CSL759: Cryptography and Computer Security

Ragesh Jaiswal

CSE, IIT Delhi

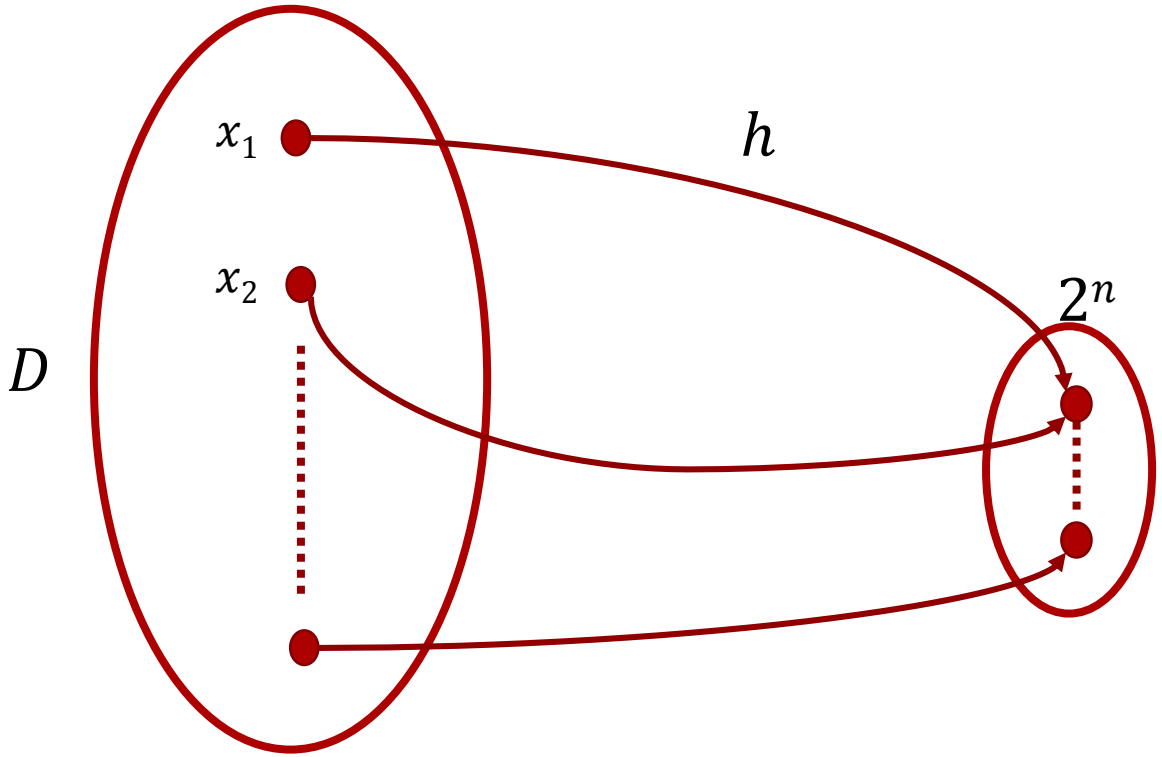
Hash Functions

Hash Functions: Introduction

- A hash function is a map $h: D \to \{0,1\}^n$ that is compressing, i.e., $|D| > 2^n$.
- Usually $|D| \gg 2^n$ and n is small.
  - Example:
    - $D = \{0,1\}^{\leq 2^{64}}$ i.e., all binary strings of length at most 2^{64}.
    - $n = 128, 160, 256$ etc.
- Examples of Cryptographic Hash Functions:

h	n
MD4	128
MD5	128
SHA1	160
SHA-256	256
SHA-512	512
WHIRLPOOL	512

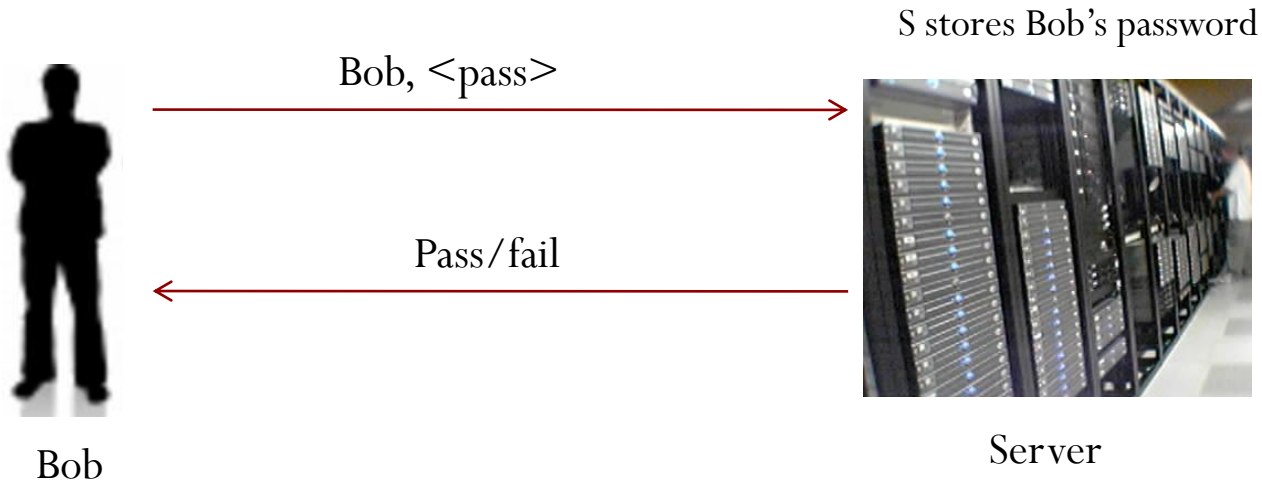
Hash Functions: Collision



Pigeonhole Principle: $h(x_1) = h(x_2), x_1 \neq x_2$

Hash Functions: Applications

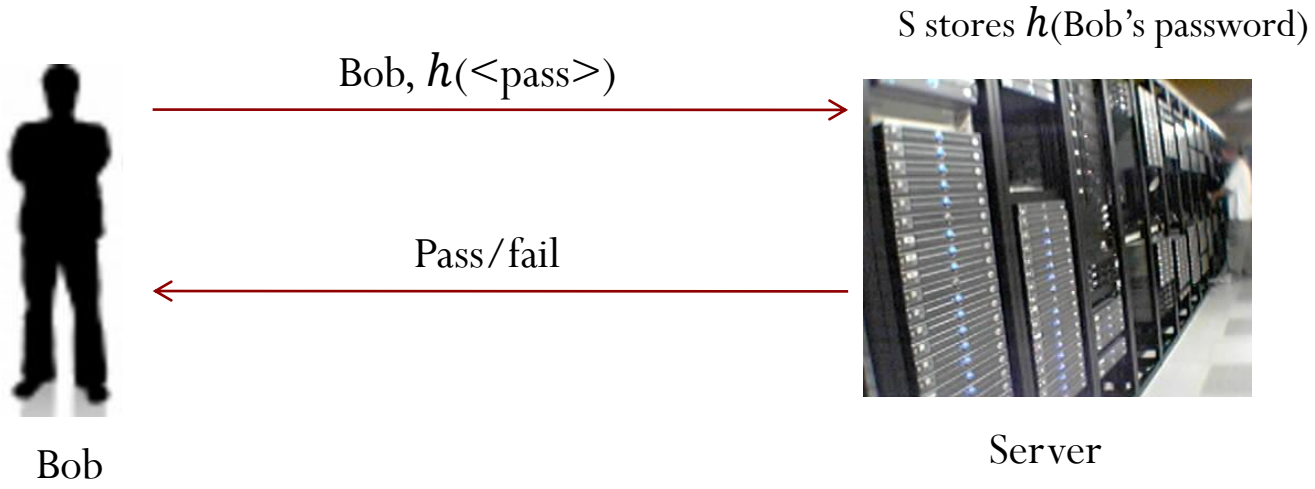
1. Password Authentication:



- Problem: If Eve hacks into the server or if the communication channel is not secure, then Eve knows the password of Bob.

Hash Functions: Applications

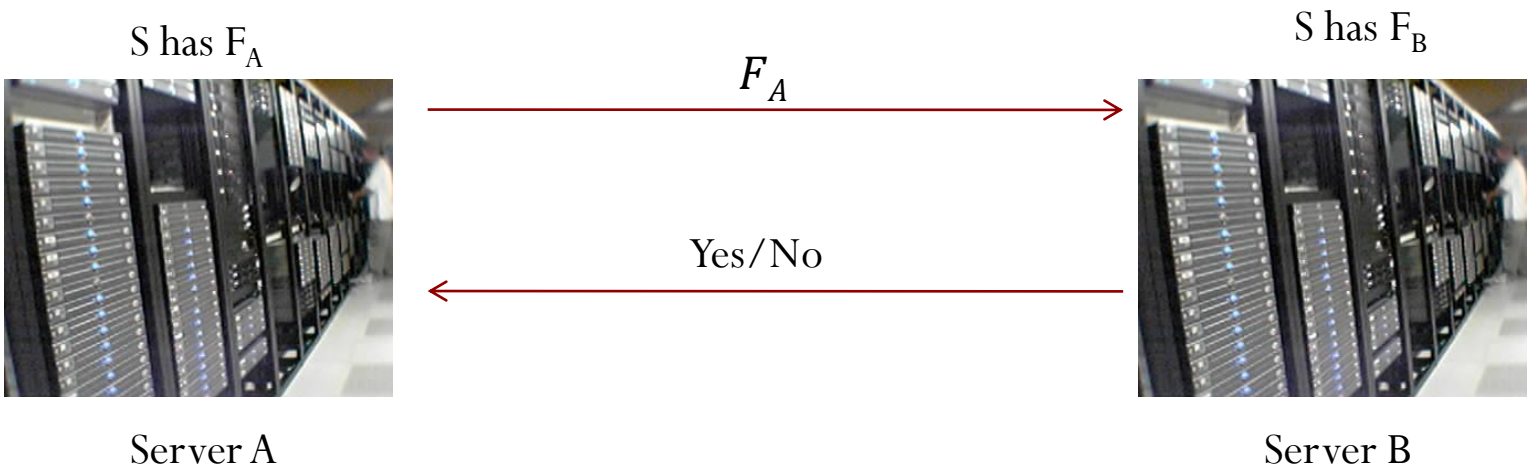
1. Password Authentication:



- Eve can only get access to h(<pass>).

Hash Functions: Applications

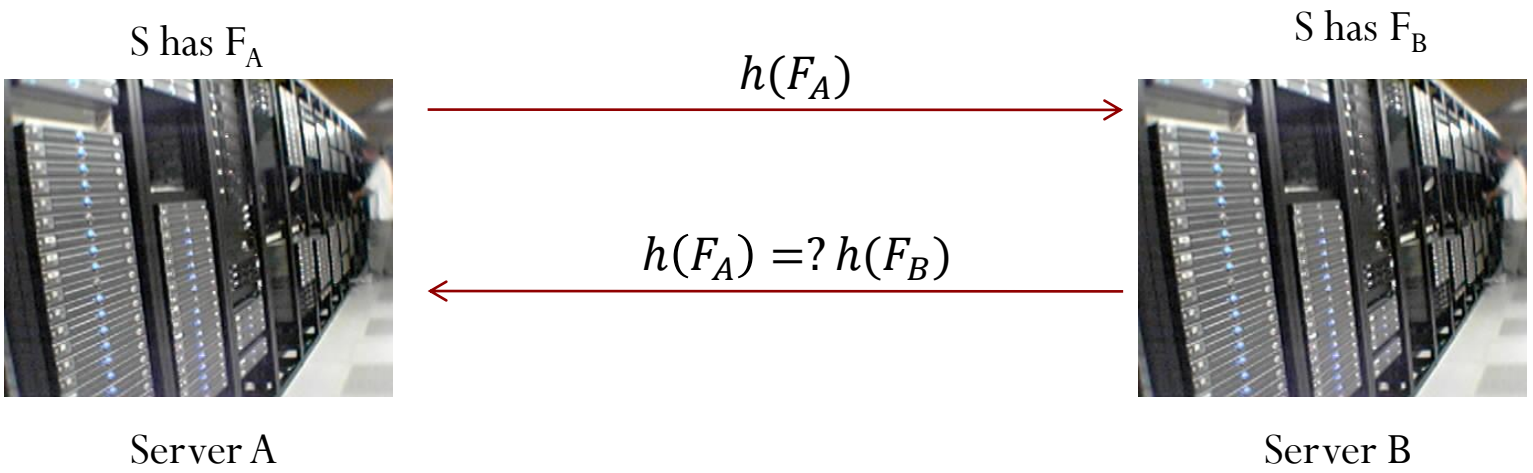
2. Comparing files by hashing:



• Problem: Files are usually very large and we would like to save communication costs/delays.

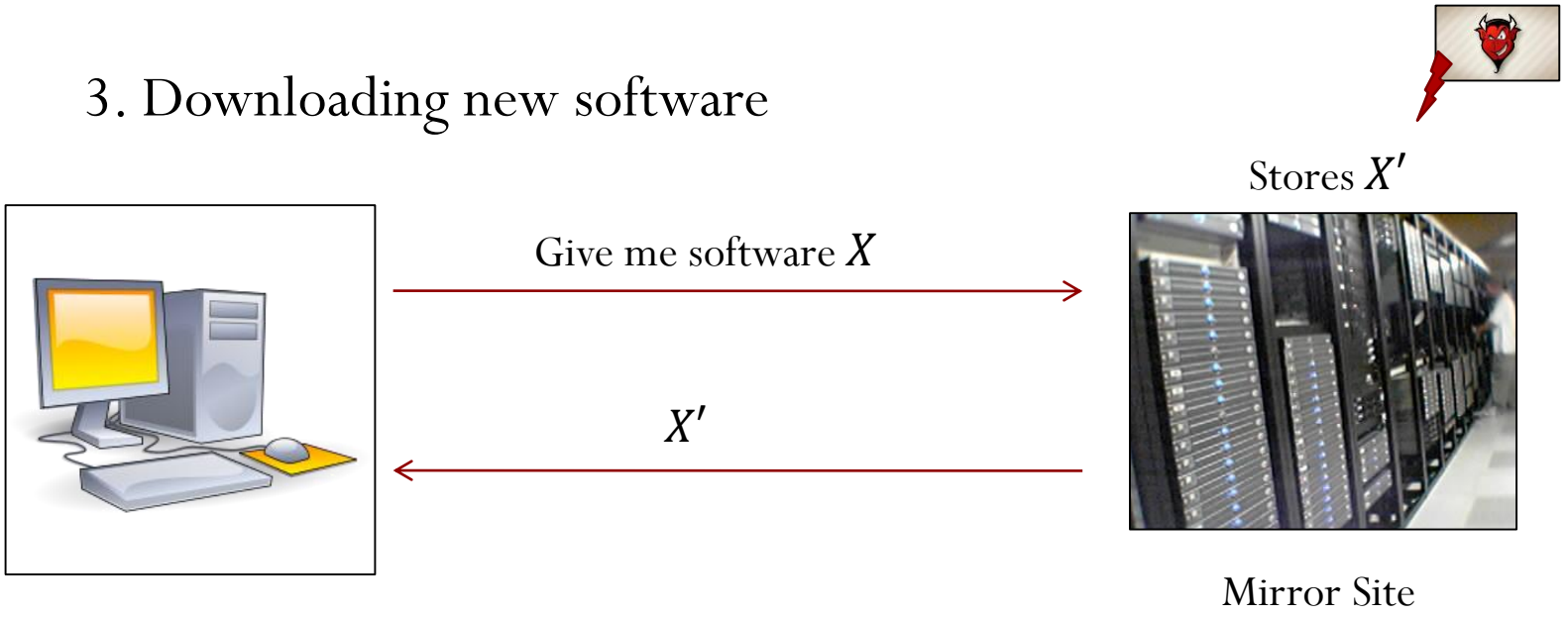
Hash Functions: Applications

2. Comparing files by hashing:



Hash Functions: Applications

3. Downloading new software



- Problem: X' could be a virus-infected version of X.

Hash Functions: Applications

3. Downloading new software

Stores X'

Give me software X

X'

$h(X)$

Mirror site

Stores $h(X)$ in read-only mode

Software site

Collision Resistance

- Password Authentication: If Eve is able to find a string S (perhaps different from $< pass >$) such that
$$h(S) = h(< pass >)$$
then the scheme breaks.
- Comparing files: If there is a different file F_S such that
$$h(FS) = h(FB)$$
the servers may agree incorrectly.
- Downloading software: If Eve can find $X' \neq X$ such that $h(X) = h(X')$, then software might cause problems.
- Collision Resistance: It is computationally infeasible to find a pair (x_1, x_2) such that $x_1 \neq x_2$ and
$$h(x_1) = h(x_2)$$
- If a hash function h is collision resistant, then the above two problems are avoided.

Collision Resistance: Discussion

- Are there functions that are collision resistant?
  - Fortunately, there are functions for which no one has been able to find a collision!
  - Example: $SHA-1\colon \{0,1\}^D \rightarrow \{0,1\}^{160}$
- Is the world drastically going to change if someone finds one or few collision for SHA-1?
  - Not really. Suppose the collision has some very specific structure, then we may avoid such structures in the strings on which the hash function is applied.
  - On the other hand, if no one finds a collision then that is a very strong notion of security and we may sleep peacefully without worrying about maintaining complicated structures in the strings.
  - We are once again going for a very strong definition of security for our new primitive similar to Block Ciphers and Symmetric Encryption.

Collision Resistance

CR-security

CR-Security

- For a hash function $h: D \rightarrow \{0,1\}^n$, CR-security is defined using the following experiment.

- $CR_{A,h}$
  - Let the adversary A return (X_0, X_1).
  - If $((h(X_0) = h(X_1)) \wedge (X_0 \neq X_1))$, output 1 else output 0

- $Adv_{CR}(A,h) = \Pr[CR_{A,h} = 1]$

CR-Security

- For a hash function $h: D \rightarrow \{0,1\}^n$, CR-security is defined using the following experiment.
- $CR_{A,h}$
  - Let the adversary A return (X_0, X_1).
  - If $(h(X_0) = h(X_1))$ and $(X_0 \neq X_1)$, output 1 else output 0
- $Adv_{CR}(A,h) = \Pr[CR_{A,h} = 1]$

- Let $h: \{0,1\}^{256} \rightarrow \{0,1\}^{128}$ defined as
  $$h(X) = h(X_1|X_2) = AES_K(X_1) \oplus AES_K(X_2)$$
  where $K \in \{0,1\}^{128}$ is a fixed constant.
- Can you design an adversary that has a high CR-advantage?

CR-Security

- For a hash function $h: D \rightarrow \{0,1\}^n$, CR-security is defined using the following experiment.
- $CR_{A,h}$
  - Let the adversary A return (X_0, X_1).
  - If $(h(X_0) = h(X_1))$ and $(X_0 \neq X_1)$, output 1 else output 0
- $Adv_{CR}(A,h) = \Pr[CR_{A,h} = 1]$

- Let $h: \{0,1\}^{256} \rightarrow \{0,1\}^{128}$ defined as
  $$h(X) = h(X_1|X_2) = AES_K(X_1) \oplus AES_K(X_2)$$
  where $K \in \{0,1\}^{128}$ is a fixed constant.
- Can you design an adversary that has a high CR-advantage?
  - Yes. Adversary sends $(0^{128}1^{128}, 1^{128}0^{128})$.
  - $Adv_{CR}(A,h) = 1$.

Merkel Damgard (MD) Transform

MD Transform

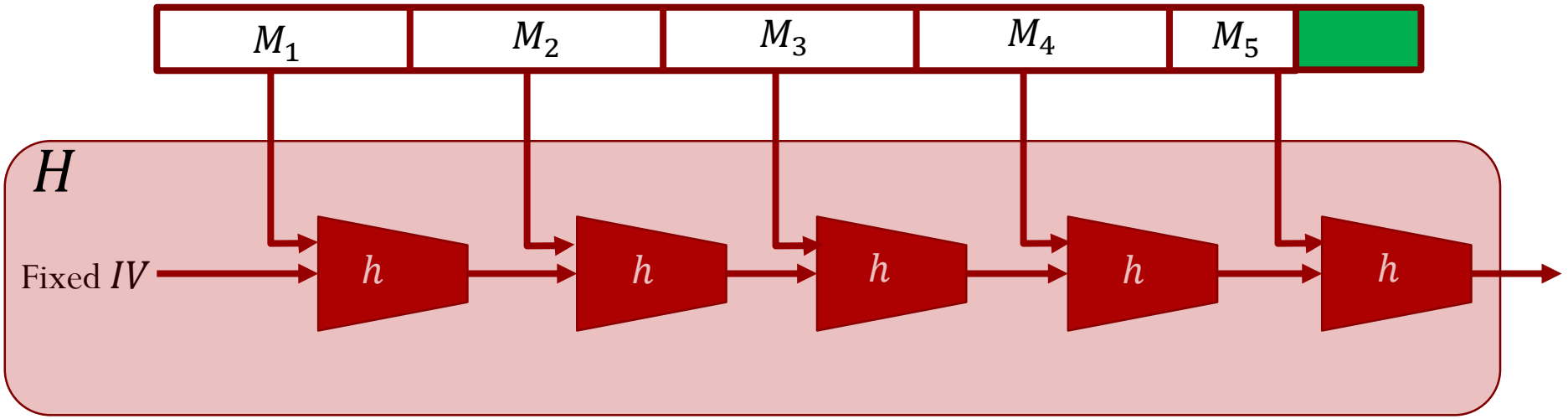
- Let us break down designing a CR-secure hash function that maps arbitrarily large bit strings to small strings, into the following two parts:
  1. Design a CR-secure hash function h for short, fixed-size messages.
  2. Use h in a **standard construction** to obtain a hash function H that hashes arbitrary long messages. Show that if h is CR-secure, then so is H.

MD Transform

- Let us break down designing a CR-secure hash function that maps arbitrarily large bit strings to small strings, into the following two parts:
  1. Design a CR-secure hash function h for short, fixed-size messages.
  2. Use h in a **standard construction** to obtain a hash function H that hashes arbitrary long messages. Show that if h is CR-secure, then so is H.
- One such standard construction of part 2 is the Merkel-Damgard (MD) Transform.

MD Transform

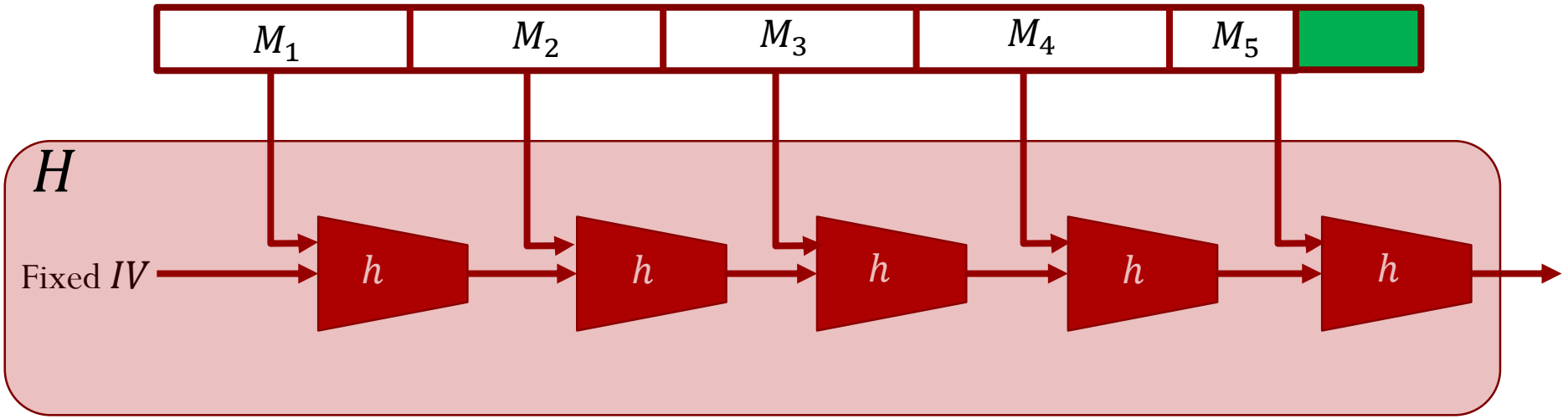
- Suppose we have a collision-resistant hash function for short strings $h: \{0,1\}^{b+n} \rightarrow \{0,1\}^n$ (e.g. SHA-1: $b = 512, n = 160$), consider a hash function $H: \{0,1\}^{\leq 2^{64}-1} \rightarrow \{0,1\}^n$ for longer strings constructed in the following manner:



Where the padding block is a string of all 0's

MD Transform

- Suppose we have a collision-resistant hash function for short strings $h: \{0,1\}^{b+n} \rightarrow \{0,1\}^n$, consider a hash function $H: \{0,1\}^{\leq 2^{64}-1} \rightarrow \{0,1\}^n$ for longer strings constructed in the following manner:

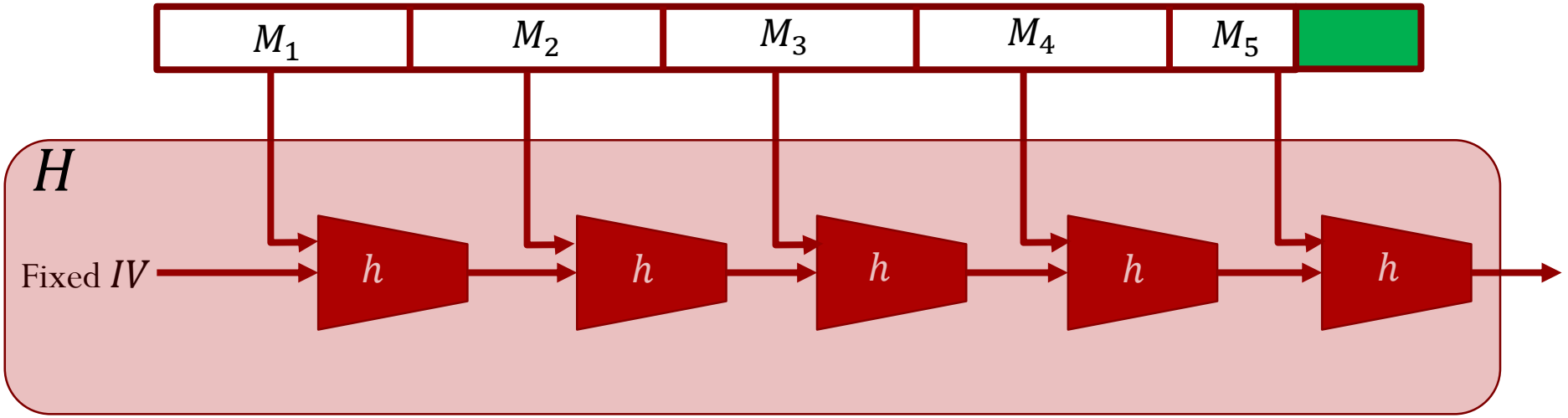


Where the padding block is a string of all 0's

- Is H a collision-resistant hash function?

MD Transform

- Suppose we have a collision-resistant hash function for short strings $h: \{0,1\}^{b+n} \rightarrow \{0,1\}^n$, consider a hash function $H: \{0,1\}^{\leq 2^{64}-1} \rightarrow \{0,1\}^n$ for longer strings constructed in the following manner:

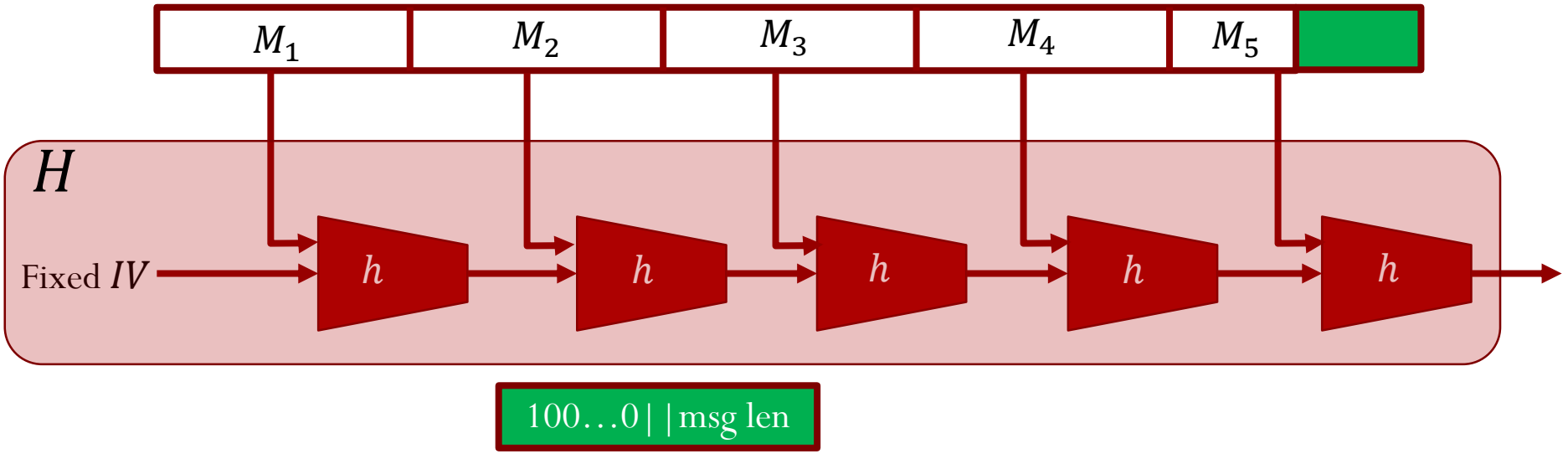


~~Where the padding block is a string of all 0's~~

- Is H a collision-resistant hash function?
  - No since $H(1) = H(100 \ldots 0)$.

MD Transform

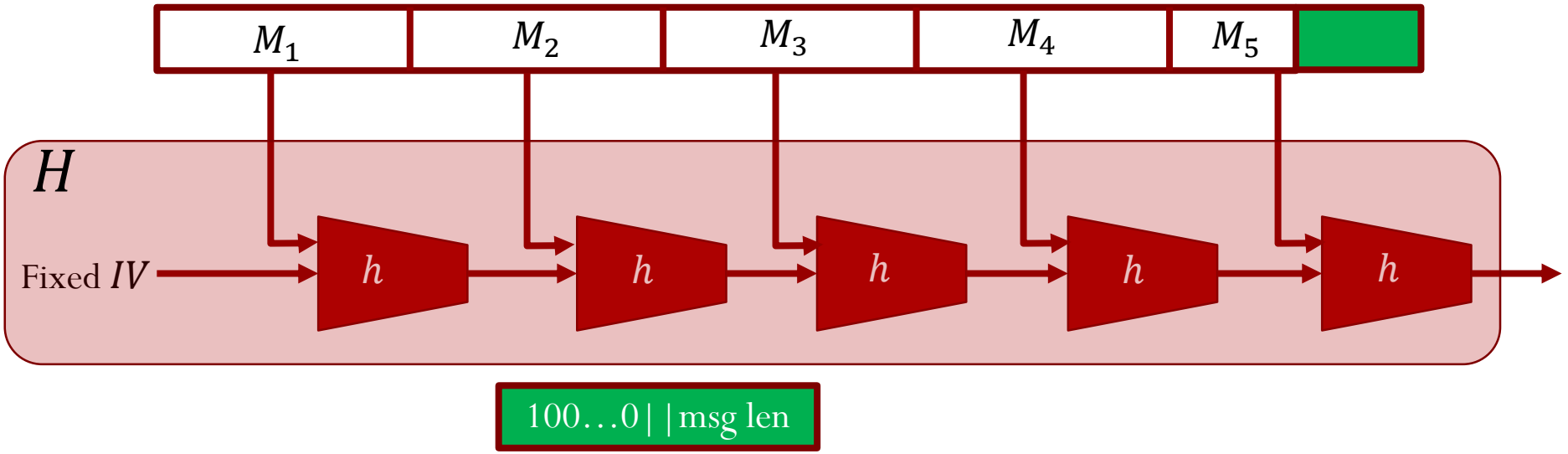
- Suppose we have a collision-resistant hash function for short strings $h: \{0,1\}^{b+n} \rightarrow \{0,1\}^n$, consider a hash function $H: \{0,1\}^{\leq 2^{64}-1} \rightarrow \{0,1\}^n$ for longer strings constructed in the following manner:



- Is H a collision-resistant hash function?

MD Transform

- Suppose we have a collision-resistant hash function for short strings $h: \{0,1\}^{b+n} \rightarrow \{0,1\}^n$, consider a hash function $H: \{0,1\}^{\leq 2^{64}-1} \rightarrow \{0,1\}^n$ for longer strings constructed in the following manner:

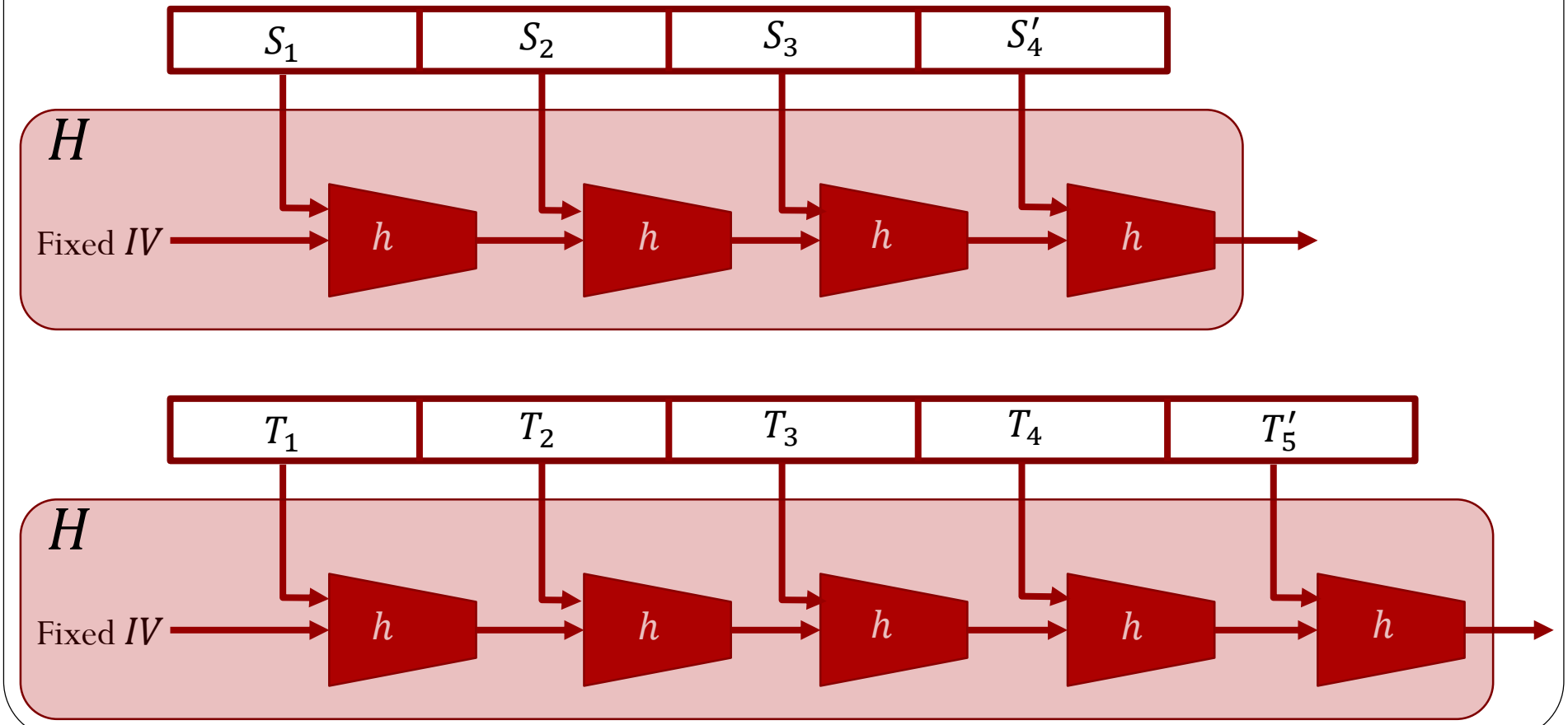


- Is H a collision-resistant hash function?
  - Yes as long as h is a collision-resistant.

MD Transform

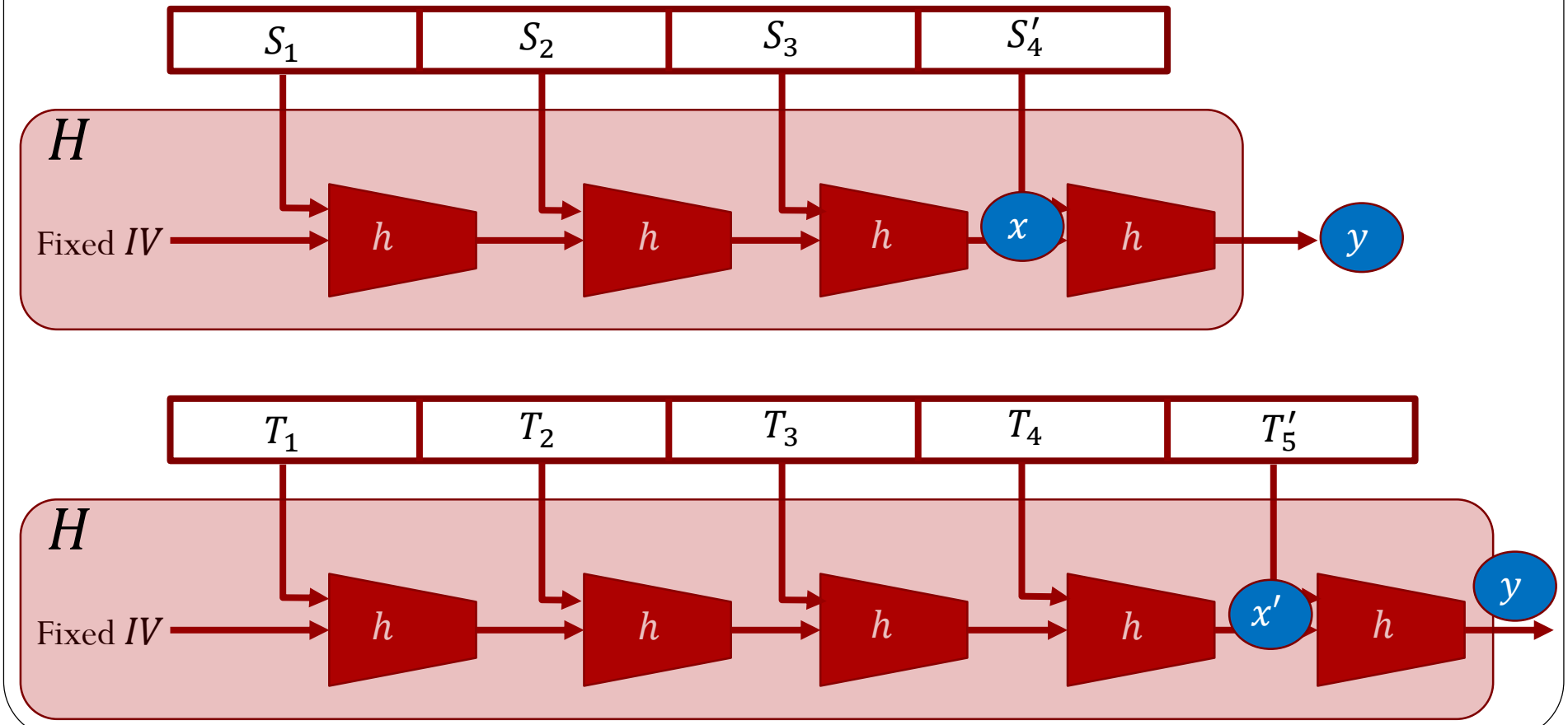
- Suppose there are two strings:
  - $S = S_1||S_2|| \dots ||S_p$
  - $T = T_1||T_2|| \dots ||T_q$

  such that $H(S) = H(T)$.



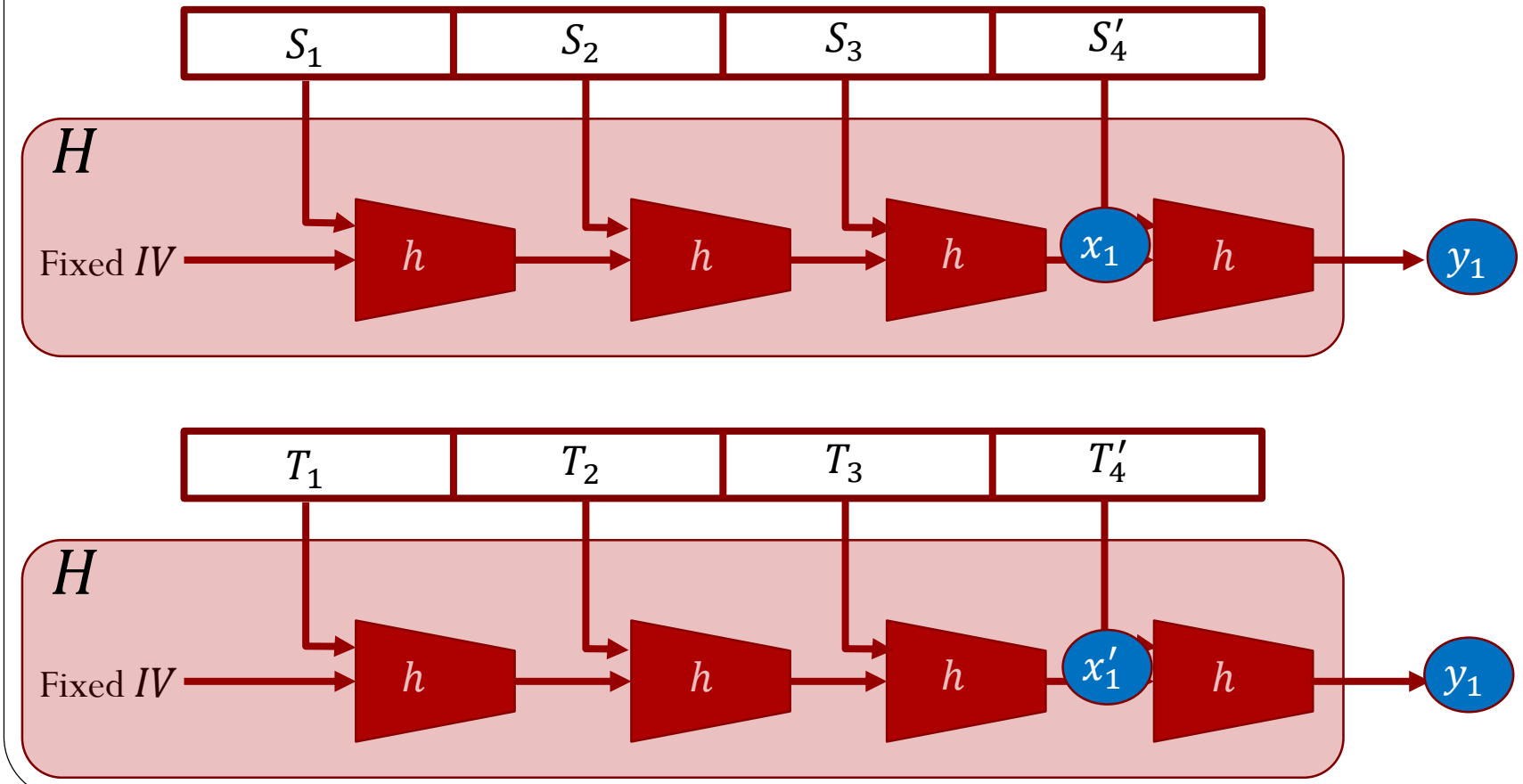
MD Transform

- If $x \neq x'$, then we have found a collision for h.



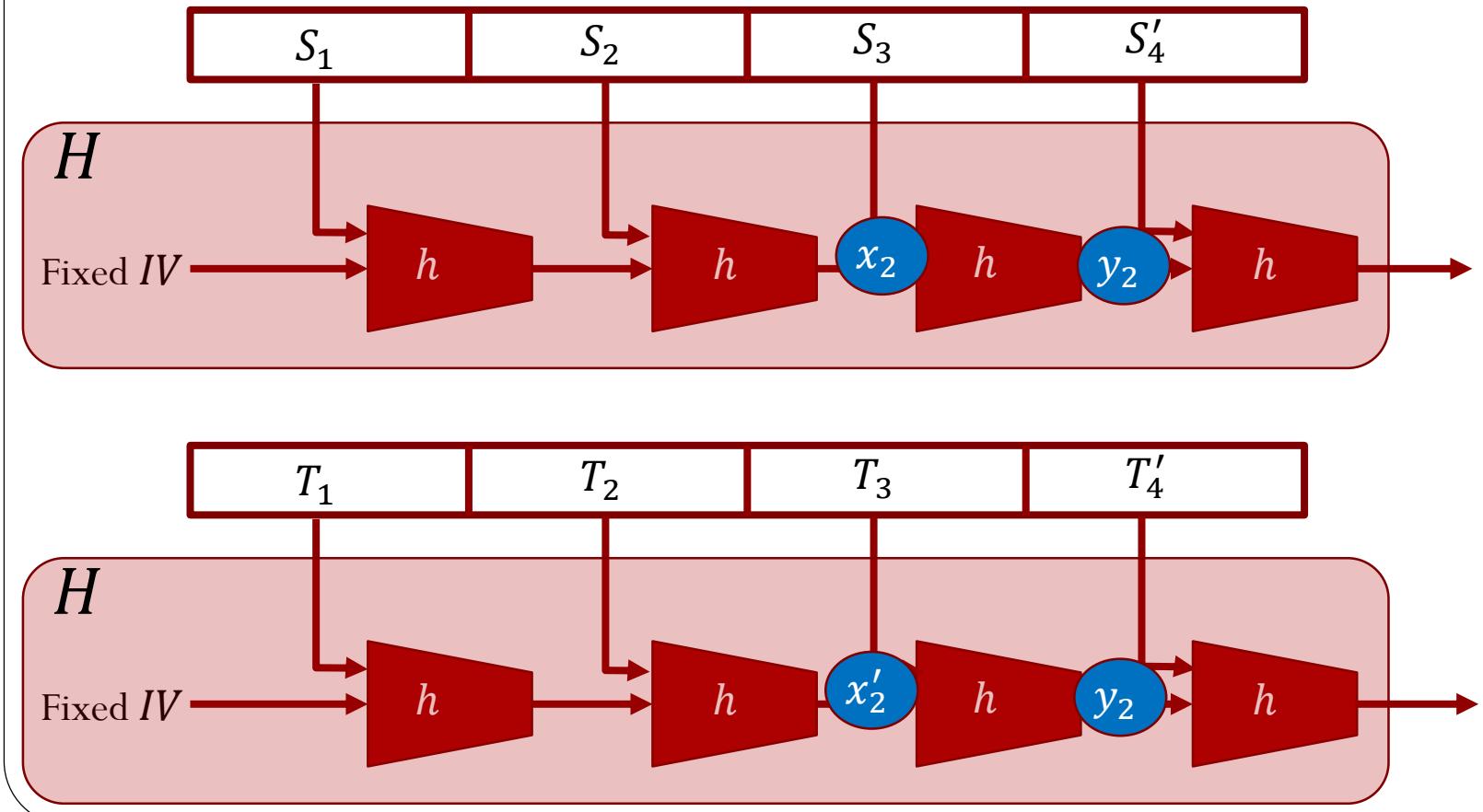
MD Transform

- If $x_1 \neq x_1'$, then we have found a collision for h.
- If $x_1 = x_1'$, then this means that the messages are of equal length.



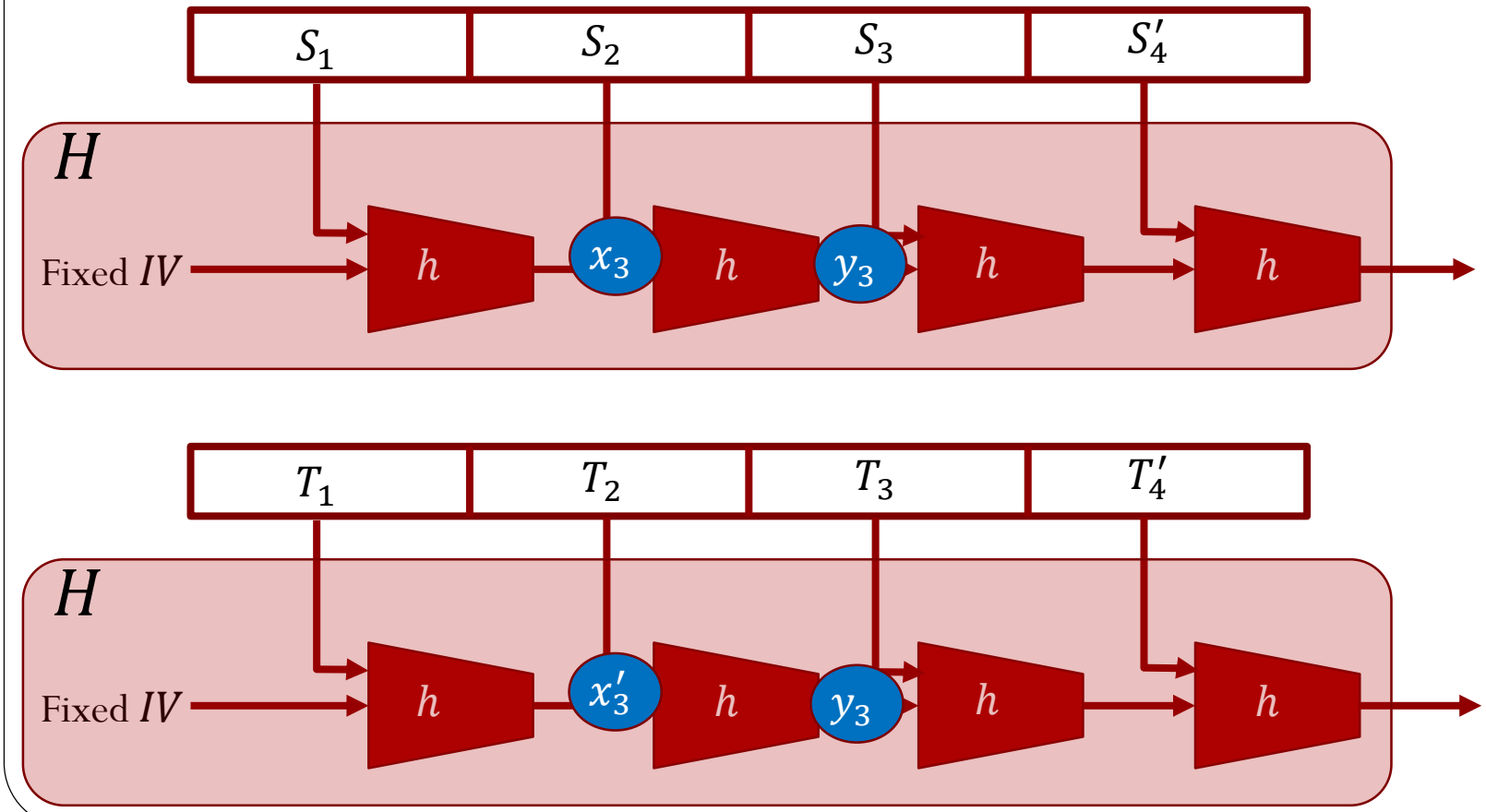
MD Transform

- Again, if $x_2 \neq x_2'$, then we have found a collision for h.
- If $x_2 = x_2'$, then shift the analysis to the previous application of h.



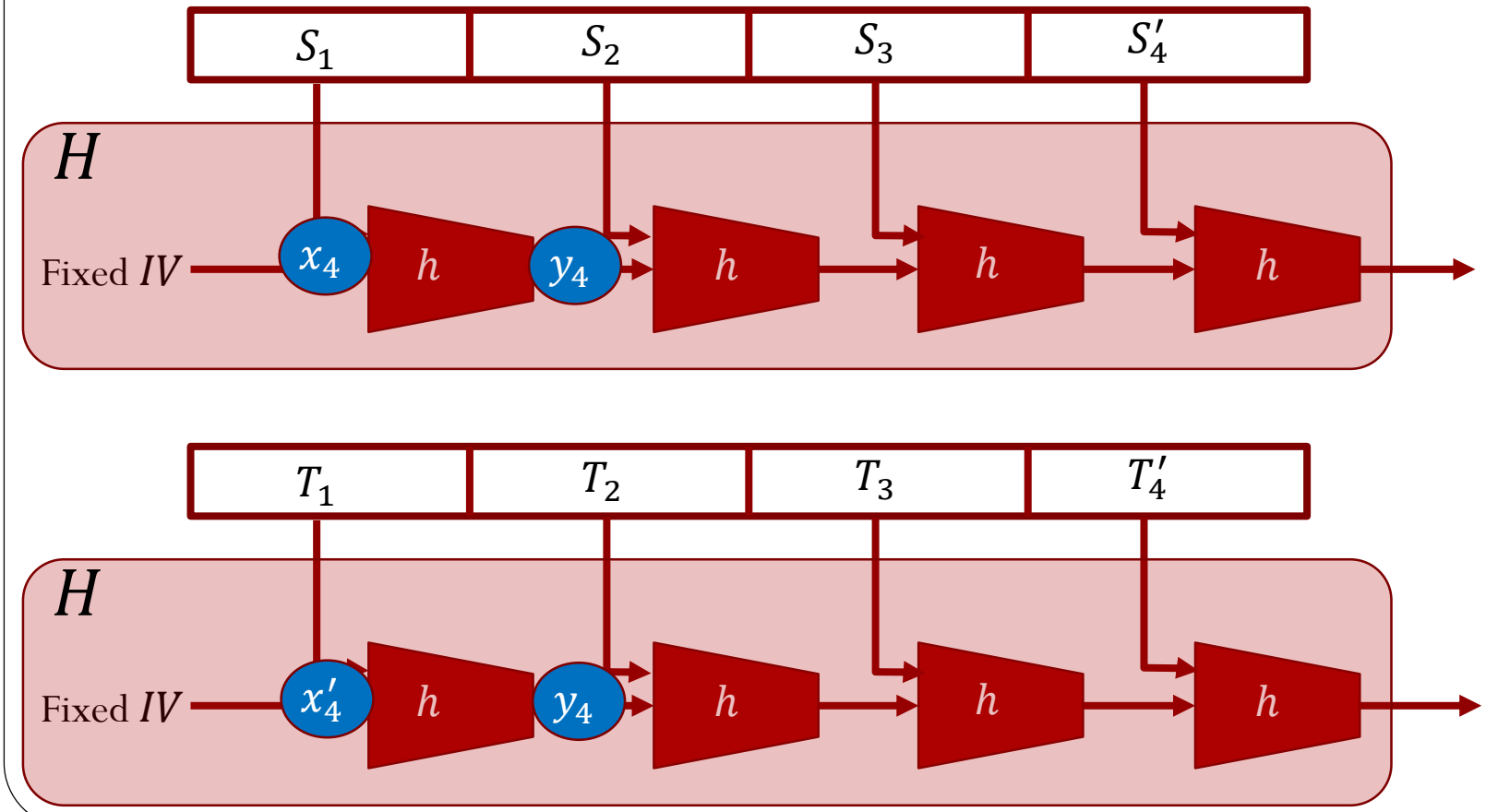
MD Transform

- Again, if $x_3 \neq x_3'$, then we have found a collision for h.
- If $x_3 = x_3'$, then shift the analysis to the previous application of h.



MD Transform

- Now $x_4 \neq x'_4$, since $S \neq T$ and we have found a collision.



Hash Function for short strings

Collision-resistant Compression Functions

- Suppose we have a block cipher $F: \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$. Consider the following candidate for a compression function $h: \{0,1\}^{k+n} \rightarrow \{0,1\}^n$ defined as:

$$h(x||v) = E_x(v)$$

- Is h collision resistant?

Collision-resistant Compression Functions

- Suppose we have a block cipher $F: \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$. Consider the following candidate for a compression function $h: \{0,1\}^{k+n} \rightarrow \{0,1\}^n$ defined as:

$$h(x||v) = E_x(v)$$

- Is h collision resistant?
  - No.
  - How do we prove this?

Collision-resistant Compression Functions

- Suppose we have a block cipher $F: \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$. Consider the following candidate for a compression function $h: \{0,1\}^{k+n} \rightarrow \{0,1\}^n$ defined as:

$$h(x||v) = E_x(v)$$

- Is h collision resistant?
  - No.
  - How do we prove this?
    - Give a collision.

Collision-resistant Compression Functions

- Suppose we have a block cipher $F: \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$. Consider the following candidate for a compression function $h: \{0,1\}^{k+n} \rightarrow \{0,1\}^n$ defined as:

$$h(x||v) = E_x(v)$$

- Is h collision resistant?
  - No.
  - How do we prove this?
    - Give a collision.
  - Pick any $s_1 = x||v$, then pick $x' \neq x$, compute $v' = E_{x'}^{-1}(E_x(v))$ and let $s_2 = x'||v'$.

Collision-resistant Compression Functions

- Suppose we have a block cipher $F: \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$. Consider the following candidate for a compression function $h: \{0,1\}^{k+n} \rightarrow \{0,1\}^n$ defined as:

$$h(x||v) = E_x(v)$$

- Is h collision resistant?
  - No.
  - How do we prove this?
    - Give a collision.
  - Pick any $s_1 = x||v$, then pick $x' \neq x$, compute $v' = E_{x'}^{-1}(E_x(v))$ and let $s_2 = x'||v'$.
  - Claim: $h(s_1) = h(s_2)$ and $s_1 \neq s_2$.

Collision-resistant Compression Functions

- Here are some examples of constructions based on clock ciphers that are believed to be collision resistant.
  - Davies-Meyer: Given a block cipher $F: \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$, the hash function $h: \{0,1\}^{k+n} \rightarrow \{0,1\}^n$ is defined as follows:

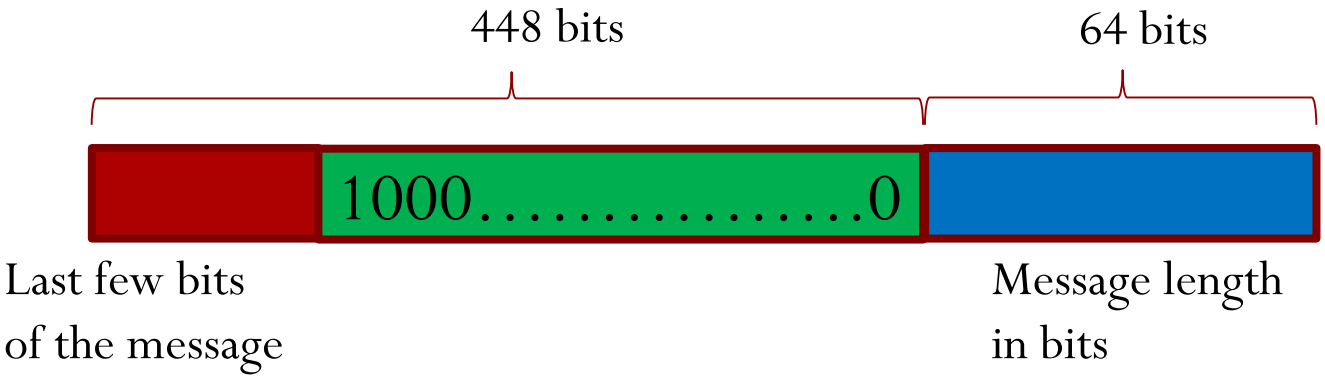
    $$h(x||v) = F_x(v) \oplus v$$

    This is used in SHA-1, SHA-2.
  - Miyaguchi-Preneel: Given a block cipher $F: \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$, the hash function $h: \{0,1\}^{n+n} \rightarrow \{0,1\}^n$ is defined as follows:
    $h(x||v) = F_x(v) \oplus v \oplus x$ and other variants. This is used in WHIRLPOOL hash function.

Hash Function Examples

Hash Function: Examples

- SHA-256: $\{0,1\}^{\leq 2^{64}-1} \rightarrow \{0,1\}^{256}$
  - Compression function: Uses Davies-Meyer construction using the SHACAL-2:$\{0,1\}^{512} \times \{0,1\}^{256} \rightarrow \{0,1\}^{256}$ block cipher.
  - Uses MD transform for longer strings.
  - Uses padding as shown below:



Birthday attack on Hash Functions

Birthday attack on hash functions

- Here is an extremely simple attack on a hash function $H: D \rightarrow \{0,1\}^n$.
  - Adversary A
    - For $i = 1$ to q
      - Let $x_i \leftarrow D$ and $y_i \leftarrow H(x_i)$
    - If there exists $i, j \in [q]$, such that $x_i \neq x_j$ and $y_i = y_j$, then output (x_1, x_2).
  - What is $Adv_{CR}(A, H)$?

Birthday attack on hash functions

- Here is an extremely simple attack on a hash function $H: D \rightarrow \{0,1\}^n$.
  - Adversary A
    - For $i = 1$ to q
      - Let $x_i \leftarrow D$ and $y_i \leftarrow H(x_i)$
    - If there exists $i, j \in [q]$, such that $x_i \neq x_j$ and $y_i = y_j$, then output (x_1, x_2).
  - What is $Adv_{CR}(A, H)$?
  - Is $Adv_{CR}(A, H) = C(q, 2^n)$?

Birthday attack on hash functions

- Here is an extremely simple attack on a hash function $H: D \rightarrow \{0,1\}^n$.
  - Adversary A
    - For $i = 1$ to q
      - Let $x_i \leftarrow D$ and $y_i \leftarrow H(x_i)$
    - If there exists $i, j \in [q]$, such that $x_i \neq x_j$ and $y_i = y_j$, then output (x_1, x_2).
  - What is $Adv_{CR}(A, H)$?
  - Is $Adv_{CR}(A, H) = C(q, 2^n)$?
    - No since y_i's are not randomly chosen strings from $\{0,1\}^n$.

Birthday attack on hash functions

- Here is an extremely simple attack on a hash function $H: D \to \{0,1\}^n$.
  - Adversary A
    - For $i = 1$ to q
      - Let $x_i \leftarrow D$ and $y_i \leftarrow H(x_i)$
    - If there exists $i, j \in [q]$, such that $x_i \neq x_j$ and $y_i = y_j$, then output (x_1, x_2).
  - What is $Adv_{CR}(A, H)$?
  - Is $Adv_{CR}(A, H) = C(q, 2^n)$?
    - No since y_i's are not randomly chosen strings from $\{0,1\}^n$.
- Suppose that the hash function H is a *regular*. This means that $\forall y \in \{0,1\}^n, |\{x \in D: H(x) = y\}| = \frac{|D|}{2^n}$, then $Adv_{CR}(A, H) = C(q, 2^n)$.

Birthday attack on hash functions

- Here is an extremely simple attack on a hash function $H: D \rightarrow \{0,1\}^n$.
- Suppose that the hash function H is a *regular*. This means that $\forall y \in \{0,1\}^n, |\{x \in D: H(x) = y\}| = \frac{|D|}{2^n}$, then $Adv_{CR}(A, H) = C(q, 2^n)$.
- Theorem: Let $H: D \rightarrow \{0,1\}^n$ be a regular hash function, then the birthday attack finds a collision in about $2^{n/2}$ trials.

Birthday attack on hash functions

- Here is an extremely simple attack on a hash function $H: D \to \{0,1\}^n$.
- Suppose that the hash function H is a *regular*. This means that $\forall y \in \{0,1\}^n, |\{x \in D: H(x) = y\}| = \frac{|D|}{2^n}$, then $Adv_{CR}(A, H) = C(q, 2^n)$.
- Theorem: Let $H: D \to \{0,1\}^n$ be a regular hash function, then the birthday attack finds a collision in about $2^{n/2}$ trials.
- What if H is not regular? How does the birthday attack behave?

Birthday attack on hash functions

- Here is an extremely simple attack on a hash function $H: D \to \{0,1\}^n$.
- Suppose that the hash function H is a *regular*. This means that $\forall y \in \{0,1\}^n, |\{x \in D: H(x) = y\}| = \frac{|D|}{2^n}$, then $Adv_{CR}(A, H) = C(q, 2^n)$.
- Theorem: Let $H: D \to \{0,1\}^n$ be a regular hash function, then the birthday attack finds a collision in about $2^{n/2}$ trials.
- What if H is not regular? How does the birthday attack behave?
  - The attack may succeed sooner.
- So, hash functions should be close to regular which seems to be the case for most popular hash functions.

Attacks against Hash Functions

Attacks on Hash Functions

Name	Output size	Birthday attack time	Best attack time	Best attack year
MD5	128	2^{64}	2^{20}	2009
RIPEMD	160	2^{80}	2^{18}	2004
SHA-1	160	2^{80}	2^{52}	2009
SHA-256	256	2^{128}	No collisions yet	

End