### BÀI 3. XÁC THỰC THÔNG ĐIỆP

Bùi Trọng Tùng, Viện Công nghệ thông tin và Truyền thông, Đại học Bách khoa Hà Nội

1

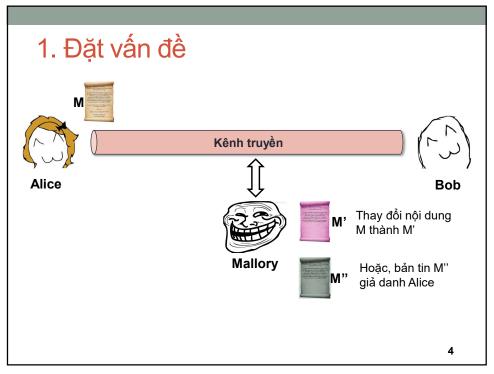
1

### Nội dung

- Các vấn đề xác thực thông điệp
- Mã xác thực thông điệp (MAC)
- · Hàm băm và hàm băm HMAC

2

# 1. ĐẶT VẤN ĐỀ



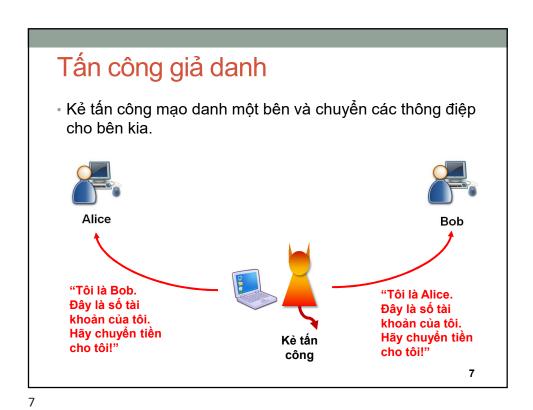
### Xác thực thông điệp

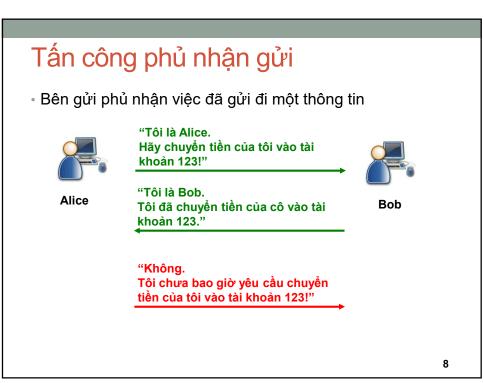
- · Bản tin phải được xác minh:
  - ⊳Nội dung toàn vẹn: bản tin không bị sửa đổi
    - √Bao hàm cả trường hợp Bob cố tình sửa đổi
  - ≻Nguồn gốc tin cậy:
    - √Bao hàm cả trường hợp Alice phủ nhận bản tin
    - √Bao hàm cả trường hợp Bob tự tạo thông báo và "vu khống" Alice tạo ra thông báo này
  - >Đúng thời điểm
  - →Các dạng tấn công điển hình vào tính xác thực: Thay thế (Substitution), Giả danh (Masquerade), tấn công phát lại (Replay attack), Phủ nhận (Repudiation)

5

5

### Tấn công thay thế Chặn thông điệp, thay đổi nội dung và chuyển tiếp cho bên kia "Tôi là Alice. Số tài khoản của tôi là 123. Hãy chuyển tiền cho tôi!" Kênh truyền Alice Bob "Tôi là Alice. Số tài khoản của tôi là 456. Hãy chuyển tiền cho tôi!" Kẻ tấn công 6





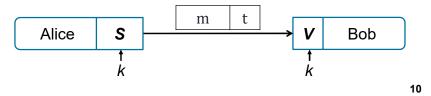
### 2. MÃ XÁC THỰC THONG ĐIỆP (MAC)

9

9

### Message Authentication Code

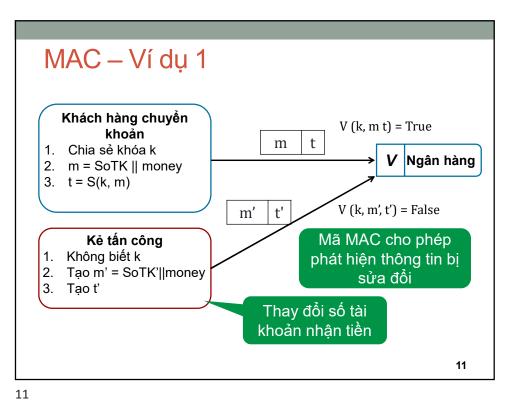
- Hai bên đã trao đổi một cách an toàn khóa mật k
- · Hàm MAC = (S, V) là một cặp thuật toán
- Sinh mã: t = S(k, m)
  - >Đầu ra: kích thước cố định, không phụ thuộc kích thước của m
- Xác minh: V(k, m, t)
  - >Nếu t = S(k, m) thì V = true
  - ≻Ngược lại V = false



### Slide 9

Rollback resistance Rollback attack

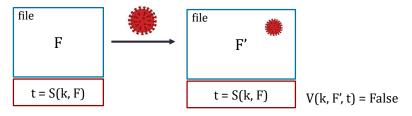
Bui Trong Tung, 21/01/2021



\_\_

### MAC – Ví dụ 2: Phần mềm Tripwire

· Khi cài đặt, tính giá trị MAC của các file cần bảo vệ



Khi máy tính khởi động, các file được kiểm tra mã MAC
 Tho phép phát hiện các file bị sửa đổi (ví dụ do nhiễm virus)

12

### Đặc tính của MAC

- Tính xác định: nếu m = m' → S(k, m) = S(k, m')
- Tính đúng đắn: V(k, m, S(k,m)) = True
- Tính hiệu quả: tính toán được với mọi m trong thời gian chấp nhận được
- Tính an toàn: nếu đối phương không biết k
  - ≻Không thể tính được tag = S(k, m)
  - ≻Không thể tìm được m sao cho S(k, m) = tag cho trước
  - ⊳Không thể tìm được m và m' sao cho S(k, m) = S(k, m')

13

13

### An toàn của MAC

thì xác suất P(b = true) ≤ ε

# Thử thách 1. Sinh khóa k 3. Tính $t_i = S(k, m_i)$ $\xrightarrow{m_1, ..., m_q}$ $\xrightarrow{t_1, ..., t_q}$ $\xrightarrow{m_t, ..., t_q}$ $\xrightarrow{m_t, ..., t_q}$ $\xrightarrow{m_t, ..., t_q}$ $\xrightarrow{m_t, ..., t_q}$

Tấn công

- 2. Chọn m<sub>1</sub>, ..., m<sub>a</sub>
- 4. Chọn m và sinh t sao cho  $(m,t) \notin \{(m_1,t_1),...,(m_q,t_q)\}$
- ↓b = {true,false}

   MAC là an toàn nếu với mọi thuật toán tấn công hiệu quả
- →kẻ tấn công không thể tạo giá trị t hợp lệ nếu không có khóa k

14

14

5. b = V(k, m, t)

### An toàn của MAC

- MAC còn an toàn không nếu tồn tại thuật toán hiệu quả cho một trong các tình huống sau:
- (1) Tìm được m\* sao cho S(?, m\*) = t với t chọn trước
- (2) Tìm được m\* sao cho S(?, m\*) = S(?, m) với m chọn trước
- (3) Tìm được m và m\* sao cho  $S(?, m^*) = S(?, m)$
- · Hoặc giá trị t có kích thước 10 bit

15

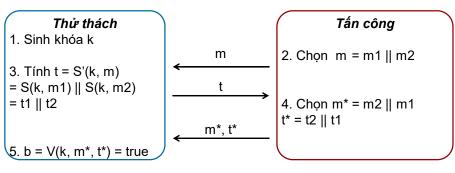
15

### Một ví dụ khác

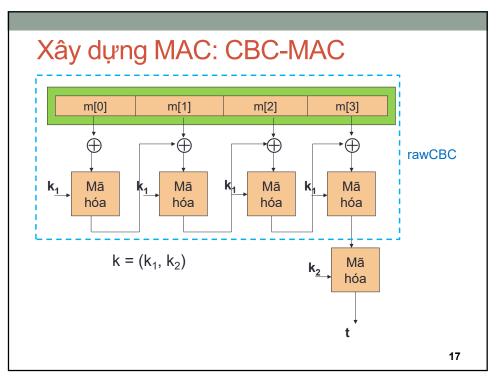
Một hàm MAC được tạo ra như sau:

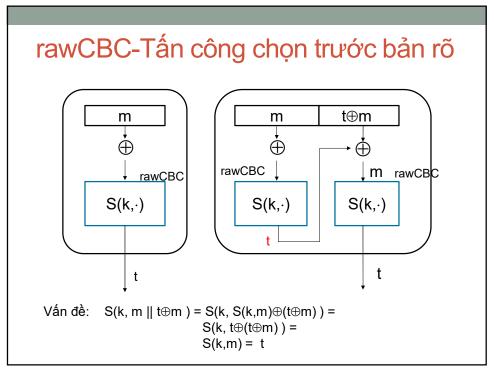
S'(k, m1 || m2) = S(k, m1) || S(k, m2)

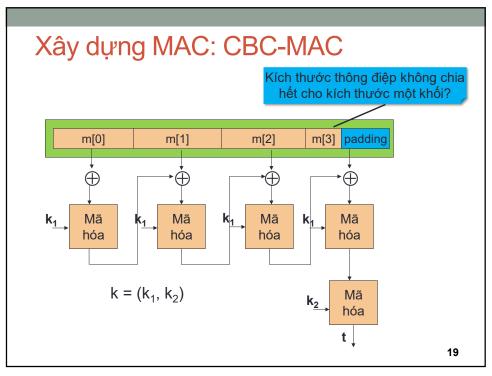
Trong đó S là hàm tạo mã MAC an toàn. S' có phải là hàm MAC an toàn hay không?

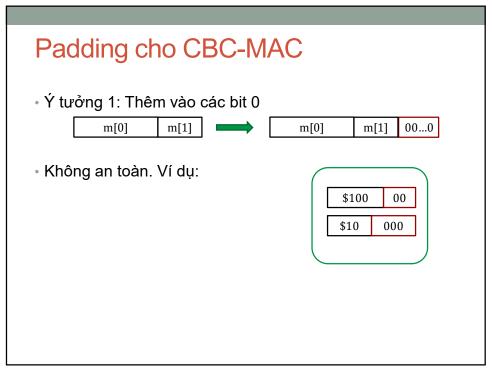


16



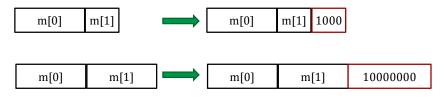






### Padding cho CBC-MAC

- Yêu cầu: M<sub>i</sub> ≠ M<sub>i</sub> thì pad(M<sub>i</sub>) ≠ pad(M<sub>i</sub>)
- Chuẩn ISO/IEC 9797-1:
  - >Sử dụng chuỗi padding bắt đầu bởi bit 1
  - Nếu kích thước thông điệp là bội số kích thước của khối, luôn thêm
     1 khối padding



21

### Độ an toàn của CBC-MAC

- Giả sử  $m_1$  và  $m_2$  là hai bản tin có mã MAC giống nhau:  $S(k, m_1) = S(k, m_2)$
- $\rightarrow$ S(k, m<sub>1</sub>||W) = S(k, m<sub>2</sub>||W) với W bất kỳ
- Kịch bản tấn công:
- 1. Kẻ tấn công nhận được  $t_x = S(k, m_x)$  với x = 1, ..., N
- 2. Tìm cặp bản tin  $(m_i, m_j)$  có  $t_i = t_j$ . Nếu không tìm thấy thực hiện lại bước 1
- 3. Chọn bản tin W và tính  $t = S(k, m_i || W)$
- 4. Thay m<sub>i</sub> || W bằng m<sub>i</sub> || W có lợi cho kẻ tấn công

### Ví dụ tấn công vào tính đụng độ

(1) Kẻ tấn công(Mr. Tung) tìm được 2 bản tin có mã MAC giống nhau:

m1: 'I will pay 1' m2: 'I will pay 2'

Chọn W = '000\$ to Mr.Tung'

m1 || W = 'I will pay 1000\$ to Mr.Tung'

m2 || W = 'I will pay 2000\$ to Mr.Tung'

(2) Đánh lừa người dùng gửi bản tin 'I will pay 1000\$ to Mr.Tung' || S(k, 'I will pay 1000\$ to Mr.Tung') cho ngân hàng

(3) Thay thế bằng 'I will pay 2000\$ to Mr.Tung' || S(k, 'I will pay 1000\$ to Mr.Tung') → Ngân hàng chấp nhận

23

23

### An toàn của CBC-MAC

- Khóa được dùng nhiều lần → giảm độ an toàn
- Néu gọi:
  - ⊳q: số bản tin được tính MAC cùng với khóa không đổi
  - ≻|X|: Số lượng giá trị có thể của t
- Xác suất tấn công thành công ≤ 2\*q² / |X|
- Để xác suất tấn công là không đáng kể (≤ 2-80) thì sau bao nhiều lần tính MAC phải đổi khóa?
- · Ví dụ: Sử dụng AES-CBC-MAC:

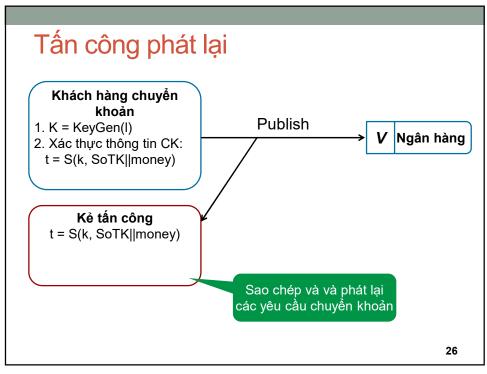
$$2*q^2/2^{128} \le 2^{-80} \rightarrow q = ?$$

### Tấn công phát lại (Replay attack)

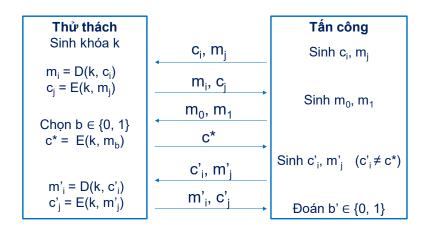
- Kẻ tấn công phát lại bản tin M đã được chứng thực trong phiên truyền thông trước đó
- Thiết kế MAC không chống tấn công phát lại
- → cần thêm các yếu tố chống tấn công phát lại trong các giao thức truyền thông sử dụng MAC
- Một số kỹ thuật chống tấn công phát lại:
  - ⊳Giá trị dùng 1 lần(nonce): S(k, m || nonce)
  - ➤ Tem thời gian: S(k, m || timestamp)

25

25



### Tấn công CCA - Nhắc lại



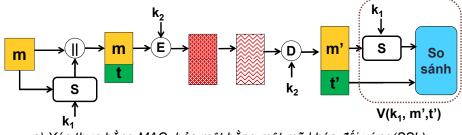
Hệ mật chống lại được tấn công CCA (độ an toàn IND-CCA)
 nếu với mọi thuật toán tấn công hiệu quả thì P(b' = b) ≤ ½ + ε

27

27

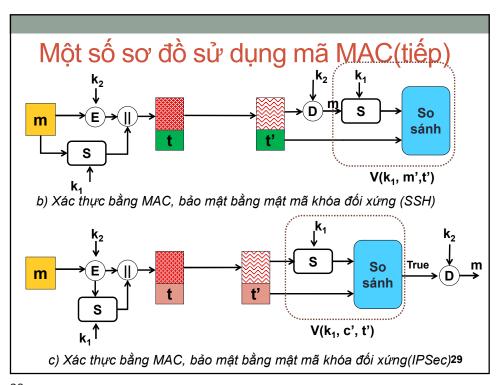
### Một số sơ đồ sử dụng mã MAC

- Các sơ đồ mật mã đã xem xét không chống lại được tấn công CCA(chosen-cipher attack)
  - Cách thức chung: kẻ tấn công sửa bẩn mã c\* thành c'<sub>i</sub> và yêu cầu giải mã
- Mật mã có xác thực: sơ đồ mật mã đảm bảo đồng thời tính bí mật và toàn vẹn
- Cách thức 1: Sử dụng mã MAC



a) Xác thực bằng MAC, bảo mật bằng mật mã khóa đối xứng(SSL)

28



### Nhận xét

### Sơ đồ a

- Xác thực toàn vẹn bản rõ
- Không xác thực toàn vẹn bản mật(không phát hiện tấn công thay thế bản mật)
- Không có thông tin về bản rõ từ MAC
- Chỉ đảm bảo an toàn CCA khi mã ở chế độ rand-CBC hoặc rand-CTR

### Sơ đồ b

- Xác thực toàn vẹn bản rõ
- Không xác thực toàn vẹn bản không phát hiện bản mật bị thay thế)
- MAC chứa thông tin bản rõ
- Có thể giảm sự an toàn mã mật

### Sơ đồ c

- Xác thực toàn ven bản rõ
- Xác thực toàn vẹn bản mật(có thể phát hiện bản mật bị thay thế)
- MAC không chứa thông tin bản rõ
- Luôn đảm bảo an toàn CCA

### Cách thức 2: Thêm dữ liệu xác thực

- · Authenticated encryption with additional data (AEAD)
- Là các phương pháp mật mã
   Hàm mã hóa E: K x M x N → C

Từ chối giải mã các bản mã không hợp lệ

Hàm giải mã D:  $K \times C \times N \rightarrow M \cup \{\bot\}$ 

- Trong đó N là một dữ liệu được thêm vào để xác thực toàn vẹn của bản tin
- Một số chuẩn:
  - > GCM: Mã hóa ở chế độ CTR sau đó tính CW-MAC
  - > CCM: Tính CBC-MAC sau đó mã hóa ở chế độ CTR (802.11i)
  - > EAX: Mã hóa ở chế độ CTR sau đó tính CMAC

31

31

3.HÀM BĂM

### Khái niệm

- · Hàm băm H: thực hiện phép biến đổi:
  - > Đầu vào: bản tin có kích thước bất kỳ
  - Đầu ra: giá trị digest h = H(m)có kích thước n bit cố định (thường nhỏ hơn rất nhiều so với kích thước bản tin đầu vào)
  - >Tính đúng đắn: đầu vào giống nhau thì đầu ra giống nhau
- Hàm băm mật mã: Không thể đoán được sự thay đổi của đầu ra khi thay đổi đầu vào
  - · Chỉ thay đổi 1 bit đầu vào, làm thay đổi hoàn toàn giá trị đầu ra
  - Ví dụ:
    - ➤ Đầu vào: "The quick brown fox jumps over the lazy dog"
    - Mã băm: 2fd4e1c67a2d28fced849ee1bb76e7391b93eb12
    - > Đầu vào: "The quick brown fox jumps over the lazy cog"
    - Dàu ra: de9f2c7fd25e1b3afad3e85a0bd17d9b100db4b3

33

33

### Một hàm băm đơn giản

- Chia thông điệp thành các khối có kích thước n-  $_{m\,=\,}$ 
  - ▶ Padding nếu cần
- Thực hiện XOR tất cả các khối → mã băm có kích thước n bit
- Tất nhiên, hàm băm này không đủ an toàn để sử dụng trong bài toán xác thực thông điệp

### Yêu cầu đối với hàm băm mật mã

- 1. Có thể áp dụng với thông điệp m với độ dài bất kỳ
- 2. Tạo ra giá trị băm h có độ dài cố định
- 3. H(m) dễ dàng tính được với bất kỳ m nào
- 4. Từ h rất khó tìm được m sao cho h = H(m): tính một chiều
- 5. Biết trước  $m_1$  rất khó tìm được  $m_2$  sao cho  $H(m_1) = H(m_2) \rightarrow$  tính chống đụng độ yếu
- 6. Rất khó tìm được cặp  $(m_1, m_2)$  sao cho  $H(m_1)=H(m_2) \rightarrow$  tính chống đụng độ mạnh

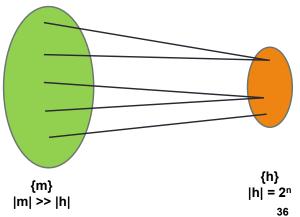
35

35

### Tấn công vào tính 1 chiều

- Mã băm có kích thước n bit
- Cho trước h có kích thước n bit → tìm được m sao cho H(m) = h

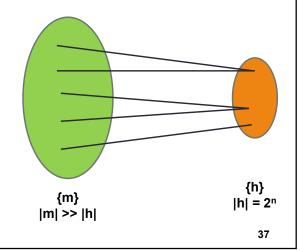
Chọn ra 2<sup>n</sup> bản tin ngẫu nhiên và tính mã băm của chúng.



## Tấn công vào tính đụng độ yếu

- · Mã băm có kích thước n bit
- Cho trước m1 → tìm được m2 sao cho H(m2) = H(m1)

Chọn ra 2<sup>n</sup> bản tin m2 ngẫu nhiên và tính mã băm của chúng.

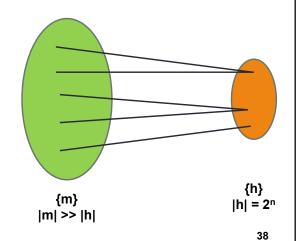


37

### Tấn công vào tính đụng độ mạnh

- · Mã băm có kích thước n bit
- Tìm được m1, m2 sao cho H(m2) = H(m1)

Chọn 2<sup>n</sup> + 1 bản tin ngẫu nhiên và tính mã băm của chúng.



### Ví dụ

- Giả sử kẻ tấn công có khả năng tính toán 1.000.000 mã băm mỗi giây. Kích thước mã băm là bao nhiêu để kẻ tấn công thực hiện 100 năm, xác suất thành công cao nhất là là 2<sup>-80</sup>
- Lời giải:
  - ▶Kích thước mã băm: n bit → số mã băm có thể là 2<sup>n</sup>
  - >Số mã băm kẻ tấn công tính được trong 100 năm là: X
  - ≻Thỏa mãn: X/2<sup>n</sup> <= 2<sup>-80</sup>

39

39

### Một số hàm băm phổ biến

- MD5
  - > Kích thước digest: 128 bit
  - Công bố thuật toán tấn công đụng độ (collision attack)
     vào 1995
  - > Đã bị tấn công thành công vào năm 2005
- SHA-1
  - >Kích thước digest: 160 bit
  - ▶Đã bị tấn công thành công vào năm 2017
  - ≻Hết han vào năm 2030
- SHA-2: 224/256/384/512 bit
- SHA-3: 224/256/384/512 bit

40

### MD5

- Bước 1: Padding dữ liệu sao cho bản tin đầu vào có độ dài L sao cho L mod 512 = 448
- Bước 2: Biểu diễn độ dài của dữ liệu ban đầu dưới dạng 64 bit. Thêm giá trị độ dài này vào khối dữ liệu.
  - Coi dữ liệu là một chuỗi các khối 512 bit: Y<sub>0</sub>, Y<sub>1</sub>, ..., Y<sub>K-1</sub>
  - Hoặc là một chuỗi các khối 32 bit : m<sub>0</sub>, m<sub>1</sub>, ...,m<sub>N</sub>
- Bước 3: Khởi tạo các giá trị hằng số A, B, C, D
  - >A = 0x67 45 23 01
  - >B = 0xEF CD AB 89
  - >C = 0x98 BA DC FE
  - >D = 0x10 32 54 76

41

### MD5

 Bước 4: Thực hiện vòng lặp xử lý các khối 512 bit

$$CV_{q+1} = F(Y_q, CV_q)$$

- >Xử lý khối dữ liệu Y<sub>q</sub>: thực hiện 4 vòng lặp. Mỗi vòng lặp áp dụng hàm nén với T[1..64] là mảng hằng số xác định trước
- ≻Cộng modulo 2³² mỗi khối với giá trị CV<sub>q</sub> để có CV<sub>q+1</sub>
- Bước 5: Kết quả xử lý khối 512 bit cuối cùng là giá trị băm của thông điệp

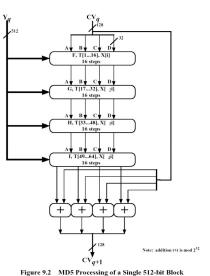


Figure 9.2 MD5 Processing of a Single 512-bit Block (MD5 Compression Function)

### Hàm nén trong MD5 (Tham khảo)

- Đầu vào:
  - CV: Khối 128 bit
  - ⊳M<sub>i</sub>: khối dữ liệu 32-bit
  - ≻T<sub>i</sub>: Hằng số
- ⊕ Cộng modulo 2<sup>32</sup>
- <<<s: dich trái s bit</p>
- ∧: AND, v: OR, ¬: NOT

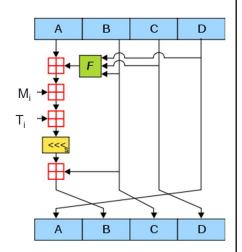
 $F1 = (B \land C) \lor (\neg B \land D)$ 

 $F2 = (B \land D) \lor (C \land \neg D)$ 

 $F3 = B \oplus C \oplus D$ 

 $F4=C \oplus (B \vee \neg D)$ 

Thực hiện vòng lặp 16 bước



43

### SHA-1

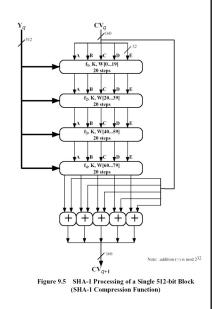
- Bước 1: Padding dữ liệu sao cho bản tin đầu vào có độ dài L sao cho L mod 512 = 448
- Bước 2: Biểu diễn độ dài của dữ liệu ban đầu dưới dạng 64 bit. Thêm giá trị độ dài này vào khối dữ liệu.
  - Coi dữ liệu là một chuỗi các khối 512 bit: Y<sub>0</sub>, Y<sub>1</sub>, ..., Y<sub>K-1</sub>
  - Hoặc là một chuỗi các khối 32 bit :  $m_0, m_1, ..., m_N$
- · Bước 3: Khởi tạo các giá trị hằng số A, B, C, D, E
  - A = 0x67 45 23 01
  - >B = 0xEF CD AB 89
  - >C = 0x98 BA DC FE
  - >D = 0x10 32 54 76
  - >E = 0xC3 D2 E1 F0

### SHA-1

 Bước 4: Thực hiện vòng lặp xử lý các khối 512 bit

$$CV_{q+1} = F(Y_q, CV_q)$$

- Xử lý khối dữ liệu Y<sub>q</sub>: thực hiện 4 vòng lặp. Mỗi vòng lặp áp dụng hàm nén với K là hằng số xác đinh trước
- Cộng modulo 2<sup>32</sup> mỗi khối với giá trị CV<sub>q</sub> để có CV<sub>q+1</sub>
- Bước 5: Kết quả xử lý khối 512 bit cuối cùng là giá trị băm của thông điệp



45

### Hàm nén trong SHA-1 (Tham khảo)

- Đầu vào:
  - CV: Khối 160 bit
  - W<sub>t</sub>: Khối dữ liệu mở rộng 32 bit
  - K₁: Hằng số
- ⊕Cộng modulo 2<sup>32</sup>
- <<<5(30): dich trái 5(30) bit</p>
- ∧: AND, v: OR, ¬: NOT

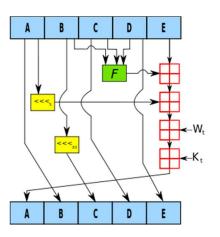
 $F1 = (B \land C) \lor (\neg B \land D)$ 

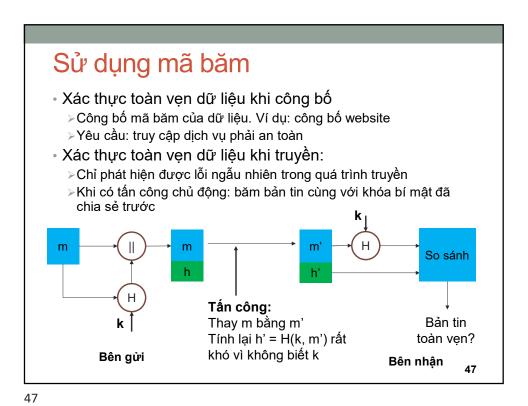
 $F2 = B \oplus C \oplus D$ 

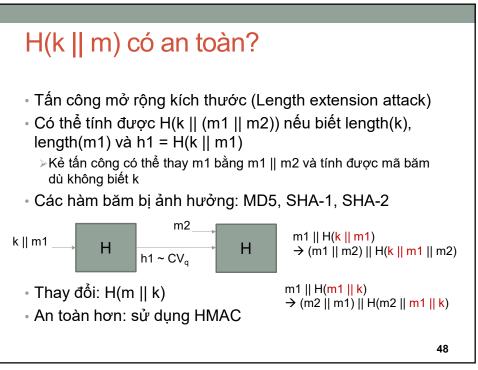
 $F3 = (B \land C) \lor (B \land D) \lor (C \land D)$ 

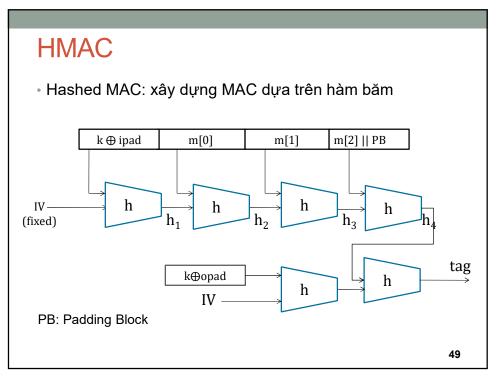
 $F4 = B \oplus C \oplus D$ 

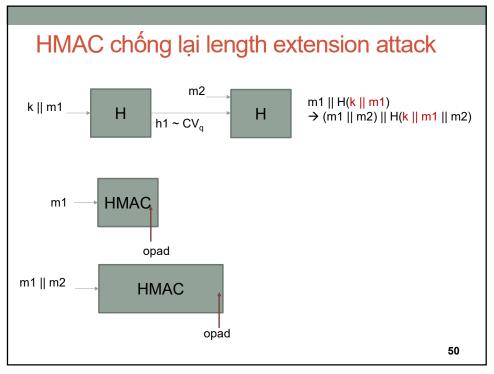
Thực hiện vòng lặp 20 bước











### HMAC và MAC

- HMAC có tính chống đụng độ chắc chắn hơn do sử dụng hàm băm
- Tốc đô tính toán của HMAC nhanh hơn
- Kích thước giá trị tag được tạo bởi HMAC lớn hơn → an toàn hơn trước các tấn công

51

51

### Tấn công tính đụng độ của hàm băm

- Đụng độ trong hàm băm: tồn tại m₁ ≠ m₂ mà
   H(m₁) = H(m₂)
- Nhắc lại: sự tồn tại các bản tin đụng độ dẫn đến các nguy cơ tấn công vào sơ đồ xác thực
- Tìm ra bản tin đụng độ: có thể thực hiện vét cạn
   Số bản tin cần tính tối thiểu là bao nhiêu sẽ chắc chắn thành công?

### Nghịch lý ngày sinh (Birthday paradox)

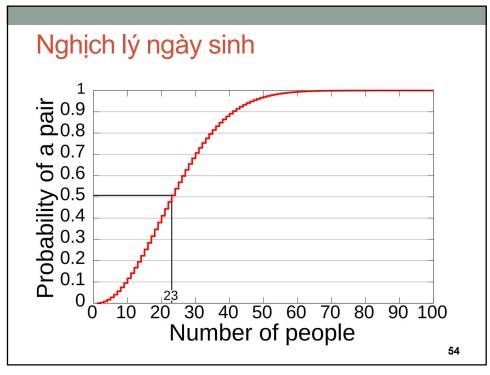
- Bài toán: Khi chọn n người bất kỳ, xác suất để có tối thiểu
   2 người có trùng ngày sinh là bao nhiêu?
- Số cách chọn ra n người bất kỳ: 365<sup>n</sup>
- Số cách chọn ra n người không có cặp nào trùng ngày sinh: 365 x 364 x ... x (365-(n-1)) = C<sup>n</sup><sub>365</sub>
- Xác suất để chọn ra n người không có cặp nào trùng ngày sinh

$$Q = \frac{365 \times 364 \times \dots \times (365 - (n-1))}{365^n}$$

- Xác suất cần tính: P = 1 Q
- n = ? để P > 0.5 (cứ 2 lần chọn thì có 1 lần thỏa mãn)

53

53



# Tấn công dựa trên nghịch lý ngày sinh (Birthday paradox attack)

- Kiểm tra  $2^{n/2}$  bản tin có xác suất tìm ra các bản tin đụng độ là  $\sim 0.5$
- Cách thức tấn công:
  - ⊳Bước 1: Chọn ra 2<sup>n/2</sup> bản tin ngẫu nhiên
  - >Bước 2: Tính mã băm/MAC cho các bản tin
  - >Bước 3: Kiểm tra sự tồn tại của các bản tin đụng độ. Nếu không có, quay lại bước 1.
  - ≻Kỳ vọng: thành công sau 2 lần thử

55

55

CASE STUDY: BẢO MẬT TRÊN IPHONE

### Bảo mật trên iPhone

- Triết lý bảo mật của Apple:
  - ▶Nếu điện thoại trong tay bạn, bạn có thể truy cập mọi thứ
  - ⊳Nếu điện thoại trong tay người khác, nó chỉ là một "cục gạch"
- Apple triển khai tất cả các cơ chế bảo mật trên chip Secure Enclave Processor (SEP)
  - > Secure Enclave Processor là TCB của thiết bị
- Mọi phần còn lại của thiết bị được coi là không tin cậy
  - ⊳Bộ nhớ là không tin cậy, vì vậy mọi dữ liệu được mã hóa
  - ▶CPU phải gửi yêu cầu Secure Enclave để giải mã
  - ▶Một vài dữ liệu bí mật chỉ có thể đọc bởi Secure Enclave

57

57

### Bảo mật trên iPhone

- Khóa chủ K<sub>phone</sub> được sử dụng để mã hóa các khóa mật mã khác
- K<sub>phone</sub> được mã hóa bởi K<sub>user</sub>, sinh từ mật khẩu của người dùng
  - >Kuser = PBKDF(password, hardcode)
  - >hardcode có kích thước 256 bit được sinh và lưu trữ trên chip SEP
- Làm cách nào kẻ tấn công có thể đánh cắp được dữ liệu mà không có mật khẩu người dùng?