CSC 591 Cryptography 10/29/2018

Lecture Hash Function

Lecturer: Alessandra Scafuro *Scribe: Shenkai Yu*

Topic/Problem

In this lecture, we will talk about the concept of Hash Function and the way to attack it. We will also talk about the Merkle Damgaard transformation.

Defintion

Hash Function:
Hash functions are simple functions that take inputs of some length and compress them into short, fixed-length outputs. The classic use of hash functions is in data structures, where they can be used to build hash tables that enable $O(1)$ lookup time when storing a set of elements.

Collision-Resistant Hash Function:
A function H is a function that takes as input a key s, which is generated by a key generation algorithm Gen, and an input string x. Then it outputs a string $H^s(x) \stackrel{def}{=} H(s,x)$. It satisfies the following properties

- **Length-Compressing:** H takes as input a key s and a string $x \in \{0,1\}^*$ and outputs a string $H^s(x) \in \{0,1\}^{l(n)}$. If H^s is defined only for inputs $x \in \{0,1\}^{l(n)'}$. Then $l'(n)$ should greater than $l(n)$.
- **Collision Resistance**: It is hard to find a collision for a hash function H^s for a randomly generated key s. More formally, collision resistance means there does not exist two different inputs $x, x' \in \{0,1\}^*$ such that $H^s(x) = H^s(x')$. More formally, for all PPT adversary $\mathcal{A}$, there exists a negligible function $\mathsf{negl}(\cdot)$ such that for all security parameters $\lambda \in \mathbb{N}$

$$\Pr[(x,x') \leftarrow \mathcal{A}(1^\lambda, H^s) : x \neq x', H^s(x) = H^s(x')] \leq \mathsf{negl}(\lambda)$$

Attack

Brute Force Attack (Running time: approximately $N = 2^n$)

1. Pick N messages $m_1, ..., m_N$
2. Compute Hash $H(m_1), ..., H(m_N)$

3. Find two different messages m_1 and m_2 such that $H(m_1) = H(m_2)$

Consider a one-bit compression hash function H such that $|H^s(m)| = n = |x| - 1$. The probability that two random elements x and x' hash to the same value is at least $\frac{1}{2^n}$. However, the probability $x = x'$ occurs with probability $\frac{1}{2^{n+1}}$. Hence the probability tha two random elements collid is at least $\frac{1}{2^n} - \frac{1}{2^{n+1}}$. Therefore, to find a collision we need to search almost all the number of message pairs in the range of H.

Birthday Attack (Running time: approximately $\sqrt{N} = 2^{\frac{n}{2}}$)

Theorem (Birthday Attack): Fix a positive integer $N = 2^n$, let $\{y_1, ..., y_q\}$ be q polynomial number of values sampled uniformly random from a set of N values.

$$Pr[Coll(q, N)] \leq \frac{q^2}{2N} = \frac{(\sqrt{N})^2}{2N} = \frac{1}{2}$$

Proof: Since we pick q random messages. The probability that two of these messages collid is at least

$$C_q^2 \cdot (\frac{1}{2^n} - \frac{1}{2^{n+1}}) \leq \frac{q^2}{2N}$$

where the number $\frac{1}{2^n} - \frac{1}{2^{n+1}}$ is computed as the probability that two random elements collid. If q is approximately equal to $\sqrt{N}$, then the theorem is proved. $\square$

Remark. Note that The above theorem shows that even the perfect hash function, can be broken in $2^{\frac{n}{2}}$

Merkle Damgaard Transformation

Definition: The Merkle Damgard transform is a common approach for extending a one-way compression function to a full-edged hash function, while maintaining the collision-resistance property of the former.

Construction: Let (Gen, h) be a hash function for inputs of length $2n$ and with output length n. Construct hash function (Gen, H) as follows:

- **Gen:** Generate hash key s.
- H**:** On input a key s and a string $x \in \{0,1\}^*$ of length $L < 2^n$, do the following:
  1. Set $B := \lceil \frac{L}{n} \rceil$. Pad x with zeros so its length is a multiple of n. Parse the padded result as the sequence of $n-$bit blocks $x_1, ..., x_B$. Set $x_{B+1} := L$, where L is encoded as an $n-$bit string.
  2. Set $z_0 := 0^n$. (This is also called the IV.)
  3. For $i = 1, ..., B+1$ compute $z_i := h^s(z_{i-1} \| x_i)$.
  4. Output z_{B+1}.

Theorem (MD Transform): If (Gen, h) is a collision resistant hash function with input length $2n$ and output length n, then the Merkle Damgard transform is a collision resistant hash function (Gen, H) for arbitrary input length and output length n.

Proof: We show that for any s, a collision in H^s yields a collision in h^s. Let x and x' be two different strings of length L and L', respectively, such that $H^s(x) = H^s(x')$. Let $x_1, ..., x_B$ be the B blocks of the padded x, and let $x'_1, ..., x'_B$ be the B' blocks of padded x'. Recall that $x_{B+1} = L$ and $x'_{B'+1} = L'$. There are two cases to consider:

1. Case $L \neq L'$. In this case, the last step of computation of $H^s(x)$ is $z_{B+1} := h^s(z_B \| L)$, and the last step of computation of $H^s(x')$ is $z'_{B'+1} := h^s(z'_{B'} \| L')$. Since $H^s(x) = H^s(x')$ it follows that $h^s(z_B \| L) = h^s(z'_{B'} \| L')$. However, $L \neq L'$ and so $z_B \| L$ and $= z'_{B'} \| L'$ are two different strings collide under same h^s.

2. Case $L = L'$. This means that $B = B'$. Let $z_0, ..., z_{B+1}$ be the values defined during the computation of $H^s(x)$. Let $I_i \stackrel{def}{=} z_{i-1} \| x_i$ denote the ith input to h^s, and set $I_{B+2} \stackrel{def}{=} z_{B+1}$. Define $I'_1, ..., I'_{B+2}$ analogously with respect to x'. Let N be the largest index for which $I_N \neq I'_N$. Since $|x| = |x'|$ but $x \neq x'$, there is an i with $x_i \neq x'_i$ and so such an N certainly exists. Because

$$I_{B+2} = z_{B+1} = H^s(x) = H^s(x') = z'_{B+1} = I'_{B+2},$$

we have $N \leq B + 1$. By maximality of N, we have $I_{N+1} = I'_{N+1}$ and in particular $z_N = z'_N$. But this means that I_N, I'_N are a collision in h^s.

$\square$