

# Digital Signatures

**PGS.TS Trần Văn Ba**

---

# Digital Signatures

## ■ Motivation

- Diffie-Hellman proposed the idea (1976)
- Simulation of the real-world into digital worlds
  - Paper contracts need signed to be valid so do electronic versions

## ■ The proofs conveyed in signatures

- Data integrity: information is original, not modified
  - Authentication: The source of the info is correct, not impersonated
-

# DS: how they work

- Digital Signature: a data string which associates a message with some originating entity.
- Digital Signature Scheme:
  - a signing algorithm: takes a message and a (private) signing key, outputs a signature
  - a verification algorithm: takes a (public) key verification key, a message, and a signature
- A DS is created based on a PK system
  - Alice signs message  $X$  by creating  $Y = D_{z_A}(X)$ , so the signed document now is  $(X, Y = D_{z_A}(X))$ .
  - Bob who receives  $(X, Y)$ , computes  $X' = E_{z_A}(Y)$  then compare if  $X = X'$  to confirm the document's validity

# DS system



$S_A$  : DS creating function

( $m$ )

Chào B, tôi là  
A. Dưới đây là  
thông tin cá  
nhân của tôi:

$$S_A(m) = s$$

Signer A

( $s$ )

hQImaw9dfDA  
WEPmj9h87on  
we1jd03n

Chào B, tôi là  
A. Dưới đây là  
thông tin cá  
nhân của tôi:

hQImaw9dfDA  
WEPmj9h87on  
we1jd03n

**true:**

- + information is original, not modified
- + The source of the info is correct, not impersonated

**false:**

- + information may has been modified
- OR
- + The source of the info is not

$$V_A(m, s)$$

Chào B, tôi là  
A. Dưới đây là  
thông tin cá  
nhân của tôi:

hQImaw9dfDA  
WEPmj9h87on  
we1jd03n

Chào B, tôi là  
A. Dưới đây là  
thông tin cá  
nhân của tôi:

hQImaw9dfDA  
WEPmj9h87on  
we1jd03n



Verifier B

Verifying function:  $V_A$

# Non-repudiation

- The signer can't deny that his/her created the document
  - Only Alice knows  $z_A$  to create  $(X, Y=D_{z_A}(X))$  but everyone else can verify
- So we say the DS scheme provides non-repudiation

# Weakness of the signature scheme mentioned so far

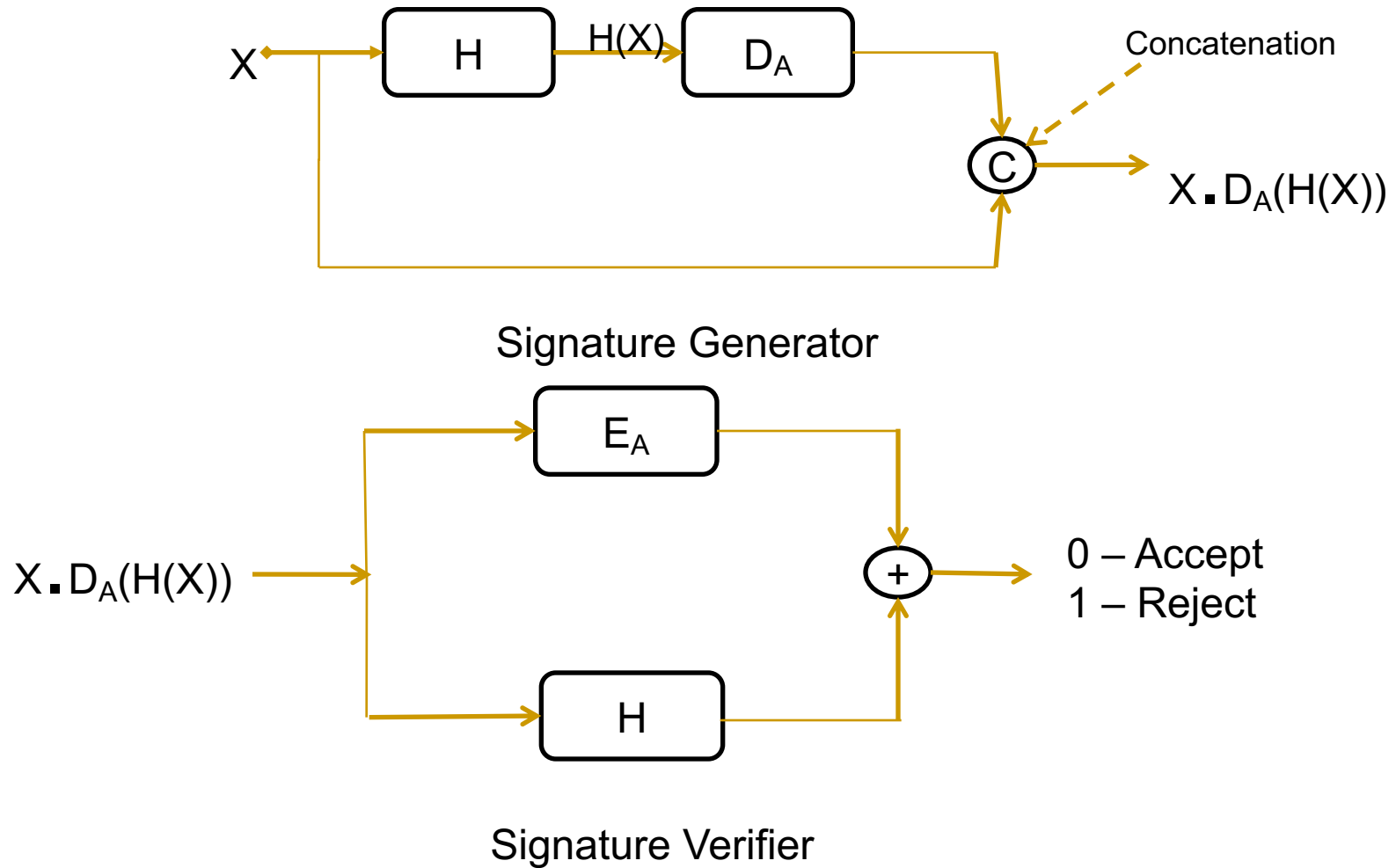
- When using a PKC to sign  $X$ ,  $X$  can be long → splitting into blocks and signs  
$$X = (X_1, X_2, X_3, \dots, X_t) \rightarrow (SA(X_1), SA(X_2), SA(X_3), \dots, SA(X_t))$$
- This creates vulnerability to attack on manipulating blocks
  - The attacker can change order of blocks, remove/ add in a few
- Slow: PKC is already slow, now is run multiple times
- Signature is long, as long as the message itself.

# Hash Functions

- A hash function  $H$  maps a message of variable length  $n$  bits to a fingerprint of fixed length  $m$  bits, with  $m < n$ .
  - This hash value is also called a digest (of the original message).
  - Since  $n > m$ , there exist many  $X$  which are map to the same digest → collision.
- Applications
  - Digital signatures
  - Message authentication



# DS schemes with hash functions



# Main properties

Given a hash function  $H: X \rightarrow Y$

- Long message  $\rightarrow$  short, fixed-length hash
- One-way property: given  $y \in Y$   
it is computationally infeasible to find a value  $x \in X$  s.t.  
 $H(x) = y$
- Collision resistance (collision-free)  
it is computationally infeasible to find any two  
distinct values  $x', x \in X$  s.t.  $H(x') = H(x)$ 
  - This property prevent against signature forgery

# Collisions

- Avoiding collisions is theoretically impossible
  - Dirichlet principle:  $n+1$  rabbits into  $n$  cages  $\rightarrow$  at least 2 rabbits go to the same cage
  - This suggest exhaustive search: try  $|Y|+1$  messages then must find a collision ( $H:X \rightarrow Y$ )
- In practice
  - Choose  $|Y|$  large enough so exhaustive search is computational infeasible.
    - $|Y|$  not too large or long signature and slow process
  - However, collision-freeness is still hard

# Birthday attack

- Can hash values be of 64 bits?
  - Look good, initially, since a space of size  $2^{64}$  is too large to do exhaustive search or compute that many hash values
  - However a birthday attack can easily break a DS with a 64-bit hash function
    - In fact, the attacker only need to create a bunch of  $2^{32}$  messages and then launch the attack with reasonably high probability for success.

# How is the attack

- Goal: given  $H$ , find  $x, x'$  such that  $H(x)=H(x')$
- Algorithm:
  - pick a random set  $S$  of  $q$  values in  $X$
  - for each  $x \in S$ , computes  $h_x = H(x)$
  - if  $h_x = h_{x'}$  for some  $x' \neq x$  then collision found:  $(x, x')$ , else fail
- The average success probability is  $\varepsilon = 1 - \exp(q(q-1)/2|Y|)$ 
  - Suppose  $Y$  has size  $2^m$ , choose  $q \approx 2^{m/2}$  then  $\varepsilon$  is almost 0.5!

# Birthday paradox

- Given a group of people, the minimum number of people

- such that two will share the same birthday with probability at least 50%

is only 23 → why “paradox”

- Computing the chance

- $1 - (1 - 1/365)(1 - 2/365) \dots (1 - 22/365) = 1 - 0.493 = 0.507$

# Birthday paradox

- Given hash function with the size of  $M$ , pickup a random set of plaintext with the size of  $q$ , then the probability of having two texts with the same hash code is  $1 - e^{\frac{-q(q-1)}{2M}}$
- Proof
  - Probability for all texts have different has code
    - $1 \times \frac{M-1}{M} \times \dots \times \frac{M-(q-1)}{M} = \left(1 - \frac{1}{M}\right) \dots \left(1 - \frac{q-1}{M}\right)$
  - Probability of having two texts with the same hash code
    - $1 - \left(1 - \frac{1}{M}\right) \dots \left(1 - \frac{q-1}{M}\right) \approx 1 - e^{\left(-\frac{1}{M}\right) \times \left(-\frac{2}{M}\right) \times \dots \times \left(-\frac{q-1}{M}\right)} = 1 - e^{\frac{-q(q-1)}{2M}}$

# Birthday paradox

- Given hash function with the size of  $M$ , pickup a random set of plaintext with the size of  $q$ , then the probability of having two texts with the same hash code is  $1 - e^{\frac{-q(q-1)}{2M}}$ 
  - With  $q \approx \sqrt{2M \ln \frac{1}{1-\varepsilon}}$ , then the probability is approximately equals to  $1 - \varepsilon$
  - With  $q \approx 1.17\sqrt{M}$ , then the probability is approximately equals to 0.5



# MAC: message authentication code

- Hash function is public and the key shared between the sender and the receiver is secret
  - Sender computes  $\text{mac1} = \text{MAC}(M, H, K)$  and sends it along with the message  $M$
  - Receiver computes  $\text{mac2} = \text{MAC}(M, H, K)$  and checks if  $\text{mac1} = \text{mac2}$  ? Yes → the message is authentic; no => reject it
- The output of MAC can not be produced without knowing the secret key
  - So, this mechanism provides data integrity and ***source authentication***