Codes and Lattices in Cryptography Mini-Course **FCT-UNL, Spring 2023**

Notes 2: Cryptography from LWE and SIS

Lecturer: João Ribeiro

Introduction

In this lecture we construct some basic cryptographic objects based on the (conjectured) hardness of two simple but widely applicable computational problems, the Short Integer Solution (SIS) problem and the Learning With Errors (LWE) problem. As we shall discuss in detail in the next lecture, the conjectured hardness of these problems is implied by the hardness of a well-studied computational problem on point lattices. It is believed that the latter problem is hard to solve by quantum computers, making SIS and LWE popular building blocks for post-quantum cryptography.

The material covered here is mostly based on the excellent surveys of Peikert [Pei15] and Regev [Reg10].

2.1 The Short Integer Solution problem

The Short Integer Solution problem was first introduced in cryptography by Ajtai [Ajt04]. Roughly speaking, it asks us to find a short integer vector in the kernel of a random q-ary matrix. More precisely, we have the following definition.

Definition 2.1 (Short Integer Solution problem) *The* Short Integer Solution *(SIS) problem parameterized by positive integers n, m, q and a real number $\beta > 0$, denoted by $\mathsf{SIS}_{n,m,q,\beta}$, corresponds to the following search problem:*

1. *Sample a uniformly random matrix $A \in \mathbb{Z}_q^{n \times m}$;*
2. *Given A, find a nonzero integer vector $z \in \mathbb{Z}^m$ such that $Az = 0 \pmod q$ and $\|z\|_2 \leq \beta$.*

Some observations are in order. First, SIS is easy to solve via gaussian elimination if no upper bound is placed on the norm of the solution. The claim is that it is hard to find such *short* solutions. Second, we must at the very least take $q > \beta$ for SIS to be (hopefully) hard, since otherwise $z = (q, 0, \ldots, 0)$ is a valid short solution.

Note also that we are not guaranteed a solution to SIS for all possible choices of parameters. However, the following theorem states that a solution is guaranteed to exist whenever m and β are chosen to be appropriately large compared to n and q.

Theorem 2.1 *If $m > n \log q$ and $\beta > \sqrt{n \log q}$, then $\mathsf{SIS}_{n,m,q,\beta}$ has at least one solution.*

Proof: This follows by a simple pigeonhole argument. Fix a matrix $A \in \mathbb{Z}_q^{n \times m}$. Without loss of generality, we can take m to be the smallest integer strictly larger than $n \log q$ (we can extend any solution z for this choice of m to larger m' by appending 0's to z). Since there are 2^m vectors in $\{0,1\}^m$ and $2^m > q^n$ by our choice of m, there exist two distinct vectors $z, z' \in \{0,1\}^m$ such that $Az = Az' \pmod q$. But then $A(z - z') = 0 \pmod q$, and so $w = z - z'$ is the desired solution with $\|w\|_2 \leq \sqrt{m} \leq \beta$. ∎

2.1.1 One-way functions from SIS

Ajtai [Ajt04] proposed a simple family of functions that is easily shown to be one-way based on the hardness of solving SIS with appropriate parameters. Consider the function $f : \mathbb{Z}_q^{n \times m} \times \{0,1\}^m \to \mathbb{Z}_q^{n \times m} \times \mathbb{Z}_q^n$ defined as

$$f(A, z) = (A, Az \pmod q). \tag{2.1}$$

It is not hard to show that this family is one-way based on SIS.

Theorem 2.2 *The family of functions defined in Equation (2.1) is one-way provided that* $\mathsf{SIS}_{n,m,q,\beta}$ *with* $m = \lceil 2n \log q \rceil \gg n \log q$ *and* $\beta = \sqrt{m}$ *is hard for PPT adversaries.*

Proof: Suppose that Adv is a PPT algorithm which when given $(A, Az \pmod q)$ for $A \leftarrow \mathbb{Z}_q^{n \times m}$ and $z \leftarrow \{0,1\}^m$ outputs $z' \in \mathbb{Z}^m$ such that $Az = Az' \pmod q$ with non-negligible probability $p = p(n)$.

We can use Adv to solve $\mathsf{SIS}_{n,m,q,\beta}$ with non-negligible probability at least $p/3$ as follows: On input $A \leftarrow \mathbb{Z}_q^{n \times m}$, sample $z \leftarrow \{0,1\}^m$ and run $\mathsf{Adv}(A, Az \pmod q)$. Suppose that Adv outputs z'. Then, it holds that

$$A(z - z') = 0 \pmod q,$$

and so $w = z - z'$ is a solution to $\mathsf{SIS}_{n,m,q,\beta}$ provided that $w \neq 0$. The choice of m and β in the theorem statement ensures that this will hold with probability at least $1/2$.

In more detail, the choice of m versus n and q ensures that there are $2^m \geq q^{2n}$ vectors $z \in \{0,1\}^m$, but only q^n possible values for $Az \pmod q$. Call $z \in \{0,1\}^m$ *bad* if the preimage of $Az \pmod q$ contains at most one vector from $\{0,1\}^m$. Note that there are at most q^n such bad z's. Therefore, the probability that our uniform sample $z \leftarrow \{0,1\}^m$ is bad is at most $q^{n-2n} = q^{-n} = \mathsf{negl}(n)$. If our sampled z is not bad and Adv then outputs z' such that $Az' = Az \pmod q$, we will have $z \neq z'$ with probability at least $1/2$, as desired. Therefore, the final success probability of our reduction can be lower bounded by $p/2 - q^{-n} = p/2 - \mathsf{negl}(n) \geq p/3$ for sufficiently large n.

Finally, we remark that for usual choices of the modulus q we can take m to be much closer to $n \log q$ and this argument will still go through. ∎

As mentioned in the previous lecture, such a one-way function is sufficient to construct several useful cryptographic primitives, such as pseudorandom generators, IND-CCA-secure symmetric-key encryption, message authentication codes, digital signatures, and commitment schemes.

Remark 2.1 The family of functions in Equation (2.1) is not only one-way but also *collision-resistant* based on the hardness of SIS: Given a random matrix A, it is infeasible to find distinct

$z, z' \in \{0,1\}^m$ such that $Az = Az' \mod q$. Such collision-resistant function families are broadly useful in cryptography.

Parameters for SIS and more structured problems. As we shall discuss in more detail later on, our current understanding of $\mathsf{SIS}_{n,m,q,\beta}$ requires us to set the modulus q to be polynomial in n. This is so that we can be confident that the associated SIS instance is hard to solve. Ajtai's original argument [Ajt04] required that $q = n^c$ for a large constant $c > 0$, but this has since been improved to the nearly-optimal $q \approx \sqrt{n}$ through a series of works culminating in [MP13].

Protocols based on SIS usually suffer from large keys. This is mostly due to the fact that we need to set $m > n \log q$ so that the associated $\mathsf{SIS}_{n,n,m,\beta}$ problem has a solution. As a result, we need to store roughly n^2 elements from $\mathbb{Z}_q$, leading to nearly quadratic keylength. Motivated by this, researchers have studied other versions of SIS, such as the ring-SIS problem, where keys can be made significantly shorter and computations can be performed much faster. However, the conjectured hardness of these problems is based on less standard assumptions. For an in-depth discussion on this topic, see Stephens-Davidowitz's lecture notes [Ste18].

2.2 The Learning With Errors problem

The Learning With Errors (LWE) problem was introduced by Regev [Reg09] in a work that was awarded the 2018 Gödel prize. At a high level, the LWE problem asks us to invert a (random) system of linear equations corrupted by short errors. The following definition presents a more formal description of the problem.

Definition 2.2 (Search Learning With Errors problem) *The* search Learning With Errors *(search-LWE) problem parameterized by positive integers n, m, q and an error distribution χ on $\mathbb{Z}_q$, denoted by* $\mathsf{Search\text{-}LWE}_{n,m,q,\chi}$*, corresponds to the following search problem:*

1. *Sample a uniformly random matrix $A \in \mathbb{Z}_q^{n \times m}$, a uniformly random secret $s \in \mathbb{Z}_q^m$, and an error vector $e = (e_1, \ldots, e_n)$ with each e_i sampled independently according to χ.*
2. *Given A and $As + e \pmod q$, find s.*

The above definition presents the search version of LWE, where we are tasked with finding the input secret s. It is also useful to consider the natural decision version of this problem, where we wish to distinguish $As + e \pmod q$ from a uniformly random vector.

Definition 2.3 (Decision Learning With Errors problem) *The* decision Learning With Errors *(decision-LWE) problem parameterized by positive integers n, m, q and an error distribution χ on $\mathbb{Z}_q$, denoted by* $\mathsf{Decision\text{-}LWE}_{n,m,q,\chi}$*, corresponds to the following distinguishing game played between an adversary and a challenger:*

1. *The challenger samples a uniformly random matrix $A \in \mathbb{Z}_q^{m \times n}$, a uniformly random secret $s \in \mathbb{Z}_q^n$, and an error vector $e = (e_1, \ldots, e_m)$ with each e_i sampled independently according to χ.*

2. *The challenger also samples a challenge bit b uniformly from $\{0,1\}$. If $b = 0$, then the challenger sends $(A, As + e \pmod q)$ to the adversary. Otherwise, if $b = 1$ then the challenger samples $u \leftarrow \mathbb{Z}_q^m$ and sends (A, u) to the adversary.*

3. *The adversary outputs b' and wins if $b' = b$.*

Search-LWE is at least as hard as Decision-LWE (in fact, both problems have essentially equivalent hardness), and both problems are easily solved via Gaussian elimination if there is no error (i.e., $e = 0$ with probability 1). Our claim is that (for appropriate parameters) these problems are hard to solve when some small errors are introduced (again, according to an appropriate error distribution).

Usual parameters and error distribution for LWE. As was done for SIS, it is instructive to discuss the usual parameter and error distribution regime for LWE. The modulus q is usually taken to be polynomial in n. The error distribution is usually taken to be a mean-zero gaussian distribution rounded to the nearest integer and then reduced modulo q (a so-called *discrete gaussian*) with variance $q/\mathsf{poly}(n)$. One property that is often exploited in LWE-based cryptographic schemes is that the error vector has norm $\ll q$ with high probability, and so can be reliably rounded-off in crucial steps of the protocol.

Similarly to SIS, LWE-based protocols have large keys because we need to set $m > n \log q$. This has led researchers to consider more structured versions of LWE, such as the ring-LWE problem [SSTX09, LPR10], which yield more efficient cryptographic protocols under less standard hardness assumptions.

We note also that LWE and ring-LWE are sometimes also studied with other error distributions, such as binary errors sampled uniformly at random from $\{0,1\}$ [MP13].

2.2.1 Public-key encryption from LWE

The conjectured hardness of Decision-LWE can be used to construct, among many other "cryptomania" primitives, public-key encryption schemes with post-quantum security.[1] We will discuss Regev's PKE scheme [Reg09], which is secure provided that Decision-LWE (with an appropriate parameterization) is hard to solve by PPT adversaries.

We proceed to describe the key generation, encryption, and decryption procedures of Regev's PKE scheme for single-bit messages. For this scheme to be correct (i.e., for us to be able to correctly decrypt ciphertexts), we need to use an error distribution χ that generates short error vectors with high probability. We discuss this in more detail below, and remark that usual instantiations of $\mathsf{Decision\text{-}LWE}_{n,m,q,\chi}$ believed to be hard have this property. For the security proof we will additionally require that $m \geq (n+1)\log q$, which is also fine from a hardness perspective.

- **Key generation:** Sample a secret $s \leftarrow \mathbb{Z}_q^n$, a matrix $A \leftarrow \mathbb{Z}_q^{m \times n}$, and an error vector $e = (e_1, \ldots, e_m)$ with each e_i sampled independently according to an error distribution χ over

[1] Other high-profile applications of LWE include fully homomorphic encryption [BV11] and indistinguishability obfuscation [JLS21].

$\mathbb{Z}_q$. Let $w = As + e \pmod q \in \mathbb{Z}_q^m$. Then, we set the secret key sk and public key pk as

$$sk = s \quad \text{and} \quad pk = A' = [A \mid w] \in \mathbb{Z}_q^{m \times (n+1)}.$$

- **Encryption:** To encrypt a bit $b \in \{0,1\}$ using $pk = A'$, sample $x \leftarrow \{0,1\}^m$ and compute the ciphertext
$$c = \mathsf{Enc}(b, pk) = x^T A' + \left(0^n, b \cdot \left\lceil \frac{q}{2} \right\rfloor\right) \pmod q,$$
where $\lceil z \rfloor$ denotes the nearest integer to z with ties broken arbitrarily.

- **Decryption:** To decrypt c using the secret key $sk = s$, we first compute
$$\begin{aligned} c \begin{bmatrix} s \\ -1 \end{bmatrix} &= x^T(As - w) - b \cdot \left\lceil \frac{q}{2} \right\rfloor \pmod q \\ &= -x^T e - b \cdot \left\lceil \frac{q}{2} \right\rfloor \pmod q. \end{aligned}$$
Since $x \in \{0,1\}^m$, we have that
$$|x^T e| \le \|e\|_1.$$
Therefore, if $\|e\|_1 < q/4$ then we can recover b by checking whether the computation above yields a value closer to 0 or $q/2$.
For example, this property of the error vector is satisfied with high probability if we follow the standard approach of taking χ to be a discrete gaussian with standard deviation $q/\mathsf{poly}(n)$. In this case, the resulting standard deviation of e would be $\sqrt{m} \cdot q/\mathsf{poly}(n) \ll q/4$, which yields the desired property.

Regev showed the following result.

Theorem 2.3 *The PKE scheme defined above is computationally secure if* $\mathsf{Decision\text{-}LWE}_{n,m,q,\chi}$ *is hard to solve by PPT adversaries and* $m \ge (n+1)\log q$.

Proof: Following [Pei15], we present an informal outline via a *hybrid argument* – a standard proof technique in cryptography. Recall that the goal is to show that no PPT adversary Adv can win the following distinguishing game except with probability at most $1/2 + \mathsf{negl}(n)$ for some negligible function negl:

1. The challenger samples keys $(sk, pk) \leftarrow \mathsf{Gen}(1^n)$ and a bit $b \leftarrow \{0,1\}$, and sends the public key pk and the ciphertext $c = \mathsf{Enc}(b, pk)$ to the adversary Adv;
2. Adv outputs $b' \leftarrow \mathsf{Adv}(1^n, c, pk)$ and wins if $b' = b$.

The first hybrid H_0 corresponds exactly to this distinguishing game played the challenger and Adv. The second hybrid H_1 is the same game as H_0 with the exception that the public key $pk = A' \in \mathbb{Z}_q^{(m+1)\times n}$ is now sampled uniformly at random from $\mathbb{Z}_q^{(m+1)\times n}$ and independently of sk.

The final hybrid H_2 is like H_1 with the exception that the ciphertext c is also sampled uniformly at random and independently of everything else.

Note that in H_2 the adversary's input is independent of the message bit b. Therefore, the winning probability for any adversary in H_2 is exactly $1/2$. As a result, if we show that the "transcripts" of H_0 and H_2 are indistinguishable to the eyes of any PPT algorithm, then this implies that no PPT adversary Adv can win H_0 with probability better than $1/2 + \mathsf{negl}(n)$ (because otherwise the winning event of Adv could be used to distinguish between H_0 and H_2). We argue this in steps, first by showing that the transcripts of H_0 and H_1 are indistinguishable, and then that those of H_1 and H_2 are indistinguishable.

We now show that the transcripts of H_0 and H_1 are computationally indistinguishable, provided that $\mathsf{Decision\text{-}LWE}_{n,m,q,\chi}$ is hard to solve by PPT adversaries. To do this, we describe how we can transform a PPT algorithm D that correctly distinguishes between the transcripts of H_0 and H_1 with probability at least $1/2 + f(n)$ for a non-negligible function f into a PPT algorithm D' that correctly solves $\mathsf{Decision\text{-}LWE}_{n,m,q,\chi}$ with the same probability. Consider the following PPT algorithm D' for $\mathsf{Decision\text{-}LWE}_{n,m,q,\chi}$ which runs D as a subroutine. The algorithm D' receives as input a uniformly random matrix $A \leftarrow \mathbb{Z}_q^{m \times n}$ and a vector $w \in \mathbb{Z}_q^m$. Recall that in the "real" case we have that $w = As + e \pmod q$ for a uniformly random $s \leftarrow \mathbb{Z}_q^n$ and a short error vector e, while in the "ideal" case we have that $w \leftarrow \mathbb{Z}_q^m$ independently of A. The algorithm D' arranges (A, w) into the augmented matrix

$$A' = [A \mid w],$$

plays the PKE distinguishing game with A' as the public key, and calls D on the resulting transcript. Observe that if we are in the real case where $w = As + e \pmod q$, then this game corresponds exactly to H_0, while if we are in the ideal case where $w \leftarrow \mathbb{Z}_q^m$, then the game corresponds exactly to H_1. Therefore, if D correctly guesses whether the transcript comes from H_0 or H_1, then D' also correctly guesses whether it is in the real or ideal case, thus solving $\mathsf{Decision\text{-}LWE}_{n,m,q,\chi}$. This implies that H_0 and H_1 are computationally indistinguishable.

Finally, we claim that the transcripts of H_1 and H_2 are indistinguishable, even to computationally-unbounded algorithms. The only difference between the two games lies in how the ciphertext is computed. In H_1 the ciphertext is computed as $c = x^T A' + (0^n, b \cdot \lceil \frac{q}{2} \rfloor)$ with $A' \leftarrow \mathbb{Z}_q^{m \times (n+1)}$ and $x \leftarrow \{0,1\}^m$, while in H_2 it is sampled uniformly at random from $\mathbb{Z}_q^n$. The desired claim follows directly from the fact that the joint distribution

$$(A', x^T A')$$

is very close to the uniform distribution over $\mathbb{Z}_q^{m \times (n+1)} \times \mathbb{Z}_q^{n+1}$ in *total variation distance* whenever $m \geq (n+1) \log q$. This is equivalent to the statement that no computationally-unbounded algorithm can distinguish between samples from these two distributions, and hence between transcripts of H_1 and H_2. ∎

Remark 2.2 The fact that the joint distribution $(A', x^T A')$ is indistinguishable from the uniform distribution over $\mathbb{Z}_q^{m \times (n+1)} \times \mathbb{Z}_q^{n+1}$ whenever $m \geq (n+1) \log q$, used in the proof of Theorem 2.3 above, is a special case of the powerful *Leftover Hash Lemma*. For simplicity, we avoid presenting a formal argument here. We will get to see a more careful treatment and further applications of the Leftover

Hash Lemma later on in the broader context of randomness extraction for information-theoretic cryptography.

2.3 Notes and additional reading

For an extensive discussion of real-world attacks on LWE and related problems, see [ACD+18, ADH+19].

References

[ACD+18] Martin R. Albrecht, Benjamin R. Curtis, Amit Deo, Alex Davidson, Rachel Player, Eamonn W. Postlethwaite, Fernando Virdia, and Thomas Wunderer. Estimate all the {LWE, NTRU} schemes! In Dario Catalano and Roberto De Prisco, editors, *Security and Cryptography for Networks*, pages 351–367, Cham, 2018. Springer International Publishing.

[ADH+19] Martin R. Albrecht, Léo Ducas, Gottfried Herold, Elena Kirshanova, Eamonn W. Postlethwaite, and Marc Stevens. The general sieve kernel and new records in lattice reduction. In Yuval Ishai and Vincent Rijmen, editors, *Advances in Cryptology – EUROCRYPT 2019*, pages 717–746, Cham, 2019. Springer International Publishing.

[Ajt04] Miklós Ajtai. Generating hard instances of lattice problems. *Quaderni di Matematica*, 3:1–32, 2004. Preliminary version in STOC 1996.

[BV11] Zvika Brakerski and Vinod Vaikuntanathan. Efficient fully homomorphic encryption from (standard) LWE. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science (FOCS 2011)*, pages 97–106, 2011.

[JLS21] Aayush Jain, Huijia Lin, and Amit Sahai. Indistinguishability obfuscation from well-founded assumptions. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 60–73, New York, NY, USA, 2021. Association for Computing Machinery.

[LPR10] Vadim Lyubashevsky, Chris Peikert, and Oded Regev. On ideal lattices and learning with errors over rings. In Henri Gilbert, editor, *Advances in Cryptology – EUROCRYPT 2010*, pages 1–23, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[MP13] Daniele Micciancio and Chris Peikert. Hardness of SIS and LWE with small parameters. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology – CRYPTO 2013*, pages 21–39, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[Pei15] Chris Peikert. A decade of lattice cryptography. Cryptology ePrint Archive, Paper 2015/939, 2015. https://eprint.iacr.org/2015/939.

[Reg09] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *J. ACM*, 56(6), sep 2009. Preliminary version in STOC 2005.

[Reg10] Oded Regev. The learning with errors problem (invited survey). In *2010 IEEE 25th Annual Conference on Computational Complexity*, pages 191–204, 2010.

[SSTX09] Damien Stehlé, Ron Steinfeld, Keisuke Tanaka, and Keita Xagawa. Efficient public key encryption based on ideal lattices. In Mitsuru Matsui, editor, *Advances in Cryptology – ASIACRYPT 2009*, pages 617–635, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[Ste18] Noah Stephens-Davidowitz. Ring-SIS and ideal lattices, 2018. Available at https://people.csail.mit.edu/vinodv/6876-Fall2018/RingSISclass.pdf.