Public-Key Cryptographic Primitives Provably as Secure as Subset Sum

Vadim Lyubashevsky[1⋆], Adriana Palacio[2], and Gil Segev[3⋆⋆]

[1] Tel-Aviv University, Tel-Aviv, Israel
vlyubash@cs.ucsd.edu

[2] Bowdoin College, Brunswick, ME USA
apalacio@bowdoin.edu

[3] Weizmann Institute of Science, Rehovot, Israel
gil.segev@weizmann.ac.il

Abstract. We propose a semantically-secure public-key encryption scheme whose security is polynomial-time equivalent to the hardness of solving random instances of the subset sum problem. The subset sum assumption required for the security of our scheme is weaker than that of existing subset-sum based encryption schemes, namely the lattice-based schemes of Ajtai and Dwork (STOC'97), Regev (STOC'03, STOC'05), and Peikert (STOC'09). Additionally, our proof of security is simple and direct. We also present a natural variant of our scheme that is secure against key-leakage attacks, and an oblivious transfer protocol that is secure against semi-honest adversaries.

1 Introduction

Since the early days of modern cryptography, the presumed intractability of the subset sum problem has been considered an interesting alternative to hardness assumptions based on factoring and the discrete logarithm problem. The appeal of the subset sum problem stems from the fact that it is simple to describe, and computing the subset sum function requires only a few addition operations. Another attractive feature is that the subset sum problem seems to be rather different in nature from number-theoretic problems. In fact, while there are polynomial-time quantum algorithms that break virtually all number-theoretic cryptographic assumptions [Sho97], there are currently no known quantum algorithms that perform better than classical ones on the subset sum problem.

The *subset sum problem*, $SS(n, M)$, is parameterized by two integers n and M. An instance of $SS(n, M)$ is created by picking a uniformly random vector $\mathbf{a} \in \mathbb{Z}_M^n$, a uniformly random vector $\mathbf{s} \in \{0,1\}^n$, and outputting $\mathbf{a}$ together with

⋆ Research supported by the Israel Science Foundation and a European Research Council (ERC) Starting Grant.

⋆⋆ Research supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities.

$T = \mathbf{a} \cdot \mathbf{s} \bmod M$. The problem is to find $\mathbf{s}$, given $\mathbf{a}$ and T. The hardness of breaking $SS(n, M)$ depends on the ratio between n and $\log M$, which is usually referred to as the *density* of the subset sum instance. When $n/\log M$ is less than $1/n$ or larger than $n/\log^2 n$, the problem can be solved in polynomial time [LO85,Fri86,FP05,Lyu05,Sha08]. However, when the density is constant or even as small as $O(1/\log n)$, there are currently no algorithms that require less than $2^{\Omega(n)}$ time. It is also known that the subset sum problem can only get harder as its density gets closer to one [IN96].

Starting with the Merkle-Hellman cryptosystem [MH78], there have been many proposals for constructions of public-key encryption schemes that were somewhat based on subset sum. Unfortunately, all of these proposals have subsequently been broken (see [Odl90] for a survey). While efforts to build subset-sum based public-key encryption schemes were met with little success, Impagliazzo and Naor were able to construct provably-secure primitives such as universal one-way hash functions, pseudorandom generators and bit-commitment schemes, based on the subset sum problem, that remain secure until this day [IN96]. The main difference between the public-key constructions and the "minicrypt" constructions in [IN96] is that the latter could be proved secure based on random instances of the standard subset sum problem, whereas the former modified the subset sum instances in order to allow decryption. Unfortunately, these modifications always seemed to introduce fatal weaknesses.

A provably-secure cryptosystem based on subset sum was finally constructed by Ajtai and Dwork [AD97], who showed that their scheme is as hard to break as solving worst-case instances of a lattice problem called the "unique shortest vector problem." The reduction of subset sum to breaking their scheme is then obtained via the classic reduction from random subset sum to the unique shortest vector problem [LO85,Fri86]. While the Ajtai-Dwork and the subsequent lattice-based cryptosystems [Reg03,Reg05,Pei09] are as hard to break as the average-case subset sum problem, these schemes are based on subset sum in a somewhat indirect way, and this causes their connection to the subset sum problem to not be as tight as possible.

In this work, we present a cryptosystem whose security is *equivalent* to the hardness of the $SS(n, q^n)$ problem, where q is a positive integer of magnitude $\tilde{O}(n)$. Compared to the lattice-based cryptosystems, the subset sum assumption required for the security of our scheme is weaker, and the proof of security is much simpler. We direct the reader to Section 1.2 for a more in-depth comparison between our scheme and the lattice-based ones.

In addition to our semantically-secure public-key encryption scheme, we present a semi-honest oblivious transfer protocol based on the same hardness assumption. We also show that a natural variant of our encryption scheme is resilient to key-leakage attacks (as formalized by Akavia et al. [AGV09]), but under slightly stronger assumptions than our basic cryptosystem.

1.1 Our Contributions and Techniques

Semantically-secure public-key encryption. Our main contribution is a semantically secure public-key encryption scheme whose security is based directly on the hardness of the subset sum problem. The construction of our scheme is similar in spirit to the cryptosystem of Alekhnovich based on the Learning Parity with Noise (LPN) problem [Ale03], and that of Regev based on the Learning With Errors (LWE) problem [Reg05]. Both of the aforementioned schemes are built from the assumption that for a randomly chosen matrix $\mathbf{A} \in \mathbb{Z}_q^{m \times n}$, a random vector $\mathbf{s} \in \mathbb{Z}_q^n$, and some "small" noise vector $\mathbf{c} \in \mathbb{Z}_q^m$, the distribution $(\mathbf{A}, \mathbf{As} + \mathbf{c})$ is computationally indistinguishable from the uniform distribution over $\mathbb{Z}_q^{m \times (n+1)}$. To construct our scheme, we show that the subset sum problem can be made to look very similar to the LWE problem. Then the main ideas (with a few technical differences) used in constructing cryptosystems based on LWE [Reg05,GPV08,Pei09] can be transferred over to subset sum.

Consider instances of the subset sum problem $SS(n, q^m)$ where q is some small integer. If $\mathbf{a}$ is a vector in $\mathbb{Z}_{q^m}^n$ and $\mathbf{s}$ is a vector in $\{0,1\}^n$, then $\mathbf{a} \cdot \mathbf{s} \bmod q^m$, written in base q, is equal to $\mathbf{As} + \mathbf{c} \bmod q$, where $\mathbf{A} \in \mathbb{Z}_q^{m \times n}$ is a matrix whose i-th column corresponds to $\mathbf{a}_i$ written in base q, and $\mathbf{c}$ is a vector in $\mathbb{Z}_q^m$ that corresponds to the carries when performing "grade-school" addition. For example, let $q = 10$, $m = n = 3$, $\mathbf{a} = (738, 916, 375)$, and $\mathbf{s} = (0, 1, 1)$. Then

$$\mathbf{a} \cdot \mathbf{s} \bmod 10^3 = 916 + 375 \bmod 10^3 = 291,$$

which can be written as addition in base q as follows:

$$\begin{bmatrix} 7 & 9 & 3 \\ 3 & 1 & 7 \\ 8 & 6 & 5 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 9 \\ 1 \end{bmatrix}$$

where all operations are performed over $\mathbb{Z}_q$.

The key observation is that the magnitude of the entries in the carries vector $[0\ 1\ 0]^T$ is at most $n - 1$, and so if $q \gg n$, then $\mathbf{As} + \mathbf{c} \bmod q \approx \mathbf{As} \bmod q$. In fact, the elements of the vector $\mathbf{c}$ are distributed normally around $n/2$ with standard deviation $\sqrt{n}$. In the instantiation of our scheme described in Section 3, we generate the elements in $\mathbf{A}$ from the range $[-\frac{q-1}{2}, \frac{q-1}{2}]$ and so the entries in the carries vector are normally distributed around 0 with standard deviation $\sqrt{n}$. Readers familiar with the cryptosystems based on LWE [Reg05,GPV08,Pei09] should recognize the resemblance of the carry vector $\mathbf{c}$ to the noise vector in the LWE-based schemes. The main difference is that in the latter the noise vector is chosen independently at random, whereas in our scheme, the carries vector $\mathbf{c}$ occurs "naturally" and is completely determined by the matrix $\mathbf{A}$ and the vector $\mathbf{s}$. The fact that the "noise" vector is not random is of no consequence to us, since it was shown by Impagliazzo and Naor that distinguishing $(\mathbf{a}, \mathbf{a} \cdot \mathbf{s} \bmod q^m)$ from uniform is as hard as recovering $\mathbf{s}$ [IN96]. Thus the distribution $(\mathbf{A}, \mathbf{As} + \mathbf{c} \bmod q)$, which is just the base q representation of the previous distribution, is also computationally indistinguishable from uniform, based on the hardness of subset sum. The following theorem summarizes our main result:

Theorem 1.1. *For any integer $q > 10n\log^2 n$, there exists a semantically secure cryptosystem encrypting k bits whose security is polynomial-time equivalent to the hardness of solving the $SS(n, q^{n+k})$ problem.*

Leakage-resilient public-key encryption. We show that a natural variant of our encryption scheme is resilient to any non-adaptive leakage of $L(1-o(1))$ bits of its secret key, where L is the length of the secret key (see Appendix A.2 for the formal definition of non-adaptive key-leakage attacks). In this paper we deal with the non-adaptive setting of key leakage, and note that this notion of leakage is still very meaningful as it captures many realistic attacks in which the leakage does not depend on the parameters of the encryption scheme. For example, it captures the cold boot attacks of Halderman et al. [HSH+08], in which the leakage depends only on the properties of the hardware devices that are used for storing the secret key. We note that although Naor and Segev [NS09] presented a generic and rather simple construction that protects any public-key encryption scheme from non-adaptive leakage attacks, we show that for our specific scheme an even simpler modification suffices.

Oblivious transfer. We use our original encryption scheme to construct an oblivious transfer (OT) protocol that provides security for the receiver against a cheating sender and security for the sender against an honest-but-curious receiver. Our protocol is an instance of a natural construction used by Gertner et al. [GKM+00], based on ideas of Even et al. [EGL82,EGL85], to show that public-key encryption with a certain property implies two-message semi-honest OT. The property is roughly that public keys can be sampled "separately of private keys," while preserving the semantic security of the encryption. Pseudorandomness of subset sum implies that our encryption scheme satisfies this property.

1.2 Comparisons with Lattice-Based Schemes

To the best of our knowledge, the only other cryptosystems based on subset sum are those that are based on the worst-case hardness of the approximate unique shortest vector problem (uSVP_γ) [AD97,Reg03,Reg05,Pei09]. The cryptosystems of Regev [Reg03] and Peikert [Pei09] are both based on the hardness of $\text{uSVP}_{n^{1.5}}$ (the latter is based on uSVP via a reduction in [LM09]). What this means is that an algorithm that breaks these cryptosystems can be used to find the shortest vector in any lattice whose shortest vector is a factor of $n^{1.5}$ shorter than the next shortest vector that is not a multiple of it.

A reduction from the random subset sum problem to uSVP_γ was given in [LO85,Fri86]. The exact parameter γ depends on the density of the subset sum instance. The smaller the density, the larger the γ can be, and the easier the uSVP_γ problem becomes. The reduction from an instance of $SS(n, M)$ to uSVP_γ is as follows:

Given an instance of $SS(n, M)$ consisting of a vector $\mathbf{a} \in \mathbb{Z}_M^n$ and an element $T \in \mathbb{Z}_M$, we define the lattice $\mathcal{L}$ as

$$\mathcal{L} \;=\; \{\mathbf{x} \in \mathbb{Z}^{n+1} : [\mathbf{a}||-T] \cdot \mathbf{x} \bmod M = 0\}.$$

Notice that the vector $\mathbf{x} = [\mathbf{s}||1]$ is in $\mathcal{L}$ for the $\mathbf{s}$ for which $\mathbf{a} \cdot \mathbf{s} \bmod M = T$, so the ℓ_2 norm of the shortest vector is approximately $\sqrt{n}$. The next shortest non-parallel vector is the vector that meets the Minkowski bound of $\sqrt{n+1} \cdot det(L)^{\frac{1}{n+1}} \approx \sqrt{n} M^{1/n}$, which is a factor $M^{1/n}$ larger than the shortest vector. Therefore solving $\text{uSVP}_{n^{1.5}}$ allows us to solve instances of $SS(n, M)$ where $M \approx n^{1.5n}$.

The cryptosystem that we construct in this paper is based on the hardness of $SS(n, M)$ where $M \approx n^n$. In order to have a lattice scheme based on the same subset sum assumption, it would need to be based on uSVP_n. The construction of such a scheme is currently not known and would be considered a breakthrough.

We want to point out that we are not claiming that just because our scheme is based on a weaker instance of subset sum, it is somehow more secure than the lattice-based schemes. All we are claiming is that the connection of our scheme to the subset sum problem is better. In terms of security, the lattice-based schemes based on LWE [Reg05,Pei09] and our scheme are actually very similar because the LWE and subset sum problems can both be viewed as average-case instances of the "bounded distance decoding" problem, with essentially the same parameters but different distributions. Unfortunately, we do not know of any tight reduction between the two distributions, so there is no clear theoretical connection between LWE and subset sum.

In practice, though, there may be some advantages of our scheme over the lattice-based ones. The secret key in our scheme is an n-bit vector $\mathbf{s} \in \{0,1\}^n$, whereas the secret keys in lattice-based schemes are on the order of $n \log n$ bits. Also, the public key in our scheme is a matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times n}$, whereas lattice-based schemes use an $n \times n \log n$ matrix. The reason for the savings of a factor of $\log n$ in the size of both the secret and public keys in our scheme has to do with the fact that the distribution $(\mathbf{A}, \mathbf{As} + \mathbf{c})$ is indistinguishable from random, where $\mathbf{s} \in \{0,1\}^n$, based on the subset sum assumption. But in order to get a proof of security based on lattices, the vector $\mathbf{s}$ has to be chosen uniformly from $\mathbb{Z}_q^n$ (see [ACPS09] for a slight improvement), and is thus $\log n$ times longer. One can thus view our proof of security based on subset sum as justification that having $\mathbf{s}$ come from a smaller set and having the "noise" be a deterministic function of $\mathbf{A}$ and $\mathbf{s}$, is still secure.

1.3 Open Problems

Our construction of the subset sum cryptosystem involves transforming the subset sum problem into something that very much resembles the LWE problem. It would be interesting to see whether the same type of idea could be used to transform other problems into LWE-type problems upon which semantically-secure cryptosystems can be built.

Another open problem concerns weakening the computational assumption underlying the multi-bit version of our scheme. While our one-bit cryptosystem is based on the hardness of solving instances of $SS(n, q^n)$ for some $q = \tilde{O}(n)$, when simultaneously encrypting k bits using the same randomness our cryptosystem becomes equivalent to the easier $SS(n, q^{n+k})$ problem (clearly, it is possible to

encrypt k bits bit-by-bit, but this is less efficient). This is somewhat peculiar since one can simultaneously encrypt polynomially-many bits using the LWE cryptosystem without making the underlying assumption stronger [PVW08], while simultaneously encrypting $\Omega(n^2)$ bits in our scheme is completely insecure (since the $SS(n, q^{n^2})$ problem can be solved in polynomial time [LO85,Fri86]). We believe that this weakness in the subset sum construction is due to the fact that the noise in the LWE schemes is generated independently, whereas in our scheme, the "noise" is just the carry bits. It is an interesting open problem to see whether one can modify our scheme so that its security does not depend on the number of bits being simultaneously encrypted using the same randomness.

Another interesting open problem concerns security against leakage attacks. First, we were not able to prove the security of our scheme against *adaptive* key-leakage attacks, in which the leakage can be chosen as a function of the public key as well. Although our scheme is somewhat similar to that of Akavia et al. [AGV09], it seems that their approach for proving security against adaptive attacks does not immediately apply to our setting. Second, our leakage-resilient scheme relies on a slightly stronger assumption than our basic scheme, and it will be interesting to minimize the required computational assumption.

Finally, we leave it as an open problem to construct a CCA-secure scheme in the standard model based directly on subset sum. While there are CCA-secure encryption schemes based on lattice problems (and thus on subset sum as well) [PW08,Pei09], it would be interesting to build one directly based on subset sum that will hopefully require weaker assumptions than the lattice based ones.

2 Preliminaries

2.1 The Subset Sum Problem

The subset sum problem with parameters n and q^m, where n and m are integers and q is a positive integer such that $2^n < q^m$, is defined as follows: Given n numbers $a_1, \ldots, a_n \in \mathbb{Z}_{q^m}$ and a target $T \in \mathbb{Z}_{q^m}$, find a subset $S \subseteq \{1, \ldots, n\}$ such that $\sum_{i \in S} a_i = T \bmod q^m$. This can be viewed as the problem of inverting the function $f_{\mathbf{a}} : \{0,1\}^n \to \mathbb{Z}_{q^m}$ defined as

$$f_{\mathbf{a}}(s_1, \ldots, s_n) = \sum_{i=1}^{n} s_i a_i \bmod q^m \,,$$

where $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{Z}_{q^m}^n$ is its index (i.e., this is a collection of functions, where a function is sampled by choosing its index $\mathbf{a}$ uniformly at random).

We denote by $SS(n, q^m)$ the subset sum problem with parameters n and q^m. Using the above notion, the hardness of the subset sum problem is the assumption that $\{f_{\mathbf{a}}\}_{\mathbf{a} \in \mathbb{Z}_{q^m}^n}$ is a collection of one-way functions. We now state two properties of the subset sum problem that were proved by Impagliazzo and Naor [IN96] and are used in analyzing the security of our constructions. The first property is that subset sum instances with larger moduli are not harder than

subset sum instances with smaller moduli. The second property is that if the subset sum is a one-way function, then it is also a pseudorandom generator. In the following two statements, we fix n, m and q as above.

Lemma 2.1 ([IN96]). *For any integers i and j such that $i < j$, if $q^{m+i} > 2^n$, then the hardness of $SS(n, q^{m+j})$ implies the hardness of $SS(n, q^{m+i})$.*

Lemma 2.2 ([IN96]). *The hardness of $SS(n, q^m)$ implies that the distributions $(\mathbf{a}, f_{\mathbf{a}}(\mathbf{s}))$ and $(\mathbf{a}, t)$ are computationally indistinguishable, where $\mathbf{a} \in \mathbb{Z}_{q^m}^n$, $\mathbf{s} \in \{0,1\}^n$, and $t \in \mathbb{Z}_{q^m}$ are chosen independently and uniformly at random.*[4]

2.2 Notation

We represent vectors by bold-case letters and all vectors will be assumed to be column vectors. Unless stated otherwise, all scalar and vector operations are performed modulo q. For simplicity, we will assume that q is odd, but our results follow for all q with minimal changes. We represent elements in $\mathbb{Z}_q$ by integers in the range $[-(q-1)/2, (q-1)/2]$. For an element $e \in \mathbb{Z}_q$, its length, denoted by $|e|$ is the absolute value of its representative in the range $[-(q-1)/2, (q-1)/2]$. For a vector $\mathbf{e} = (e_1, \ldots, e_m) \in \mathbb{Z}_q^m$, we define $\|\mathbf{e}\|_\infty = \max_{1 \le i \le m} |e_i|$.

We now present some notation that is convenient for describing the subset sum function. For a matrix $\mathbf{A} \in \mathbb{Z}_q^{m \times n}$ and a vector $\mathbf{s} \in \{0,1\}^n$, we define $\mathbf{A} \odot \mathbf{s}$ as the vector $\mathbf{t}^T = (t_0, \ldots, t_{m-1})$ such that $|t_i| \le (q-1)/2$ for every $1 \le i \le m$, and

$$\sum_{i=0}^{m-1} t_i q^i \equiv \left(\sum_{j=0}^{n-1} s_j \sum_{i=0}^{m-1} A_{i,j} q^i \right) \bmod q^m \ .$$

In other words, we interpret the n columns of $\mathbf{A}$ as elements in $\mathbb{Z}_{q^m}$ represented in base q, and sum all the elements in the columns j where $s_j = 1$. The result is an element in $\mathbb{Z}_{q^m}$, which we write in base q using coefficients between $-(q-1)/2$ and $(q-1)/2$. We then write the coefficients of the base q representation as an m-dimensional vector $\mathbf{t}$. It will sometimes be more convenient to consider the subset sum of the numbers represented by the rows of $\mathbf{A}$, and to this effect we naturally define $\mathbf{r}^T \odot \mathbf{A} = \left(\mathbf{A}^T \odot \mathbf{r}\right)^T$.

3 The Encryption Scheme

In this section we present our main contribution: a public-key encryption scheme that is based directly on the hardness of the subset sum problem. Given a security parameter n, we set $q(n)$ to be some number greater than $10n \log^2 n$, let $k \in \mathbb{N}$ be the number of bits we want to encrypt, and define the following encryption scheme:

[4] Impagliazzo and Naor [IN96] only prove their result for q's that are prime or a power of 2, but their results extend to all q.

- **Key generation:** On input 1^n sample $\mathbf{A}' \in \mathbb{Z}_q^{n \times n}$ and $\mathbf{s}_1, \ldots, \mathbf{s}_k \in \{0,1\}^n$ independently and uniformly at random. For every $1 \leq i \leq k$ let $\mathbf{t}_i = \mathbf{A}' \odot \mathbf{s}_i$, and let $\mathbf{A} = [\mathbf{A}' || \mathbf{t}_1 || \cdots || \mathbf{t}_k]$. Output $\mathsf{pk} = \mathbf{A}$ and $\mathsf{sk} = (\mathbf{s}_1, \ldots, \mathbf{s}_k)$.
- **Encryption:** On input a message $\mathbf{z} \in \{0,1\}^k$, sample $\mathbf{r} \in \{0,1\}^n$ uniformly at random, and output the ciphertext $\mathbf{u}^T = \mathbf{r}^T \odot \mathbf{A} + (\frac{q-1}{2})[0^n || \mathbf{z}^T]$.
- **Decryption:** On input a ciphertext $\mathbf{u}^T = [\mathbf{v}^T || w_1 || \cdots || w_k]$ where $\mathbf{v} \in \mathbb{Z}_q^n$ and $w_1, \ldots, w_k \in \mathbb{Z}_q$, for every $1 \leq i \leq k$ compute $y_i = \mathbf{v}^T \mathbf{s}_i - w_i$. If $|y_i| < q/4$ then set $z_i = 0$ and otherwise set $z_i = 1$. Output $\mathbf{z}^T = (z_1, \ldots, z_k)$.

The intuition for the semantic security of the scheme is fairly simple. Because the vectors $\mathbf{t}_i$ are subset sums of the numbers represented by the columns of $\mathbf{A}'$, the public key $\mathbf{A}$ is computationally indistinguishable from random. Therefore, to an observer, the vector $\mathbf{r}^T \odot \mathbf{A}$, which is a subset sum of numbers represented by the rows of $\mathbf{A}$, is again computationally indistinguishable from uniform. The formal proof is in Section 3.1.

The intuition for decryption is based on the fact that $\mathbf{A}' \odot \mathbf{s}_i \approx \mathbf{A}'\mathbf{s}_i$ and $\mathbf{r}^T \odot \mathbf{A} \approx \mathbf{r}^T \mathbf{A}$. For simplicity, assume that $\mathbf{A}' \odot \mathbf{s}_i = \mathbf{A}'\mathbf{s}_i$ and $\mathbf{r}^T \odot \mathbf{A} = \mathbf{r}^T \mathbf{A}$. Then it is not hard to see that

$$|\mathbf{v}^T \mathbf{s}_i - w_i| = \left| (\mathbf{r}^T \mathbf{A}')\mathbf{s}_i - \left(\mathbf{r}^T (\mathbf{A}' \mathbf{s}_i) + \frac{q-1}{2} z_i \right) \right| = \frac{q-1}{2} z_i \,,$$

and we recover z_i. Because the subset sum function does not quite correspond to a vector/matrix multiplication, decryption will recover $\frac{q-1}{2} z_i + \text{error}$. What we will need to show is that this error term is small enough so that we can still tell whether z_i was 0 or 1. The proof is in Section 3.2.

3.1 Proof of Security

Our scheme enjoys a rather simple and direct proof of security. The proof consists of two applications of the pseudorandomness of the subset sum function, which by Lemma 2.2 is implied by the hardness of the subset sum problem. Informally, the first application allows us to replace the values $\mathbf{A}' \odot \mathbf{s}_1, \ldots, \mathbf{A}' \odot \mathbf{s}_k$ in the public key with k vectors that are sampled independently and uniformly at random. Then, the second application allows us to replace the value $\mathbf{r}^T \odot \mathbf{A}$ in the challenge ciphertext with an independently and uniformly chosen vector. In this case, the challenge ciphertext is statistically independent of the encrypted message and the security of the scheme follows. More formally, the following theorem establishes the security of the scheme:

Theorem 3.1. *Assuming the hardness of the $SS(n, q^{n+k})$ problem, where n is the security parameter and k is the plaintext length, the above public-key encryption scheme is semantically secure.*

Proof. We show that for any two messages $\mathbf{m}_0, \mathbf{m}_1 \in \{0,1\}^k$, the ensembles $(\mathsf{pk}, \mathcal{E}_{\mathsf{pk}}(\mathbf{m}_0))$ and $(\mathsf{pk}, \mathcal{E}_{\mathsf{pk}}(\mathbf{m}_1))$ are computationally indistinguishable. In fact, we prove an even stronger statement, namely that $(\mathbf{A}, \mathbf{r}^T \odot \mathbf{A})$ is computationally

indistinguishable from $(\mathbf{M}, \mathbf{v})$, where $\mathbf{M} \in \mathbb{Z}_q^{n \times (n+k)}$ and $\mathbf{v} \in \mathbb{Z}_q^{n+k}$ are sampled independently and uniformly at random. This, in turn, implies that for every $b \in \{0,1\}$, the distribution $(\mathsf{pk}, \mathcal{E}_{\mathsf{pk}}(\mathbf{m}_b))$ is computationally indistinguishable from a distribution that perfectly hides the message $\mathbf{m}_b$. Therefore, any probabilistic polynomial-time adversary attacking the scheme will have a negligible cpa-advantage.

The hardness of the $SS(n, q^n)$ problem, Lemmas 2.1 and 2.2, and a standard hybrid argument imply that the distributions $(\mathbf{A}', \mathbf{A}' \odot \mathbf{s}_1, \ldots, \mathbf{A}' \odot \mathbf{s}_k)$ and $(\mathbf{A}', \mathbf{b}_1, \ldots, \mathbf{b}_k)$, where $\mathbf{b}_1, \ldots, \mathbf{b}_k \in \mathbb{Z}_q^n$ are sampled independently and uniformly at random, are computationally indistinguishable. Letting $\mathbf{M} = [\mathbf{A}' || \mathbf{b}_1 || \cdots || \mathbf{b}_k]$, it then follows that the distributions $(\mathbf{A}, \mathbf{r}^T \odot \mathbf{A})$ and $(\mathbf{M}, \mathbf{r}^T \odot \mathbf{M})$, are computationally indistinguishable. Now, the hardness of the $SS(n, q^{n+k})$ problem and Lemma 2.2 imply that the latter distribution is computationally indistinguishable from $(\mathbf{M}, \mathbf{v})$, where $\mathbf{v} \in \mathbb{Z}_q^{n+k}$ is sampled uniformly at random, independently of $\mathbf{M}$. This concludes the proof of the theorem. □

3.2 Proof of Correctness

We will use the following bound due to Hoeffding [Hoe63] throughout our proof.

Lemma 3.2 (Hoeffding Bound). *Let $X_1, \ldots, X_n$ be independent random variables in the range $[a, b]$ and let $X = X_1 + \ldots + X_n$. Then*

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

The next lemma shows that the carries during the subset sum operation $\mathbf{r}^T \odot \mathbf{A}$ are distributed with mean 0 and their absolute value is bounded (with high probability) by $\sqrt{n} \log n$. In addition, the carries are almost independent of each other. The slight dependency comes from the fact that a carry element can cause the following carry to increase by 1.

Lemma 3.3. *For any $n, m \in \mathbb{N}$ and $\mathbf{r} \in \{0,1\}^n$,*

$$Pr_{\mathbf{A} \xleftarrow{\$} \mathbb{Z}_q^{n \times m}} \left[\|\mathbf{r}^T \odot \mathbf{A} - \mathbf{r}^T \mathbf{A}\|_\infty < \sqrt{n} \log n \right] = 1 - n^{-\omega(1)} .$$

Furthermore, the vector $\mathbf{r}^T \odot \mathbf{A} - \mathbf{r}^T \mathbf{A}$ can be written as a sum of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{Z}_q^m$ where all the coordinates of $\mathbf{x}$ are independently distributed with mean 0, while all the coordinates of $\mathbf{y}$ have absolute value at most 1 (but could be dependent among themselves).

Proof. Computing $\mathbf{r}^T \odot \mathbf{A}$ can be done via the following algorithm, where $\mathbf{a}_i$ is the i-th column of $\mathbf{A}$:

$c_0 = 0$
for $i = 0$ to $m - 1$
  $b_i = (c_i + \mathbf{r}^T \mathbf{a}_i) \bmod q$
  $c_{i+1} = \left\lceil \frac{c_i + \mathbf{r}^T \mathbf{a}_i}{q} \right\rfloor$
output $\mathbf{b}^T = (b_0, \ldots, b_{m-1})$

Notice that this algorithm is just performing addition in base q, where all the coefficients are between $-(q-1)/2$ and $(q-1)/2$. The difference $\mathbf{r}^T \odot \mathbf{A} - \mathbf{r}^T\mathbf{A}$ is simply the "carries" c_i. Note that the only dependency among the c_i's is that c_{i+1} slightly depends on c_i. We can rewrite the above algorithm by writing each c_i as $x_i + y_i$ such that all the x_i's are independent among themselves, whereas the y_i's could be dependent but are very small.

$x_0 = 0\,;\ y_0 = 0$
for $i = 0$ to $m-1$
  $b_i = (x_i + y_i + \mathbf{r}^T\mathbf{a}_i) \bmod q$
  $x_{i+1} = \left\lceil \frac{\mathbf{r}^T\mathbf{a}_i}{q} \right\rfloor$
  $y_{i+1} = \left\lceil \frac{x_i + y_i + \mathbf{r}^T\mathbf{a}_i}{q} \right\rfloor - \left\lceil \frac{\mathbf{r}^T\mathbf{a}_i}{q} \right\rfloor$
output $\mathbf{b}^T = (b_0, \ldots, b_{m-1})$

Observe that in the second algorithm, the x_i's are completely independent among themselves. We now bound the absolute value of the x_i's. Each vector $\mathbf{a}_i$ consists of numbers uniformly distributed between $-(q-1)/2$ and $(q-1)/2$. Applying the Hoeffding bound (Lemma 3.2), we obtain that

$$Pr[|\mathbf{r}^T\mathbf{a}_i| \geq q\sqrt{n}\log n] \leq 2e^{-2\log^2 n}\,.$$

Therefore, with probability $1 - n^{-\omega(1)}$, $|x_i| \leq \sqrt{n}\log n$ for all $0 \leq i \leq m-1$. Also notice that by symmetry, $E[x_i] = 0$. By induction, we will now show that $|y_i| \leq 1$. This is true for y_0, and assume it is true for y_i. Then,

$$|y_{i+1}| = \left| \left\lceil \frac{x_i + y_i + \mathbf{r}^T\mathbf{a}_i}{q} \right\rfloor - \left\lceil \frac{\mathbf{r}^T\mathbf{a}_i}{q} \right\rfloor \right| \leq \left| \left\lceil \frac{x_i + y_i}{q} \right\rfloor + 1 \right| \leq 1\,,$$

where the last inequality follows because $|x_i| \leq \sqrt{n}\log n < q/2 - 1$ and $|y_i| \leq 1$, and so $\left\lceil \frac{x_i + y_i}{q} \right\rfloor = 0$. □

Lemma 3.4. *For any* $\mathbf{r}, \mathbf{s} \in \{0,1\}^n$,

$$Pr_{\mathbf{A} \xleftarrow{\$} \mathbb{Z}_q^{n \times n}} \left[\|(\mathbf{r}^T \odot \mathbf{A})\mathbf{s} - \mathbf{r}^T\mathbf{A}\mathbf{s}\|_\infty < n\log^2 n \right] = 1 - n^{-\omega(1)}\,.$$

Proof. Using Lemma 3.3, we can rewrite $\mathbf{r}^T \odot \mathbf{A}$ as $\mathbf{r}^T\mathbf{A} + \mathbf{x}^T + \mathbf{y}^T$ where each element of $\mathbf{x}$ is independently distributed around 0 with magnitude at most $\sqrt{n}\log n$, and each element of $\mathbf{y}$ has magnitude at most 1. Multiplying by $\mathbf{s}$, we obtain $(\mathbf{r}^T \odot \mathbf{A})\mathbf{s} - \mathbf{r}^T\mathbf{A}\mathbf{s} = \mathbf{x}^T\mathbf{s} + \mathbf{y}^T\mathbf{s}$.

Because $\|\mathbf{y}\|_\infty \leq 1$, we have $|\mathbf{y}^T\mathbf{s}| \leq n$. By the Hoeffding bound (Lemma 3.2), we obtain that

$$Pr[|\mathbf{x}^T\mathbf{s}| \geq n\log^2 n] \leq 2e^{-\frac{\log^2 n}{2}}\,,$$

and the lemma is proved. □

Theorem 3.5. *Decryption succeeds with probability* $1 - n^{-\omega(1)}$.

Proof. The encryption of a message $\mathbf{z}$ is the vector $\mathbf{u}^T = \mathbf{r}^T \odot \mathbf{A} + (\frac{q-1}{2})(0^n || \mathbf{z}^T)$. To decrypt bit i, we write $\mathbf{u}^T = [\mathbf{v}^T || w_1 || \dots || w_k]$ and compute $\mathbf{v}^T \mathbf{s}_i - w_i$. Observe that $\mathbf{v}^T$ is equal to $\mathbf{r}^T \odot \mathbf{A}' + (0^{n-1} || \nu)$ and $w_i = \mathbf{r}^T \mathbf{t}_i + \frac{q-1}{2} z_i + \eta$, where ν, η are carries whose magnitudes are less than n (actually, we can show that with high probability $\nu, \eta < \sqrt{n} \log n$, but the looser bound suffices here). Therefore, if s_n is the last element of $\mathbf{s}_i$, then

$$\begin{aligned}
\mathbf{v}^T \mathbf{s}_i - w_i &= (\mathbf{r}^T \odot \mathbf{A}' + (0^{n-1} || \nu)) \mathbf{s}_i - \left(\mathbf{r}^T \mathbf{t}_i + \frac{q-1}{2} z_i + \eta \right) \\
&= (\mathbf{r}^T \odot \mathbf{A}') \mathbf{s}_i + \nu s_n - \left(\mathbf{r}^T (\mathbf{A}' \odot \mathbf{s}_i) + \frac{q-1}{2} z_i + \eta \right).
\end{aligned}$$

We will now show that $\frac{q-1}{2} z_i$ is the dominant term in the second equation. Thus, if $z_i = 0$, the result will be close to 0, and if $z_i = 1$, the result will be close to $-(q-1)/2$. We will show this by bounding the magnitude of the other terms.

$$\begin{aligned}
&|(\mathbf{r}^T \odot \mathbf{A}') \mathbf{s}_i + \nu s_n - \mathbf{r}^T (\mathbf{A}' \odot \mathbf{s}_i) - \eta| \\
&\qquad \le \left| (\mathbf{r}^T \odot \mathbf{A}') \mathbf{s}_i - \mathbf{r}^T \mathbf{A}' \mathbf{s}_i - \mathbf{r}^T (\mathbf{A}' \odot \mathbf{s}_i) + \mathbf{r}^T \mathbf{A}' \mathbf{s}_i \right| + |\nu s_n| + |\eta| \\
&\qquad \le \left| (\mathbf{r}^T \odot \mathbf{A}') \mathbf{s}_i - \mathbf{r}^T \mathbf{A}' \mathbf{s}_i \right| + \left| \mathbf{r}^T (\mathbf{A}' \odot \mathbf{s}_i) - \mathbf{r}^T \mathbf{A}' \mathbf{s}_i \right| + 2n \\
&\qquad \le n \log^2 n + n \log^2 n + 2n \,,
\end{aligned}$$

where the last inequality follows from applying Lemma 3.4 twice to bound $\left| (\mathbf{r}^T \odot \mathbf{A}') \mathbf{s}_i - \mathbf{r}^T \mathbf{A}' \mathbf{s}_i \right|$ and $\left| \mathbf{r}^T (\mathbf{A}' \odot \mathbf{s}_i) - \mathbf{r}^T \mathbf{A}' \mathbf{s}_i \right|$. So if $z_i = 0$, we will have

$$|\mathbf{v}^T \mathbf{s}_i - w_i| \le 2n \log^2 n + 2n < q/4$$

with probability $1 - n^{-\omega(1)}$, and we will decrypt to 0. If $z_i = 1$, we will decrypt to 1 since

$$|\mathbf{v}^T \mathbf{s}_i - w_i| \ge (q-1)/2 - 2n \log^2 n - 2n > q/4 \,. \quad \square$$

4 Security Against Key-Leakage Attacks

In this section we prove that a natural variant of the scheme described in Section 3 is resilient to any non-adaptive leakage of $L(1 - o(1))$ bits, where L is the length of the secret key (see Appendix A.2 for the formal definition of non-adaptive key-leakage attacks). Given a security parameter n and a leakage parameter $\lambda = \lambda(n)$, set $q = O\left(\left(n + \frac{\lambda}{\log n} \right) n \log^2 n \right)$, $T = \sqrt{q}$, and $m \ge (\lceil n \log q \rceil + \lambda + \omega \log n) / \log T$. Consider the following encryption scheme:

- **Key generation:** On input 1^n sample $\mathbf{A}' \in \mathbb{Z}_q^{n \times m}$ and $\mathbf{s} \in \{-(T-1)/2, \dots, (T-1)/2\}^m$ uniformly and independently at random, and let $\mathbf{A} = [\mathbf{A}' || \mathbf{A}'\mathbf{s}]$. Output $\mathsf{pk} = \mathbf{A}$ and $\mathsf{sk} = \mathbf{s}$.
- **Encryption:** On input a bit b, sample $\mathbf{r} \in \{0,1\}^n$ uniformly at random, and output the ciphertext $\mathbf{u}^T = \mathbf{r}^T \odot \mathbf{A} + (\frac{q-1}{2})[0^m || b]$.

– **Decryption:** On input a ciphertext $\mathbf{u}^T = [\mathbf{v}^T||w]$ where $\mathbf{v} \in \mathbb{Z}_q^m$ and $w \in \mathbb{Z}_q$, compute $y = \mathbf{v}^T\mathbf{s} - w$. If $|y| < q/4$ then output 0. Otherwise, output 1.

The main idea underlying the scheme is that the min-entropy of the secret key is $m \log T \geq \lceil n \log q \rceil + \lambda + \omega \log n$, and thus even given any leakage of λ bits it still has average min-entropy at least $\lceil n \log q \rceil + \omega \log n$. Since the leakage is independent of the public key, we can apply the leftover hash lemma and argue that $\mathbf{A} = [\mathbf{A}'||\mathbf{A}'\mathbf{s}]$ is statistically close to uniform, even given the leakage.

We note that in this scheme, unlike in the scheme presented in Section 3, we use matrix-vector multiplication instead of the subset sum operation in forming the public key. The proof of correctness in this case is similar to that presented in Section 3. Specifically, a generalization of Lemma 3.4 shows that for every $\mathbf{r} \in \{0,1\}^n$ and $\mathbf{s} \in \{-(T-1)/2, \ldots, (T-1)/2\}^m$, with overwhelming probability over the choice of $\mathbf{A} \xleftarrow{\$} \mathbb{Z}_q^{n \times m}$ it holds that $\|(\mathbf{r}^T \odot \mathbf{A})\mathbf{s} - \mathbf{r}^T\mathbf{A}\mathbf{s}\|_\infty < \sqrt{Tmn}\log^2 n + Tm$. As in the proof of Theorem 3.5, this implies that $\mathbf{v}^T\mathbf{s} - w = \gamma + \frac{q-1}{2}z$, where $|\gamma| \leq \sqrt{Tmn}\log^2 n + (T+2)m$. Therefore, we need to set q to be an integer such that $q/4 > \sqrt{Tmn}\log^2 n + (T+2)m$. By setting roughly $q = \left(n + \frac{\lambda}{\log n}\right) n \log^2 n$ (ignoring a small leading constant) and $T = \sqrt{q}$, we can base the security of the scheme on the hardness of the $SS(n, q^m)$ problem, where

$$q^m = q^{\frac{n \log q + \lambda + \omega \log n}{\log T}} = q^{2\left(n + \frac{\lambda + \omega \log n}{\log q}\right)} = \left(\left(n + \frac{\lambda}{\log n}\right) n \log^2 n\right)^{2n} \cdot 4^{\lambda + \omega \log n} .$$

The following theorem establishes the security of the scheme:

Theorem 4.1. *Assuming the hardness of the $SS(n, q^{m+1})$ problem for $q = q(n)$ and $m = m(n)$ as above, the scheme is semantically secure against non-adaptive $\lambda(n)$-key-leakage attacks, where n is the security parameter.*

Proof. We show that for any efficiently computable leakage function f mapping secret keys into $\{0,1\}^\lambda$, the ensembles $(\mathsf{pk}, \mathcal{E}_{\mathsf{pk}}(0), f(\mathsf{sk}))$ and $(\mathsf{pk}, \mathcal{E}_{\mathsf{pk}}(1), f(\mathsf{sk}))$ are computationally indistinguishable. In fact, we prove a stronger statement, namely that $(\mathbf{A}, \mathbf{r}^T \odot \mathbf{A}, f(\mathbf{s}))$ is computationally indistinguishable from $(\mathbf{M}, \mathbf{v}, f(\mathbf{s}))$, where $\mathbf{M} \in \mathbb{Z}_q^{n \times (m+1)}$, $\mathbf{v} \in \mathbb{Z}_q^{m+1}$ are sampled independently, uniformly at random.

Lemma A.1 guarantees that the average min-entropy of $\mathbf{s}$ given $f(\mathbf{s})$ is at least $m \log T - \lambda \geq n \log q + \omega \log n$. The leftover hash lemma (when adapted to the notion of average min-entropy – see Lemma A.3) then implies that the statistical distance between the distributions $(\mathbf{A}', \mathbf{A}'\mathbf{s}, f(\mathbf{s}))$ and $(\mathbf{A}', \mathbf{t}, f(\mathbf{s}))$, where $\mathbf{t} \in \mathbb{Z}_q^n$ is sampled uniformly at random, is negligible in n. Letting $\mathbf{M} = [\mathbf{A}'||\mathbf{t}]$ and noting that applying a deterministic function cannot increase the statistical distance between distributions, it follows that the statistical distance between $(\mathbf{A}, \mathbf{r}^T \odot \mathbf{A}, f(\mathbf{s}))$ and $(\mathbf{M}, \mathbf{r}^T \odot \mathbf{M}, f(\mathbf{s}))$, where $\mathbf{M} \in \mathbb{Z}_q^{n \times (m+1)}$ is sampled uniformly at random, is negligible. Now, the hardness of the $SS(n, q^{m+1})$ problem implies that the latter distribution is computationally indistinguishable from $(\mathbf{M}, \mathbf{v}, f(\mathbf{s}))$, where $\mathbf{v} \in \mathbb{Z}_q^{m+1}$ is sampled uniformly at random, independently of $\mathbf{M}$. This concludes the proof of the theorem. □

5 Oblivious Transfer Protocol

In this section we present an oblivious transfer (OT) protocol based on subset sum that provides security for the receiver against a cheating sender, and security for the sender against an honest-but-curious receiver. (See Appendix A.3 for the formal definition of OT.) Our protocol is an instance of a construction proposed by Gertner et al. [GKM+00], based on protocols by Even et al. [EGL82,EGL85], to show that a special property of public-key encryption is sufficient for the construction of two-message semi-honest OT. Informally, the property is that it is possible to efficiently sample a string pk with a distribution indistinguishable from that of a properly generated public key, while preserving the semantic security of the encryption $\mathcal{E}_{\mathsf{pk}}$. Our cryptosystem satisfies this property, by pseudorandomness of subset sum. For the sake of self-containment, however, we provide direct proofs of our OT protocol's correctness and security.

5.1 OT Based on Subset Sum

Our oblivious transfer protocol is a simple application of our encryption scheme. We denote by $\mathcal{G}$, $\mathcal{E}$ and $\mathcal{D}$, respectively, the key-generation, encryption and decryption algorithms of the public-key encryption scheme described in Section 3. The receiver with inputs $1^n, b$ first sends a properly generated public key pk_b and a uniformly random fake public key $\mathsf{pk}_{1-b} \in \mathbb{Z}_q^{n\times(n+k)}$. The sender with inputs $1^n, \mathbf{z}_0, \mathbf{z}_1$ uses each key pk_i to encrypt its input $\mathbf{z}_i$ and replies with the ciphertexts $\mathbf{u}_0^T, \mathbf{u}_1^T$. The receiver can then retrieve $\mathbf{z}_b$ by decrypting $\mathbf{u}_b^T$, using the secret key corresponding to pk_b. Details follow.

Let $n, k \in \mathbb{N}$, $b \in \{0,1\}$, and $\mathbf{z}_0, \mathbf{z}_1 \in \{0,1\}^k$

Receiver $R(1^n, b)$:	$(\mathsf{pk}_b, \mathsf{sk}_b) \xleftarrow{\$} \mathcal{G}(1^n)$; $\mathsf{pk}_{1-b} \xleftarrow{\$} \mathbb{Z}_q^{n\times(n+k)}$; Send $\mathsf{pk}_0, \mathsf{pk}_1$
Sender $S(1^n, \mathbf{z}_0, \mathbf{z}_1)$:	$\mathbf{u}_0^T \leftarrow \mathcal{E}_{\mathsf{pk}_0}(\mathbf{z}_0)$; $\mathbf{u}_1^T \leftarrow \mathcal{E}_{\mathsf{pk}_1}(\mathbf{z}_1)$; Send $\mathbf{u}_0^T, \mathbf{u}_1^T$
Receiver R:	$\mathbf{z}_b \leftarrow \mathcal{D}_{\mathsf{sk}_b}(\mathbf{u}_b^T)$; Return $\mathbf{z}_b$

5.2 Proofs of Correctness and Security

We now show that correctness follows from correctness of the cryptosystem.

Theorem 5.1. *If the sender and receiver both follow the protocol, then the former outputs nothing and the latter outputs* $\mathbf{z}_b$ *with probability* $1 - n^{-\omega(1)}$.

Proof. Since pk_b is a properly generated public key corresponding to secret key sk_b, $\mathbf{u}_b^T$ is a valid encryption of message $\mathbf{z}_b$ under pk_b, and the receiver computes the decryption of $\mathbf{u}_b^T$ using sk_b, the proof follows from Theorem 3.5. □

Security for the receiver is proved based on the pseudorandomness of subset sum. A properly generated public key is indistinguishable from a uniformly random element in $\mathbb{Z}_q^{n\times(n+k)}$. Therefore, for any input bit, the receiver's message consists of two elements from computationally indistinguishable distributions.

It follows that the distribution of the receiver's message when the input is 0 is computationally indistinguishable from the distribution when the input is 1. The precise statement of this result is the following.

Theorem 5.2. *Assuming the hardness of the $SS(n, q^n)$ problem, where n is the security parameter, the above OT protocol is secure for the receiver.*

Proof. Let $R(1^n, b)$ denote the message sent by the honest receiver with inputs $1^n, b$. We show that the ensembles $R(1^n, 0)$ and $R(1^n, 1)$ are computationally indistinguishable.

As in the proof of Theorem 3.1, the hardness of the $SS(n, q^n)$ problem implies that the distributions pk_0 and pk_1 are computationally indistinguishable. This implies that ensembles $R(1^n, 0)$ and $R(1^n, 1)$ are indistinguishable as well. □

The protocol is not secure against malicious receivers. Indeed, a malicious receiver can properly generate two key pairs $\mathsf{pk}_0, \mathsf{sk}_0$ and $\mathsf{pk}_1, \mathsf{sk}_1$, and then use the secret keys to decrypt both ciphertexts $\mathbf{u}_0^T, \mathbf{u}_1^T$. The protocol is, however, secure for the sender against honest-but-curious receivers, as we now show.

Theorem 5.3. *Assuming the hardness of the $SS(n, q^{n+k})$ problem, where n is the security parameter and k is the length of the sender's input messages, the above OT protocol is secure for the sender against an honest-but-curious receiver.*

Proof. Let $R(1^n, b)$ denote the message sent by the honest receiver with inputs $1^n, b$, and $S(1^n, \mathbf{z}_0, \mathbf{z}_1, R(1^n, b))$ denote the reply of the honest sender with inputs $1^n, \mathbf{z}_0, \mathbf{z}_1$. We show that the ensembles $(S(1^n, \mathbf{z}_0, \mathbf{z}_1, R(1^n, 0)), R(1^n, 0))$ and $(S(1^n, \mathbf{z}_0, 0^k, R(1^n, 0)), R(1^n, 0))$ are computationally indistinguishable, and the ensembles $(S(1^n, \mathbf{z}_0, \mathbf{z}_1, R(1^n, 1)), R(1^n, 1))$ and $(S(1^n, 0^k, \mathbf{z}_1, R(1^n, 1)), R(1^n, 1))$ are computationally indistinguishable.

In the proof of Theorem 3.1, we showed that for any $\mathbf{m}_0, \mathbf{m}_1 \in \{0,1\}^k$ the ensembles $(\mathsf{pk}, \mathcal{E}_{\mathsf{pk}}(\mathbf{m}_0))$ and $(\mathsf{pk}, \mathcal{E}_{\mathsf{pk}}(\mathbf{m}_1))$ are computationally indistinguishable. This is true when pk is a properly generated public key and also when pk is a random element in $\mathbb{Z}_q^{n \times (n+k)}$. Therefore, the ensembles $(\mathsf{pk}_{1-b}, \mathcal{E}_{\mathsf{pk}_{1-b}}(\mathbf{z}_{1-b}))$ and $(\mathsf{pk}_{1-b}, \mathcal{E}_{\mathsf{pk}_{1-b}}(0^k))$ are computationally indistinguishable. Hence for $b \in \{0,1\}$ the ensembles $(S(1^n, \mathbf{z}_b, \mathbf{z}_{1-b}, R(1^n, b)), R(1^n, b))$ and $(S(1^n, \mathbf{z}_b, 0^k, R(1^n, b)), R(1^n, b))$ are computationally indistinguishable. This completes the proof. □

Acknowledgements

We would like to thank Phong Nguyen for pointing out a mistake in our connection between uSVP and subset sum in Section 1.2, and Petros Mol for pointing out that the result of Impagliazzo and Naor stated in Lemma 2.2 generalizes for all q. The first author would also like to thank Richard Lindner, Markus Rückert, and Michael Schneider for useful conversations.

References

[ACPS09] B. Applebaum, D. Cash, C. Peikert, and A. Sahai. Fast cryptographic primitives and circular-secure encryption based on hard learning problems. In *CRYPTO*, 2009.

[AD97] M. Ajtai and C. Dwork. A public-key cryptosystem with worst-case/average-case equivalence. In *STOC*, 1997. An improved version is described in ECCC 2007.

[AGV09] A. Akavia, S. Goldwasser, and V. Vaikuntanathan. Simultaneous hardcore bits and cryptography against memory attacks. In *TCC*, 2009.

[Ale03] M. Alekhnovich. More on average case vs approximation complexity. In *FOCS*, 2003.

[Cré87] C. Crépeau. Equivalence between two flavours of oblivious transfers. In *CRYPTO*, 1987.

[DORS08] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM J. Computing*, 38(1), 2008.

[EGL82] S. Even, O. Goldreich, and A. Lempel. A randomized protocol for signing contracts. In *CRYPTO*, 1982.

[EGL85] S. Even, O. Goldreich, and A. Lempel. A randomized protocol for signing contracts. *Communications of the ACM*, 28(6), 1985.

[FP05] A. Flaxman and B. Przydatek. Solving medium-density subset sum problems in expected polynomial time. In *STACS*, 2005.

[Fri86] A. Frieze. On the Lagarias-Odlyzko algorithm for the subset sum problem. *SIAM Journal on Computing*, 15, 1986.

[GKM+00] Y. Gertner, S. Kannan, T. Malkin, O. Reingold, and M. Viswanathan. The relationship between public key encryption and oblivious transfer. In *FOCS*, 2000.

[GMW87] O. Goldreich, S. Micali, and A. Wigderson. How to play a mental game - a completeness theorem for protocols with honest majority. In *STOC*, 1987.

[Gol04] O. Goldreich. *Foundations of Cryptography - Volume 2 (Basic Applications)*. Cambridge University Press, 2004.

[GPV08] C. Gentry, C. Peikert, and V. Vaikuntanathan. Trapdoors for hard lattices, and new cryptographic constructions. In *STOC*, 2008.

[Hai08] I. Haitner. Semi-honest to malicious oblivious transfer – The black-box way. In *TCC*, 2008.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301), 1963.

[HSH+08] J. A. Halderman, S. D. Schoen, N. Heninger, W. Clarkson, W. Paul, J. A. Calandrino, A. J. Feldman, J. Appelbaum, and E. W. Felten. Lest we remember: Cold boot attacks on encryption keys. In *USENIX Security*, 2008.

[IN96] R. Impagliazzo and M. Naor. Efficient cryptographic schemes provably as secure as subset sum. *Journal of Cryptology*, 9(4), 1996.

[Kil88] J. Kilian. Founding cryptography on oblivious transfer. In *STOC*, 1988.

[LM09] V. Lyubashevsky and D. Micciancio. On bounded distance decoding, unique shortest vectors, and the minimum distance problem. In *CRYPTO*, 2009.

[LO85] J. C. Lagarias and A. M. Odlyzko. Solving low density subset sum problems. *Journal of the ACM*, 32, 1985.

[LPS09] V. Lyubashevsky, A. Palacio, and G. Segev. Public-key cryptographic primitives provably as secure as subset sum. *ePrint*, 2009.
[Lyu05] V. Lyubashevsky. The parity problem in the presence of noise, decoding random linear codes, and the subset sum problem. In *RANDOM*, 2005.
[MH78] R.C. Merkle and M.E. Hellman. Hiding information and signatures in trapdoor knapsacks. *IEEE Trans. on Inf. Theory*, IT-24, 1978.
[NS09] M. Naor and G. Segev. Public-key cryptosystems resilient to key leakage. In *CRYPTO*, 2009.
[Odl90] A. Odlyzko. The rise and fall of knapsack cryptosystems. In *Symposia of Applied Mathematics*, 1990.
[Pei09] C. Peikert. Public-key cryptosystems from the worst-case shortest vector problem. In *STOC*, 2009.
[PVW08] C. Peikert, V. Vaikuntanathan, and B. Waters. A framework for efficient and composable oblivious transfer. In *CRYPTO*, 2008.
[PW08] C. Peikert and B. Waters. Lossy trapdoor functions and their applications. In *STOC*, 2008.
[Rab81] M. O. Rabin. How to exchange secret keys by oblivious transfer. In *Technical Report TR-81*. Harvard Aiken Computation Laboratory, 1981.
[Reg03] O. Regev. New lattice based cryptographic constructions. In *STOC*, 2003.
[Reg05] O. Regev. On lattices, learning with errors, random linear codes, and cryptography. In *STOC*, 2005.
[Sha08] A. Shallue. An improved multi-set algorithm for the dense subset sum problem. In *ANTS*, 2008.
[Sho97] P. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5), 1997.
[Yao86] A. C. Yao. How to generate and exchange secrets. In *FOCS*, 1986.

A Cryptographic Definitions

Do to space constraints, the well-known definition of semantically-secure public-key encryption is presented in the full version of our paper [LPS09].

A.1 Randomness Extraction

We say that two variables are ϵ-*close* if their statistical distance is at most ϵ. The *min-entropy* of a random variable X is $\mathrm{H}_\infty(X) = -\log(\max_x \Pr[X = x])$.

Dodis et al. [DORS08] formalized the notion of *average min-entropy* that captures the remaining unpredictability of a random variable X conditioned on the value of a random variable Y, formally defined as follows:

$$\widetilde{\mathrm{H}}_\infty(X|Y) = -\log\left(E_{y\leftarrow Y}\left[2^{-\mathrm{H}_\infty(X|Y=y)}\right]\right) .$$

The average min-entropy corresponds exactly to the optimal probability of guessing X, given knowledge of Y. The following bound on average min-entropy was proved in [DORS08]:

Lemma A.1 ([DORS08]). *If Y has 2^r possible values and Z is any random variable, then*

$$\widetilde{\mathrm{H}}_\infty(X|(Y,Z)) \geq \mathrm{H}_\infty(X|Z) - r .$$

A main tool in our constructions in this paper is a strong randomness extractor. The following definition naturally generalizes the standard definition of a strong extractor to the setting of average min-entropy:

Definition A.2 ([DORS08]). *A function* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is an* average-case (k,ϵ)-strong extractor *if for all random variables* X *and* I *such that* $X \in \{0,1\}^n$ *and* $\widetilde{\mathrm{H}}_\infty(X|I) \geq k$ *it holds that*

$$\mathrm{SD}\left((\mathsf{Ext}(X,S),S,I),(U_m,S,I)\right) \leq \epsilon\,,$$

where S *is uniform over* $\{0,1\}^d$.

Dodis et al. proved the following lemma stating that any strong extractor is in fact also an average-case strong extractor:

Lemma A.3 ([DORS08]). *For any* $\delta > 0$*, if* Ext *is a (worst-case)* $(m - \log(1/\delta),\epsilon)$*-strong extractor, then* Ext *is also an average-case* $(m,\epsilon+\delta)$*-strong extractor.*

A.2 Key-Leakage Attacks

We follow the framework introduced by Akavia et al. [AGV09] and recall their notion of a key-leakage attack. Informally, an encryption scheme is secure against key-leakage attacks if it is semantically secure even when the adversary obtains sensitive leakage information. This is modeled by allowing the adversary to submit any function f and receive $f(\mathsf{sk})$, where sk is the secret key, as long as the output length of f is bounded by a predetermined parameter λ.

Akavia et al. defined two notions of key-leakage attacks: adaptive attacks and non-adaptive attacks. In an adaptive key-leakage attack, the adversary is allowed to choose the leakage function after seeing the public key, and in a non-adaptive key-leakage attack the adversary has to choose the leakage function in advance. In this paper we deal with the non-adaptive setting, and note that this notion of leakage is still very meaningful as it captures many realistic attacks in which the leakage does not depend on the parameters of the encryption scheme. For example, it captures the cold boot attacks of Halderman et al. [HSH+08], in which the leakage depends only on the properties of the hardware devices that are used for storing the secret key.

Formally, for a public-key encryption scheme $(\mathcal{G},\mathcal{E},\mathcal{D})$ we denote by sk_n and pk_n the sets of secret keys and public keys that are produced by $\mathcal{G}(1^n)$. That is, $\mathcal{G}(1^n) : \{0,1\}^* \to \mathsf{sk}_n \times \mathsf{pk}_n$ for every $n \in \mathbb{N}$. The following defines the notion of a non-adaptive key-leakage attack:

Definition A.4 (non-adaptive key-leakage attacks). *A public-key encryption scheme* $(\mathcal{G},\mathcal{E},\mathcal{D})$ *is* semantically secure against non-adaptive $\lambda(n)$-key-leakage attacks *if for any collection* $\mathcal{F} = \left\{f_n : \mathsf{sk}_n \to \{0,1\}^{\lambda(n)}\right\}_{n\in\mathbb{N}}$ *of efficiently computable functions and any two messages* m_0 *and* m_1*, the distributions* $(\mathsf{pk}, \mathcal{E}_{\mathsf{pk}}(m_0), f_n(\mathsf{sk}))$ *and* $(\mathsf{pk}, \mathcal{E}_{\mathsf{pk}}(m_1), f_n(\mathsf{sk}))$ *are computationally indistinguishable, where* $(\mathsf{sk},\mathsf{pk}) \xleftarrow{\$} \mathcal{G}(1^n)$.

A.3 Oblivious Transfer

Oblivious transfer is a cryptographic primitive, introduced by Rabin [Rab81], which has been shown to be sufficiently strong to enable any multiparty computation [Yao86,GMW87,Kil88]. There are several equivalent formulations of OT in the literature. We use the version of Even, et al. [EGL85] known as *1-out-of-2 oblivious transfer*, and refer to it as simply OT. Crépeau [Cré87] showed that this variant is equivalent to the original definition of oblivious transfer.

A 1-out-of-2 oblivious transfer is a two-party protocol in which a sender has two secret strings $\mathbf{z}_0, \mathbf{z}_1$ and a receiver has a secret bit b. At the end of the interaction, the receiver learns $\mathbf{z}_b$ but has no information about $\mathbf{z}_{1-b}$, and the sender learns nothing about b. General OT guarantees security even in the face of cheating parties who deviate from the prescribed protocol. *Honest* OT, on the other hand, guarantees security only against honest-but-curious parties. These are parties that follow the protocol, but keep a record of all intermediate results and may perform any computation to extract additional information from this record, once the protocol ends. Any honest OT protocol can be transformed into a general OT protocol, using either black-box techniques [Hai08], or using zero-knowledge proofs to force parties to behave in an honest-but-curious manner [Gol04]. The formal definition of OT follows.

Definition A.5. *Oblivious Transfer (OT) is a two-party protocol involving a sender S with inputs 1^n and $\mathbf{z}_0, \mathbf{z}_1 \in \{0,1\}^k$, where k is a constant, and a receiver R with inputs 1^n and $b \in \{0,1\}$. S and R are polynomial-time randomized algorithms such that if both follow the protocol, then the former outputs nothing and the latter outputs $\mathbf{z}_b$ (with overwhelming probability). We consider the following security requirements:*

Security for the receiver: *Let $R(1^n, b)$ denote the message sent by the honest receiver with inputs $1^n, b$. Then the ensembles $\{R(1^n, 0)\}_{n\in\mathbb{N}}$ and $\{R(1^n, 1)\}_{n\in\mathbb{N}}$ are computationally indistinguishable.*

Security for the sender: *Let $S(1^n, \mathbf{z}_0, \mathbf{z}_1, m)$ denote the message sent by the honest sender with inputs $1^n, \mathbf{z}_0, \mathbf{z}_1$ when the (possibly cheating, polynomial time) receiver's message is m. Then for every $\mathbf{z}_0, \mathbf{z}_1 \in \{0,1\}^k$ and every polynomial-length message $m \in \{0,1\}^*$, either the ensembles $\{S(1^n, \mathbf{z}_0, \mathbf{z}_1, m), m\}_{n\in\mathbb{N}}$ and $\{S(1^n, \mathbf{z}_0, 0^k, m), m\}_{n\in\mathbb{N}}$ **or** the ensembles $\{S(1^n, \mathbf{z}_0, \mathbf{z}_1, m), m\}_{n\in\mathbb{N}}$ and $\{S(1^n, 0^k, \mathbf{z}_1, m), m\}_{n\in\mathbb{N}}$ are computationally indistinguishable.*

Security against honest-but-curious (a.k.a. "semi-honest") receivers relaxes the second condition above to consider only a receiver that faithfully follows the protocol, but keeps a record of all intermediate results and may perform any computation, after the protocol is completed, to extract additional information from this record.