Chosen-Ciphertext Security from Subset Sum*

Sebastian Faust[1], Daniel Masny[1]**, and Daniele Venturi[2]***

[1] Horst-Görtz Institute for IT Security and Faculty of Mathematics, Ruhr-Universität Bochum, Bochum, Germany
[2] Department of Computer Science, Sapienza University of Rome, Rome, Italy

Abstract. We construct a public-key encryption (PKE) scheme whose security is polynomial-time equivalent to the hardness of the Subset Sum problem. Our scheme achieves the standard notion of indistinguishability against chosen-ciphertext attacks (IND-CCA) and can be used to encrypt messages of arbitrary polynomial length, improving upon a previous construction by Lyubashevsky, Palacio, and Segev (TCC 2010) which achieved only the weaker notion of semantic security (IND-CPA) and whose concrete security decreases with the length of the message being encrypted.
At the core of our construction is a trapdoor technique which originates in the work of Micciancio and Peikert (Eurocrypt 2012).

Keywords: Public-key cryptography, chosen-ciphertext security, subset sum problem.

1 Introduction

Public-Key Encryption (PKE) is perhaps the most basic application of public-key cryptography [10]. Intuitively a PKE scheme allows Alice to encrypt a message M for Bob, given just Bob's public key pk; the received ciphertext C can be decrypted by Bob using the secret key sk corresponding to pk.

Security of a PKE scheme can be formulated in different ways, depending on the assumed adversarial capabilities. The most basic and natural notion is that of indistinguishability against chosen-plaintext attacks (IND-CPA, a.k.a. semantic security) [14]; here we demand that a passive (computationally bounded) adversary only given pk should not be able to distinguish the encryption of two (adversarially chosen) messages M_0, M_1.

Whilst already sufficient for some applications, IND-CPA security is not enough to deal with active adversaries. Hence, researchers have put forward

** Supported by DFG Research Training Group GRK 1817/1.
*** Supported by the European Commission (Directorate General Home Affairs) under the GAINS project HOME/2013/CIPS/AG/4000005057, and by the European Unions Horizon 2020 research and innovation programme under grant agreement No 644666.

stronger security notions. The de-facto standard notion of security for PKE is that of indistinguishability against chosen-ciphertext attacks [29] (IND-CCA), where we now demand that an active (computationally bounded) adversary given pk should not be able to distinguish the encryption of two (adversarially chosen) messages M_0, M_1 even given access to an oracle decrypting arbitrarily chosen ciphertexts.[3]

By now we dispose of many PKE schemes satisfying IND-CCA security under a variety of assumptions, including factoring [15], decisional and computational Diffie-Hellman [8,6], and learning parity with noise [18].

THE SUBSET SUM ASSUMPTION. Since its introduction, the Subset Sum problem has been considered a valid alternative to number-theoretic assumptions. In its basic computational version, the Subset Sum problem $\mathrm{SS}(n, q)$ (parametrized by integers q and n) asks to find a secret vector $\mathbf{s} \in \{0,1\}^n$ given a vector $\mathbf{a} \in \mathbb{Z}_q^n$ together with the target value $T := \langle \mathbf{a} \cdot \mathbf{s} \rangle \bmod q$, where both $\mathbf{a}$ and $\mathbf{s}$ are chosen uniformly at random, and $\langle \cdot, \cdot \rangle$ denotes the inner product. The hardness of $\mathrm{SS}(n, q)$ depends on the so-called *density*, which is defined by the ratio $\delta := n/\log q$. In case $\delta < 1/n$ or $\delta > n/\log^2 n$, the problem can be solved in polynomial time [20,13,12,21,32]. In case δ is $o(1)$ or even as small as $O(1/\log n)$, the problem is considered to be hard. The best classical algorithm for solving Subset Sum is due to [19], and takes sub-exponential time for solving instances with $\delta = o(1)$ and time $2^{(\ln 2/2 + o(1))n/\log\log n}$ for instances with $\delta = O(1/\log n)$.

One nice feature of the Subset Sum problem is its believed hardness against quantum attacks. At the time of writing, the best quantum attack—due to Bernstein et al. [3]—on Subset Sum requires complexity $2^{(0.241+o(1))n}$ to solve a random instance of the problem.

PKE FROM SUBSET SUM. The first PKE scheme based on the hardness of Subset Sum was constructed in the seminal work of Ajtai and Dwork [2], who presented a scheme whose semantic security is as hard to break as solving worst-case instances of a lattice problem called "the unique shortest vector problem" (uSVP). It is well known that Subset Sum can be reduced to uSVP [20,13].

A disadvantage of the scheme in [2] (and its extensions [30,31,27]) is that they are based on Subset Sum only in an indirect way (i.e., via a non-tight reduction to uSVP). This limitation was overcome by the work of Lyubashevsky, Palacio, and Segev [22] that proposed a new PKE scheme achieving IND-CPA security with a simple and direct reduction to solving random instances of the Subset Sum problem.

More precisely, the security of the scheme in [22] is based on the assumption that a random instance $(\mathbf{a}, T)$ of the Subset Sum problem is indistinguishable from uniform. Such a decisional variant of the problem was shown to be equivalent to the above introduced computational version (i.e., to the task of recovering $\mathbf{s}$) by Impagliazzo and Naor [16].

[3] Clearly, the decryption oracle cannot be queried on the challenge ciphertext.

1.1 Our Contributions and Techniques

The work of [22] left as an explicit open problem to construct a PKE scheme achieving IND-CCA security with a direct reduction to the hardness of Subset Sum.

Contributions. In this paper we present a new PKE scheme resolving the above open problem. Previous to our work, the only known PKE schemes with IND-CCA security from Subset Sum were the ones based on uSVP [27,28] (which are not directly based on the hardness of Subset Sum). An additional advantage of our scheme is that it can be used to encrypt an arbitrary polynomial number of bits; this stands in sharp contrast with the scheme of [22], whose concrete security starts to decrease when encrypting messages of length longer than $n \log n$ (where n is, as usual, the Subset Sum dimension).[4] The theorem below summarizes our main result.

Theorem 1 (Main result, informal). *For prime*[5] $p = \Theta(m^2 n \log^2 n)$ *and* $q = p^m$ *there exists a PKE scheme with IND-CCA security based on the hardness of* $\mathrm{SS}(n, q)$.

Techniques. Our scheme (similar to the one of [22]) is based on the decisional variant of $\mathrm{SS}(n, p^m)$, where p is a prime and m is an integer. The main observation (also made in [22]) is that, in case $q = p^m$, the target value $T := \langle \mathbf{a} \cdot \mathbf{s} \rangle \bmod p^m$ written in base p is equal to $\mathbf{A}\mathbf{s} + e(\mathbf{A}, \mathbf{s})$ where $\mathbf{A} \in \mathbb{Z}_p^{m \times n}$ is a matrix whose i-th column corresponds to the i-th element of vector $\mathbf{a}$ written in base p, and $e(\mathbf{A}, \mathbf{s})$ is a vector in $\mathbb{Z}_p^m$ (function of $\mathbf{A}$ and $\mathbf{s}$) which corresponds to the carries when performing "grade-school" addition. This particular structure resembles the structure of an instance of the learning with errors (LWE) problem [31], with the important difference that the noise term is "deterministic" and, in fact, completely determined by the matrix $\mathbf{A}$ and the vector $\mathbf{s}$.

We use the above similarity between LWE and Subset Sum to construct our new PKE scheme, using a trapdoor technique due to Micciancio and Peikert [24]. Essentially our scheme relies on a tag-based trapdoor function, where the trapdoor is associated with a hidden tag. Whenever the function is evaluated w.r.t. the hidden tag, the trapdoor disappears and the function is hard to invert; for all other tags the function can be inverted efficiently given the trapdoor. Using the leftover hash lemma, one can switch the hidden tag without the adversary noticing.

The above technique allows us to prove that our PKE scheme achieves a weaker (tag-based) CCA notion. This means that each ciphertext is associated with a tag τ, and in the security game the adversary has to commit in advance

[4] In particular, for message length n^2 the scheme of [22] can be broken in polynomial time.

[5] In the original version, we have proposed a smaller modulus that is even. We are not able to prove security for the original choices of parameters. For applying the leftover hash lemma we need a prime modulus p greater than 2 and we need a larger modulus to ensure that our decryption does not leak the secret key.

to the tag τ^* which will be associated with the challenge ciphertext.[6] In the security proof we first switch the tag associated with the hidden trapdoor with the challenge tag (using the trapdoor technique outlined above). Now, the simulator is not able to decrypt a message related to the challenge tag which allows us to argue about indistinguishability of the PKE scheme.

It is well known that the above weak tag-based CCA notion can be generically enhanced to full-fledged IND-CCA security using a one-time signature scheme [17]. This allows us to conclude Theorem 1.

EFFICIENCY. Let ℓ be the length of the messages to be encrypted, and denote by n, p and m the parameters of the Subset Sum problem. The secret key of our PKE scheme consists of a binary matrix of dimension $n \times m$; the public key consists of 3 matrices of elements in $\mathbb{Z}_p$, with dimensions (respectively) $m \times n$, $n \times n$, and $\ell \times n$. A ciphertext consists of 3 vectors of elements in $\mathbb{Z}_p$, with dimensions (respectively) m, n, and ℓ.

1.2 Related Work

Pioneered by Merkle and Hellman [23], the first construction of PKE schemes based on Subset Sum were based on instances of the problem with special structure. All these constructions have been subsequently broken. (See [26] for a survey.)

In a seminal paper, Impagliazzo and Naor [16] presented constructions of universal one-way hash functions, pseudorandom generators and bit commitment schemes based on the hardness of random instances of Subset Sum.

Besides constructing PKE schemes, [22] additionally presents an oblivious transfer protocol with security against malicious senders and semi-honest receivers. The Subset Sum problem has also recently been used to solve the problem of outsourced pattern matching [11] in the cloud setting.

2 Preliminaries

For two distributions $\mathcal{D}$ and $\mathcal{D}'$ over Ω, $\mathcal{D}(x)$ is the probability assigned to $x \in \Omega$ and $\Delta[\mathcal{D}, \mathcal{D}'] := \frac{1}{2} \sum_{x \in \Omega} |\mathcal{D}(x) - \mathcal{D}'(x)|$ is the statistical distance between $\mathcal{D}$ and $\mathcal{D}'$. We denote with $x \leftarrow X$ that x is sampled according to the distribution X. If X is a set, then this denotes that x is sampled uniformly at random from X. $\lfloor \cdot \rceil_2 : \mathbb{Z}_p \to \mathbb{Z}_2$ is the rounding function defined by $\lfloor x \rceil_2 := \lfloor x \cdot \frac{2}{p} \rceil$.

Vectors and matrices are denoted in boldface. For two vectors $\mathbf{u}, \mathbf{v}$, with $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$, the inner product between $\mathbf{u}$ and $\mathbf{v}$ is defined as $\langle \mathbf{u}, \mathbf{v} \rangle := \sum_{i=1}^{n} u_i \cdot v_i$. We represent elements in $\mathbb{Z}_p$ by integers in the range $[-(p-1)/2; (p-1)/2]$. For an element $v \in \mathbb{Z}_p$, its length, denoted by $|v|$ is the absolute value of its representative in the range $[-(p-1)/2; (p-1)/2]$. For a vector $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{Z}_p^n$, we define $\|\mathbf{v}\|_\infty := \max_{1 \le i \le n} |v_i|$.

[6] Decryption queries for the challenge tag τ^* are disallowed.

We say that a function ν is negligible in the security parameter n, if it is asymptotically smaller than the inverse of any polynomial in n, i.e. $\nu(n) = n^{-\omega(1)}$. An algorithm A is probabilistic polynomial-time (PPT) if A is randomized, and for any input $x, r \in \{0,1\}^*$ the computation of $\mathsf{A}(x;r)$ (i.e., A with input x and random coins r) terminates in at most $\text{poly}(|x|)$ steps.

In the following, we will need to bound the norm on the inner product between vectors. For this, we will use the Hoeffding bound.

Lemma 1 (Hoeffding Bound). *For $n \in \mathbb{N}$, $c \in \mathbb{R}$, let $X_1, \ldots, X_n$ be independent random variables in range $[a,b]$, and $X := \sum_{i=1}^{n} X_i$, then*

$$\Pr[|X - \mathbb{E}[X]| \geq c] \leq 2e^{-\frac{2c^2}{n(a-b)^2}}.$$

2.1 Subset Sum

Traditionally a Subset Sum $\text{SS}(n,q)$ instance is defined as $\mathbf{a} := (a_1, \ldots, a_n)$ and a target $T := \langle \mathbf{a}, \mathbf{s} \rangle \bmod q$, where the goal is to recover $\mathbf{s} \in \{0,1\}^n$. For a modulus $q = p^m$, Lyubashevsky, Palacio, and Segev [22] gave an alternative description which shows its similarities with the LWE problem more clearly. First they define matrix $\mathbf{A} \in \mathbb{Z}_p^{m \times n}$, where $a_{j,i} := \left\lfloor \frac{a_i}{p^{j-1}} \right\rfloor \bmod p$. Thus,

$$\mathbf{A} \odot \mathbf{s} := \sum_{i=1}^{n} s_i \cdot \left(\sum_{j=1}^{m} a_{j,i} p^{j-1} \right) \bmod p^m = \sum_{i=1}^{n} s_i \cdot a_i \bmod p^m = \langle \mathbf{a}, \mathbf{s} \rangle \bmod p^m$$

where $\mathbf{s} \in \{0,1\}^n$. Notice that when $\mathbf{As}$ is the matrix vector multiplication mod$p$, then $\mathbf{A} \odot \mathbf{s} = \mathbf{As} + e(\mathbf{A}, \mathbf{s}) \bmod p \in \mathbb{Z}_p^m$. Here $e(\mathbf{A}, \mathbf{s})_1 := 0$, and for $1 < j \leq m$ the j-th component of $e(\mathbf{A}, \mathbf{s})$ is given by

$$e(\mathbf{A}, \mathbf{s})_j := \left\lfloor \frac{\sum_{i=1}^{n} s_i a_{j-1,i}}{p} \right\rfloor + c_j \mod p,$$

for carry c_j which is recursively defined by $c_2 := 0$ and

$$c_j := \left\lfloor \frac{(\sum_{i=1}^{n} s_i a_{j-1,i}) \bmod p + e(\mathbf{A}, \mathbf{s})_{j-1}}{p} \right\rfloor \mod p.$$

Since c_j is small, and moreover it is the only part of $e(\mathbf{A}, \mathbf{s})_j$ which depends on $e(\mathbf{A}, \mathbf{s})_{j-1}$, one has that $e(\mathbf{A}, \mathbf{s})_j - c_j$ is bound by the Hoeffding bound. This implies an overall bound on $e(\mathbf{A}, \mathbf{s})_j$:

Lemma 2 ([22] Lemma 3.3). *For any $n, m \in \mathbb{N}$, prime modulus p and $\mathbf{s} \in \{0,1\}^n$, there exists a negligible function $\nu : \mathbb{N} \to [0,1]$ such that*

$$\Pr_{\mathbf{A} \leftarrow \mathbb{Z}_p^{m \times n}} [\|e(\mathbf{A}, \mathbf{s})\|_\infty \geq \sqrt{n} \log n] \leq \nu(n).$$

The main difference between Subset Sum and LWE is that error term $e(\mathbf{A}, \mathbf{s})$ is uniquely determined given $\mathbf{A}$ and $\mathbf{s}$ where as in case of LWE, error e is sampled from a discrete Gaussian distribution independent of $\mathbf{A}$, $\mathbf{s}$.

The Subset Sum assumption. A $\mathrm{SS}(n, p^m)$ instance has the following distribution:

$$\mathcal{D}_{\mathrm{SS}(n,p^m)} := \{(\mathbf{A}, \mathbf{A} \odot \mathbf{s}) \mid \mathbf{A} \leftarrow \mathbb{Z}_p^{m \times n}, \mathbf{s} \leftarrow \{0,1\}^n\}.$$

The challenge is to distinguish $\mathcal{D}_{\mathrm{SS}(n,p^m)}$ from a uniform $(\mathbf{A}, \mathbf{b}) \in \mathbb{Z}_p^{m \times n} \times \mathbb{Z}_p^m$. The advantage of an algorithm A in breaking the $\mathrm{SS}(n, p^m)$ assumption is

$$\mathbf{Adv}_{\mathrm{SS}(n,p^m)}(\mathsf{A}) = |\Pr[\mathsf{A}(\mathbf{A}, \mathbf{b}) = 1] - \Pr[\mathsf{A}(\mathbf{A}', \mathbf{b}') = 1]|,$$

where $(\mathbf{A}, \mathbf{b}) \leftarrow \mathcal{D}_{\mathrm{SS}(n,p^m)}$ and $(\mathbf{A}', \mathbf{b}') \leftarrow \mathbb{Z}_p^{m \times n} \times \mathbb{Z}_p^m$. It was shown by Impagliazzo and Naor [16] that this decisional version of Subset Sum is as hard as recovering the hidden vector $\mathbf{s}$.

Re-randomizing Subset Sum. We use a technique introduced by Lyubashevsky [21] allowing to re-randomize a Subset Sum sample. This technique is based on the leftover hash lemma:

Lemma 3 (Leftover hash lemma). *For prime p, $2m \geq n+1+\omega(\log n + 1)/\log p$ and polynomial ℓ, there exists a negligible function $\nu : \mathbb{N} \to [0,1]$ such that the statistical distance*

$$\Delta\big[(\mathbf{A}, \mathbf{RA}, \mathbf{a}, \mathbf{Ra}), (\mathbf{A}, \mathbf{C}, \mathbf{a}, \mathbf{c})\big] \leq \nu(n),$$

for $\mathbf{A} \leftarrow \mathbb{Z}_p^{m \times n}$, $\mathbf{R} \leftarrow [-\sqrt{p}/2, \sqrt{p}/2]^{\ell \times m}$, $\mathbf{a} \leftarrow \mathbb{Z}_p^n$, $\mathbf{C} \leftarrow \mathbb{Z}_p^{\ell \times n}$, $\mathbf{c} \leftarrow \mathbb{Z}_p^\ell$.

A Subset Sum sample $(\mathbf{A}, \mathbf{b}) \leftarrow \mathcal{D}_{\mathrm{SS}(n,p^m)}$ can now be re-randomized to $(\mathbf{RA}, \mathbf{Rb})$ where $\mathbf{RA}$ is statistically close to uniform given $\mathbf{A}, \mathbf{b}$ and $\mathbf{Rb}$. Note that $(\mathbf{RA}, \mathbf{Rb})$ is not $\mathrm{SS}(n, p^m)$-distributed anymore.

Given this re-randomization technique, we are able to construct a tag-based trapdoor function [24] and a PKE scheme whose hardness is independent of the amount of simultaneously encrypted bits. Of major significance is the fact that, after re-randomization, the noise is still bounded:

Lemma 4 ([22] Lemma 3.4). *For any $n, m \in \mathbb{N}$, prime modulus p, $\mathbf{s} \in \{0,1\}^n$ and $\mathbf{r} \in [-\sqrt{p}/2, \sqrt{p}/2]^m$, there exists a negligible function $\nu : \mathbb{N} \to [0,1]$ such that*

$$\Pr_{\mathbf{A} \leftarrow \mathbb{Z}_p^{m \times n}}[\mathbf{r} \cdot e(\mathbf{A}, \mathbf{s}) \geq \sqrt{pnm}\log^2 n + \sqrt{p}m] \leq \nu(n).$$

This bound will be crucial to show the correctness of our proposed PKE. For the security of our scheme, we need a stronger statement.

Lemma 5. *For any $n, m \in \mathbb{N}$ and $\mathbf{r} \leftarrow [-\sqrt{p}/2, \sqrt{p}/2]^m$, there exists a negligible function $\nu : \mathbb{N} \to [0,1]$ such that*

$$\Pr_{\mathbf{r}}\left[\exists \mathbf{e} \in \mathbb{Z}^m, \|\mathbf{e}\|_\infty \leq \sqrt{n}\log n : |\mathbf{re}| \geq \frac{\sqrt{np}\log n}{2}(\sqrt{m}\log n + m)\right] \leq \nu(n).$$

Proof. We first assume $\|\mathbf{e}\|_\infty = 1$. For any $\mathbf{r} \in \mathbb{Z}_p^m$,

$$\sum_{j=1}^{m} |r_j| \geq \mathbf{re} \geq -\sum_{j=1}^{m} |r_j|.$$

For simplicity, we will give a bound for $\sum_{j=1}^{m} r_j$, for $\mathbf{r} \leftarrow [1, \sqrt{p}/2]^m$, which is sufficient. The expectation of this sum is $\frac{m}{2}(\sqrt{p}/2 + 1)$. Hence, by the Hoeffding bound (Lemma 1)

$$\Pr_{\mathbf{r}}\left[\left|\sum_{j=1}^{m} r_j - \frac{m}{2}\left(\frac{\sqrt{p}}{2}+1\right)\right| \geq \frac{\sqrt{pm}}{2}\log n\right] \leq 2e^{-\frac{(\sqrt{pm}\log n)^2}{2m\left(\frac{\sqrt{p}}{2}-1\right)^2}} \leq 2e^{-2\log^2 n},$$

where the sum is carried out in $\mathbb{R}$. Thus, there is a negligible function $\nu(n)$ such that

$$\Pr_{\mathbf{r}}\left[\exists \mathbf{e} \in \mathbb{Z}_p^m, \|\mathbf{e}\|_\infty \leq \sqrt{n}\log n : |\mathbf{re}| \geq \frac{\sqrt{np}\log n}{2}(\sqrt{m}\log n + m)\right] \leq \nu(n).$$

2.2 Tag-Based Encryption

The main motivation behind the concept of tag-based encryption (TBE) comes from the fact that it is possible to transform an identity-based encryption scheme into an IND-CCA secure PKE scheme [5,4]. Kiltz [17] showed that these transformations already work starting from TBE.

A TBE scheme with tag-space $\mathcal{T}$, message-space $\mathcal{M}$, and security parameter n, consists of the following three PPT algorithms $\mathrm{TBE} = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$.

$\mathsf{Gen}(1^n)$**:** Outputs a secret key sk and a public key pk.
$\mathsf{Enc}(pk, \tau, M)$**:** Outputs a ciphertext C for $M \in \mathcal{M}$, and tag $\tau \in \mathcal{T}$.
$\mathsf{Dec}(sk, \tau, C)$**:** Outputs the decrypted message M of ciphertext C with respect to tag $\tau \in \mathcal{T}$, or an invalid symbol $\perp$.

For correctness, we require that for any τ, M and $(sk, pk) \leftarrow \mathsf{Gen}(1^n)$:

$$\mathsf{Dec}(sk, \tau, \mathsf{Enc}(pk, \tau, M)) = M$$

holds with overwhelming probability. As for security, we define the following selective-tag weak CCA game $\mathrm{G}_{\mathrm{TBE}}$ [17]:

1. Adversary A picks a tag $\tau^* \in \mathcal{T}$.
2. Run $(sk, pk) \leftarrow \mathsf{Gen}(1^n)$. Adversary A receives public key pk and gets permanent access to an oracle which outputs $\mathsf{Dec}(sk, \tau, C)$ upon input requests of the form $\mathsf{QueryDec}(C, \tau)$ for all $\tau \neq \tau^*$, and $\perp$ otherwise.
3. A chooses M_0 and M_1 from $\mathcal{M}$ and receives $C \leftarrow \mathsf{Enc}(pk, \tau^*, M_u)$ for $u \leftarrow \{0, 1\}$.
4. Finally A outputs u' and $\mathrm{G}_{\mathrm{TBE}}$ outputs 1 iff $u' = u$.

The advantage of an adversary A in game $\mathrm{G}_{\mathrm{TBE}}$ is defined as

$$\mathbf{Adv}_{\mathrm{TBE}}(\mathsf{A}) := \left| \Pr[\mathrm{G}_{\mathrm{TBE}}(\mathsf{A}) = 1] - \frac{1}{2} \right|,$$

and a TBE scheme is called secure against selective-tag weak CCA adversaries, if for all PPT A there exists a negligible function $\nu : \mathbb{N} \to [0,1]$ such that $\mathbf{Adv}_{\mathrm{TBE}}(\mathsf{A}) \leq \nu(n)$.

Given an exponential tag-space, there is a transformation from a TBE scheme satisfying the above notion to an IND-CCA secure PKE; the transformation requires a one-time signature scheme or a message authentication code plus a commitment [17].

We embed the tags in our proposed TBE using a full-rank differences (FRD) encoding $\mathcal{H}$ [7,1]. This means that $\mathcal{H} : \mathbb{Z}_2^n \to \mathbb{Z}_2^{n \times n}$, $\tau \mapsto \mathbf{H}_\tau$ and $\forall \tau \neq \tau' \in \mathbb{Z}_2^n$ $\mathbf{H}_\tau - \mathbf{H}_{\tau'}$ has full rank.

3 A Subset Sum Based TBE

For security parameter n, let $p = \Theta(m^2 n \log^2 n)$ for prime $p > 2$, and $m = \Theta(n)$ for appropriate constant factors. The following three algorithms describe our $\mathrm{TBE} = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ based on $\mathrm{SS}(n, p^m)$ with tag space $\mathcal{T} := \mathbb{Z}_2^n \setminus \{\mathbf{0}\}$ (where $\mathbf{0}$ is the all-zero vector of length n) and message space $\mathcal{M} := \{0,1\}^\ell$.

$\mathsf{Gen}(1^n)$**:** Sample $\mathbf{R} \leftarrow [-\sqrt{p}/2, \sqrt{p}/2]^{n \times m}$ and $\mathbf{A} \leftarrow \mathbb{Z}_p^{m \times n}$, $\mathbf{C} \leftarrow \mathbb{Z}_p^{\ell \times n}$. Define $\mathbf{B} := \mathbf{RA}$. The private and public key are defined as

$$sk := \mathbf{R}, \quad pk := (\mathbf{A}, \mathbf{B}, \mathbf{C}) \in \mathbb{Z}_p^{m \times n} \times \mathbb{Z}_p^{n \times n} \times \mathbb{Z}_p^{\ell \times n}.$$

$\mathsf{Enc}(pk, \tau, M)$**:** Pick $\mathbf{R}' \leftarrow [-\sqrt{p}/2, \sqrt{p}/2]^{n \times m}$, $\mathbf{R}'' \leftarrow [-\sqrt{p}/2, \sqrt{p}/2]^{\ell \times m}$, $\mathbf{s} \leftarrow \{0,1\}^n$ and define

$$\begin{aligned}
\mathbf{c}_0 &:= \mathbf{As} + e(\mathbf{A}, \mathbf{s}) && \in \mathbb{Z}_p^m \\
\mathbf{c}_1 &:= \left(\mathbf{B} + \frac{p-1}{2} \cdot \mathbf{H}_\tau \right) \mathbf{s} + \mathbf{R}' \cdot e(\mathbf{A}, \mathbf{s}) && \in \mathbb{Z}_p^n \\
\mathbf{c}_2 &:= \mathbf{Cs} + \mathbf{R}'' \cdot e(\mathbf{A}, \mathbf{s}) + \frac{p-1}{2} \cdot M && \in \mathbb{Z}_p^\ell
\end{aligned}$$

where $\mathbf{H}_\tau$ is the matrix representation of τ.

$\mathsf{Dec}(sk, \tau, C)$**:** Compute

$$\hat{\mathbf{s}} := \left\lfloor (\mathbf{R}\ \mathbf{I}) \cdot \begin{pmatrix} -\mathbf{c}_0 \\ \mathbf{c}_1 \end{pmatrix} \right\rceil_2 .$$

and $\mathbf{s} = \mathbf{H}_\tau^{-1} \hat{\mathbf{s}}$. If $\|\mathbf{c}_0 - \mathbf{As}\|_\infty \geq \sqrt{n} \log n$ or $\|\mathbf{c}_1 - \left(\mathbf{B} + \frac{p-1}{2} \cdot \mathbf{H}_\tau\right) \mathbf{s}\|_\infty \geq \frac{p}{8}$ output $\perp$. Otherwise output message $M = \lfloor \mathbf{c}_2 - \mathbf{Cs} \rceil_2$.

3.1 Correctness

The correctness of the scheme follows basically from the bounds on the noise of re-randomized Subset Sum instances. Given these bounds, the noise will be smaller than $p/4$ such that it will be rounded away by the rounding function $\lfloor \cdot \rceil_2$.

Theorem 2 (Correctness). *Let $p = O(m^2 n \log^2 n)$ for prime $p > 2$, $m = \Theta(n)$, and $\ell \in O(n^c)$ for some constant c. Then for any $\tau \in \mathbb{Z}_2^n$, $M \in \{0,1\}^\ell$, there exists a negligible function $\nu : \mathbb{N} \to [0,1]$ such that*

$$\Pr_{(sk,pk) \leftarrow \mathsf{Gen}(1^n)} [\mathsf{Dec}(sk, \tau, \mathsf{Enc}(pk, \tau, M)) \neq M] \leq \nu(n).$$

Proof. Given a ciphertext $C = (\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2)$, in case Dec successfully reconstruct $\mathbf{s}$, the decryption algorithm computes

$$\lfloor \mathbf{c}_2 - \mathbf{C}\mathbf{s} \rceil_2 = \left\lfloor \mathbf{R}'' \cdot e(\mathbf{A}, \mathbf{s}) + \frac{p-1}{2} \cdot M \right\rceil_2 = \lfloor \mathbf{R}'' \cdot e(\mathbf{A}, \mathbf{s}) \rceil_2 + M.$$

By Lemma 2, $\|\mathbf{c}_0 - \mathbf{A}\mathbf{s}\|_\infty \leq \sqrt{n} \log n$ and by Lemma 4 $\|\mathbf{R}'' \cdot e(\mathbf{A}, \mathbf{s})\|_\infty \leq \sqrt{pnm} \log^2 n + \sqrt{p}m < \frac{p}{8}$ with overwhelming probability over $\mathbf{A} \leftarrow \mathbb{Z}_p^{m \times n}$ (for appropriately chosen constants). Hence, $\|\mathbf{R}'' \cdot e(\mathbf{A}, \mathbf{s})\|_\infty < \frac{p}{8}$ and Dec outputs M.

For the same reason $\|(\mathbf{R}' - \mathbf{R}) \cdot e(\mathbf{A}, \mathbf{s})\|_\infty < \frac{p}{4}$ holds with overwhelming probability over $\mathbf{A} \leftarrow \mathbb{Z}_q^{m \times n}$ (for appropriately chosen constants). Therefore Dec reconstructs

$$\left\lfloor (\mathbf{R}\ \mathbf{I}) \cdot \begin{pmatrix} -\mathbf{c}_0 \\ \mathbf{c}_1 \end{pmatrix} \right\rceil_2 = \left\lfloor \left(\frac{p-1}{2} \cdot \mathbf{H}_\tau \right) \mathbf{s} + (\mathbf{R}' - \mathbf{R}) e(\mathbf{A}, \mathbf{s}) \right\rceil_2 = \left\lfloor \left(\frac{p-1}{2} \cdot \mathbf{H}_\tau \right) \mathbf{s} \right\rceil_2.$$

Note that for every entry s' of $\left(\frac{p-1}{2} \cdot \mathbf{H}_\tau\right) \mathbf{s} \in \mathbb{Z}_p^n$ holds: s' is close to 0, close to $\frac{p-1}{2}$, if its corresponding entry of $\mathbf{H}_\tau \mathbf{s} \in \mathbb{Z}_2^n$ is 0, 1 respectively, where close means at most additive term $\Theta(n)$ away. Therefore, the outcome of the rounding function $\lfloor \cdot \rceil_2$ is not influenced by the noise term, which is sufficiently smaller than $\frac{p}{4}$ for appropriately chosen constants. Hence, Dec computes $\hat{\mathbf{s}} = \left\lfloor \frac{p-1}{2} \cdot \mathbf{H}_\tau \mathbf{s} \right\rceil_2 = \mathbf{H}_\tau \mathbf{s}$. This results in the correct reconstruction of $\mathbf{s} = \mathbf{H}_\tau^{-1} \hat{\mathbf{s}} = \mathbf{H}_\tau^{-1} \mathbf{H}_\tau \mathbf{s}$.

3.2 Proof of Security

The intuition behind the security proof is that $\mathbf{B} = \mathbf{R}\mathbf{A}$ is statistically indistinguishable from $\mathbf{B}' = \mathbf{R}\mathbf{A} - \frac{p-1}{2}\mathbf{H}_{\tau^*}$. But when $\mathbf{B}'$ is used as part of the public key, ciphertexts with tag τ^* can not be decrypted using Dec anymore. During the proof, we will show that there are ciphertexts for τ^* which are at least as hard to decrypt as solving $\mathrm{SS}(n, p^m)$. Given any algorithm guessing the message encrypted in such a ciphertext and therefore breaking the security of TBE, there will be also an algorithm solving $\mathrm{SS}(n, p^m)$.

Theorem 3 (CCA Security). *Let $p = \Theta(m^2 n \log^2 n)$ for prime $p > 2$, and $m = \Theta(n)$ for appropriate constant factors. If the* $\mathrm{SS}(n, p^m)$ *assumption holds (which corresponds to density $\delta \in O(1/\log n)$), then the proposed* TBE *scheme is secure against* selective-tag weak CCA *adversaries. In particular, for every PPT algorithm* A *there exist a PPT algorithm* D *and a negligible function $\nu : \mathbb{N} \to [0, 1]$ such that:*

$$\mathbf{Adv}_{\mathrm{TBE}}(\mathsf{A}) \leq \mathbf{Adv}_{\mathrm{SS}(n,p^m)}(\mathsf{D}) + \nu(n).$$

Proof. We construct an algorithm D which will distinguish $\mathrm{SS}(n, p^m)$ from uniform invoking a successful adversary A in game $\mathrm{G}_{\mathrm{TBA}}$. If D receives a $\mathrm{SS}(n, p^m)$ instance D will simulate game $\mathrm{G}_{\mathrm{TBA}}$ and a successful A will guess b correctly with probability $\frac{1}{2} + \mathbf{Adv}_{\mathrm{TBE}}(\mathsf{A}) > \frac{1}{2} + \nu(n)$. When D receives a uniform input, D will simulate a game in which the challenge ciphertext is independent of message M_u, and hence independent of u. Therefore guess u' of A will be correct (i.e., $u' = u$) with probability $\frac{1}{2}$.

In the following, we describe algorithm D interacting with A and afterwards we analyse its success probability.

1. D receives a $\mathrm{SS}(n, p^m)$ challenge $(\mathbf{A}, \mathbf{b})$ and invokes A which will send a tag $\tau^* \in \mathcal{T}$.
2. D samples $\mathbf{R}' \leftarrow [-\sqrt{p}/2, \sqrt{p}/2]^{n \times m}$, $\mathbf{R}'' \leftarrow [-\sqrt{p}/2, \sqrt{p}/2]^{\ell \times m}$ and sets $pk = (\mathbf{A}, \mathbf{B} := \mathbf{R}'\mathbf{A} - \frac{p-1}{2}\mathbf{H}_{\tau^*}, \mathbf{C} := \mathbf{R}''\mathbf{A})$ which is by Lemma 3 statistically close to the output distribution of public keys of Gen. The public key pk is given to A.
   Thus, D uses $\mathbf{R}'$ to respond to $\mathsf{QueryDec}(C, \tau)$ queries as follows: If $\tau = \tau^*$ output $\perp$. Otherwise D uses $\mathsf{Dec}(\mathbf{R}', \tau, C)$ to reconstruct:

$$\hat{\mathbf{s}} := \left\lfloor (\mathbf{R}' \; \mathbf{I}) \cdot \begin{pmatrix} -\mathbf{c}_0 \\ \mathbf{c}_1 \end{pmatrix} \right\rceil_2 .$$

   For a properly distributed ciphertext C, the entries of $\left(\frac{p-1}{2} \cdot \mathbf{H}_\tau - \frac{p-1}{2}\mathbf{H}_{\tau^*}\right)\mathbf{s}$ are either close to 0 or close to $\frac{p-1}{2}$ and the noise terms are small.

$$\hat{\mathbf{s}} = \left\lfloor \left(\tfrac{p-1}{2} \cdot \mathbf{H}_\tau - \tfrac{p-1}{2}\mathbf{H}_{\tau^*}\right)\mathbf{s} \right\rceil_2 = (\mathbf{H}_\tau - \mathbf{H}_{\tau^*})\mathbf{s}.$$

   $\mathbf{s}$ is reconstructed by computing $\mathbf{s} = (\mathbf{H}_\tau - \mathbf{H}_{\tau^*})^{-1}\hat{\mathbf{s}}$.
   Let C be any ciphertext that passes the sanity check, i.e. $\|\mathbf{c}_0 - \mathbf{A}\mathbf{s}\|_\infty < \sqrt{n}\log n$ and $\|\mathbf{c}_1 - \left(\mathbf{B} + \frac{p-1}{2} \cdot \mathbf{H}_\tau\right)\mathbf{s}\|_\infty < \frac{p}{8}$. Clearly, if $C \in \mathbb{Z}_p^{m+n+\ell}$ does not pass this sanity check, Dec always outputs $\perp$, which is independent of secret key $\mathbf{R}'$. By Lemma 5, for any choice of $\mathbf{c}_0 - \mathbf{A}\mathbf{s}$ with $\|\mathbf{c}_0 - \mathbf{A}\mathbf{s}\|_\infty < \sqrt{n}\log n$, $\|\mathbf{R}'(\mathbf{c}_0 - \mathbf{A}\mathbf{s})\|_\infty < \frac{q}{8}$ holds with overwhelming probability over the choice of $\mathbf{R}'$. Then the noise term is sufficiently small, i.e. $\|\mathbf{c}_1 - \left(\mathbf{B} + \frac{p-1}{2} \cdot \mathbf{H}_\tau\right)\mathbf{s} - \mathbf{R}'(\mathbf{c}_0 - \mathbf{A}\mathbf{s})\|_\infty < \frac{q}{4}$, and thus,

$$\left\lfloor (\mathbf{R}' \; \mathbf{I}) \cdot \begin{pmatrix} -\mathbf{c}_0 \\ \mathbf{c}_1 \end{pmatrix} \right\rceil_2 = \left\lfloor \left(\frac{p-1}{2} \cdot \mathbf{H}_\tau - \frac{p-1}{2}\mathbf{H}_{\tau^*}\right)\mathbf{s} \right\rceil_2 = (\mathbf{H}_\tau - \mathbf{H}_{\tau^*})\mathbf{s}$$

reconstructs $\hat{\mathbf{s}}$ correctly for any $C \in \mathbb{Z}_p^{m+n+\ell}$ that passes the sanity check. This ensures that the output of $\mathsf{QueryDec}(C, \tau)$ depends only on $\mathbf{B}$ and $\mathbf{A}$, but not on $\mathbf{R}'$. Further, if $\hat{\mathbf{s}}$ is reconstructed correctly, Dec outputs the correct message M.

3. A sends M_0 and M_1. Now D samples $u \leftarrow \{0,1\}$, sets $C^* := (\mathbf{b}, \mathbf{R}'\mathbf{b}, \mathbf{R}''\mathbf{b} + \frac{p-1}{2} M_u)$, and sends C^* to A.
4. Finally A outputs u' and D outputs 1 iff $u' = u$.

When $\mathbf{b} = \mathbf{A}\mathbf{s} + e(\mathbf{A}, \mathbf{s})$, the challenge ciphertext C^* is a proper ciphertext for public key pk and randomness $\mathbf{s}$:

$$\mathbf{c}_0 := \mathbf{b} = \mathbf{A}\mathbf{s} + e(\mathbf{A}, \mathbf{s})$$
$$\mathbf{c}_1 := \mathbf{R}'\mathbf{b} = \mathbf{R}'\mathbf{A}\mathbf{s} + \mathbf{R}'e(\mathbf{A}, \mathbf{s}) = \left(\mathbf{B} + \frac{p-1}{2} \cdot \mathbf{H}_\tau\right)\mathbf{s} + \mathbf{R}'e(\mathbf{A}, \mathbf{s})$$
$$\mathbf{c}_2 := \mathbf{R}''\mathbf{b} + \frac{p-1}{2} M_u = \mathbf{C}\mathbf{s} + \mathbf{R}''e(\mathbf{A}, \mathbf{s}) + \frac{p-1}{2} M_u.$$

Note that, by Lemma 3, there is enough entropy in $\mathbf{R}'$, $\mathbf{R}''$ such that $\mathbf{B}$, $\mathbf{R}'e(\mathbf{A}, \mathbf{s})$ and $\mathbf{C}$, $\mathbf{R}''e(\mathbf{A}, \mathbf{s})$ are independent. In this case B outputs 1 with roughly probability $\frac{1}{2} + \mathbf{Adv}_{\mathrm{TBE}}(\mathsf{A})$.

In the other case, i.e. when $\mathbf{A}, \mathbf{b} \leftarrow \mathbb{Z}_p^{m \times n} \times \mathbb{Z}_p^m$, we know that $\mathbf{c}_2 := \mathbf{R}''\mathbf{b} + \frac{p-1}{2} M_u$ is uniform and independent of $\mathbf{A}, \mathbf{C}, \mathbf{c}_0$ and $\mathbf{c}_1$ by Lemma 3. Therefore C^* is independent of u and for any output u' of A:

$$\Pr_{u \leftarrow \{0,1\}}[u = u'] = \frac{1}{2}.$$

Summing up, D outputs 1 for a $\mathrm{SS}(n, p^m)$ instance with roughly probability $\frac{1}{2} + \mathbf{Adv}_{\mathrm{TBE}}(\mathsf{A})$, and it outputs 1 otherwise with probability $\frac{1}{2}$. This implies

$$\mathbf{Adv}_{\mathrm{SS}(n,p^m)}(\mathsf{D}) = \mathbf{Adv}_{\mathrm{TBE}}(\mathsf{A}) - \nu(n),$$

for a negligible function ν, concluding the proof.

4 Conclusions and Open Problems

We presented a construction of a new PKE scheme with a simple and direct security proof based on the hardness of random instances of the Subset Sum problem. Our scheme achieves IND-CCA security and its concrete security does not depend on the length of the messages being encrypted. This resolves the main open problems from the previous work by Lyubashevsky, Palacio, and Segev [22].

Similarly to one of the constructions in [22], it is not hard to see that actually our PKE scheme achieves the stronger notion of IND-CCA security against non-adaptive leakage attacks.[7] We leave it as an open problem to construct a

[7] Since the latter notion is a very weak form of leakage resilience, we preferred to not work out the details.

PKE scheme with IND-CCA security against fully adaptive leakage attacks. An approach towards answering this question would be to construct a hash proof system [9] based on Subset Sum, as this would directly yield a leakage-resilient IND-CCA secure PKE [25].

It would also be interesting to construct PKE schemes with additional properties (always based on Subset Sum), such as circular security, key-dependent message security, and security against related-key attacks.

References

1. Agrawal, S., Boneh, D., Boyen, X.: Efficient lattice (H)IBE in the standard model. In: EUROCRYPT. pp. 553–572 (2010)
2. Ajtai, M., Dwork, C.: A public-key cryptosystem with worst-case/average-case equivalence. In: ACM STOC. pp. 284–293 (1997)
3. Bernstein, D.J., Jeffery, S., Lange, T., Meurer, A.: Quantum algorithms for the subset-sum problem. In: PQCrypto. pp. 16–33 (2013)
4. Boneh, D., Canetti, R., Halevi, S., Katz, J.: Chosen-ciphertext security from identity-based encryption. SIAM J. Comput. 36(5), 1301–1328 (2007)
5. Canetti, R., Halevi, S., Katz, J.: Chosen-ciphertext security from identity-based encryption. In: EUROCRYPT. pp. 207–222 (2004)
6. Cash, D., Kiltz, E., Shoup, V.: The twin Diffie-Hellman problem and applications. J. Cryptology 22(4), 470–504 (2009)
7. Cramer, R., Damgård, I.: On the amortized complexity of zero-knowledge protocols. In: CRYPTO. pp. 177–191 (2009)
8. Cramer, R., Shoup, V.: A practical public key cryptosystem provably secure against adaptive chosen ciphertext attack. In: CRYPTO. pp. 13–25 (1998)
9. Cramer, R., Shoup, V.: Universal hash proofs and a paradigm for adaptive chosen ciphertext secure public-key encryption. In: EUROCRYPT. pp. 45–64 (2002)
10. Diffie, W., Hellman, M.E.: New directions in cryptography. IEEE Transactions on Information Theory 22(6), 644–654 (1976)
11. Faust, S., Hazay, C., Venturi, D.: Outsourced pattern matching. In: ICALP. pp. 545–556 (2013)
12. Flaxman, A., Przydatek, B.: Solving medium-density subset sum problems in expected polynomial time. In: STACS. pp. 305–314 (2005)
13. Frieze, A.M.: On the Lagarias-Odlyzko algorithm for the subset sum problem. SIAM J. Comput. 15(2), 536–539 (1986)
14. Goldwasser, S., Micali, S.: Probabilistic encryption. J. Comput. Syst. Sci. 28(2), 270–299 (1984)
15. Hofheinz, D., Kiltz, E., Shoup, V.: Practical chosen ciphertext secure encryption from factoring. J. Cryptology 26(1), 102–118 (2013)
16. Impagliazzo, R., Naor, M.: Efficient cryptographic schemes provably as secure as subset sum. J. Cryptology 9(4), 199–216 (1996)
17. Kiltz, E.: Chosen-ciphertext security from tag-based encryption. In: TCC. pp. 581–600 (2006)
18. Kiltz, E., Masny, D., Pietrzak, K.: Simple chosen-ciphertext security from low-noise LPN. In: PKC. pp. 1–18 (2014)
19. Kirchner, P., Fouque, P.: An improved BKW algorithm for LWE with applications to cryptography and lattices. In: CRYPTO. pp. 43–62 (2015)

20. Lagarias, J.C., Odlyzko, A.M.: Solving low-density subset sum problems. J. ACM 32(1), 229–246 (1985)
21. Lyubashevsky, V.: The parity problem in the presence of noise, decoding random linear codes, and the subset sum problem. In: APPROX. pp. 378–389 (2005)
22. Lyubashevsky, V., Palacio, A., Segev, G.: Public-key cryptographic primitives provably as secure as subset sum. In: TCC. pp. 382–400 (2010)
23. Merkle, R.C., Hellman, M.E.: Hiding information and signatures in trapdoor knapsacks. IEEE Transactions on Information Theory 24(5), 525–530 (1978)
24. Micciancio, D., Peikert, C.: Trapdoors for lattices: Simpler, tighter, faster, smaller. In: EUROCRYPT. pp. 700–718 (2012)
25. Naor, M., Segev, G.: Public-key cryptosystems resilient to key leakage. SIAM J. Comput. 41(4), 772–814 (2012)
26. Odlyzko, A.M.: The rise and fall of knapsack cryptosystems. In: Symposia of Applied Mathematics. pp. 75–88 (1990)
27. Peikert, C.: Public-key cryptosystems from the worst-case shortest vector problem. In: ACM STOC. pp. 333–342 (2009)
28. Peikert, C., Waters, B.: Lossy trapdoor functions and their applications. SIAM J. Comput. 40(6), 1803–1844 (2011)
29. Rackoff, C., Simon, D.R.: Non-interactive zero-knowledge proof of knowledge and chosen ciphertext attack. In: CRYPTO. pp. 433–444 (1991)
30. Regev, O.: New lattice based cryptographic constructions. In: ACM STOC. pp. 407–416 (2003)
31. Regev, O.: On lattices, learning with errors, random linear codes, and cryptography. J. ACM 56(6) (2009)
32. Shallue, A.: An improved multi-set algorithm for the dense subset sum problem. In: Algorithmic Number Theory (ANTS-VIII). pp. 416–429 (2008)