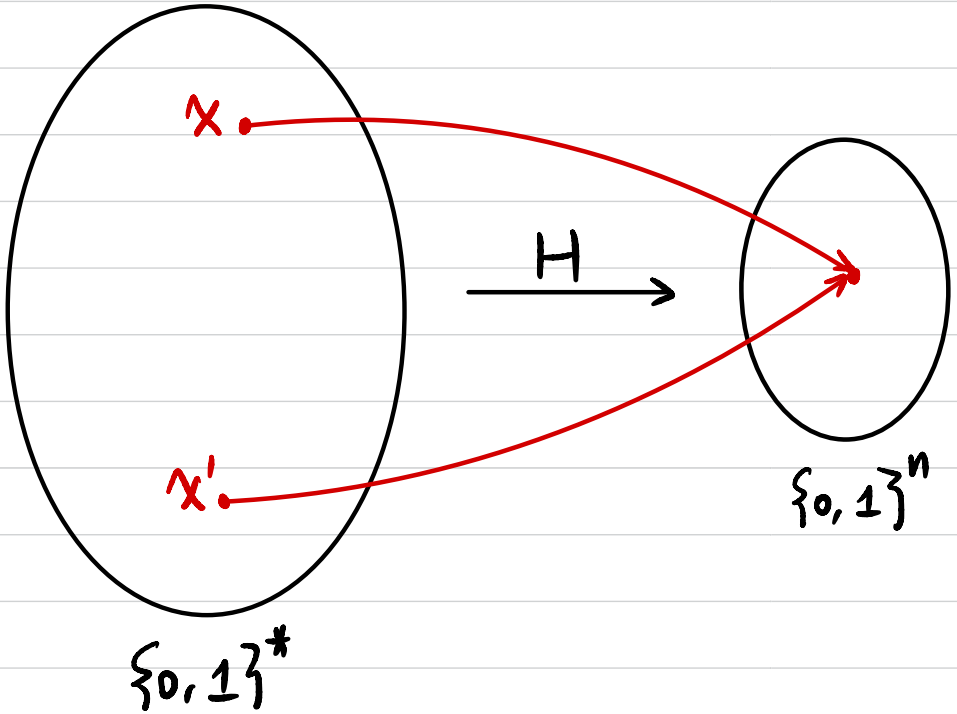
CSCI 1510

This Lecture:

- Collision-Resistant Hash Function (continued)
- Merkle-Damgård Transform
- Hash-and-MAC
- Applications of Hash Functions

Cryptographic Hash Function

$H: \{0,1\}^* \longrightarrow \{0,1\}^n$



Collision-Resistant Hash Function (CRHF):

It's computationally hard to find $x, x' \in \{0,1\}^*$ s.t.

$x \neq x'$, $H(x) = H(x')$ (collision)

Collision-Resistant Hash Function (CRHF)

- **Syntax:**

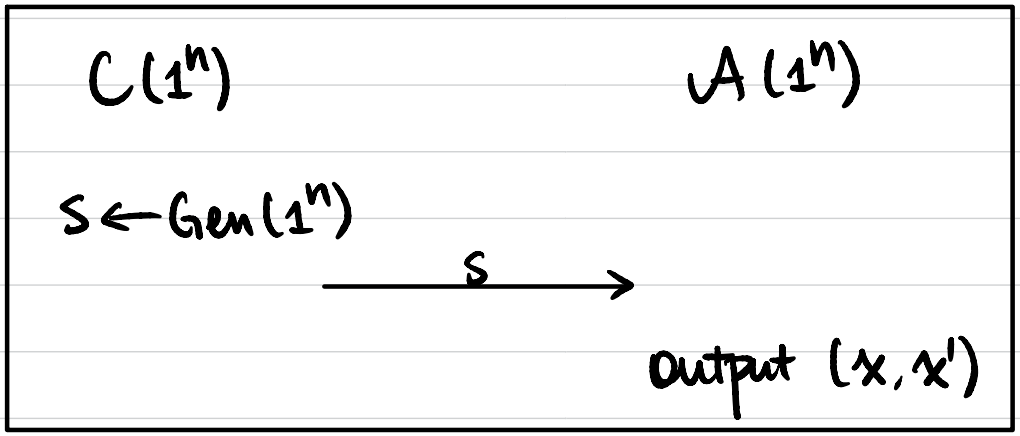
  A hash function is defined by a pair of PPT algorithms (Gen, H):

  - $Gen(1^n)$: output s
  - $H^s(x)$: $x \in \{0,1\}^*$, output $h \in \{0,1\}^{\ell(n)}$

- **Security**

  A hash function (Gen, H) is **collision-resistant** if

  $\forall$ PPT $\mathcal{A}$, $\exists$ negligible function $\varepsilon(\cdot)$ s.t. $\Pr[x \neq x' \wedge H^s(x) = H^s(x')] \leq \varepsilon(n)$.



Birthday Problem / Paradox

There are q students in a class.

Assume each student's birthday is a random $y_i \xleftarrow{\$} [365]$

What's the probability of a collision?

$q = 366 \Rightarrow$ prob. $= 1$

$q = 23 \Rightarrow$ prob. $\simeq 50\%$

$q = 70 \Rightarrow$ prob. $\simeq 99.9\%$

$y_i \xleftarrow{\$} [N]$

$q = N+1 \Rightarrow$ prob. $= 1$

$q = \sqrt{N} \Rightarrow$ prob. $\simeq 50\%$

If security parameter $n = 128$, $\ell = ?$

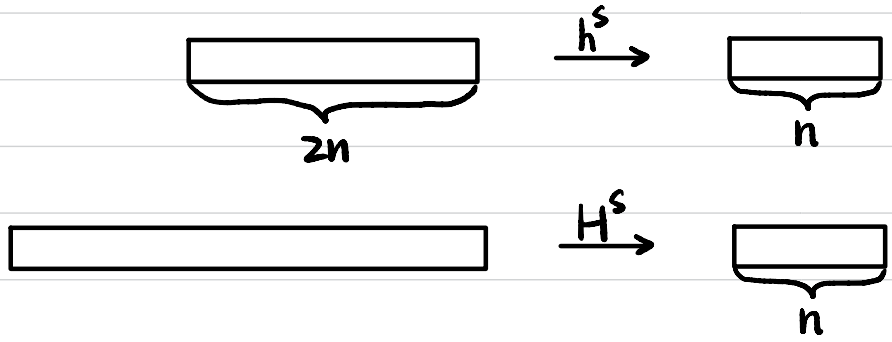
$\ell = 256$ $\quad H: \{0,1\}^* \to \{0,1\}^{256}$

$q = \sqrt{2^{256}} = 2^{128}$

Domain Extension: Merkle-Damgård Transform

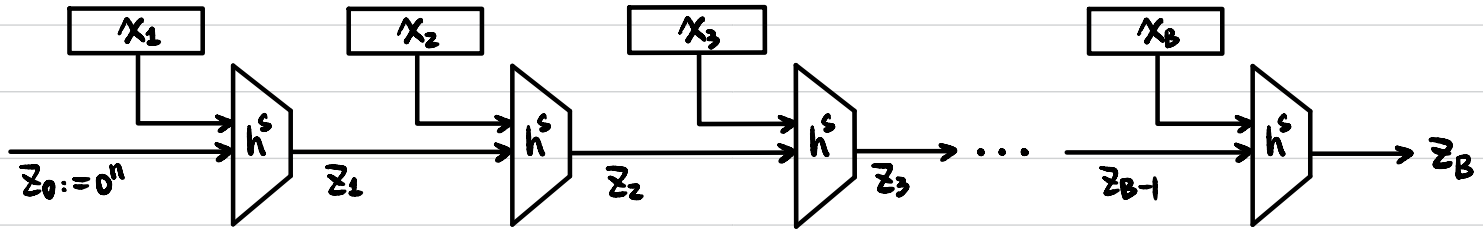
Given a CRHF (Gen, h) from $\{0,1\}^{2n}$ to $\{0,1\}^n$.

Construct a CRHF (Gen, H) from $\{0,1\}^*$ to $\{0,1\}^n$.



① Assume $|x|$ is a multiple of n

② Parse $x = x_1 \| x_2 \| \cdots \| x_B$, $x_i \in \{0,1\}^n$ $\forall i \in [B]$



$z_0 := 0^n$ $\quad z_i := h^s(z_{i-1} \| x_i)$ $\forall i \in [B]$ $\quad H^s(x) := z_B$

Is this a CRHF for arbitrary-length messages (multiple of n)? No!

Step 1: Assume $(\widetilde{Gen}, \tilde{h})$ is a CRHF from $\{0,1\}^{2n}$ to $\{0,1\}^n$

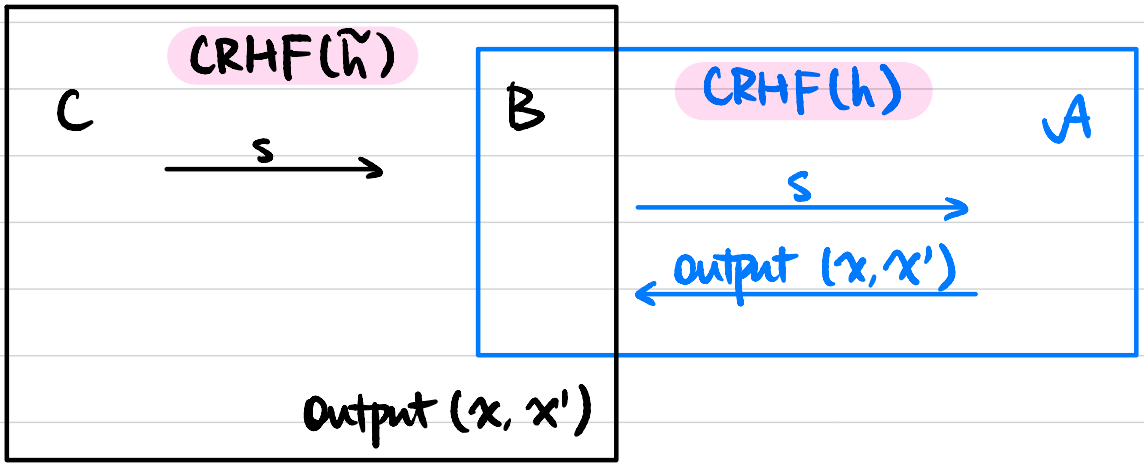
We construct (Gen, h) from $\{0,1\}^{2n}$ to $\{0,1\}^n$ as follows:

- $Gen(1^n)$: same as $\widetilde{Gen}(1^n)$.
- $h^s(x) := \tilde{h}^s(x) \oplus \tilde{h}^s(0^{2n})$

Step 2: If $(\widetilde{Gen}, \tilde{h})$ is a CRHF, then so is (Gen, h).

~~Proof~~ Assume not, then $\exists$ PPT $\mathcal{A}$ that breaks the collision resistence of (Gen, h).

We construct a PPT B to break the collision resistence of $(\widetilde{Gen}, \tilde{h})$.



$$x \neq x' \wedge h^s(x) = h^s(x')$$

$$\Downarrow$$

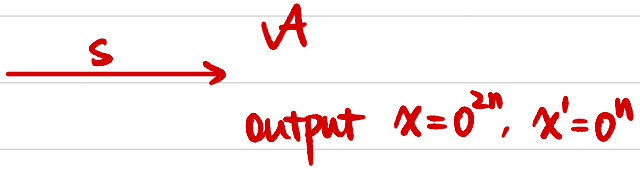
$$\tilde{h}^s(x) \oplus \tilde{h}^s(0^{2n}) = \tilde{h}^s(x') \oplus \tilde{h}^s(0^{2n})$$

$$\Downarrow$$

$$\tilde{h}^s(x) = \tilde{h}^s(x')$$

$\Rightarrow (x, x')$ is a collision for $\tilde{h}^s$.

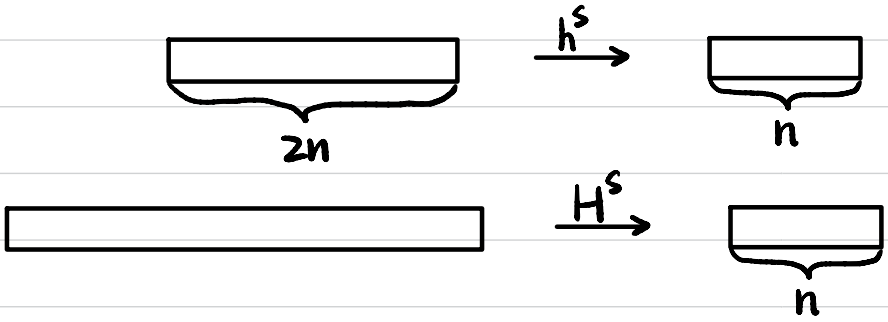
Step 3: (Gen, H) instantiated with (Gen, h) is not a CRHF for arbitrary-length messages.



Domain Extension: Merkle-Damgård Transform

Given a CRHF (Gen, h) from $\{0,1\}^{2n}$ to $\{0,1\}^n$.

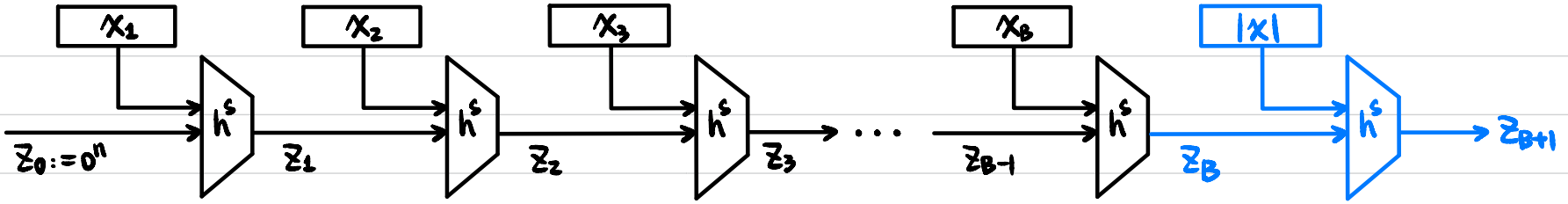
Construct a CRHF (Gen, H) from $\{0,1\}^*$ to $\{0,1\}^n$.



$H^s(x)$: $x \in \{0,1\}^*$

① Pad x with $100\cdots0$ to a multiple of n $\rightarrow \tilde{x}$

② Parse $\tilde{x} = x_1 \| x_2 \| \cdots \| x_B$, $x_i \in \{0,1\}^n$ $\forall i \in [B]$



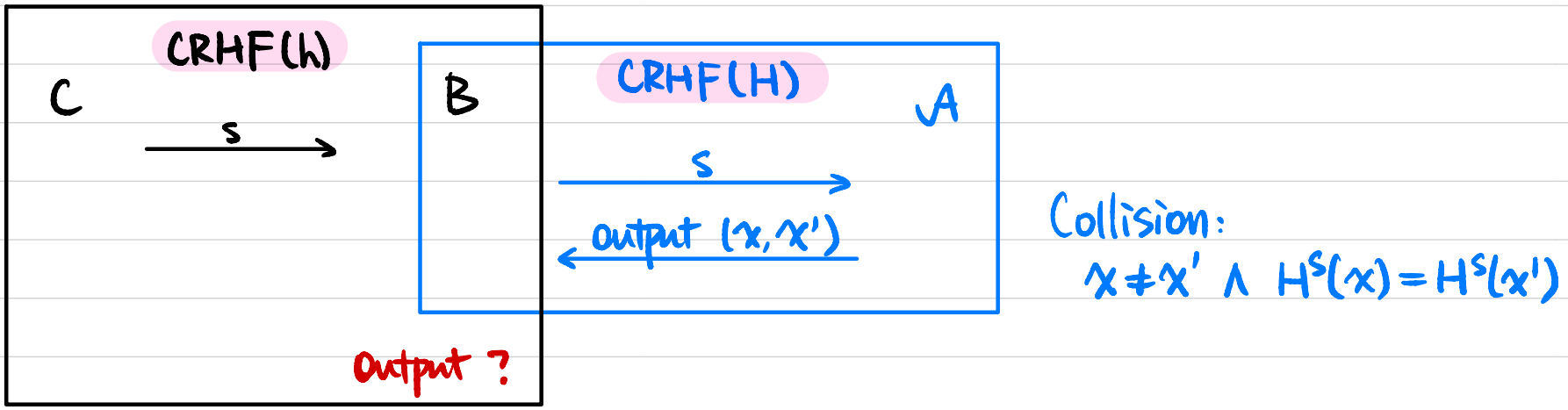
bit representation of $|x|$

$z_0 := 0^n$ $\quad z_i := h^s(z_{i-1} \| x_i)$ $\forall i \in [B]$ $\quad z_{B+1} := h^s(z_B \| \langle |x| \rangle)$ $\quad H^s(x) := z_{B+1}$

Thm If (Gen, h) is CRHF, then so is (Gen, H).

Thm If (Gen, h) is CRHF, then so is (Gen, H).

Proof Assume not, then $\exists$ PPT $\mathcal{A}$ that breaks the collision resistence of (Gen, H).
We construct a PPT B to break the collision resistence of (Gen, h).



Case 1: $z_B \| |x| \neq z'_{B'} \| |x'|$

$$h^s(z_B \| |x|) = h^s(z'_{B'} \| |x'|)$$

$$m := z_B \| |x| \qquad m' := z'_{B'} \| |x'|$$

Case 2: $z_B \| |x| = z'_{B'} \| |x'|$, $B = B'$

$$\exists i \in [B] \text{ s.t. } x_i \neq x'_i \wedge z_i = z'_i$$

$$m := z_{i-1} \| x_i \qquad m' := z'_{i-1} \| x'_i$$

B outputs (m, m') as a collision for h^s.

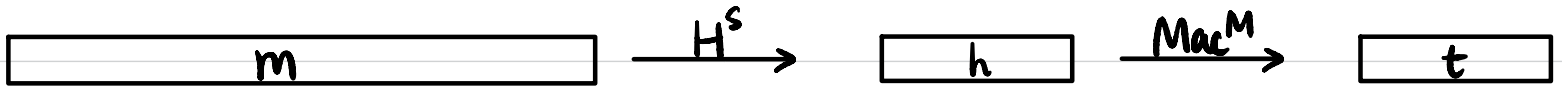
Hash-and-MAC

Secure MAC for fixed-length messages
+
CRHF for arbitrary-length inputs
$\Rightarrow$ Secure MAC for arbitrary-length messages

Let $\Pi^M = (Gen^M, Mac^M, Vrfy^M)$ be a secure MAC for messages of length n.
Let $\Pi^H = (Gen^H, H)$ be a CRHF for arbitrary-length inputs with output length n.

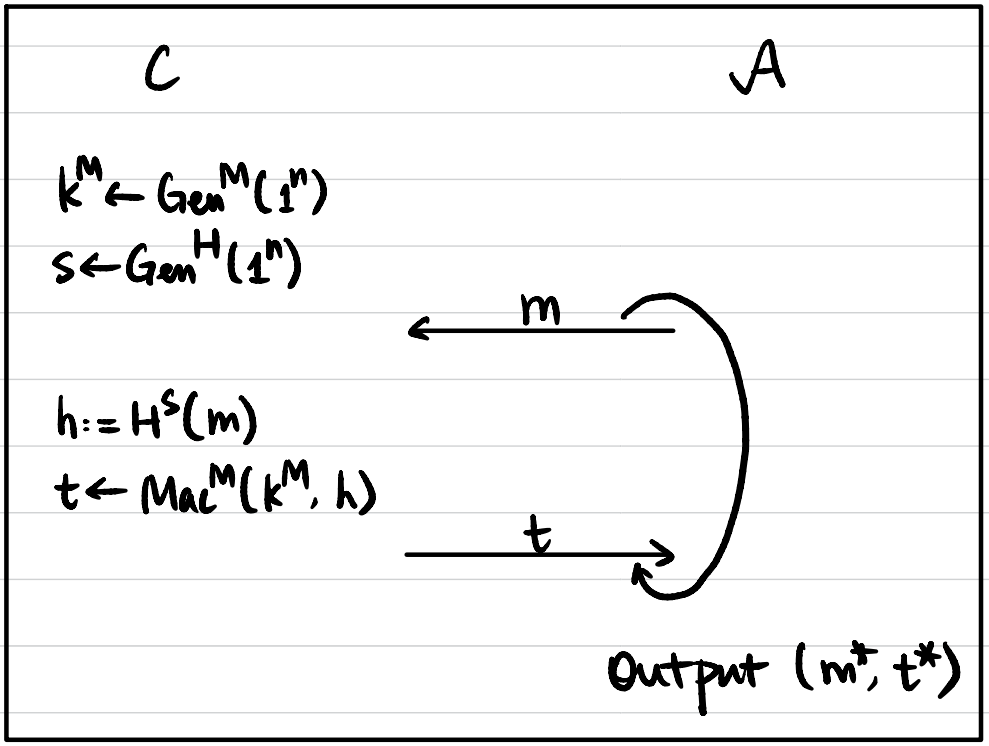
Construct $\Pi = (Gen, Mac, Vrfy)$:

- $Gen(1^n)$: $k^M \leftarrow Gen^M(1^n)$. $s \leftarrow Gen^H(1^n)$. Output $k = (k^M, s)$
- $Mac(k, m)$: $m \in \{0,1\}^*$. Parse $k = (k^M, s)$
  $h := H^s(m)$, $t \leftarrow Mac^M(k^M, h)$. Output t.
- $Vrfy(k, (m, t))$: Parse $k = (k^M, s)$
  $h := H^s(m)$. $b := Vrfy^M(k^M, (h, t))$. Output b.

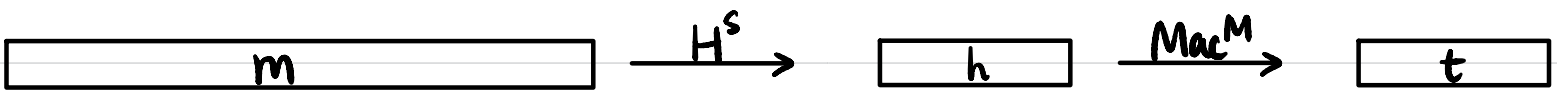


Thm If Π^M is a secure MAC and Π^H is CRHF, then Π is a secure MAC.

Thm If Π^M is a secure MAC and Π^H is CRHF, then Π is a secure MAC.



$Q := \{m \mid m \text{ queried by } \mathcal{A}\}$



Step 1: $\forall$ PPT $\mathcal{A}$, $\Pr[\exists m \in Q \text{ s.t. } m \neq m^* \wedge H^s(m) = H^s(m^*)] \leq \text{negl}(n)$.
(follows from collision-resistance of H)

Step 2: Assume $H^s(m) \neq H^s(m^*)$ $\forall m \in Q$, then unforgeability follows from MAC security.

Applications of Hash Functions

- **Deduplication**

$H(D_1) \rightarrow h_1$

$H(D_2) \rightarrow h_2$

unique identifier

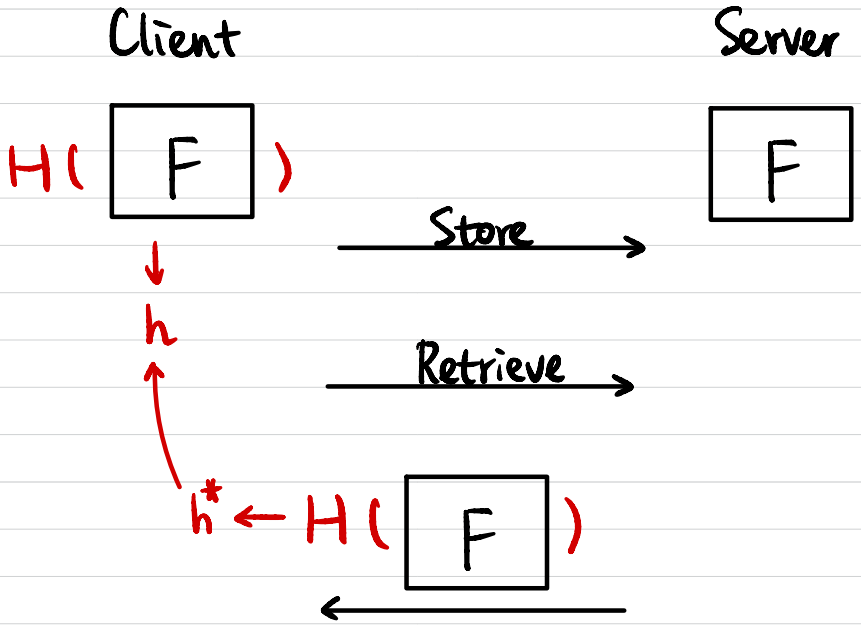
If $h_1 \neq h_2 \Rightarrow D_1 \neq D_2$

If $h_1 = h_2 \Rightarrow D_1 = D_2$ Why?

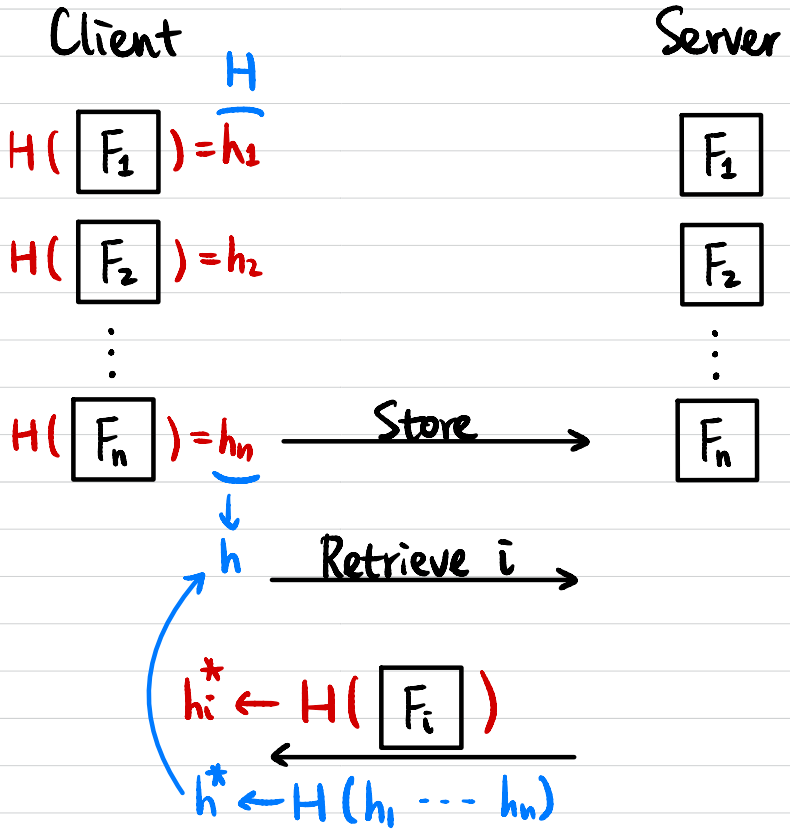
Virus Scan $H(F) \stackrel{?}{=} H(F^*)$

Video Deduplication $H(V_1) \stackrel{?}{=} H(V_2)$

Applications of Hash Functions



Is the file changed?



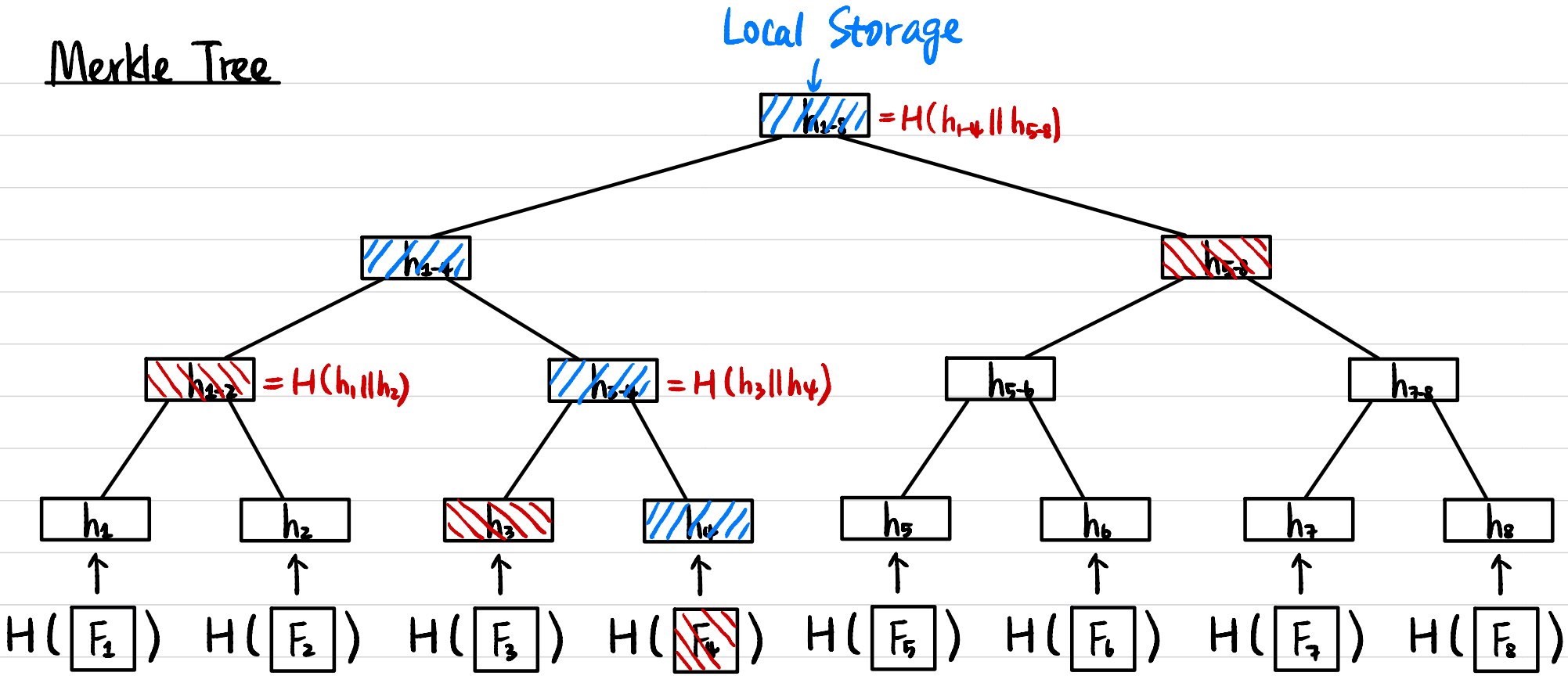
Is the file changed?

Goal:

① Client's storage doesn't grow with n. — O(1)

② Verification doesn't grow with n. — O(log n)

Merkle Tree



$$H^s : \{0,1\}^* \to \{0,1\}^n$$

$$MT_t^s(F_1 \| \cdots \| F_t) \to \{0,1\}^n$$

How does verification work?

Thm If (Gen, H) is a CRHF, then (Gen, MT_t) is a CRHF for any fixed $t = 2^k$.