TRƯỜNG ĐẠI HỌC SÀI GÒN

KHOA CÔNG NGHỆ THÔNG TIN

Tiểu luận học phần Seminar chuyên đề

TÌM HIỂU HÀM BĂM VÀ CHỮ KÝ SỐ

Sinh viên thực hiện:

Vũ Thị Hồng Xương 3118410492

GVHD: TS. PHAN TẤN QUỐC

Thành phố Hồ Chí Minh, tháng 5 năm 2022

NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Thành phố Hồ Chí Minh, ngày tháng 5 năm 2022

Giảng viên hướng dẫn

MỤC LỤC

Nội dung

MỤC LỤC HÌNH ẢNH

LỜI MỞ ĐẦU

Ngày nay trong mọi hoạt động của con người thông tin đóng một vai trò quan trọng không thể thiếu. Xã hội càng phát triển nhu cầu trao đổi thông tin giữa các thành phần trong xã hội ngày càng lớn. Mạng máy tính ra đời đã mang lại cho con người rất nhiều lợi ích trong việc trao đổi và xử lý thông tin một cách nhanh chóng và chính xác. Chính từ những thuận lợi này đã đặt ra cho chúng ta một câu hỏi, liệu thông tin đi từ nơi gửi đến nơi nhận có đảm bảo tuyệt đối an toàn không. Thông tin được lưu giữ, truyền dẫn trên mạng lưới thông tin công cộng có thể bị nghe trộm, chiếm đoạt hoặc ăn cắp dẫn đến sự tổn thất. Đặc biệt là đối với những số liệu của hệ thống ngân hàng, hệ thống thương mại, cơ quan quản lý của chính phủ hoặc thuộc lĩnh vực quân sự được lưu giữ và truyền dẫn trên mạng. Nếu như thông tin đưa lên mạng máy tính không an toàn thì hiệu suất làm việc sẽ bị ảnh hưởng rất nhiều. Trước các yêu cầu cần thiết đó, việc mã hoá thông tin sẽ đảm bảo an toàn cho thông tin tại nơi lưu trữ cũng như khi thông tin được truyền trên mạng.

Chữ ký số ra đời để giải quyết vấn về an toàn thông tin trên mạng máy tính. Với đặc điểm là đơn giản cho người sử dụng mà vẫn đảm bảo được tính bảo mật, kỹ thuật sử dụng chữ ký số là một trong những kỹ thuật được sử dụng phổ biến, đa dạng trong hầu hết các lĩnh vực, nhất là Tài chính, Ngân hang, Kế toán… Do đó, tôi đã chọn đề tài tìm hiểu về hàm băm và chữ ký số.

Chương 1: Giới thiệu

Chữ ký số là một cơ chế mật mã hóa được sử dụng để kiểm tra độ chân thực và tính toàn vẹn của dữ liệu số. Có thể xem nó như là một phiên bản kỹ thuật số của các chữ ký bằng tay thông thường, nhưng với mức độ phức tạp và bảo mật cao hơn.

Nói một cách đơn giản, chúng ta có thể mô tả chữ ký số như một mã được đính kèm với tin nhắn hoặc tài liệu. Sau khi được tạo, mã này đóng vai trò như một bằng chứng cho thấy tin nhắn không bị giả mạo trong quá trình chuyển từ người gửi sang người nhận.

Trong bài tiểu luận này tôi sẽ tìm hiểu về hàm băm và chữ ký số, nguyên tắc và xây dựng một số hàm băm và chữ ký số, bài tiểu luận gồm 4 chương và có bố cục như sau:

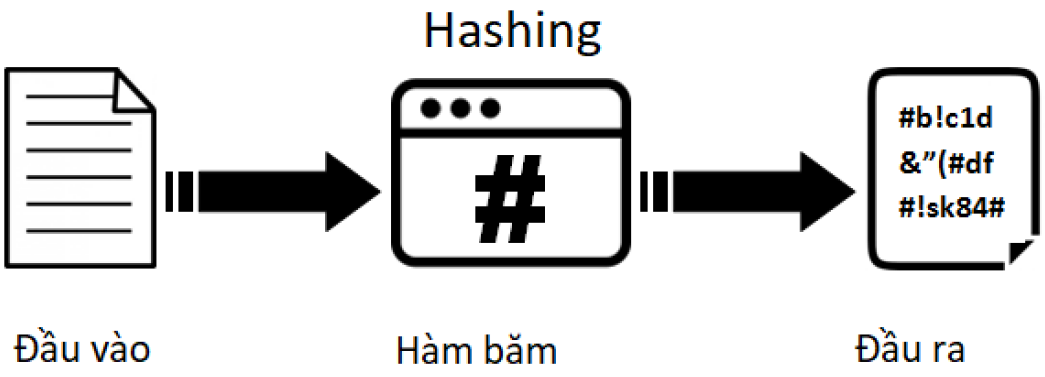
Chương 1: Giới thiệu
Chương 2: Tìm hiểu hàm băm
Chương 3: Tìm hiểu chữ ký số
Chương 4: Kết luận
Mặc dù khái niệm bảo mật thông tin liên lạc bằng mật mã đã có từ thời cổ đại, hệ thống chữ ký số đã trở thành hiện thực vào những năm 1970 - nhờ vào sự phát triển của Mật mã khóa công khai (PKC). Vì vậy, để tìm hiểu chữ ký số hoạt động như thế nào, ta cần hiểu những điều cơ bản về hàm băm và mật mã khóa công khai.

Chương 2. Tìm hiểu hàm băm

2.1. Giới thiệu:

Hàm băm (Hash function) là một hàm toán học chuyển đổi một thông điệp đầu vào có độ dài bất kỳ thành một dãy bit có độ dài cố định (tuỳ thuộc vào thuật toán băm). Dãy bit này được gọi là thông điệp rút gọn (message digest) hay giá trị băm (hash value), đại diện cho thông điệp ban đầu.

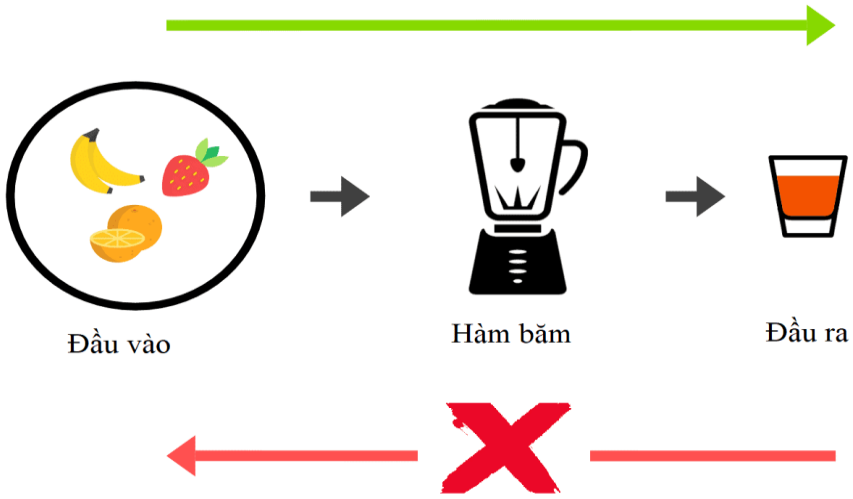
Hàm băm (hash function) là hàm một chiều mà nếu đưa một lượng dữ liệu bất kì qua hàm này sẽ cho ra một chuỗi có độ dài cố định ở đầu ra.



Hình 2.1. Mô tả hàm Băm.

2.1.1. Tính chất cơ bản của hàm Hash

- **Tính một chiều:** không thể suy ra dữ liệu ban đầu từ kết quả, điều này tương tự như việc bạn không thể chỉ dựa vào một dấu vân tay lạ mà suy ra ai là chủ của nó được.



Hình 2.2. Mô tả tính một chiều của hàm Băm (ảnh: cryptoviet.com)

- **Tính duy nhất**: xác suất để có một vụ va chạm (hash collision), tức là hai thông điệp khác nhau có cùng một kết quả hash là cực kỳ nhỏ

2.1.2. Danh sách các hàm băm mật mã học

Thuật toán	Kích thước đầu ra	Kích thước trạng thái trong	Kích thước khối	Độ dài	Kích thước world	Xung đột
HAVAL	256/224/192/160/128	256	1024	64	32	Có
MD2	128	384	128	Không	8	khả năng lớn
MD4	128	128	512	64	32	Có
MD5	128	128	512	64	32	Có
PANAMA	256	8736	256	No	32	Có lỗi
RIPEMD	128	128	512	64	32	Có
RIPEMD-128/256	128/256	128/256	512	64	32	Không
RIPEMD-160/320	160/320	160/320	512	64	32	Không

SHA-0	160	160	512	64	32	Không
SHA-1	160	160	512	64	32	Có lỗi
SHA-256/224	256/224	256	512	64	32	Không
SHA-512/384	512/384	512	1024	128	64	Không
Tiger(2)-192/160/128	192/160/128	192	512	64	64	Không
VEST-4/8 (hash mode)	160/256	256/384	8	80/128	1	Không
VEST-16/32 (hash mode)	320/512	512/768	8	160/256	1	Không
WHIRLPOOL	512	512	512	256	8	Không

Bảng 1. Một số hàm băm mật mã (wikipedia).

Trong các hàm băm trên hàm SHA-1 là một trong những hàm được sử dụng rộng rãi nhất ở Việt Nam.

2.2. Các yêu cầu đối với hàm băm

Mục đích sử dụng hàm băm là nhận được đặc trưng của file tin tức.

Để chứng tỏ lợi ích đối với việc xác thực tin tức, hàm băm H cần phải có các đặc tính sau:

1. Được ứng dụng với khối số liệu có độ dài bất kỳ.

2. Đưa ở đầu ra một giá trị có độ dài cố định

3. Giá trị H(x) phải cần tính toán tương đối dễ dàng đối với bất kỳ giá trị x đã cho nào đó, cần phải cho khả năng thực hiện cả bằng thiết bị và cả bằng chương trình.

4. Đối với bất kỳ giá trị h nào, cần phải trong thực tế không có khả năng tính x, để H(x) = h. Đặc tính như thế đôi khi còn gọi là đặc tính một chiều.

5. Đối với x đã cho bất kỳ, thực tế cần phải không có khả năng tính $y \neq x$, để cho H(x) = H(y). Đặc tính như thế đôi khi người ta còn gọi là chống va chạm yếu.

6. Cần phải trong thực tiễn không có khả năng tính cặp giá trị bất kỳ khác nhau x và y, đối với chúng có H(x) = H(y). Đặc tinh như thế đôi khi còn gọi là chống va chạm mạnh.

Ba đặc tính đầu tiên được nêu ở trên, bảo đảm khả năng thực tiễn ứng dụng hàm băm để xác thực tin tức. Đặc tính thứ bốn bảo đảm tính một chiều: dễ dàng nhận được mã hàm băm h trên cơ sở của tin tức đã cho, nhưng thực tiễn không có khả năng khôi phục lại tin tức, khi chỉ có h phù hợp. Đặc tính thứ năm bảo đảm rằng không có người tham gia nào có thể tìm được một bản tin khác, có cùng một kết quả mã hàm băm như tin tức đã cho, điều này ngăn ngừa khả năng làm giả tin tức, trong trường hợp khi thực hiện mã hoá hàm băm Đặc tính thứ sáu xác định độ bền của hàm băm tới một lớp tấn công

cụ thể, được biết dưới tên cuộc tấn công dựa trên bài toán nghịch lí về ngày sinh nhật (birthday attack).

2.3. Nguyên tắc xây dựng hàm băm

Có hai phương án chính để xây dựng hàm băm:

- Xây dựng trên cơ sở mật mã đối xứng (Hash Functions Based on Block Ciphers).
- Xây dựng các hàm hash riêng biệt (Dedicated hash functions).

Xây dựng trên cơ sở mật mã đối xứng

Một trong các phương pháp hiệu quả để xây dựng hàm hash là xây dựng trên cơ sở hàm mật mã khối.

Giả sử E là hàm mật mã khối an toàn với kích thước khối là m bít và bản tin đầu vào được chia ra n khối m bít $M = (M_1, M_2,\ldots, M_n)$. Hàm tính giá trị băm có thể viết dưới dạng sau:

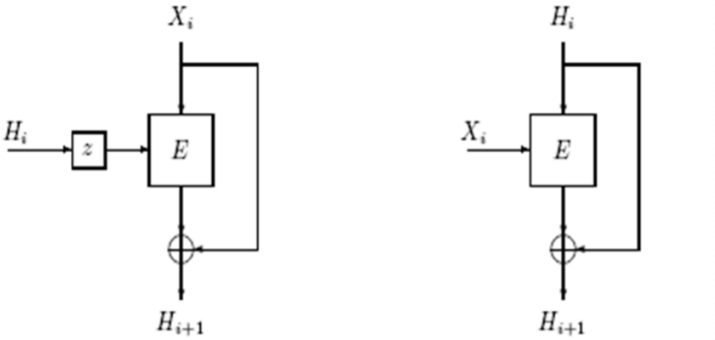
$$H_i = E(H_{i-1}, M_i)$$

- $i = 1, 2, \ldots, n$.
- H_0 là giá trị ban đầu đặc biệt.
- Giá trị băm của hàm $H = H_n$.

Có hai phương án chính để xây dựng các sơ đồ hàm hash, đó là các phương án được đưa ra bởi Mytyas-Meyer-Oseas và Davies-Meyer.

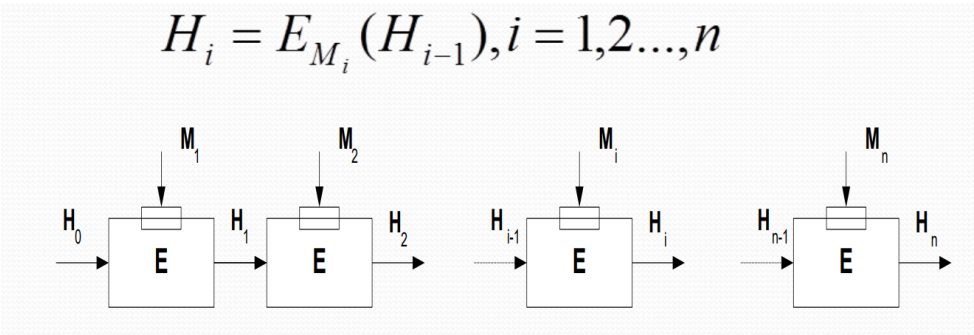
– $H_{i+1} = f(H_i, X_i) = E_{s(H_i)}(X_i) \oplus X_i$ (Matyas-Meyer-Oseas);

– $H_{i+1} = f(H_i, X_i) = E_{X_i}(H_i) \oplus H_i$ (Davies-Meyer).



Hình 2.3. Sơ đồ hàm hash bởi Mytyas-Meyer-Oseas và Davies-Meyer (ảnh: thầy Nguyễn Hiếu Minh – HVKTQS).

Sơ đồ hàm Hash Rabin có thể viết dưới dạng công thức:



Hình 2.4. Sơ đồ hàm Hash Rabin (ảnh: thầy Nguyễn Hiếu Minh – HVKTQS).

Hàm một chiều có thể xây dựng trên cơ sở hàm mật mã khối bền vững E, ví dụ có thể xây dựng F theo công thức sau:

$$F(X) = X \oplus E_k(X)$$

2.4. Thuật toán Băm

2.4.1. Hàm băm mật mã MD5

a. Giới thiệu

MD5 (Message-Digest algorithm 5) là một hàm băm mã hóa giá trị băm đầu vào có độ dài bất kì, đầu ra 128 bit, được phát triển bởi Ron Rivest tại đại học MIT. Giải thuật gồm 5 bước thao tác trên khối 512 bits.

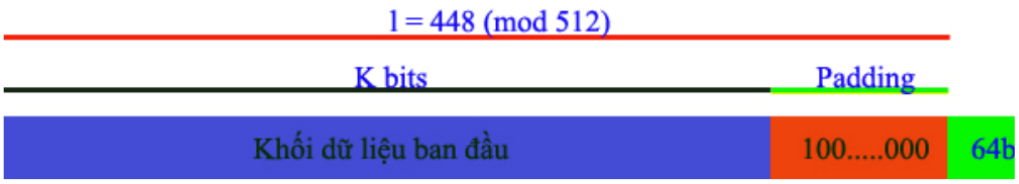
MD5 từng được xem là một chuẩn trên Internet, được sử dụng rộng rãi trong các chương trình an ning mạng, kiểm tra tính nguyên vẹn của tập tin.

b. Thuật toán MD5

Các bước thực hiện:

Bước 1: Bổ sung các bit làm đầy

- Được bổ sung các bit để độ dài đồng dư với 448 mod 512.
- Việc bổ sung luôn luôn thực hiện (thậm chí với độ dài mong muốn).
- Số lượng bit bổ xung nằm trong khoảng 1-512 bit
- Các bit bổ sung gồm 1 bit “1” và các bit 0 theo sau: 10…0

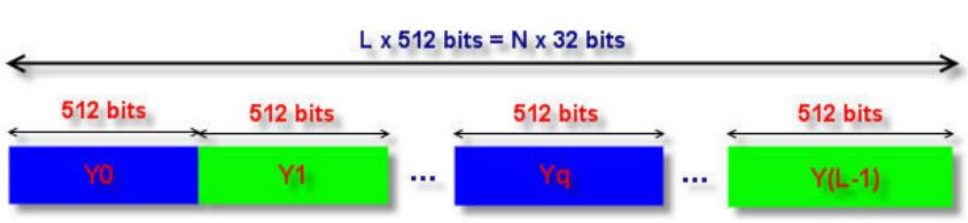


Hình 2.5. Bổ sung bit làm đầy trong thuật toán MD5 (ảnh: thầy Phạm Nguyên Khang, Đại học Cần thơ)

Bước 2: Bổ sung giá trị độ dài

- Ghi độ dài tin tức theo mod 2^{64} .
- Độ dài của khối dữ liệu ban đầu được biểu diễn dưới dạng nhị phân 64-bit và được thêm vào cuối chuỗi nhị phân kết quả của bước 1
- Nếu độ dài của khối dữ liệu ban đầu > 264, chỉ 64 bits thấp được sử dụng, nghĩa là giá trị được thêm vào bằng K mod 264
- Kết quả có được từ 2 bước đầu là một khối dữ liệu có độ dài là bội số của 512. Khối dữ liệu được biểu diễn:
    - Bằng một dãy L khối 512-bit $Y_0, Y_1,\ldots, Y_{L-1}$
    - Bằng một dãy N từ (word) 32-bit M_0, M_1, M_{N-1}.

  Vậy N = L x 16 (32 x 16 = 5)



Hình 2.6. Bổ sung giá trị độ dài trong thuật toán MD5 (ảnh: thầy Phạm Nguyên Khang, Đại học Cần thơ)

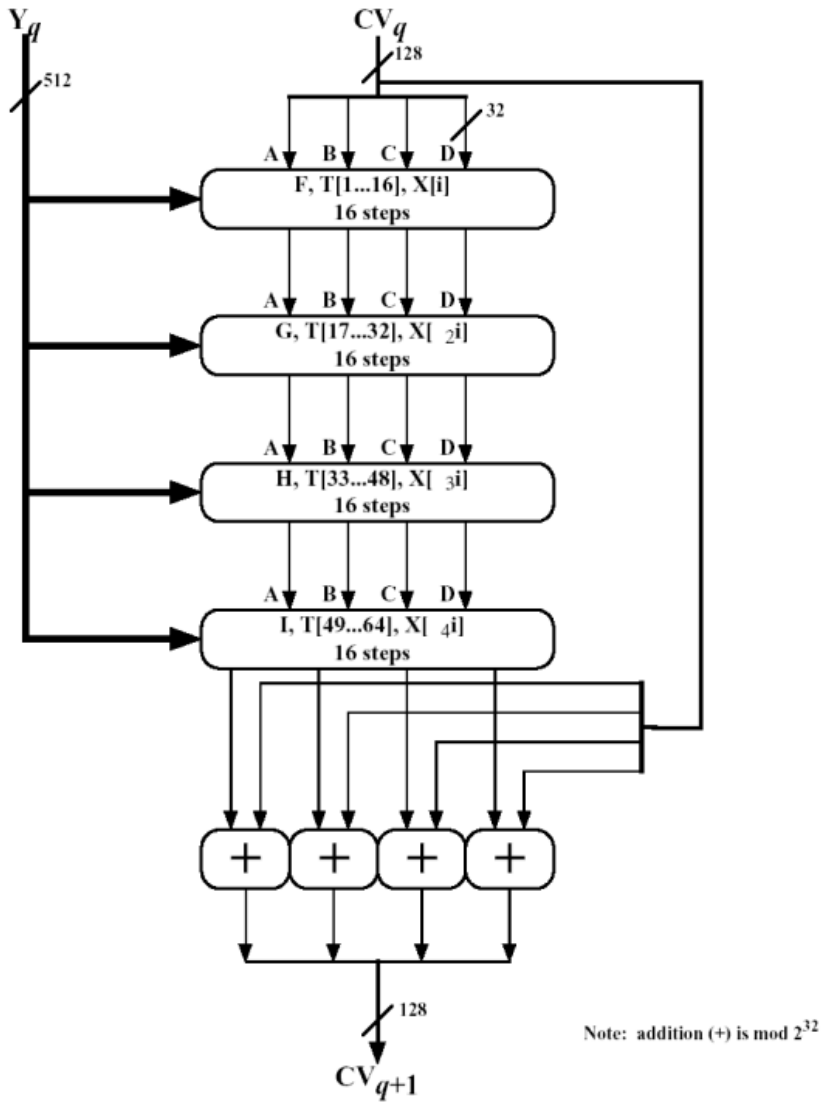
Bước 3: khởi tạo bộ đệm MD (MD buffer)

- Một bộ đệm 128-bit được dùng lưu trữ các giá trị băm trung gian và kết quả. Bộ đệm được biểu diễn bằng 4 thanh ghi 32- bit với các giá trị khởi tạo ở dạng little-endian (byte có trọng số nhỏ nhất trong từ nằm ở địa chỉ thấp nhất) như sau:
    - A = 67 45 23 01
    - B = EF CD AB 89
    - C = 98 BA DC FE
    - D = 10 32 54 76

- Các giá trị này tương đương với các từ 32-bit sau:
  - A = 01 23 45 67
  - B = 89 AB CD EF
  - C = FE DC BA 98
  - D = 76 54 32 10

Bước 4: xử lý các khối dữ liệu 512- bit

- Trọng tâm của giải thuật là hàm nén (compression function) gồm 4 "vòng" xử lý. Các vòng này có cấu trúc giống nhau nhưng sử dụng các hàm luận lý khác nhau gồm F, G, H và I
  - $F(X,Y,Z) = X \wedge Y \vee \neg X \wedge Z$
  - $G(X,Y,Z) = X \wedge Z \vee Y \wedge \neg Z$
  - H(X,Y,Z) = X xor Y xor Z
  - $I(X,Y,Z) = Y \text{ xor } (X \vee \neg Z)$
- Mảng 64 phần tử được tính theo công thức: $T[i] = 2^{32} \times abs(sin(i))$, i được tính theo radian.
- Kết quả của 4 vòng được cộng (theo modulo 2^{32} với đầu vào CV_q để tạo CV_{q+1}



Hình 2.7. Xử lý khối dữ liệu 512 bit (ảnh: thầy Phạm Nguyên Khang, Đại học Cần thơ).

Bước 5: Xuất kết quả Sau khi xử lý hết L khối 512-bit, đầu ra của lần xử lý thứ L là giá trị băm 128 bits.

Giải thuật MD5 được tóm tắt như sau:

- $CV_0 = IV$
- $CV_{q+1} = SUM_{32}[CV_q, RF_I(Y_q, RF_H(Y_q,RF_G(Y_q,RF_F(Y_q,CV_q))))]$
- $MD = CV_{L-1}$

Với các tham số

- IV: bộ đệm gồm 4 thanh ghi ABCD Yq: khối dữ liệu thứ q gồm 512 bits
- L: số khối 512-bit sau khi nhồi dữ liệu
- CV_q: đầu ra của khối thứ q sau khi áp dụng hàm nén
- RF_x: hàm luận lý sử dụng trong các "vòng" (F,G,H,I)
- MD: message digest – giá trị băm
- SUM_{32}: cộng modulo 2^{32}

Mỗi vòng thực hiện 16 bước, mỗi bước thực hiện các phép toán để cập nhật giá trị buffer ABCD, mỗi bước được mô tả như sau

- $A \leftarrow B + ((A + F(B,C,D) + X[k] + T[i]) <<< s)$
- A,B,C,D: các từ của thanh ghi
- F: một trong các hàm F,G,H,I
- <<< s : dịch vòng trái s bits
- $M_i \sim X[k]$: từ 32-bit thứ k của khối dữ liệu 512 bits.k=1..15
- $K_i \sim T[i]$: giá trị thứ i trong bảng T.
- +: phép toán cộng modulo 2^{32}

2.4.2. Thuật toán hàm băm SHA-1

a. Giới thiệu hàm băm SHA-1

Năm 1990, Ron Rivest đã sáng tạo ra hàm băm MD4. Sau đó năm 1992, ông cải tiến MD4 và phát triển một hàm băm khác: MD5. Năm 1993, Cơ quan An ninh Quốc gia Hoa Kỳ/Cục An ninh Trung ương (NSA) đã công bố,một hàm băm rất giống với MD5 được gọi là SHA. Vào năm 1995, sau việc khắc phục những lỗ hổng kỹ thuật, NSA đã thay đổi SHA trở thành một hàm băm mật mã khác gọi là SHA-1.

SHA-1 (Sercue Hash Algorithm) là thuật toán cũng được xây dựng trên thuật toán MD4, đang được sử dụng rộng rãi. Thuật toán SHA-1 tạo ra chuỗi mã băm có chiều dài cố định 160 bit từ chuỗi bit dữ liệu đầu vào x có chiều dài tùy ý.

b. Thuật toán SHA-1

I**nput**: thông điệp với độ dài tối đa 2^{64} bits

Output: thông điệp rút gọn (message digest) có độ dài 160 bits

Giải thuật gồm 5 bước trên khối 512 bits

Bước 1: Nhồi dữ liệu

- Thông điệp được nhồi thêm các bit sao cho độ dài L mod 512 luôn đồng dư là 448.
- Thông điệp luôn luôn được nhồi thêm các bit.
- Số bit nhồi thêm phải nằm trong khoảng 1-512.
- Phần thêm vào cuối dữ liệu gồm 1 bit “1” và theo sau là các bit 0: 10…0.

Bước 2: Thêm độ dài*:*

- Độ dài khối dữ liệu ban đầu sẽ được biểu diễn dưới dạng nhị phân 64 bit và được thêm cuối chuỗi nhị phân mà ta thu được ở bước 1.
- Độ dài được biểu diễn dưới dạng nhị phân 64 bit không dấu

- Kết quả thu được từ 2 bước là một khối dữ liệu có độ dài là bội số của 512. (Với cứ 512 bit là một khối dữ liệu)

Bước 3: Khởi tạo bộ đệm MD (MD buffer)

Một bộ đệm 160 bit được dùng để lưu trữ các giá trị băm trung gian và kết quả. Bộ đệm được biểu diễn bằng 5 thanh ghi 32-bit với các giá trị khởi tạo ở dạng big-endian (buyte có trọng số lớn nhất trong từ nằm ở địa chỉ thấp nhất) và có 2 bộ đệm. 5 thanh ghi của bộ đệm đầu tiên được đánh đặt tên là A, B,C,D,E và tương tự cho bộ đệm thứ 2 là H_0, H_1, H_2, H_3, H_4. Có giá trị như sau (Theo dạng Hex):

H_0=67452301

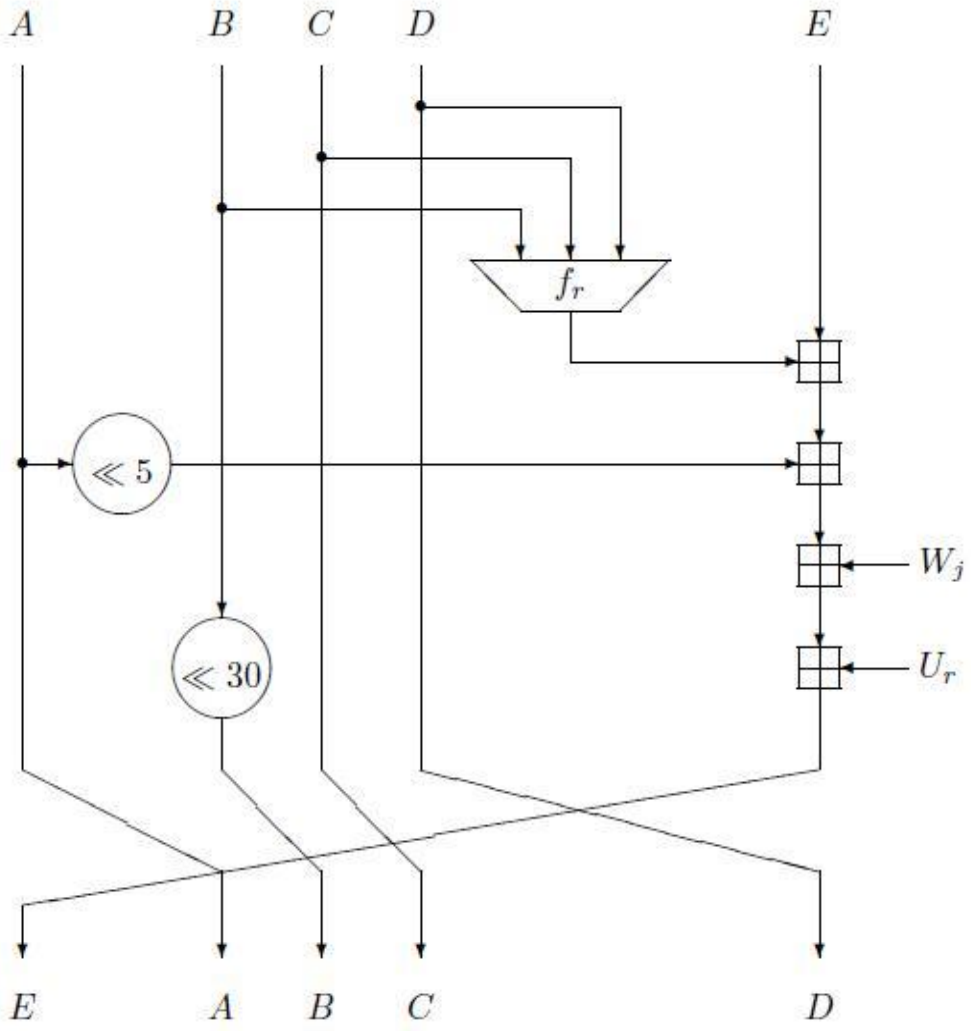
H_2=EFCDAB89

H_3= 98BADCFE

H_4= 10325476

H_5= C3D2E1F0

Bước 4: Xử lý các khối dữ liệu 512 bit



Hình 2.8. Thuật toán SHA-1 (ảnh: thầy Phạm Nguyên Khang, Đại học Cần thơ).

- Trọng tâm của giải thuật bao gồm 4 vòng lặp thực hiện tất cả 80 bước.
- 4 vòng lặp có cấu trúc như nhau, chỉ khác nhau ở hàm logic F_t..
- Mỗi vòng có đầu vào gồm khối 512-bit hiện thời và một bộ điệm 160 bit A, C, B, D, E. Các thao tác sẽ cập nhật giá trị bộ đệm.
- Chia khối dữ liệu đã nhồi thêm (cuối bước 2) thành 16 nhóm (mỗi nhóm gồm 32 bit) và đặt theo thứ tự là: W_0,W_1,...W_{15}.

- Mở rộng từ 16 nhóm 32bit lên đến 80 nhóm 32 bit bằng vòng lặp

  For 16 to 79 let

  W_t= S^1 (W_{t-3} XOR W_{t-8} XOR W_{t-14} XOR W_{t-16})
- Gán A=H_0, B=H_1, C=H_2, D=H_3, E=H_4.
- Mỗi vòng lặp sử dụng theo công thức chung với 1 hằng số K_t= (0≤ t ≤ 79) như sau:

  For t= 0 to 79 do

  TEMP= S^5(A)+F_t(B,C,D)+ E + W_t + K_t

  E=D; D=C; C= S^{30}(B); B=A; A= TEMP

  Với:

  K_t= 5A827999 (0≤ t ≤ 19)

  K_t= 6ED9EBA1(20≤ t ≤ 39)

  K_t= 8F1BBCDC(40≤ t ≤ 59)

  K_t= CA62C1D6(60≤ t ≤ 79).

- Đầu ra của 4 vòng (bước 80) được cộng với giá trị của bộ đệm để tạo ra 1 chuỗi kết quả dài 160 bit.

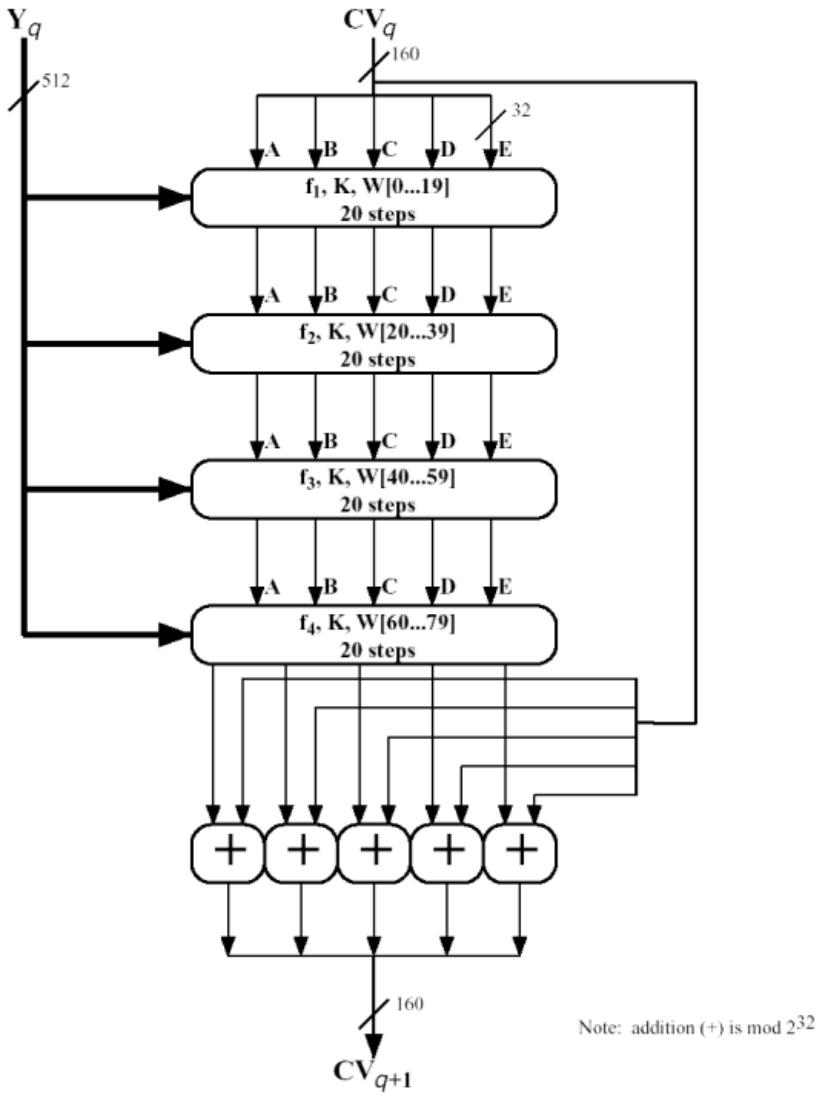
  H_0 = H_0 + A

  H_1 = H_1 + B

  H_2 = H_2 + C

  H_3 = H_3 + D

  H_4 = H_4 + E



Hình 2.9. Xử lý khối dữ liệu trong SHA-1 (ảnh: thầy Phạm Nguyên Khang, Đại học Cần thơ).

Bước 5: Xuất kết quả

Sau khi thao tác trên toàn bộ N khối dữ liệu (blocks). Kết quả của khối thứ N là chuỗi băm 160 bit.

H= $H_0H_1H_2H_3H_4$

Giải thuật được tóm tắt như sau:

- CV 0 = IV
- CVq+1 = SUM32(CV q, ABCDE q)
- MD = CV L

Với

- IV = giá trị khởi tạo của bộ đệm ABCDE
- ABCDE q = đầu ra của hàm nén trên khối thứ q
- L = số khối 512-bit của thông điệp
- SUM32 = phép cộng modulo 232 trên từng từ (32 bits) của đầu vào
- MD = giá trị băm

Giải thuật thực hiện tất cả 80 bước, mỗi bước được mô tả như sau:

- $A \leftarrow E + f(t, B, C, D) + S^5(A) + W_t + K_t$
- $B \leftarrow A$
- $C \leftarrow S^{30}(B)$
- $D \leftarrow C$
- $E \leftarrow D$

Trong đó

- A,B,C,D,E = các từ trong bộ đệm
- t = số thứ tự của bước
- F(t,B,C,D) = làm logic tại bước t
- S^k = dịch vòng trái k bits
- W_t = từ thứ t của khối dữ liệu
- K_t = hằng số
- + = phép cộng modulo 2^{32}

2.4.3. So sánh MD5 và SHA-1

Khả năng chống lại tấn công brute-force:

- Để tạo ra thông điệp có giá trị băm cho trước, cần 2128 thao tác với MD5 và 2160 với SHA-1
- Để tìm 2 thông điệp có cùng giá trị băm, cần 264 thao tác với MD5 và 280 với SHA

Khả năng chống lại thám mã (cryptanalysis): cả 2 đều có cấu trúc tốt

Tốc độ:

- Cả hai dựa trên phép toán 32 bit, thực hiện tốt trên các kiến trúc 32 bit
- SHA-1 thực hiện nhiều hơn 16 bước và thao tác trên thanh ghi 160 bit nên tốt độ thực hiện chậm hơn

Tính đơn giản: cả hai đều được mô tả đơn giản và dễ dàng cài đặt trên phần cứng và phần mềm

2.5. Ứng dụng của hàm Băm Hash

- **Xác thực mật khẩu**

Mật khẩu thường không được lưu dưới dạng văn bản rõ (clear text), mà ở dạng tóm tắt. Để xác thực một người dùng, mật khẩu do người đó nhập vào được băm ra bằng hàm Hash và so sánh với kết quả băm được lưu trữ.

- **Xác thực thông điệp (Message authentication – Thông điệp tóm tắt - message digests)**

Giá trị đầu vào(tin nhắn, dữ liệu...) bị thay đổi tương ứng giá trị băm cũng bị thay đổi. Do vậy nếu 1 kẻ tấn công phá hoại, chỉnh sửa dữ liệu thì server có thể biết ngay lập tức.

- **Bảo vệ tính toàn vẹn của tập tin, thông điệp được gửi qua mạng**

Hàm băm mật mã có tính chất là hàm 1 chiều. Từ khối dữ liệu hay giá trị đầu vào chỉ có thể đưa ra 1 giá trị băm duy nhất. Như chúng ta đã biết đối với tính chất của hàm 1 chiều. Một người nào đó dù bắt được giá trị băm họ cũng không thể suy ngược lại giá trị, đoạn tin nhắn băm khởi điểm.

Ví dụ: việc xác định xem một file hay một thông điệp có bị sửa đổi hay không có thể thực hiện bằng cách so sánh tóm tắt được tính trước và sau khi gửi (hoặc một sự kiện bất kỳ nào đó). Còn có thể dùng tóm tắt thông điệp làm một phương tiện đáng tin cậy cho việc nhận dạng file.

Hàm băm thường được dùng trong bảng băm nhằm giảm chi phí tính toán khi tìm một khối dữ liệu trong một tập hợp. Giá trị băm đóng vai trò gần như một khóa để phân biệt các khối dữ liệu.

Chương 3. Tìm hiểu chữ ký số

3.1. Tổng quan chữ ký số

3.1.1. Khái niệm

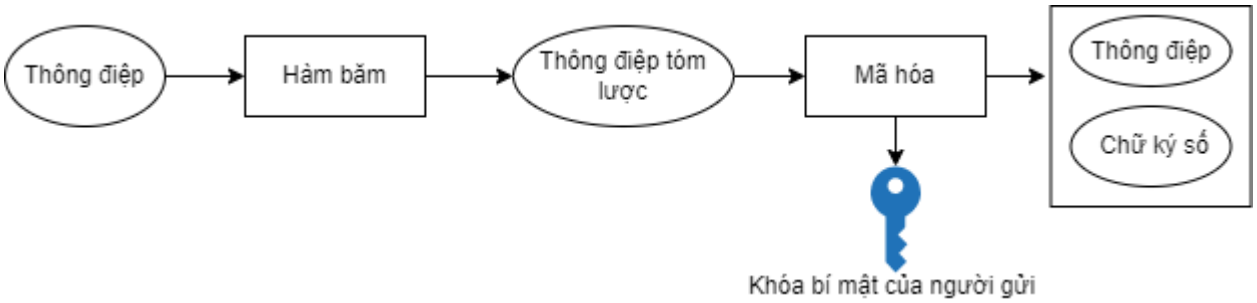
Hiện nay có nhiều định nghĩa về chữ ký số theo khía cạnh và quan điểm nghiên cứu khác nhau. Tuy nhiên, theo Nghị Định 130/2018/NĐ-CP, ngày 27/9/2018 của Chính phủ Việt Nam, chữ ký số được định nghĩa là một loại chữ ký điện tử, được tạo bằng sự chuyển đổi thông điệp dữ liệu sử dụng một hệ thống mật mã không đối xứng, theo đó người có được thông điệp dữ liệu ban đầu và khóa công khai của người ký có thể xác định được chính xác:

- Việc biến đổi nêu trên được tạo ra bằng đúng khóa bí mật tương ứng với khóa công khai trong cùng một cặp khóa.
- Sự toàn vẹn nội dung của thông điệp dữ liệu kể từ khi thực hiện việc biến đổi nêu trên.

Cho đến nay, chữ ký số đã có những bước phát triển mạnh mẽ và trở thành bộ phận cấu thành quan trọng của ngành mật mã học. Hiện nay, đã có nhiều các lược đồ chữ ký số khác nhau được nghiên cứu và phát triển dựa trên các chuẩn chữ ký số và các lược đồ chữ ký số phổ biến như GOST R34.10-94, GOST R34.10-2012, RSA, Rabin, Schnorr, EC-Schnorr,... Gần 40 năm qua, dựa vào các tiêu chí khác nhau có thể chia các lược đồ chữ ký số thành nhiều loại như chữ ký số nhóm, chữ ký số tập thể, chữ ký số đại diện, chữ ký số mù,... Chữ ký số cũng đã được ứng dụng rộng rãi trong thực tiễn và đã được đưa thành chuẩn và được triển khai ở hơn 40 quốc gia trên thế giới như FIPS của Mỹ, GOST của Liên bang Nga,…

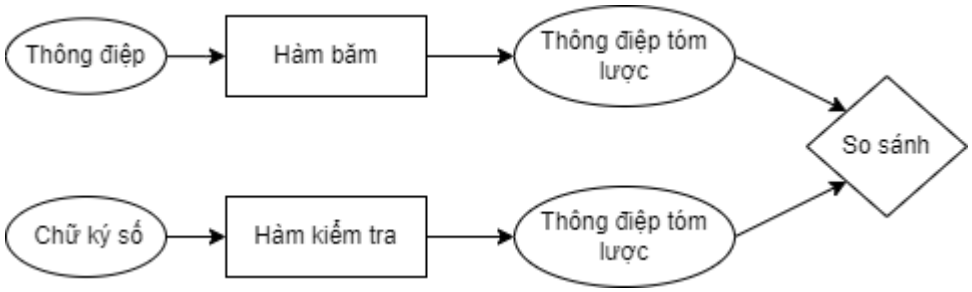
Chữ ký số bao gồm 3 thành phần: thủ tục tạo khóa, thủ tục tạo chữ ký và thủ tục kiểm tra chữ ký.

- Thủ tục tạo ra chữ ký là hàm tính toán chữ ký trên cơ sở khóa riêng và dữ liệu cần ký.



Hình 3.1. Sơ đồ tạo chữ ký trong chữ ký số.

- Thủ tục kiểm tra chữ ký là hàm kiểm tra xem chữ ký đã cho có đúng với khóa công khai của người ký hay không.



Hình 3.2. Sơ đồ kiểm tra chữ ký trong chữ ký số.

Ví dụ chữ ký số:

Ta có thể mô phỏng trực quan một hệ mật mã khoá công khai như sau : Bảo muốn gửi cho Ly một thông tin mật mà Bảo chỉ có Ly mới có thể đọc được. Để làm được điều này, Ly gửi cho Bảo một chiếc hộp có khóa đã mở sẵn (Khóa công khai) và giữ lại chìa khóa. Bảo nhận chiếc hộp, cho vào đó một tờ giấy viết thư chứa thông tin Bảo muốn gửi cho

Ly và khóa lại (chỉ cần sập chốt lại, sau khi sập chốt khóa ngay cả Bảo cũng không thể mở lại được, không đọc lại hay sửa thông tin trong thư được nữa). Sau đó Bảo gửi chiếc hộp lại cho Ly. Ly mở hộp với chìa khóa của mình và đọc thông tin trong thư. Trong ví dụ này, chiếc hộp với khóa mở đóng vai trò khóa công khai, chiếc chìa khóa chính là khóa bí mật.

3.1.2. Chức năng của chữ ký số

Xác thực được nguồn gốc thông điệp: Tùy thuộc vào từng thông điệp mà có thể thêm các thông tin nhận dạng như tên tác giả, nhãn thời gian,…

Tính toàn vẹn của thông điệp: Khi có sự thay đổi bất kỳ vô tình hay cố ý lên thông điệp thì giá trị hàm băm sẽ bị thay đổi và kết quả kiểm tra sẽ cho kết quả không đúng hay nói rằng thông điệp không toàn vẹn.

Chống từ chối thông điệp: Vì chỉ có chủ thông điệp mới có khóa riêng để ký lên thông điệp nên người ký không thể chối bỏ thông điệp của mình

3.1.3. Ứng dụng của chữ ký số

Phạm vi ứng dụng của chữ ký số rất rộng, gồm nhiều lĩnh vực, như:

- Ký số trong thư điện tử cho phép khách hàng xác định chính xác người gửi;
- Sử dụng chữ ký số thực hiện việc ký các văn bản xác nhận khi đầu tư chứng khoán trực tuyến, bán hàng trực tuyến, thanh toán trực tuyến, chuyển tiền trực tuyến;
- Ký số trong hợp đồng kinh tế mà không cần gặp mặt trực tiếp;
- Ký số trong kê khai, nộp thuế trực tuyến, khai báo hải quan và thông quan trực tuyến…
- Trong các cơ quan Nhà nước, ứng dụng chữ ký số là một yếu tố không thể thiếu để xây dựng Chính phủ điện tử và cải cách thủ tục hành chính.

- Trong các doanh nghiệp, chữ ký số là công cụ hữu hiệu trong giao dịch với các cơ quan nhà nước thông qua các dịch vụ công trực tuyến, giao dịch với các đối tác và khách hàng của mình.

Việc ứng dụng chữ ký số giúp tiết kiệm chi phí (chi phí mua giấy in, mực in, chi phí và thời gian gửi văn bản); giảm thiểu sức lao động trong công tác quản lý, bảo mật dữ liệu cá nhân và dữ liệu chuyên môn; giảm thời gian, tiết kiệm chi phí đi lại của người dân và doanh nghiệp; quan trọng nhất là minh bạch hóa thông tin, làm thay đổi phương pháp, tác phong công tác, phương thức làm việc...

3.2. Chữ ký số RSA

Giải thuật mã hóa RSA được 3 nhà khoa học Ronald Rivest, Adi Shamir và Leonard Adleman phát minh năm vào 1977. Độ an toàn của RSA dựa trên tính khó của việc phân tích số nguyên rất lớn.

RSA sử dụng một cặp khóa: Khóa công khai (Public key) dùng để mã hóa; Khóa riêng (Private key) dùng để giải mã, chỉ khóa riêng cần giữ bí mật, khóa công khai có thể công bố rộng rãi.

Tạo khóa:

Người gửi A chọn cặp số nguyên tố đủ lớn p và q, tính $n = p \times q$.

Tính $\Phi(n) = (p-1) \times (q-1)$.

Chọn số nguyên e thỏa mãn $\gcd(e, \phi(n)) = 1$.

Người gửi A tính số nguyên d, thỏa mãn phương trình $d \equiv e^{-1} \bmod \phi(n)$.

Ta có, d là khóa riêng của người gửi A, còn (n, e) là khóa công khai.

Tạo chữ ký:

Để tạo ra chữ ký của bản tin $m \in Z_n$, người gửi A tính giá trị:

$$c = m^d \bmod n$$

Kiểm tra chữ ký:

Để kiểm tra chữ ký c có phải của người gửi a không, người nhận phải kiểm tra bằng thủ tục sau:

$$Verify_{(n,e)}(m,c) = \text{TRUE nếu } m = c^e \bmod n$$

Quá trình tạo chữ ký và kiểm tra chữ ký giống với quá trình mã hoá và giải mã của hệ mật RSA chỉ khác là quá trình tạo chữ ký thì người ký dùng khóa riêng còn quá trình kiểm tra thì người nhận dùng khóa công khai.

Ví dụ:

Tạo khóa:

Người gửi chọn 2 số nguyên tố lớn ngẫu nhiên là 23 và 41.

n = 23*41 = 943

Φ(n) = 22*40 = 880

Chọn e = 7 vì UCLN(7,880) = 1.

Tính $d \equiv e^{-1} \bmod \phi(n)$, bằng cách dùng thuật toán Euclide, tìm số tự nhiên x sao cho:

$$d = \frac{x * \Phi(n) + 1}{e}$$

$$\Rightarrow d = 1 + 880x$$

$$\Rightarrow d = 503, x = 4$$

Khóa công khai: n = 943, e = 7

Khóa bí mật: d = 503

Kiểm tra chữ ký và giải mã:

Biểu diễn thông tin cần gửi thành số m (0<= m <= n-1), cho thông tin cần gửi m = 35

Người gửi tính giá trị c:

$$c = m^d \bmod n$$

$$\Rightarrow c = 35^{503} \bmod 943 = 545$$

Người gửi gửi giá trị c cho người nhận, người nhậ sẽ giải mã bằng cách tính giá trị m:

$$m = c^e \bmod n$$

$$\Rightarrow m = 545^7 \bmod 943 = 35$$

Người nhận đã giải mã được giá trị m và thông tin nhận được là m = 35.

3.3. Chữ ký số Schnorr

Lược đồ chữ ký số Schnorr [56] được phát triển dựa trên bài toán DLP và được mô tả như sau:

Tham số sử dụng: tham số miền là (p, q, g); số nguyên tố lớn ngẫu nhiên q; số nguyên tố lớn ngẫu nhiên p thoả mãn (p-1) chia hết cho q; phần tử sinh $g = h^{(p-1)/q} \bmod p$ với h bất kỳ và (1 < h < p-1) , chọn lại nếu g=1; Khoá bí mật $d \in Z_q$; thông điệp cần ký M.

Tạo khoá:

Tính khóa công khai $\rho = g^d \bmod p$;

Tạo chữ ký: Tính các thành phần của chữ ký

Người ký chọn giá trị ngẫu nhiên với $(1 < k < q)$ và tính $c = g^k \bmod p$;

Tính thành phần đầu tiên $r = H(M, c) \bmod q$, nếu $r = 0$ thì chọn lại k;

Tính $s = (k - rd) \bmod q$, nếu $s = 0$ thì quay lại phần chọn k và tính c;

Đầu ra là cặp (r, s) là chữ ký số trên thông điệp M.

Xác thực chữ ký:

Tính mod $c' = g^s \rho^r \bmod q$ và $r' = H(M, c')$;

Nếu $r' = r$ thì chữ ký được chấp nhận, ngược lại không được chấp nhận.

3.4. Chuẩn chữ ký số DSS

Thiết lập tham số hệ thống:

1. Chuẩn DSS chọn hàm băm là SHA-1
2. Chọn một số nguyên tố q có độ dài N bit. N phải có độ dài nhỏ hơn hoặc bằng giá trị đầu ra của hàm băm.
3. Chọn một số nguyên tố p có độ dài L bit, sao cho p - 1 là một bội số của q.
4. Chọn phần tử $g \in Z_p$ có bậc là q.

Hình thành khóa riêng và khóa công khai:

Phát sinh một số ngẫu nhiên x $g \in Z_q$ (x là khóa riêng) và tính khóa công khai:

$$y = g^x \pmod p$$

Khóa công khai bao gồm (p, q, g, y), x là khóa riêng.

Tạo chữ ký:

Để ký lên bức điện m $\in$ { 0,1}, người gửi A tạo số ngẫu nhiên k $\in Z_q$ và hình thành nên cặp (s,r), ở đây:

$w \leftarrow s^{-1} \pmod q$,

$u_1 \leftarrow H(m)w \pmod q$,

$u2 \leftarrow rw \pmod q$,

$Verify_{(p, q, g, y)}(m,(r,s))$ = TRUE, nếu $r = (g^{u1}y^{u2} \bmod p) \pmod q$.

3.5. Chuẩn chữ ký số ECDSS

Hình thành khóa:

Tham số cơ sở của một chữ ký số dựa trên đường cong elliptic là một bộ các thành phần được ký hiệu: D = (q, FR, a, b, [DomainParameterSeed], G, n, h). Trong đó:

1. Một số mốt nguyên q là kích thước của trường hữu hạn F_q, có thể chọn q = p, với p là số nguyên tố lẻ, hoặc $q = 2^m$ với m là số nguyên dương.
2. FR: $E(F_p)$ hoặc $E(F_{2m})$.
3. [DomainParameterSeed] là một chuỗi bit mầm có chiều dài tối thiểu là 160 bit.
4. Hai phần tử a, b $\in F_q$ là hai hệ số trong phương trình của đường cong elliptic E được định nghĩa trên F_q.
5. Hai phần tử x_G, $y_G \in F_q$ là tọa độ của điểm cơ sở bậc nguyên tố G thuộc $E(F_q)$.
6. Bậc n của điểm G, với $n > 2^{160}$ và $n > \sqrt[4]{q}$
7. Phần phụ đại số $h = \frac{\# E\ (\mathrm{Fq})}{n}$

Người gửi phải phát sinh một cặp khóa phù hợp với đường cong elliptic lựa chọn, gồm có 1 khóa riêng d (một số nguyên được lựa chọn ngẫu nhiên trong khoảng [1,n-1]) và một khóa công khai Q (ở đây Q = dG)

Tạo chữ ký:

Để ký lên bức điện m $\in$ { 0,1}, người gửi A thực hiện các bước sau:

1. Tính e ← H(m), ở đây H(m) là một hàm băm như SHA-1
2. Chọn số ngẫu nhiên k trong khoảng [1,n-1]
3. Tính r ← x_1 (mod n), ở đây (x_1, y_1) = kG. Nếu r = 0 thì quay lại bước 2.
4. Tính s ← k^{-1}(e+rd)(mod n). Nếu s = 0 thì quay lại bước 2.

Chữ ký bản tin m là cặp (r,s)

Kiểm tra chữ ký:

Người nhận B muốn kiểm tra chữ ký (r,s) có đúng là người gửi A đã ký lên m hay không thì thực hiện các bước sau:

1. Kiểm tra r và s là các số nguyên trong khoảng [1,n-1] thì chuyển sang bước 2, ngược lại thì chữ ký không hợp lệ.
2. Tính e ← H(m).
3. Tính w ← s^1(mod n).
4. Tính u_1 ← ew (mod n) và u2 ← rw(mod n).
5. Tính (x_1, y_1) ← u_1G + u_2Q.

Verify$_{(Q)}$(m, (r,s)) = TRUE, nếu r = x_1(mod n).

Chương 4. Kết luận

Qua thời gian nghiên cứu và tìm hiểu về đề tài "Tìm hiểu hàm băm và chữ ký số" tôi đã trình bày các khái niệm, các bước tạo và xây dựng hàm băm và chữ ký số. Bài tiểu luận đã trình bày và tìm hiểu được các nội dung sau:

1.Tổng quan về hàm băm, tính chất, nguyên tắc xây dựng hàm băm và các thuật toán băm: SHA-1, MD5.

2. Tổng quan về chữ ký số, chức năng của chữ ký số, lược đồ chữ ký số RSA, Schnorr, chuẩn chữ ký số DSS.

Định hướng phát triển tiếp theo của tiểu luận: tìm hiểu, nghiên cứu khai thác rộng và sâu hơn các tri thức về lý thuyết các dạng chữ ký số phổ biến, xây dựng chương trình thực nghiệm áp dụng lý thuyết chữ ký số, liên hệ các kiến thức liên quan để áp dụng vào thực tiễn.

Sau khi tìm hiểu đề tài tiểu luận và các kiến thức đã học trên lớp bản thân tôi đã có thêm kiến thức về hàm băm và chữ ký số, cách nghiên cứu, tìm tài liệu, góp ý xây dựng để hoàn thành đề tài tìm hiểu. Qua đó tôi nhận thấy vấn đề hàm băm và chữ ký số ứng dụng rất nhiều trong thực tiễn, thúc đẩy tôi tích cực học tập và nghiên cứu nhiều hơn. Mở rộng kiến thức cho tôi trong quá trình học tôi chân thành cảm ơn thầy đã nhiệt tình hướng dẫn và giúp đỡ tôi trong suốt thời gian qua.

Tài liệu tham khảo

[1] Giáo trình "Hàm Băm", Phạm Nguyên Khang, Đại học Cần thơ.

[2] Giáo trình "An Toàn Bảo Mật Hệ Thống Thông Tin", Chương 4: Các ký thuật mã hóa thông tin, TS Hoàng Xuân Dậu, Học viên Công nghệ Bưu Chính Viễn thông.

[3] Matt Bishop, *Introduction to Computer Security*, Prentice Hall, 2004.

[4] William Stallings, *Cryptography and Network Security*, Prentice Hall, 2010.

[5] David Kim, Michael G. Solomon, Fundamentals of Information Systems Security, Jones & Bartlettlearning, 2012.

[6] Giáo trình "Hàm băm và chữ ký số", Nguyễn Hiếu Minh, Học viện Kỹ thuật Quân sự.

[7] Lưu Hồng Dũng, Hoàng Thị Mai, Nguyễn Hữu Mộng (2015), "Nghiên cứu về một dạng lược đồ chữ ký số mới được xây dựng trên bài toán phân tích một số". Kỷ yếu Hội nghị Quốc gia lần thứ VIII về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR), Hà Nội, ngày 9-10/7/2015.

[8] Dolmatov (2013), "Digital Signature Algorithm draft-dolmatov-gost34-10- 2012-00 ", Cryptocom, Ltd.

[9] Giới thiệu về một số hàm băm và ứng dụng trong một số sản phẩm mật mã dân sự (https://nacis.gov.vn/nghien-cuu-trao-doi/-/view-content/213743/gioi-thieu-ve-mot-so-ham-bam-va-ung-dung-trong-mot-so-san-pham-mat-ma-dan-su) truy cập ngày 5/5/2022.

[10] Đẩy mạnh ứng dụng chữ ký số trong lĩnh vực quân sự, quốc phòng (http://www.mod.gov.vn/) truy cập ngày 6/5/2022.