# 9 L9: Hash

Hash 是一種 compression function。長的 input 經過 hash 之後會變成短的 ouput。
Hash 也被稱為 fingerprint / hash value / digest。

因為是壓縮，所以會存在一些碰撞 (collision)。
Collision: a pair of distinct items $x, x'$ for which Hash(x) = Hash(x')。

## 9.1 Syntax

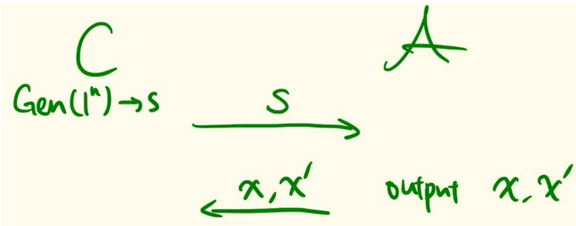
**Definition 11**
A hash function (with output length $l(n)$) is a pair of PPT algorithm $(\mathrm{Gen}, \mathrm{H})$
- $\mathrm{Gen}$: takes a security parameter $1^n$ and outputs a key $s$.
- $\mathrm{H}$: takes input as a key $s$ and a string $x \in \{0,1\}^*$, and outputs a string $\mathrm{H}^s(x) \in \{0,1\}^{l(n)}$.

If $x \in \{0,1\}^{l'(n)}$ and $l'(n) > l(n)$, then we say it's a fixed-length input hash.

**Definition 12** (Collision resistant)
A hash function $\Pi = (\mathrm{Gen}, \mathrm{H})$ is collision resistant if $\forall$ PPT adversaries $A$, there is a negligible function $\mathrm{negl}$ such that

$$\Pr[\text{Hash-coll}_{A,\Pi}(n) = 1] \leqslant \mathrm{negl}(n)$$

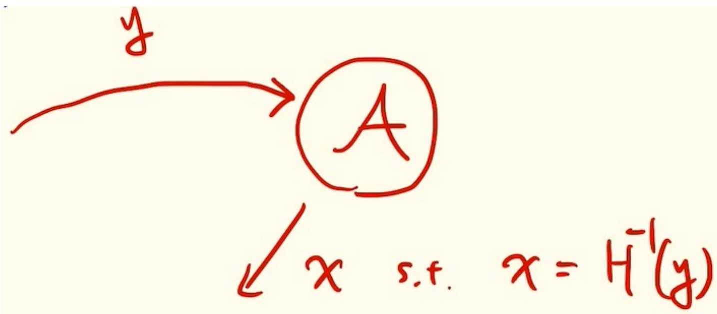
$\text{Hash-coll}_{A,\Pi}(n)$ 的 scenario 如下：




首先 challenger $C$ 會先產生一個 key $s$ 給 adversary $A$。$A$ 可以通過自己的計算，試圖猜出一個 pair $(x, x')$，再傳給 $C$。若 $\mathrm{H}^s(x) = H^s(x')$ 且 $x \neq x'$，則 output 1，也就是 $\text{Hash-coll}_{A,\Pi}(n) = 1$。

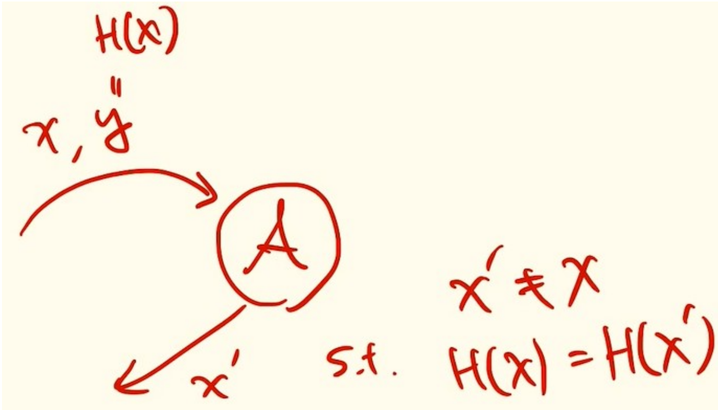
**§ Preimage resistant (one-wayness)**

Adversary 會拿到一個 key $s$（下圖省略）和一個 hash function $\mathrm{H}$ 的 output $y$，並且當 adversary 給出 $x = \mathrm{H}^{-1}(y)$ 時，視為破解成功。



### § Second-preimage resistant

Adversary 會拿到一個 key $s$ ( 下圖省略 ) 和一對 value — hash function $\mathrm{H}$ 的 input $x$ 和 output $y = H(x)$，並且當 adversary 給出 $x' \neq x$ 且 $\mathrm{H}(x') = \mathrm{H}(x)$ 時，視為破解成功。




若要以上述情況作為 security 的定義，則兩者成功的機率都是要 $\leqslant \mathrm{negl}$。

**Quiz**

Compare the security notions of hash function. 比較各種 hash function 的安全定義的難易程度 ( 易、難、無法比較 )
(Ex:Second-preimage resistant is harder than collision resistant.)

## 9.2 SIS

### § Short Integer Solution (SIS) Problem

$\mathbb{Z}_q^n$: n-dimensional vectors modulo $q$ (e.g. $q \approx\ n^3$)

Goal: find non-trivial small (ex: $\{0,1\}$) $z_1, z_2, \ldots, z_m \in \mathbb{Z}$ such that

$$z_1 \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{13} \end{bmatrix} + z_2 \begin{bmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{23} \end{bmatrix} + z_m \begin{bmatrix} a_{m1} \\ a_{m2} \\ \vdots \\ a_{m3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{Z}_q^n$$

Remark:

- $z_1, z_2, \ldots, z_m = 0 \quad \Rightarrow \quad$ “trival”
- $z_1, z_2, \ldots, z_m \notin \{0,1\} \quad \Rightarrow \quad$ “easy”

### § SIS-based Hash Function

Rewrite the forementioned definition of SIS problem.

$\mathbb{Z}_q^n$: n-dimensional vectors modulo $q$ (e.g. $q \approx\ n^3$)

Goal: find non-trivial small (ex: $\{0,1\}$) $z_1, z_2, \ldots, z_m \in \mathbb{Z}$ such that

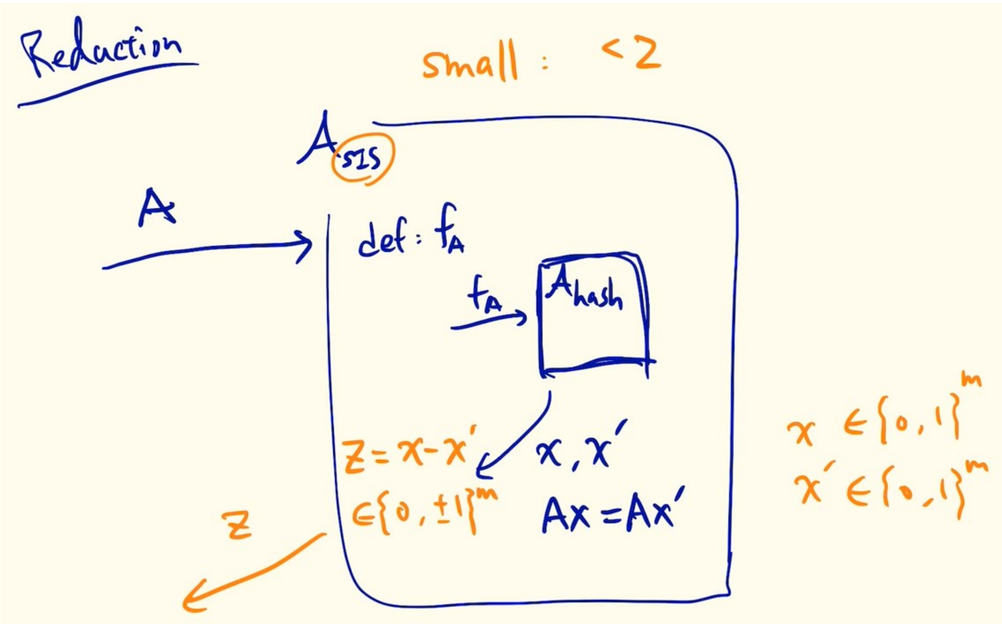
$$\mathbf{Az} = \begin{bmatrix} a_1 & a_2 & \ldots & a_m \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix} = \mathbf{0} \in \mathbb{Z}_q^n$$

**Construction of Hash**

Set $m > n \log q$ (for compression).
Define $f_{\mathbf{A}} : \{0,1\}^m \to \mathbb{Z}_q^n$ as $f_{\mathbf{A}}(x) = \mathbf{Ax}$

Collision: $\mathbf{x}, \mathbf{x}' \in \{0,1\}^m$ where $\mathbf{x} \neq \mathbf{x}'$ and $\mathbf{Ax} = \mathbf{Ax}'$

**§ Collision Resistant SIS-based Hash**




**Quiz**

We do not formally write down the security proof, and only provide proof intuition.

(i) Please show the assumption ($\Pr[success] \leqslant \text{negl}$), aka SIS, which is used in the proof.
(ii) Please complete the proof with probability analysis.

## 9.3 Arbitrary-length Hash Function

前面介紹的 SIS-based hash function 和現實中使用的 hash function 都是 fixed-length compression function。我們可以透過 Merkel-Damgard transformation 來做到 arbitrary-length hash function。

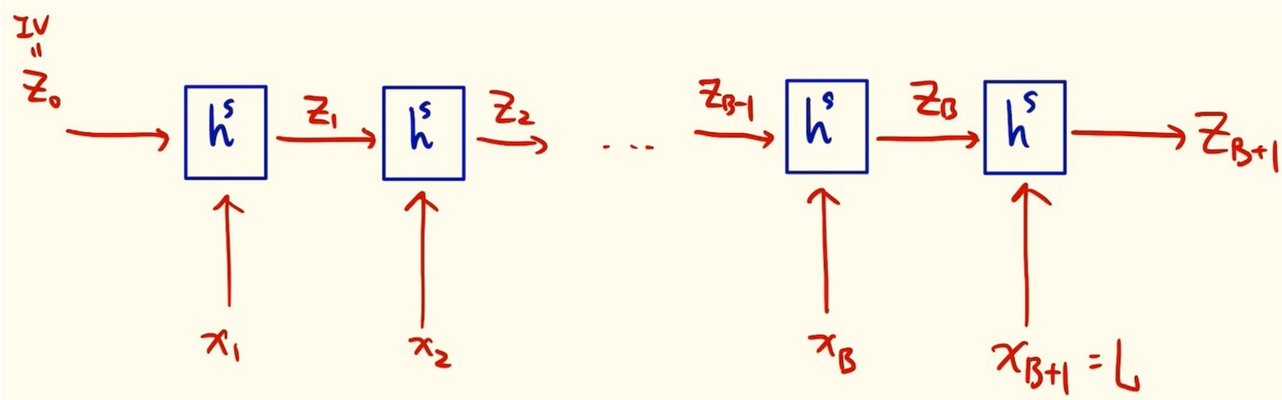
**§ Merkle-Damgard Transformation**

和 CBC-MAC 的概念相似。

Let $(\mathrm{Gen}', \mathrm{h})$ be a fixed input length hash. $\mathrm{h} : \{0,1\}^{2n} \to 0, 1^n$
Let $(\mathrm{Gen}, \mathrm{H})$ be a fixed input length hash. $\mathrm{H} : \{0,1\}^{*} \to 0, 1^n$

Use $(\mathrm{Gen}, \mathrm{h})$ to build $(\mathrm{Gen}, \mathrm{H})$

- $\mathrm{Gen}$: run $\mathrm{Gen}'(1^n) \to s$ (key)
- $\mathrm{H}$: on input $s$ and a string $x \in \{0,1\}^*$ of length $L$ where $L < 2^n$.
    - (i) Set $B = \lceil \frac{L}{n} \rceil$ ($B$: number of blocks)
      Pad $x$ with $0$s, so length will be a multiple of n.
      Parse $x$ to $x_1, x_2, \dots, x_B$, and set $x_{B+1} = L$
    - (ii) Set $z_0 = 0^n$ as IV
    - (iii) For $i = 1, 2, \dots, B+1$, compute $z_i = \mathrm{h}^s(z_{i-1}||x_i)$




    - (iv) Output $z_{B+1}$ as the hash value of $x$.

**Quiz**

We found some cute trick in Merkle-Damgard transformation:

(i) The purpose of $L$? (Hint: related to collision)
(ii) Suppose the fixed-length hash is $h : 0, 1^{n+1} \to 0, 1^n$
How to build an arbitrary length has from the above?

## § Security of Merkel-Damgard Transformation

**Theorem 9**

If $(\mathrm{Gen}', \mathrm{h})$ is collision resistant, then $(\mathrm{Gen}, \mathrm{h})$ is collision resistant.

***Proof***

For any $s$, a collision in $\mathrm{H}^s$ yields a collision $h^s$.
Assume two distinct strings $(x, x')$ of length $(L, L')$ such that $\mathrm{H}^s(x) = \mathrm{H}^s(x')$.

Let $x_1, x_2, \dots, x_B$ are the blocks of padded $x$ and $x_{B+1} = L$,
and $x'_1, x'_2, \dots, x'_B$ are the blocks of padded $x'$ and $x_{B'+1} = L'$.

Case 1: $L \neq L'$
In the last step of $\mathrm{H}^s(x)$ (resp. $\mathrm{H}^s(x')$),
$z_{B+1} = \mathrm{h}^s(z_B||L)$ (resp. $z'_{B'+1} = \mathrm{h}^s(z'_{B'}||L')$)
Assume $\mathrm{H}^s(x) = \mathrm{H}^s(x')$
$\Rightarrow \quad h^s(z_B||L) = h^s(z'_{B'}||L')$ which is a collision in $\mathrm{h}^s$

Case 2: $L = L'$ (implies $B = B'$)
Let $I_i \stackrel{\text{def}}{=} z_{i-1}||x_i$. ($i$-th input of $\mathrm{h}^s$) ($I'_i$, resp.)
Set $I_{B+2} \stackrel{\text{def}}{=} z_{B+1}$.

Assume $\mathrm{H}^s(x) = \mathrm{H}^s(x')$. Let $N$ be the largest index for $I_N \neq I'_N$.
Since $|x| = |x'|$, but $x \neq x'$, there must exist an $i$ with $x_i \neq x'_i$.
$I_{B+2} = z_{B+1} = \mathrm{H}^s(x) = \mathrm{H}^s(x') = z'_{B+1} = I'_{B+2}$
$\Rightarrow \quad N \leqslant B+1 \Rightarrow \quad I_{N+1} = I'_{N+1}$
For this $N$, $I_N, I'_N$ are collision in $\mathrm{h}^s$.

$\square$