_q \times \mathcal{R}_q)$ and $U(\mathcal{R}'_{q'} \times \mathcal{R}'_{q'})$ respectively.

(2) We assume that (u, v) and (u', v') look like independent.

The Heuristic 1 can be explained by the pseudorandomness of Ring-LWE ciphertexts assuming the hardness of Ring-LWE. The Heuristic 2 is reasonable because the key generation and encryption of $RLWE_{n,q,\sigma}$ and $RLWe'_{n',q',\sigma',\tau}$ are independent, despite the fact that two plaintexts that they correspond to are strongly correlated.

Ciphertexts modulo p: For uninfected ciphertexts (u, v), each coefficient of $(u \bmod p)$ and $(v \bmod p)$ follows the same distribution D_1 over $\mathbb{Z}_p$ that

$$D_1(i) = \begin{cases} \dfrac{l+1}{q}, & i < w; \\ \dfrac{l}{q}, & i \geq w. \end{cases}$$

However, for infected ciphertexts $(\bar{u}, \bar{v})$, the ciphertext modulo p equals an extension (by θ_{ext}) of (u', v') and thus the distributions of coefficients of $\bar{u} \bmod p$ and $\bar{v} \bmod p$ could be different. We denote by $\mathbf{ext}(\alpha) \in \mathbb{Z}_p^k$ the p-adic expansion of

$\alpha \in \mathbb{Z}_{q'}$ and by $D_2^{(k)}$ the distribution of $\mathbf{ext}(\alpha)$ over $\mathbb{Z}_p^k$ where $\alpha \hookleftarrow U(\mathbb{Z}_{q'})$. Then, for $i_0, \ldots, i_{k-1} \in \mathbb{Z}_p$, we have

$$D_2^{(k)}(i_0 i_1 \cdots i_{k-1}) = \begin{cases} \dfrac{1}{q'}, & \sum_j i_j p^j < q'; \\ 0, & \sum_j i_j p^j \geq q'. \end{cases}$$

Let D_1^k be the distribution over $\mathbb{Z}_p^k$ where each coordinate follows D_1 independently. On the one hand, according to the probability mass functions of D_1 and $D_2^{(k)}$, we conclude that D_1^k and $D_2^{(k)}$ are indeed different, which may be used to check the existence of such backdoor. On the other hand, notice that

$$\begin{aligned}\Delta(D_1^k; D_2^{(k)}) &\leq \Delta(D_1^k; U(\mathbb{Z}_p^k)) + \Delta(U(\mathbb{Z}_p^k); D_2^{(k)}) \\ &\leq k \cdot \Delta(D_1; U(\mathbb{Z}_p)) + \frac{p^k - q'}{p^k} \\ &= \frac{kw(p-w)}{pq} + \frac{p^k - q'}{p^k}, \end{aligned} \qquad (3)$$

thus D_1^k and $D_2^{(k)}$ could be close when k is small, $q \gg p$ and $q' \approx p^k$, which implies a way to select parameters for high-quality backdoors.

Ciphertexts modulo q: In Step 4 of the klepto encryption, the black-box calculates certain small terms Δ_u and Δ_v according to (u, v) and (u', v'). However, as claimed in the algorithm, such Δ_u and Δ_v do not always exist and thus the black-box may restart several times, which is different from the case in [16].

For $z \in \mathbb{Z}_q$ and $z' \in \mathbb{Z}_p$, we say (z, z') is a *bad* pair if and only if the set $\{\Delta_z \in \mathbb{Z} \,|\, ((z + \Delta_z) \bmod q) = z' \bmod p\} \bigcap [-p/2, p/2)$ is empty. The total number of bad pairs is shown in the following lemma.

LEMMA 4.1. *Let $q > p$ and $w = (q \bmod p) \in \mathbb{Z}_p$, then there are totally $w(p - w)$ bad pairs.*

Proof. Let $y = z - \left\lfloor \dfrac{p}{2} \right\rfloor \bmod q$, then (z, z') is a bad pair if and only if $\{x \in \mathbb{Z}_p \,|\, ((y + x) \bmod q) = z' \bmod p\} = \varnothing$. If $y + p \leq q$, the term $((y + x) \bmod q)$ equals $y + x$ for $x \in \mathbb{Z}_p$. When x runs over $\mathbb{Z}_p$, there is always a unique x_0 such that $y + x_0 = z' \bmod p$. Thus a bad pair (z, z') implies that $y = z - \left\lfloor \dfrac{p}{2} \right\rfloor > q - p$. Now we are to calculate the number of (y, z') corresponding to bad pairs.

Let $y = q - p + \lambda$ where $\lambda \in [1, p - 1]$, then

$$\begin{aligned} S_y &= \{((y + x) \bmod q) \bmod p \,|\, x \in \mathbb{Z}_p\} \\ &= S_\lambda^{(1)} \bigcup S_\lambda^{(2)}, \end{aligned}$$

where $S_\lambda^{(1)} = \{(q - p + \lambda) \bmod p, \ldots, (q-1) \bmod p\}$ and $S_\lambda^{(2)} = \{0, 1, \ldots, \lambda - 1\}$. The total number of (y, z') corresponding to bad pairs is

$$\begin{aligned} N &= \sum_y (p - |S_y|) \\ &= \sum_{\lambda=1}^{p-1} \left(\left| \overline{S_\lambda^{(1)}} \right| - \left| \overline{S_\lambda^{(1)}} \bigcap S_\lambda^{(2)} \right| \right) \\ &= \frac{p(p-1)}{2} - \sum_{\lambda=1}^{p-1} \left| \overline{S_\lambda^{(1)}} \bigcap S_\lambda^{(2)} \right|, \end{aligned}$$

where $\overline{S_\lambda^{(1)}} = \mathbb{Z}_p \backslash S_\lambda^{(1)} = \{w, \ldots, w + \lambda - 1\}$.

For any $i \in \mathbb{Z}_p$, if $i \in [0, w)$, there are exactly $(w - i)$ $\overline{S_\lambda^{(1)}} \bigcap S_\lambda^{(2)}$'s containing i; if $i \in [w, p - 1]$, there are exactly $(p - 1 - i)$ $\overline{S_\lambda^{(1)}} \bigcap S_\lambda^{(2)}$'s containing i. Therefore, it leads to that

$$\begin{aligned} \sum_{\lambda=1}^{p-1} \left| \overline{S_\lambda^{(1)}} \bigcap S_\lambda^{(2)} \right| &= \sum_{i=0}^{w-1} (w - i) + \sum_{i=w}^{p-1} (p - 1 - i) \\ &= \frac{w(w-1)}{2} + \frac{(p-w)(p-w-1)}{2}, \end{aligned}$$

and we immediately obtain that $N = w(p - w)$. We now complete the proof. □

As a direct corollary, we give the expected number of repetitions of the klepto encryption.

COROLLARY 4.1. *The expected number of repetitions of klepto encryption is (almost)*

$$\left(1 - \frac{w(p-w)}{pq}\right)^{-2n} - 1,$$

if $(\theta_{\text{ext}}(u'), \theta_{\text{ext}}(v'))$ is (almost) uniform over $U(\mathcal{R}_p \times \mathcal{R}_p)$.

It implies that the klepto encryption terminates after few repetitions when q is sufficiently large.

Now we are to compare (u, v) and $(\bar{u}, \bar{v})$ by studying each of their coefficients. For $i \in \mathbb{Z}_q$, let N_i denote the number of (z, z') such that $i = z + \Delta_z \bmod q$ and $i = z' \bmod p$ for some $\Delta_z \in [-p/2, p/2)$. It is easy to verify the following facts:

- $\forall\, i \in \mathbb{Z}_q$, $N_i \leq p$;
- $\forall\, i \in \mathbb{Z}_q \bigcap \left(\dfrac{p}{2}, q - \dfrac{p}{2}\right]$, $N_i = p$;
- $\sum_{i=0}^{q-1} N_i = pq - w(p - w)$. (By Lemma 4.1)

Let D_z and $D_{\bar{z}}$ be the distributions of a random coefficient of (u, v) and $(\bar{u}, \bar{v})$, respectively. If we replace (u'', v'') in Step 4 of the klepto encryption with $(\widehat{u'}, \widehat{v'}) \hookleftarrow U(\mathcal{R}_p \times \mathcal{R}_p)$, then we obtain a new pair, denoted by $(\hat{u}, \hat{v})$. Let $D_{\hat{z}}$ be the distribution of a random coefficient of $(\hat{u}, \hat{v})$, then

TABLE 1. Experimental measure of decryption failures. DFRR and DFRA are the abbreviations of 'Decryption failure rate for legitimate receivers' and 'Decryption failure rate for attackers', respectively. We use LMR to denote the proportion of leaked message.

Parameter		$\left\lfloor q'^{\frac{n'}{n}} \right\rfloor$	DFRR (%)	DFRA (%)	LMR
(n, q, σ)	(n', q', σ', τ)				
$\left(512,\ 12289,\ \frac{12.18}{\sqrt{2\pi}}\right)$	$(128, 8^4, 11/\sqrt{2\pi}, 1)$	8	93.82	88.63	0.25
	$(128, 9^4, 11/\sqrt{2\pi}, 1)$	9	92.96	100	0.25
	$(128, 7681, 11/\sqrt{2\pi}, 1)$	10	90.47	100	0.25
	$(128, 9^4, 11/\sqrt{2\pi}, 2)$	9	92.84	38.79	0.5
	$(128, 7681, 11/\sqrt{2\pi}, 2)$	10	90.21	79.15	0.5
	$(128, 9473, 11/\sqrt{2\pi}, 2)$	10	90.27	98.33	0.5
	$(128, 10^4, 11/\sqrt{2\pi}, 2)$	10	91.01	99.15	0.5
$\left(1024,\ 2^{32}-1,\ \frac{8}{\sqrt{2\pi}}\right)$	$(256, 7681, 11.31/\sqrt{2\pi}, 1)$	10	100	99.08	0.25
	$(256, 10^4, 11.31/\sqrt{2\pi}, 1)$	10	100	99.99	0.25
	$(512, 12289, 12.18/\sqrt{2\pi}, 1)$	111	100	97.13	0.5
	$(512, 111^2, 12.18/\sqrt{2\pi}, 1)$	111	100	97.31	0.5
	$(256, 11^4, 11.31/\sqrt{2\pi}, 2)$	11	100	97.85	0.5
	$(256, 15361, 11.31/\sqrt{2\pi}, 2)$	12	100	99.12	0.5
	$(512, 160^2, 12.18/\sqrt{2\pi}, 2)$	160	100	98.76	1
	$(512, 25601, 12.18/\sqrt{2\pi}, 2)$	161	100	98.79	1
	$(512, 161^2, 12.18/\sqrt{2\pi}, 2)$	161	100	99.22	1
$\left(1024,\ 12289,\ \frac{8}{\sqrt{2\pi}}\right)$	$(256, 7681, 11.31/\sqrt{2\pi}, 1)$	10	100	99.38	0.25
	$(256, 10^4, 11.31/\sqrt{2\pi}, 1)$	10	100	99.99	0.25
	$(256, 13313, 11.31/\sqrt{2\pi}, 2)$	11	99.99	92.91	0.5
	$(256, 11^4, 11.31/\sqrt{2\pi}, 2)$	11	100	98.07	0.5
	$(256, 15361, 11.31/\sqrt{2\pi}, 2)$	12	99.99	99.21	0.5
	$(256, 12^4, 11.31/\sqrt{2\pi}, 2)$	12	100	100	0.5

$$\Delta(D_z; D_{\bar{z}}) \in \Delta(D_z; D_{\hat{z}}) + [-\Delta(D_{\hat{z}}; D_{\bar{z}}), \Delta(D_{\hat{z}}; D_{\bar{z}})].$$

It is easy to check that $\Delta(D_{\hat{z}}; D_{\bar{z}}) = 0$ when $q' = p^k$. Intuitively, $\Delta(D_{\hat{z}}; D_{\bar{z}})$ is supposed to be small when $q' \approx p^k$. Moreover, it can be verified that, when $q > w(p - w)$,

$$\begin{aligned} |D_z(i) - D_{\hat{z}}(i)| &= \left| \frac{N_i}{pq - w(p-w)} - \frac{1}{q} \right| \\ &\leq \max\left\{ \frac{p}{pq - w(p-w)} - \frac{1}{q}, \frac{1}{q} \right\}, \end{aligned}$$

then it follows that when $q > w(p - w)$ and $q' \approx p^k$,

$$\begin{aligned} \Delta(D_z; D_{\bar{z}}) \approx\ & \Delta(D_z; D_{\hat{z}}) \\ &= \frac{1}{2} \sum_{i \in I} \left| \frac{p}{pq - w(p-w)} - \frac{1}{q} \right| \\ &\quad + \frac{1}{2} \sum_{i \notin I} \left| \frac{N_i}{pq - w(p-w)} - \frac{1}{q} \right| \\ &\in \frac{(q-p)w(p-w)}{2q(pq - w(p-w))} + \left[0, \frac{p}{2q}\right] \end{aligned} \tag{4}$$

where $I = \left(\frac{p}{2}, q - \frac{p}{2}\right]$.

Remark 2. For an ideal case where $w = 0$ and $q' = p^k$, following the deductions of Equations (3) and (4), we know that to distinguish infected and uninfected ciphertexts modulo q and p is computationally hard under the assumed hardness of Ring-LWE. In this case, the SETUP totally follows Definition 2.1. In a real scheme, q is fixed as a public parameter thus the attacker may not be able to ensure $w = 0$. However, it is easy to choose q' such that $q' = p^k$ and $w(p - w) \ll pq$, for which infected and uninfected ciphertexts modulo q and p also seem to be indistinguishable.

4.4. Middle terms in decryption

In Section 4.3, the decryption key is not involved in distinguishing infected and uninfected ciphertexts yet. Indeed, given a legitimate decryption key, one would be able to know more information contained in the ciphertexts, for example, the middle term, i.e.

$$M = v - us - \left\lfloor \frac{q}{2} \right\rfloor m \in \mathcal{R}_q$$

where s is the secret key and (u, v) is a ciphertext of the message m.

TABLE 2. Experimental measure of the closeness between infected and uninfected ciphertexts. The comparisons among the distributions over $\mathbb{Z}_q$ are not provided for $q = 2^{32} - 1$, because $2^{32} - 1$ is too large for us to generate accurate statistics.

Parameter		$\Delta_{rk}^{(q)}$	$\Delta_{ru}^{(q)}$	$\Delta_{ku}^{(q)}$	$\Delta_{rk}^{(p)}$	$\Delta_{ru}^{(p)}$	$\Delta_{ku}^{(p)}$
(n, q, σ)	(n', q', σ', τ)						
$\left(512, 12289, \dfrac{12.18}{\sqrt{2\pi}}\right)$	$(128, 7681, 11/\sqrt{2\pi}, 1)$	0.0608	0.0139	0.0600	0.2323	0.0252	0.2319
	$(128, 7681, 11/\sqrt{2\pi}, 2)$	0.0606	0.0138	0.0603	0.2313	0.0252	0.2319
	$(128, 9473, 11/\sqrt{2\pi}, 2)$	0.0251	0.0138	0.0214	0.0668	0.0249	0.0592
	$(128, 8^4, 11/\sqrt{2\pi}, 1)$	0.0184	0.0139	0.0139	0.0227	0.0163	0.0159
	$(128, 9^4, 11/\sqrt{2\pi}, 1)$	0.0184	0.0140	0.0138	0.0285	0.0204	0.0199
	$(128, 9^4, 11/\sqrt{2\pi}, 2)$	0.0184	0.0138	0.0138	0.0289	0.0198	0.0207
	$(128, 10^4, 11/\sqrt{2\pi}, 2)$	0.0187	0.0139	0.0139	0.0353	0.0248	0.0248
$\left(1024, 2^{32} - 1, \dfrac{8}{\sqrt{2\pi}}\right)$	$(256, 7681, 11.31/\sqrt{2\pi}, 1)$				0.2318	0.0177	0.2319
	$(512, 12289, 12.18/\sqrt{2\pi}, 1)$				0.0211	0.0139	0.0152
	$(256, 15361, 11.31/\sqrt{2\pi}, 2)$				0.2591	0.0252	0.2592
	$(512, 25601, 12.18/\sqrt{2\pi}, 2)$				0.0348	0.0201	0.0267
	$(256, 10^4, 11.31/\sqrt{2\pi}, 1)$				0.0249	0.0176	0.0178
	$(512, 111^2, 12.18/\sqrt{2\pi}, 1)$				0.0197	0.0140	0.0140
	$(256, 11^4, 11.31/\sqrt{2\pi}, 2)$				0.0299	0.0212	0.0210
	$(512, 160^2, 12.18/\sqrt{2\pi}, 2)$				0.0284	0.0200	0.0198
	$(512, 161^2, 12.18/\sqrt{2\pi}, 2)$				0.0285	0.0200	0.0201
$\left(1024, 12289, \dfrac{8}{\sqrt{2\pi}}\right)$	$(256, 7681, 11.31/\sqrt{2\pi}, 1)$	0.0604	0.0098	0.0600	0.2320	0.0177	0.2319
	$(256, 13313, 11.31/\sqrt{2\pi}, 2)$	0.0266	0.0099	0.0246	0.0939	0.0212	0.0915
	$(256, 15361, 11.31/\sqrt{2\pi}, 2)$	0.0654	0.0099	0.0653	0.2592	0.0254	0.2592
	$(256, 10^4, 11.31/\sqrt{2\pi}, 1)$	0.0131	0.0098	0.0098	0.0250	0.0176	0.0178
	$(256, 11^4, 11.31/\sqrt{2\pi}, 2)$	0.0133	0.0098	0.0099	0.0300	0.0215	0.0213
	$(256, 12^4, 11.31/\sqrt{2\pi}, 2)$	0.0131	0.0097	0.0097	0.0357	0.0254	0.0253

As mentioned before, the middle term M equals $(re + e_2 - e_1 s)$ for a legitimate ciphertext and $(re + e_2 - e_1 s) + \Delta_v - \Delta_u s$ for an infected ciphertext, where $r, e, e_1, e_2 \hookleftarrow D_{\sigma,\mathcal{R}}$ and $\|\Delta_u\|_\infty, \|\Delta_v\|_\infty \le \dfrac{p}{2}$. For simplicity, we assume that all coefficients of a middle term independently follow the same distribution denoted by D_M (resp. $D_{\overline{M}}$) for uninfected ciphertexts (resp. infected ciphertexts). From the theoretical aspect, we do not prove the statistical or computational indistinguishability between D_M and $D_{\overline{M}}$. Indeed this is a non-trivial problem. A possible solution is to design delicate perturbations Δ_u and Δ_v. Moreover, one may be able to give a rigorous proof when smudging technique [22] or some other ciphertext sanitization technique [23] is applied. However, under such setting, the schemes would be impractical. To formally confirm the indistinguishability between D_M and $D_{\overline{M}}$ given the decryption key is left as a future work. From the practical aspect, we will compare the statistics of D_M and $D_{\overline{M}}$ experimentally in the next section. Indicated by experimental results, when p is small, it is not easy to distinguish the middle terms yielded by infected and uninfected ciphertexts in practice.

5. PRACTICAL IMPLEMENTATION

We implemented the klepto scheme in Sage [24] and ran experiments to observe the impact of the backdoor from the aspects discussed in Section 4. Three sets of public Ring-LWE parameters we discussed are $(n, q, \sigma) = (512, 12289, 12.18/\sqrt{2\pi})$, $(1024, 2^{32} - 1, 8/\sqrt{2\pi})$ and $(1024, 12289, 8/\sqrt{2\pi})$. The first two tuples were discussed in [21] and [25], respectively and claimed to provide at least 128-bits security, and the third one is adapted from the second one by decreasing the modulus like [15]. For each public parameter tuple, different backdoor parameter tuples (n', q', σ', τ) were discussed and compared. For each parameter set, we generated 100 random instances and encrypted 100 random plaintexts for each instance so that 10 000 ciphertexts were collected totally.

Table 1 shows how the backdoor parameter affects decryption failures for legitimate receivers and attackers. Experimental results confirm that for fixed (n, q, σ), smaller $p = \left\lceil q'^{\frac{n'}{n}} \right\rceil$ leads to less decryption failures for legitimate receivers. The proportion of leaked message equals $\dfrac{\tau n'}{n}$, thus a direct approach to

TABLE 3. Experimental measure of $\mathbb{E}_M$, $\mathbb{E}_{\overline{M}}$, σ_M and $\sigma_{\overline{M}}$.

Parameter		$\left[q'^{\frac{n'}{n}}\right]$	$\mathbb{E}_M$	$\mathbb{E}_{\overline{M}}$	σ_M	$\sigma_{\overline{M}}$
(n, q, σ)	(n', q', σ', τ)					
$\left(512, 12289, \frac{12.18}{\sqrt{2\pi}}\right)$	$(128, 8^4, 11/\sqrt{2\pi}, 1)$	8	0.4949	5.2413	752.0630	795.3722
	$(128, 9^4, 11/\sqrt{2\pi}, 1)$	9	0.0578	0.3553	755.5527	806.6366
	$(128, 7681, 11/\sqrt{2\pi}, 1)$	10	0.5509	−2.8801	755.9129	821.8440
	$(128, 9^4, 11/\sqrt{2\pi}, 2)$	9	0.5805	0.5908	757.5418	810.0109
	$(128, 7681, 11/\sqrt{2\pi}, 2)$	10	−0.1230	−0.4453	757.1177	822.7273
	$(128, 9473, 11/\sqrt{2\pi}, 2)$	10	−0.0263	−4.7967	752.7451	818.1350
	$(128, 10^4, 11/\sqrt{2\pi}, 2)$	10	−0.2817	0.9695	755.7126	820.5229
$\left(1024, 2^{32}-1, \frac{8}{\sqrt{2\pi}}\right)$	$(256, 7681, 11.31/\sqrt{2\pi}, 1)$	10	0.1186	−2.3709	461.4615	549.4083
	$(256, 10^4, 11.31/\sqrt{2\pi}, 1)$	10	−0.0850	−0.2645	460.7307	548.3159
	$(512, 12289, 12.18/\sqrt{2\pi}, 1)$	111	−0.0910	−0.1882	461.8267	3312.2316
	$(512, 111^2, 12.18/\sqrt{2\pi}, 1)$	111	0.0283	−0.4126	461.2856	3307.7249
	$(256, 11^4, 11.31/\sqrt{2\pi}, 2)$	11	−0.0428	0.0240	460.8434	562.8343
	$(256, 15361, 11.31/\sqrt{2\pi}, 2)$	12	−0.1359	−1.1515	460.8889	581.9961
	$(512, 160^2, 12.18/\sqrt{2\pi}, 2)$	160	0.2552	1.1382	461.9073	4766.3459
	$(512, 25601, 12.18/\sqrt{2\pi}, 2)$	161	0.1939	−0.1191	461.0203	4755.4355
	$(512, 161^2, 12.18/\sqrt{2\pi}, 2)$	161	0.2985	−0.7573	459.8787	4758.0793
$\left(1024, 12289, \frac{8}{\sqrt{2\pi}}\right)$	$(256, 7681, 11.31/\sqrt{2\pi}, 1)$	10	−0.1168	−1.3590	461.1512	549.2173
	$(256, 10^4, 11.31/\sqrt{2\pi}, 1)$	10	−0.0759	0.0225	460.5395	548.0335
	$(256, 13313, 11.31/\sqrt{2\pi}, 2)$	11	0.1015	0.2012	460.4020	561.4611
	$(256, 11^4, 11.31/\sqrt{2\pi}, 2)$	11	0.1423	−0.0885	459.6836	560.9952
	$(256, 15361, 11.31/\sqrt{2\pi}, 2)$	12	0.1803	−4.6646	459.9769	581.8082
	$(256, 12^4, 11.31/\sqrt{2\pi}, 2)$	12	0.1378	−3.0988	461.2178	583.7178

leak more message is to increase τ or n'. For larger τ, Equation (1) shows that q' should be increased accordingly to ensure correct decryption for attacker, which may weaken the security of the backdoor scheme, but it does not seem to increase p too much. For larger n', the backdoor seems more secure, but p would be increased a lot, which significantly affects the correct decryption for legitimate receivers when q is not so large. From experimental results, we also conclude that it is more convenient to add backdoors to Ring-LWE encryption scheme with very large q, because large q allows more backdoor parameter tuples for stealing different proportions of message.

We also measured experimentally the closeness between infected and uninfected ciphertexts via statistical distances following the discussions in Section 4.3. We assume that all coefficients of uninfected (resp. infected) ciphertexts follow the same distribution D_z (resp. $D_{\bar{z}}$), and measure them together to collect more samples. Let $k = n/n'$ and $p = \left[q'^{\frac{1}{k}}\right]$. For a ciphertext (u, v) where $u = (u_0, \ldots, u_{n-1})$ and $v = (v_0, \ldots, v_{n-1})$, we generate $2n'$ values $a_0, \ldots, a_{n'-1}$, $b_0, \ldots, b_{n'-1} \in \mathbb{Z}_{p^k}$ where $a_i = \sum_{j=0}^{k-1} p^j (u_{ik+j} \bmod p)$ and $b_i = \sum_{j=0}^{k-1} p^j (v_{ik+j} \bmod p)$. We also assume that these $2n'$ values follow the same distribution denoted by D'_z (resp. $D'_{\bar{z}}$) when (u, v) is uninfected (resp. infected). Let $\Delta_{rk}^{(q)} = \Delta(D_z; D_{\bar{z}})$ and $\Delta_{rk}^{(p)} = \Delta(D'_z; D'_{\bar{z}})$. For better comparisons, we also considered $\Delta_{ru}^{(q)} = \Delta(D_z; U(\mathbb{Z}_q))$, $\Delta_{ku}^{(q)} = \Delta(D_{\bar{z}}; U(\mathbb{Z}_q))$ and $\Delta_{ru}^{(p)} = \Delta(D'_z; U(\mathbb{Z}_{p^k}))$ and $\Delta_{ku}^{(p)} = \Delta(D'_{\bar{z}}; U(\mathbb{Z}_{p^k}))$. All experimental results are listed in Table 2.

We compare experimental results with our estimations in Equations (3) and (4). Let $\delta_1 = \frac{w(p-w)}{pq}$ where $w = (q \bmod p)$ and $\delta_2 = \frac{p^k - q'}{p^k}$ that are closely related to statistical distances between ciphertext distributions. We observe that $\Delta_{rk}^{(p)}$ and $\Delta_{ku}^{(p)}$ are approximately equal to δ_2 when δ_2 is not very small (e.g. $\delta_2 > 0.18$), and these two measures are less than 0.04 when $\delta_2 = 0$. That suggests that to hide the backdoor well, the attacker should choose parameters such that $\delta_2 = 0$ or very small. We also notice that $\Delta_{ru}^{(p)}$ may exceed the upper bound implicit in Equation (3), i.e. $\Delta_{ru}^{(p)} \leq k\delta_1$. There are two possible causes for this phenomenon: (1) the inherent difference between Ring-LWE ciphertext distribution and uniform distribution, and (2) the experimental error. Furthermore, when $q' \neq p^k$, $\Delta_{rk}^{(q)}$ may be relatively larger, which is implicit in the approximation in Equation (4).

To compare the middle terms with respect to uninfected and infected ciphertexts, we considered the distributions D_M

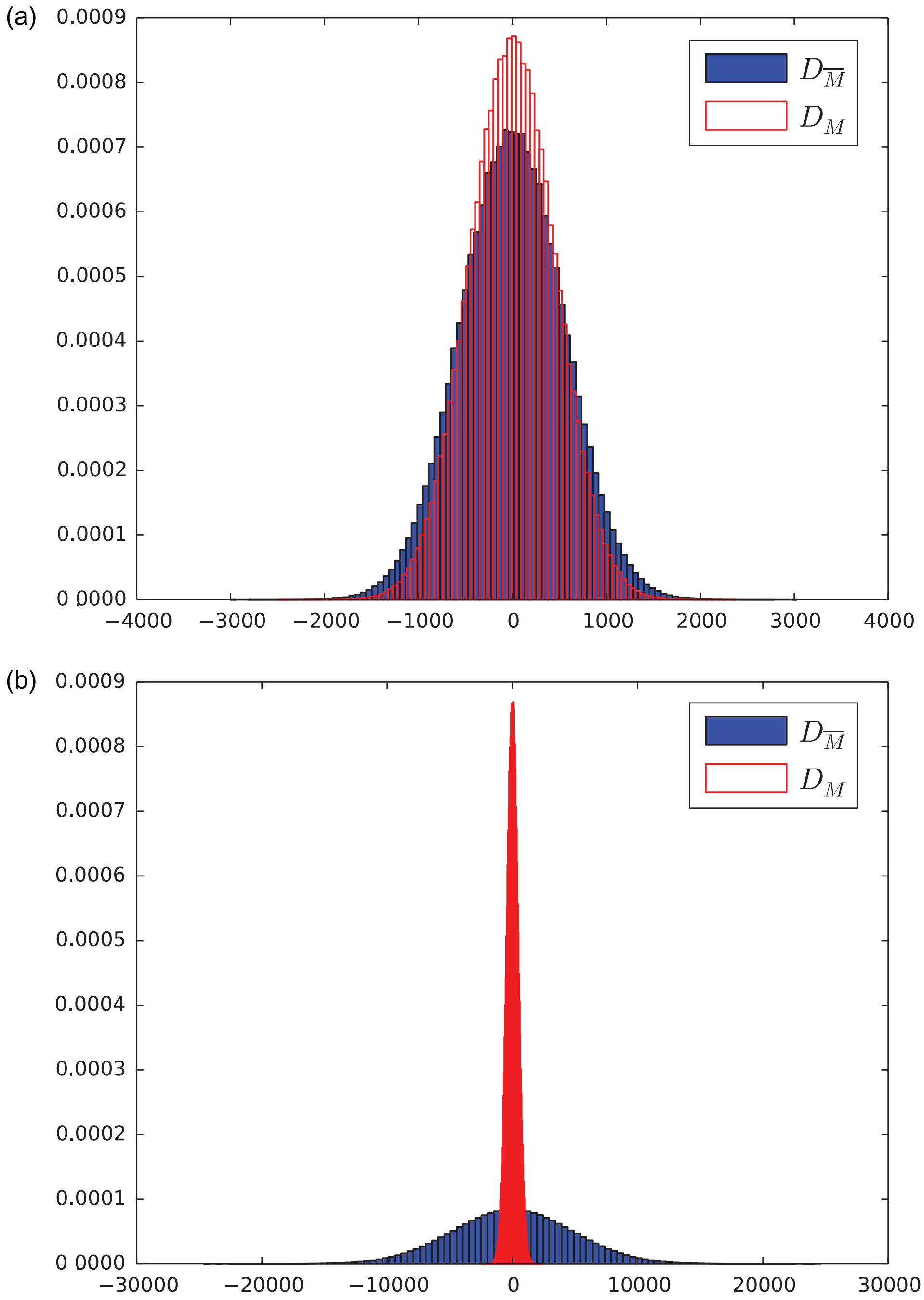

FIGURE 1. Comparisons between D_M and $D_{\overline{M}}$ for different p's when $(n, q, \sigma) = (1024, 2^{32} - 1, 8/\sqrt{2\pi})$. (**a**) $(n', q', \sigma', \tau) = (256, 10^4, 11.31/\sqrt{2\pi}, 1)$, $\left\lfloor q'^{\frac{n'}{n}} \right\rceil = 10$ and (**b**) $(n', q', \sigma', \tau) = (512, 161^2, 12.18/\sqrt{2\pi}, 2)$, $\left\lfloor q'^{\frac{n'}{n}} \right\rceil = 161$.

and $D_{\overline{M}}$ that are defined in Section 4.4. First, we experimentally measured the statistics of D_M and $D_{\overline{M}}$. For $X \hookleftarrow D_M$ and $\overline{X} \hookleftarrow D_{\overline{M}}$, let $\mathbb{E}_M$ and $\mathbb{E}_{\overline{M}}$ be the expectation of X and $\overline{X}$, and σ_M and $\sigma_{\overline{M}}$ be the standard deviation, respectively. Table 3 illustrates the experimental results. Since both $\mathbb{E}_M$ and $\mathbb{E}_{\overline{M}}$ are close to 0, coefficients of middle terms should be of symmetry. It also can be observed that $\sigma_{\overline{M}}$ is usually larger than σ_M and the difference seems to depend on the magnitudes of Δ_u, Δ_v, i.e. $p = \left\lfloor q'^{\frac{n'}{n}} \right\rceil$. For better illustrations, we also plot Figure 1 to show the comparison between D_M and $D_{\overline{M}}$ corresponding to small and large p respectively. Indeed D_M and $D_{\overline{M}}$ behave different: the middle term with respect to an infected ciphertext tends to be of smaller size. However, as p decreases, D_M seems to converge to $D_{\overline{M}}$. Thus, when p is small, to detect the backdoor may still require sufficiently many ciphertexts even though the decryption key is available.

6. CONCLUSION

In this paper, we propose a construction of backdoor for Ring-LWE encryption scheme and study it theoretically and experimentally. As indicated by our analysis and experiments, it seems practical to modify Ring-LWE scheme into a klepto variant in such a way, especially for the scheme with large modulus. Therefore, we believe that black-box implementations of Ring-LWE cryptographic algorithms are potentially dangerous and not supposed to be accepted easily. Our analysis in Section 4 can be used for detecting such backdoors preliminarily. It would be meaningful to exploit other cryptanalysis and tools to give an elaborative backdoor detection. We leave it as future work.

FUNDING

This research was supported by The National Key Research and Development Program of China (Project no. 2017YFA0303903) and 973 National Program on Key Basic Research Project of China (Project no. 2013CB834205).

ACKNOWLEDGEMENTS

We thank Léo Ducas and Prof. Xiaoyun Wang for helpful discussions and comments. We also would like to thank the anonymous reviewers for their constructive comments.

REFERENCES

[1] Young, A. and Yung, M. (1996) The Dark Side of ‘Black-Box’ Cryptography, or: Should We Trust Capstone? *Advances in Cryptology – CRYPTO 1996*, Santa Barbara, CA, USA, August 18–22, pp. 89–103. Springer, Berlin, Heidelberg.

[2] Young, A. and Yung, M. (1996) Cryptovirology: Extortion-Based Security Threats and Countermeasures. *IEEE Symp. Security and Privacy*, Oakland, CA, USA, May 6–8, pp. 129–140. IEEE Computer Society, Washington, D.C., USA.

[3] Young, A. and Yung, M. (1997) Kleptography: Using Cryptography Against Cryptography. *Advances in Cryptology – EUROCRYPT 1997*, Konstanz, Germany, May 11–15, pp. 62–74. Springer, Berlin, Heidelberg.

[4] Young, A. and Yung, M. (2004) *Malicious Cryptography: Exposing Cryptovirology*. Wiley, Hoboken, NJ, USA.

[5] Hoffstein, J., Pipher, J. and Silverman, J.H. (1998) NTRU: A Ring-Based Public Key Cryptosystem. *Proc. ANTS 1998,* Portland, OR, USA, June 21–25, pp. 267–288. Springer, Berlin, Heidelberg.

[6] Stehlé, D. and Steinfeld, R. (2011) Making NTRU as Secure as Worst-Case Problems Over Ideal Lattices. *Advances in Cryptology – EUROCRYPT 2011*, Tallinn, Estonia, May 15–19, pp. 27–47. Springer, Berlin, Heidelberg.

[7] Yu, Y., Xu, G. and Wang, X. (2017) Provably Secure NTRU Instances over Prime Cyclotomic Rings. *Public-Key Cryptography – PKC 2017*, Amsterdam, The Netherlands, March 28–31, 2017, Proceedings, Part I, pp. 409–434. Springer, Berlin, Heidelberg.

[8] Yu, Y., Xu, G. and Wang, X. (2017). Provably Secure NTRUEncrypt Over more General Cyclotomic Rings. Cryptology ePrint Archive, Report 2017/304, Available at https://eprint.iacr.org/2017/304.

[9] Regev, O. (2005) On Lattices, Learning with Errors, Random Linear Codes, and Cryptography. *Proc. STOC 2005*, Baltimore, MD, USA, May 22–24, pp. 84–93. ACM, New York City, USA.

[10] Lyubashevsky, V., Peikert, C. and Regev, O. (2010) On Ideal Lattices and Learning with Errors over Rings. *Advances in Cryptology – EUROCRYPT 2010*, French Riviera, May 30–June 3, pp. 1–23. Springer, Berlin, Heidelberg.

[11] Chen, Y. and Nguyen, P.Q. (2011) BKZ 2.0: Better Lattice Security Estimates. *Advances in Cryptology – ASIACRYPT 2011*, Seoul, South Korea, December 4–8, pp. 1–20. Springer, Berlin, Heidelberg.

[12] Coppersmith, D. and Shamir, A. (1997) Lattice Attacks on NTRU. *Advances in Cryptology – EUROCRYPT 1997*, Konstanz, Germany, May 11–15, pp. 52–61. Springer, Berlin, Heidelberg.

[13] Howgrave-Graham, N. (2007) A Hybrid Lattice-Reduction and Meet-in-the-Middle Attack Against NTRU. *Advances in Cryptology – CRYPTO 2007*, Santa Barbara, CA, USA, August 19–23, pp. 150–169. Springer, Berlin, Heidelberg.

[14] Kirchner, P. and Fouque, P.-A. (2015) An Improved BKW Algorithm for LWE with Applications to Cryptography and Lattices. *Advances in Cryptology – CRYPTO 2015*, Santa Barbara, CA, USA, August 16–20, pp. 43–62. Springer, Berlin, Heidelberg.

[15] Alkim, E., Ducas, L., Pöppelmann, T. and Schwabe, P. (2016) Post-Quantum Key Exchange—A New Hope. *USENIX Security 2016*, Austin, TX, USA, August 10–12, pp. 327–343. USENIX Association, Berkeley, CA, USA.

[16] Kwant, R., Lange, T. and Thissen, K. (2017) Lattice Klepto: Turning Post-quantum Crypto Against Itself. *Selected Areas in Cryptography – SAC 2017*, Ottawa, Ontario, Canada, August 16–18. Springer, Berlin, Heidelberg.

[17] Micciancio, D. and Regev, O. (2007) Worst-case to average-case reductions based on Gaussian measures. *SIAM J. Comput.*, **37**, 267–302.

[18] Banaszczyk, W. (1993) New bounds in some transference theorems in the geometry of numbers. *Math. Ann.*, **296**, 625–635.

[19] Langlois, A. and Stehlé, D. (2015) Worst-case to average-case reductions for module lattices. *Designs Codes Cryptogr.*, **75**, 565–599.

[20] Lyubashevsky, V., Peikert, C. and Regev, O. (2013) A Toolkit for Ring-LWE Cryptography. *Advances in Cryptology – EUROCRYPT 2013*, Athens, Greece, May 26–30, pp. 35–54. Springer, Berlin, Heidelberg.

[21] Liu, Z., Seo, H., Roy, S.S., Großschädl, J., Kim, H. and Verbauwhede, I. (2015) Efficient Ring-LWE Encryption on

8-bit AVR Processors. *Cryptographic Hardware and Embedded Systems – CHES 2015*, Saint-Malo, France, September 13–16, pp. 663–682. Springer, Berlin, Heidelberg.

[22] Asharov, G., Jain, A., López-Alt, A., Tromer, E., Vaikuntanathan, V. and Wichs, D. (2012) Multiparty Computation with Low Communication, Computation and Interaction via Threshold FHE. *Advances in Cryptology – EUROCRYPT 2012*, Cambridge, UK, April 15–19. Proceedings, pp. 483–501. Springer, Berlin, Heidelberg.

[23] Ducas, L. and Stehlé, D. (2016) Sanitization of FHE Ciphertexts. *Advances in Cryptology – EUROCRYPT 2016*, Vienna, Austria, May 8–12. Proceedings, Part I, pp. 294–310. Springer, Berlin, Heidelberg.

[24] The Sage Development Team SageMath: A Free Open-Source Mathematics Software System. Available at https://www.sagemath.org/.

[25] Bos, J.W., Costello, C., Naehrig, M. and Stebila, D. (2015) Post-Quantum Key Exchange for the TLS Protocol from the Ring Learning with Errors Problem. *IEEE Symp. Security and Privacy*, San Jose, CA, USA, May 17–21, pp. 553–570. IEEE Computer Society, Washington, DC, USA.