Cryptography and Security Protocols **IST-UL, 2024/2025**

Collision-resistance from the Short Integer Solution problem

Lecturer: João Ribeiro

Introduction

In these notes we construct collision-resistant hash functions based on the (conjectured) hardness of the Short Integer Solution (SIS) problem. It is believed that SIS is hard to solve by quantum computers, making it a popular building block for post-quantum cryptography.

We will discuss more aspects of post-quantum cryptography later in the course. If you are curious, Peikert's excellent survey [Pei15] is a great starting point.

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 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 $n \times m$ 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 $\|z\|_2 \leq \beta$, where $\|z\|_2 = \left(\sum_{i=1}^m |z_i|^2\right)^{1/2}$.*

We say that $\mathsf{SIS}_{n,m,q,\beta}$ is hard *if for every PPT adversary $\mathcal{A}$ there exists a negligible function $\varepsilon(n \cdot m \cdot \log q)$ such that $\mathcal{A}(1^{n \cdot m \cdot \log q})$ solves the search problem above with probability at most $\varepsilon(n \cdot m \cdot \log q)$.*[1]

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 *short* solutions (i.e., solutions z with small norm $\|z\|_2$). 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.

[1] Note that for SIS the "input size" is $A \in \mathbb{Z}_q^{n \times m}$, which has description length $n \cdot m \cdot \log q$. We usually take m and q to be polynomial in n, and so the input size is also polynomial in n.

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 1 *If $m > n \log q$ and $\beta > \sqrt{n \log q}$, then $\mathsf{SIS}_{n,m,q,\beta}$ has at least one solution with probability 1 over the choice of the matrix A.*

Proof: This follows by a 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 $m > n \log q$, 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

$$\|w\|_2 = \|z - z'\|_2 \leq \sqrt{m} \leq \beta.$$

The first inequality uses the fact that $z_i - z'_i \in \{-1, 0, 1\}$, and so $|z_i - z'_i|^2 \leq 1$, for all i. ■

2 Collision-resistant compression functions from SIS

In the previous lecture we defined hash functions, which for a given security parameter n map bitstrings of arbitrary length to bitstrings of length $\ell(n)$ that depends only on n. On the way to building collision-resistant hash functions, we will first build collision-resistant *compression functions*. A compression function is just a hash function (Gen, H) that is only defined for inputs $x \in \{0,1\}^{\ell'(n)}$, for some input length $\ell'(n) > \ell(n)$. We can define collision-resistance for compression functions analogously to how we did it for hash functions.

Ajtai [Ajt04] proposed a simple family of compression functions that is easily shown to be collision-resistant based on the hardness of solving SIS with appropriate parameters. Given an $n \times m$ matrix $A \in \mathbb{Z}_q^{n \times m}$, consider the associated hash function $h_A : \{0,1\}^m \to \mathbb{Z}_q^n$ given by

$$h_A(z) = Az \pmod q. \tag{1}$$

Note that this function is only compressing when $m > n \log q$.

For simplicity, suppose that $n \cdot m \cdot \log q$ is polynomial in n (this is the case for the relevant regime of parameters). We consider the (keyed) compression function (Gen, H) where $\mathsf{Gen}(1^n)$ samples $A \in \mathbb{Z}_q^{n \times m}$ uniformly at random, and $H(A, z) = h_A(z)$ for $z \in \{0,1\}^m$.

The following theorem formally states the collision-resistance of (Gen, H).

Theorem 2 *The keyed compression function (Gen, H) defined above is collision-resistant provided that $\mathsf{SIS}_{n,m,q,\beta}$ with $m > n \log q$ and $\beta = \sqrt{m}$ is hard.*

Proof: Suppose not. Then, there exists a PPT adversary $\mathcal{A}$ and a polynomial $p(n)$ such that that on input a uniformly random $A \in \mathbb{Z}_q^{n \times m}$, $\mathcal{A}$ outputs distinct $z, z' \in \{0,1\}^m$ such that $h_A(z) = h_A(z')$ with probability at least $1/p(n)$, for infinitely many n's. Fix such an n.

A collision $h_A(z) = h_A(z')$ is equivalent to $A(z - z') = 0 \pmod q$. Since $z \neq z'$, we also have $z - z' \neq 0$. Furthermore, $\|z - z'\|_2 \leq \sqrt{m} = \beta$. Therefore, the PPT adversary $\mathcal{A}'$ that, given A, runs $\mathcal{A}(A)$ and outputs $z - z'$ if $\mathcal{A}$ finds a collision solves $\mathsf{SIS}_{n,m,q,\beta}$ with probability at least $1/p(n)$ too, a contradiction. ■

3 The Merkle-Damgård transform

Recommended reading: Katz-Lindell, Section 5.2.

We saw above how to construct collision-resistant compression functions from SIS. Recall that our end goal is to construct collision-resistant *hash* functions, which map arbitrary inputs to bitstrings of length $\ell(n)$ depending only on the security parameter n.

In this section, we present an elegant way of extending the domain of any collision-resistant compression function (even a compression function that only compresses by 1 bit) while maintaining collision-resistance, called the *Merkle-Damgård transform*. This method was originally introduced in Ralph Merkle's 1979 PhD thesis, and shown to be sound independently by Merkle [Mer89] and Damgård [Dam89].

The methodology we present (construct a hopefully fast compression function, and then apply the Merkle-Damgård transform) underlies many practical constructions of cryptographic hash functions. From a more theoretical perspective, it also tells us that compressing even just by 1 bit is as hard as compressing by an arbitrary amount.

For simplicity, suppose that (Gen, H) is a compression function mapping inputs of length $2n$ to outputs of length n. Consider (Gen', H') defined on inputs $x \in \{0,1\}^L$ of any length $L < 2^n$:

1. $\mathsf{Gen}'(1^n)$ simply runs $s \leftarrow \mathsf{Gen}(1^n)$ (i.e., it samples a compression function H_s).
2. Divide x into $B = \lceil L/n \rceil$ blocks of length n, $x_1, \ldots, x_B$. If needed, first append 0s to x until its length is a multiple of n. Set $x_{B+1} = L$, written as an n-bit string.
3. Define $z_0 = 0^n$. For $i > 0$, define $z_i = H_s(z_{i-1} \| x_i)$.
4. Set $H'_s(x) = z_{B+1}$.

We have the following result.

Theorem 3 *If* (Gen, H) *is collision-resistant then so is* (Gen', H').

Proof: Fix any function description s. We show how given a collision for H'_s we can efficiently find a collision for H_s, which contradicts the collision-resistance of (Gen, H).

Suppose that there exist distinct $x \in \{0,1\}^L$ and $x' \in \{0,1\}^{L'}$ such that $H'_s(x) = H'_s(x')$. To compute these outputs, H'_s divides x into B blocks $x_1, \ldots, x_B$ of length n, and divides x' into B' blocks $x'_1, \ldots, x'_{B'}$ of length n. Recall that $x_{B+1} = L$ and $x'_{B'+1} = L'$.

We consider two cases:

- $L \neq L'$. In this case x and x' have different lengths, and so $x_{B+1} \neq x'_{B+1}$. The last step of the computation of $H'_s(x)$ is $H_s(z_B \| x_{B+1}) = H_s(z_B \| L)$, while the last step of the computation of $H'_s(x')$ is $H_s(z'_{B'} \| x'_{B'+1}) = H_s(z'_{B'} \| L')$. Since $L \neq L'$ we have also $z_B \| L \neq z'_{B'} \| L'$, and so these two strings are a collision for H_s.
- $L = L'$. In this case x and x' have the same length, and so $B = B'$. Let $z_1, \ldots, z_{B+1}$ and $z'_1, \ldots, z'_{B+1}$ be the values defined along the computation of $H'_s(x)$ and $H'_s(x')$, respectively. Let $I_j = z_{j-1} \| x_j$ and $I'_j = z'_{j-1} \| x'_j$ denote the corresponding inputs.

  Let N be the largest integer such that $I_N \neq I'_N$. Since $x \neq x'$ (and so $x_j \neq x'_j$ for at least one j) such an N exists. By the maximality of N we know that

  $$z_N \| x_{N+1} = I_{N+1} = I'_{N+1} = z'_N \| x'_{N+1}.$$

  In particular, $H_s(I_N) = z_N = z'_N = H_s(I'_N)$. This means that I_N and I'_N are a collision for H_s.

■

4 Parameters for SIS and more structured problems

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].

References

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

[Dam89] Ivan Bjerre Damgård. A design principle for hash functions. In Gilles Brassard, editor, *Advances in Cryptology — CRYPTO' 89 Proceedings*, pages 416–427, New York, NY, 1989. Springer New York.

[Mer89] Ralph C. Merkle. One way hash functions and DES. In Gilles Brassard, editor, *Advances in Cryptology — CRYPTO' 89 Proceedings*, pages 428–446, New York, NY, 1989. Springer New York.

[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.

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