TD 7: Hash functions

Definition 1. *A **hash function** is a pair of probabilistic polynomial-time algorithms* $(\texttt{Gen}, H)$ *satisfying the following:*

- $\texttt{Gen}$ *is a probabilistic algorithm which takes as input a security parameter* 1^n *and outputs a key* s*. We assume that* 1^n *is implicit in* s*.*
- *There exists a polynomial* l *such that* H *takes as input a key* s *and a string* $x \in \{0,1\}^*$ *and outputs a string* $H^s(x) \in \{0,1\}^{l(n)}$ *(where* n *is the value of the security parameter implicit in* s*).*

If H^s *is defined only for inputs* $x \in \{0,1\}^{l'(n)}$ *and* $l'(n) > l(n)$*, then we say that* $(\texttt{Gen}, H)$ *is a **fixed-length** hash function for inputs of length* $l'(n)$*.*

Definition 2. *The **collision-finding game** is defined as follows:*

1. *A key* s *is generated by running* $\texttt{Gen}(1^n)$
2. *The adversary* $\mathcal{A}$ *is given* s *and outputs* x, x' *(if* Π *is a fixed-length hash function for inputs of length* $l'(n)$ *then we require* $x, x' \in \{0,1\}^{l'(n)}$*).*
3. $\mathcal{A}$ *wins (i.e., it finds a collision) if and only if* $x \neq x'$ *and* $H^s(x) = H^s(x')$*.*

Definition 3. *A hash function* $\Pi = (\texttt{Gen}, H)$ *is **collision resistant** if for all probabilistic polynomial-time adversaries* $\mathcal{A}$*, we have*

$$Pr[\texttt{HashColl}_{\mathcal{A}}(\Pi)]$$

is negligible.

Exercise 1. [*Collision resistance*]

1. Let (Gen, H) be a collision-resistant hash function. Is $(Gen, \widehat{H})$ defined by $\widehat{H}^s =_{def} H^s(H^s(x))$ necessarily collision-resistant?
2. Let (Gen, H_1) and (Gen', H_2) be a collision-resistant hash functions such that $H_1 := \{0,1\}^n \to \{0,1\}^m$ and $H_1 := \{0,1\}^m \to \{0,1\}^l$. Is $(Gen, \widehat{H})$ defined by $\widehat{H}^{(s_1,s_2)} =_{def} H_2^{s_2}(H_1^{s_1}(x))$ necessarily collision-resistant?

Exercise 2. [*SIS*]

Definition 4 (Learning with Errors)**.** *Let* $\ell < k \in \mathbb{N}$*,* $n < m \in \mathbb{N}$*,* $q = 2^k$*,* $B = 2^\ell$*,* $\mathbf{A} \hookleftarrow U(\mathbb{Z}_q^{m\times n})$*. The Learning with Errors (*LWE*) distribution is defined as follows:* $D_{\mathrm{LWE},\mathbf{A}} = (\mathbf{A}, \mathbf{A}\cdot\mathbf{s} + \mathbf{e} \bmod q)$ *for* $\mathbf{s} \hookleftarrow U(\mathbb{Z}_q^n)$ *and* $\mathbf{e} \hookleftarrow U\left(\left[-\frac{B}{2}, \frac{B}{2}\right]^m \cap \mathbb{Z}^m\right)$*.*

The $LWE_{\mathbf{A}}$ *assumption* states that, given suitable parameters k, ℓ, m, n, it is computationally hard to distinguish $D_{\mathrm{LWE},\mathbf{A}}$ from the distribution $(\mathbf{A}, U(\mathbb{Z}_q^m))$.

Given a matrix $\mathbf{A} \in \mathbb{Z}_q^{m\times n}$ with $m > n \lg q$, let us define the following hash function:

$$\begin{array}{rccc} H_{\mathbf{A}}: & \{0,1\}^m & \to & \{0,1\}^n \\ & \mathbf{x} & \mapsto & \mathbf{x}^T \cdot \mathbf{A} \bmod q. \end{array}$$

1. Why finding a sufficiently "short" non-zero vector $\mathbf{z}$ such that $\mathbf{z}^T \cdot \mathbf{A} = \mathbf{0}$ is enough to distinguish $D_{\text{LWE},\mathbf{A}}$ from the distribution $(\mathbf{A}, U(\mathbb{Z}_q^m))$? Define "short".
2. Show that $H_{\mathbf{A}}$ is *collision-resistant* under the $LWE_{\mathbf{A}}$ assumption.
3. Is it still a secure hash function if we let $H_{\mathbf{A}} : \mathbf{x} \mapsto \mathbf{x}^T \cdot \mathbf{A}$? (without the reduction modulo)

Exercise 3. [*HMAC*]

1. In the Merkle-Damgård transform, the message is split into consecutive blocks, and we add as a last block the binary representation of the length of this message. Suppose that we do not add this block: does this transform still lead to a collision-resistant hash function?
2. Before HMAC was invented, it was quite common to define a MAC by $\text{Mac}_k(m) = H^s(k \parallel m)$ where H is a collision-resistant hash function. Show that this is not a secure MAC when H is constructed via the Merkle-Damgård transform.

Exercise 4. [*Pedersen's hash function*]

Pedersen's hash function is as follows:

- Given a security parameter n, algorithm `Gen` samples (G, g, q) where $G = \langle g \rangle$ is a cyclic group of cardinality q, a prime number. It then sets $g_1 = g$ and samples g_i uniformly in G for all $i \in \{2, \ldots, k\}$, where $k \geq 2$ is some parameter. Finally, it returns $(G, q, g_1, \ldots, g_k)$.
- The hash of message $M = (M_1, \ldots, M_k) \in (\mathbb{Z}/q\mathbb{Z})^k$ is $H(M) = \prod_{i=1}^k g_i^{M_i} \in G$.

1. Assume for this question that G is a subgroup of prime order q of $(\mathbb{Z}/p\mathbb{Z})^\times$, where $p = 2q + 1$ is prime. What is the compression factor in terms of k and p?

Definition 5. (Discrete Logarithm Problem (DLP)). *Given G, g, and $h \in G$ where $G = \langle g \rangle$ is a cyclic group of cardinality q, prime number. The DLP asks for $x \in \mathbb{Z} \backslash q\mathbb{Z}$ such that $g^x \equiv h \mod q$. The problem is hard if no efficient adversary can find such x with non-negligible advantage.*

2. Assume for this question that $k = 2$. Show that Pedersen's hash function is collision-resistant, under the assumption that the DLP is hard for G.
3. Same question as the previous one, with $k \geq 2$ arbitrary.