CS 442
Introduction to Cryptography

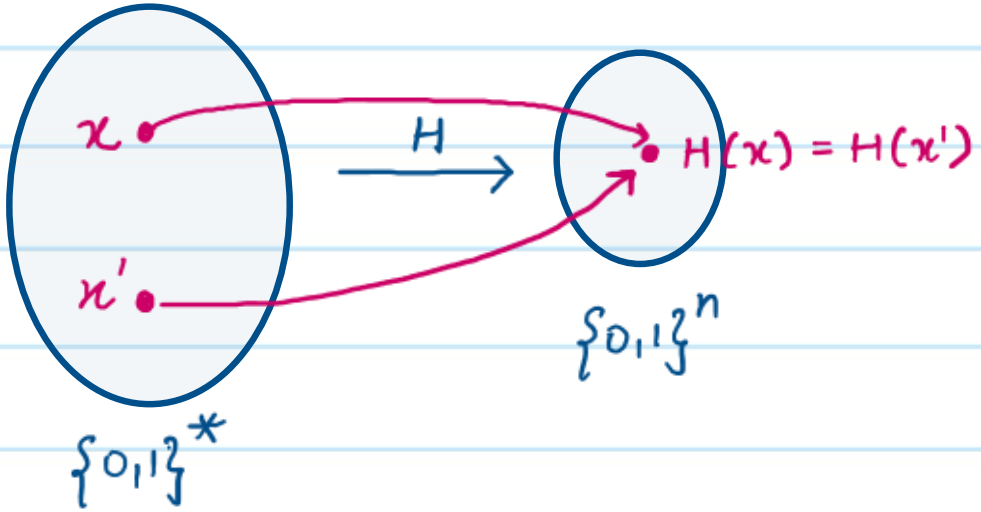
Lecture 19: Example Problems on Hash Functions

Instructor: Aarushi Goel

Spring 2026

Hash Functions

* Functions that take inputs of arbitrary length and compress them into short, fixed length outputs.
* $H: \{0,1\}^{*} \rightarrow \{0,1\}^{n}$ ($*$: arbitrary length)
* A hash function naturally has *collisions, where a collision is a pair of distinct elements x and x', such that $H(x) = H(x')$.
* A good hash function yields few collisions, because when a collision happens, two elements end up in the same row in the hash table, increasing the lookup time.



Collision - Resistant Hash Functions

* Collision-resistant hash functions (CRHF) are cryptographic versions of hash functions, where the goal is not just to minimize the number of collisions, but it should also be infeasible for any probabilistic polynomial-time algorithm to find a collision in H.
* Note that CRHFs are still compression functions, so collisions exist, but we just want such collisions to be hard to find.
* We typically consider keyed CRHFs.

$$H^K(x) = H(K, x)$$

unlike all other primitives we have discussed so far, here we don't care about hiding the key.

* It should be hard to find a collision in H^K, for randomly chosen K, even given K.

Collision-Resistant Hash Functions

Definition: A family of functions $H = \{h_k : D_k \to R_k\}_{k \in \mathcal{K}}$ is a collision-resistant hash function family (CRHF) if:

* Easy to Sample: There exists a PPT Gen, such that $k \leftarrow Gen(n)$, $k \in \mathcal{K}$.
* Compression: $|R_k| < |D_k|$.
* Easy to Evaluate: There exists a polynomial-time algorithm Eval, such that, given $x \in D_k$, $k \in \mathcal{K}$, $Eval(k, x) = h_k(x)$.
* Collision Resistance: For all non-uniform PPT adversaries A, there exists a negligible function $\mu(\cdot)$, such that

$$Pr\left[\begin{array}{l|l} x \neq x' \text{ and } & k \xleftarrow{\$} Gen(n), \\ h_k(x) = h_k(x') & (x, x') \leftarrow A(k, n) \end{array}\right] \leq \mu(n).$$

$H = \{H^k : \{0,1\}^* \to \{0,1\}^n\}_{k \in K}$ is a family of CRHF.

#1 Let $H^k : \{0,1\}^* \to \{0,1\}^n$ be a collision resistant hash function. Is the following also a collision resistant hash?

$\hat{H}^k(x) = H^k(x_1 \| x_2 \| \ldots \| x_{n-1})$, where $x = x_1 \| x_2 \| \ldots \| x_n$ & each $x_i \in \{0,1\}^n$.

NO! (if we are claiming that is not collision-resistant, then we must give an example of a collision)

Consider two inputs $x = x_1 \| x_2 \| \ldots \| x_n$ & $x' = x_1 \| x_2 \| \ldots \| x_{n-1} \| x'_n$, where $x_n \neq x'_n$.

It is easy to see that $\hat{H}^k(x) = H^k(x_1 \| x_2 \| \ldots \| x_{n-1}) = \hat{H}^k(x')$.

Since both these inputs hash to the same output, this is an example of a collision.

#2 Let $H^k: \{0,1\}^* \to \{0,1\}^n$ be a collision resistant hash function. Is the following also a collision resistant hash? $\hat{H}^k(x) = H^k(x) \oplus 1^n$.

Yes! (If we are claiming that this is collision resistant, we must explain why. Typically, the cleanest argument is via proof by contradiction).

Let us assume for the sake of contradiction that $\hat{H}^k$ is not collision resistant. This means that it is easy for some PPT adversary (given K) to find (x, x'), such that $\hat{H}^k(x) = \hat{H}^k(x')$.

$\Rightarrow \hat{H}^k(x) \oplus 1^n = \hat{H}^k(x') \oplus 1^n \quad \Rightarrow H^k(x) = H^k(x')$.

$\Rightarrow$ Such a pair (x, x') is also a collision in H^k. However, since H^k is collision resistant, no PPT adversary should be able to find such collisions in H^k, except with negligible probability. Therefore, our assumption must be wrong, and no PPT Adv should be able to easily find collisions in $\hat{H}^k$. $\Rightarrow \hat{H}^k$ is also a CRHF.

#3 Let $H: \{0,1\}^* \to \{0,1\}^n$ be a collision resistant hash function. Is the following also a collision resistant hash?

$$\hat{H}^k(x) = H^k(H^k(x))$$

Yes! We can use a similar argument as in the previous question.

Let (x, x') be such that $\hat{H}^k(x) = \hat{H}^k(x')$.

$\Rightarrow H^k(H^k(x)) = H^k(H^k(x'))$

$\Rightarrow$ either $H^k(x) = H^k(x')$, in which case x, x' is also a collision in H^k.

or if $H^k(x) \neq H^k(x')$, then $y = H^k(x)$, $y' = H^k(x')$ is a collision in H^k.

In either case, if it is easy to find a collision $\hat{H}^k$, then it is also easy to find a collision in H^k. But since H^k is collision resistant, it is not easy to find collisions in H^k. $\Rightarrow \hat{H}^k$ must also be collision resistant.

#4 Let $H: \{0,1\}^{2n-1} \to \{0,1\}^{n-1}$ be a collision resistant hash function. Is the following also a collision resistant hash?

$\hat{H}: \{0,1\}^{2n} \to \{0,1\}^{n}$, $\hat{H}^K(x_1,\ldots,x_{2n}) = \begin{cases} x_1 \| H^K(x_2,\ldots,x_{2n}) & \text{if } x_1 = 0 \\ 1^n & \text{otherwise} \end{cases}$

No!

Any pair of strings of the form $x = 1, x_2, \ldots, x_{2n}$ & $x' = 1, x'_2, \ldots, x'_{2n}$ will hash to the same output 1^n.

⇒ They are collision in $\hat{H}^K$.

⇒ Since it is easy to find collisions in $\hat{H}^K$, it is not collision-resistant.

Merkle - Damgård Transform

* Let $h_K : \{0,1\}^{2n} \to \{0,1\}^n$ be a CRHF
* We can design a CRHF $H_K : \{0,1\}^* \to \{0,1\}^n$ as follows:

1. Let $x \in \{0,1\}^L$ be an L-bit input that we want to hash using H.
2. Parse $x = x_1, \ldots, x_B$, where $B = \frac{L}{n}$ and each $x_i \in \{0,1\}^n$.
   If L is not divisible by n, we pad x with appropriately many 0s.
3. Set $IV = 0^n$ initialization vector.

Claim: If h_K is a CRHF, then so is H_K.

#5 What if we use the Merkle-Damgård transform on $\hat{H}$? Is the resulting function collision-resistant?

Yes!

Recall that in the Merkle-Damgård transform, we start with $z_0 = IV = 0^n$.

$z_1 = \hat{H}^k(z_0 \| x_1) = \hat{H}^k(0^n \| x_1) = 0 \| H^k(0^{n-1} \| x_1)$

$z_2 = \hat{H}^k(z_1 \| x_2) = \hat{H}^k(0 \| H^k(0^{n-1} \| x_1) \| x_2) = 0 \| H^k(H^k(0^{n-1} \| x_1) \| x_2)$

$\vdots$

This exactly like applying the Merkle-Damgård transform on H^k, except that each z_i is prepended with 0.

→ The proof is exactly as for Merkle-Damgård applied on H^k.

#6 Let $\{H^k : \{0,1\}^* \to \{0,1\}^n\}_k$ be a family of collision-resistant hash functions. obtained via the the Merkle-Damgård transform.

Let Gen be the corresponding key sampling algorithm.

Is the following an unforgeable MAC scheme?

→ Setup: $K \leftarrow$ Gen. Output r ⟶ public & known to the adversary

→ KeyGen: $r \xleftarrow{\$} \{0,1\}^n$. Output r. → MAC Key hidden from adversary

→ Sign(K, r, m): Output $H^K(r \| m)$.

NO!

Let H^K be obtained via a Merkle-Damgård transform on a CRHF $h^K : \{0,1\}^{2n} \to \{0,1\}^n$.

Adv

Ch

K

$K \leftarrow Gen$

m

$r \xleftarrow{\$} \{0,1\}^n$

$m \in \{0,1\}^n$

Let L be an n-bit representation of integer $2n$

$\sigma = H^K(r \| m)$

$H^K(r \| m) = h^K\left(h^K\left(h^K(0^n \| r) \| m\right) \| L\right)$

m', σ'

Let $m' = m \| L$

Let L' be an n-bit representation of integer $3n$

$\sigma' = h^K(\sigma \| L')$

It is easy to see that this is a valid forgery.