CHƯƠNG 2 : MẬT MÃ HỌC

2.1 .NHỮNG KHÁI NIỆM CƠ BẢN

- Mật mã học bao gồm hai lĩnh vực : mã hóa (cryptography) và thám mã (cryptanalysis codebreaking) trong đó:
- Mã hóa: nghiên cứu các thuật toán và phương thức để đảm bảo tính bí mật và xác thực của thông tin gồm các hệ mã mật , các hàm băm, các hệ chư ký điện số, các cơ chế phân phối, quản lý khóa và các giao thức mật mã.
- Thám mã: Nghiên cứu các phương pháp phá mã hoặc tạo mã giả gồm các phương pháp thám mã , các phương pháp giả mạo chư ký, các phương pháp tấn công ,các hàm băm và các giao thức mật mã

2.1.1. Định nghĩa mật mã

- Mã hóa (cryptography) là một ngành khoa học của các phương pháp truyền tin bảo mật. Trong tiếng Hy Lạp, “Crypto” (krypte) có nghĩa là che dấu hay đảo lộn, còn “Graphy” (grafik) có nghĩa là từ. [3]
- Văn bản gốc có thể hiểu được hay bản rõ (P-Plaintext)
- Văn bản ở dạng bí mật không thể hiểu được thì được gọi là bản mã (C-Ciphertext).
- Có 2 phương thức mã hoá cơ bản: thay thế và chuyển vị

2.1.2. Hệ mật mã

Một hệ mã mật là bộ 5 (P, C, K, E, D) thoả các điều kiện

1). P là không gian rõ: tập hữu hạn các bản rõ có thể có.

2). C là không gian mã: tập hữu hạn các bản mã có thể có.

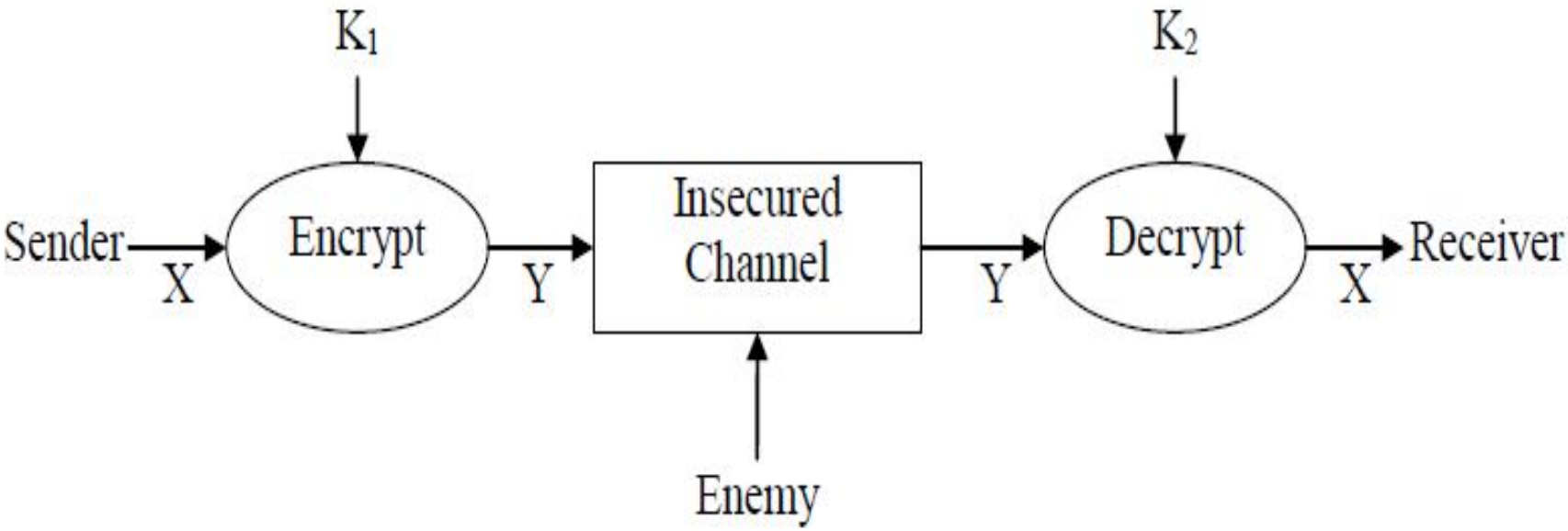
3). K là kkhông gian khoá: tập hữu hạn các khoá có thể có.

4). Đối với mỗi k ϵ K, có một quy tắc mã hoá ek ϵ E và một quy tắc giải mã tương ứng dk ϵ D.

5).Với mỗi ek: P →C và dk: C →P là những hàm mà dk(ek(x)) = x cho mọi bản rõ x ϵ P. Hàm giải mã dk() chính là ánh xạ ngược của hàm mã hóa ek

- Tính chất 4 ,5 là tính chất quan trọng nhất của mã hoá. Nếu mã hoá bằng ek và bản mã nhận được sau đó được giải mã bằng hàm dk() thì kết quả nhận được phải là bản rõ ban đầu x , hàm ek(x) phải là một đơn ánh, nếu không thì ta sẽ không giải mã được. Vì nếu tồn tại (x1 ,x2) : y = ek(x1) = ek(x2) → Bản mã Y không tồn tại.
- Trong một hệ mật bất kỳ ta luôn có |C| ≥ |P| vì mỗi quy tắc mã hoá là một đơn ánh. Khi |C| = |P| thì mỗi hàm mã hoá là một hoán vị.

2.1.3. Mô hình truyền tin cơ bản của mật mã học và luật Kirchoff



Hình 1.1: Mô hình cơ bản của truyền tin bảo mật

- Theo luật Kirchoff (1835 - 1903) (một nguyên tắc cơ bản trong mã hoá) thì: *toàn bộ cơ chế mã/giải mã trừ khoá là không bí mật đối với kẻ địch*
- *Ý nghĩa :*sự an toàn của các hệ mã mật không phải dựa vào sự phức tạp của thuật toán mã hóa sử dụng.

2.2.Sơ lược về lịch sử mật mã học

- Mật mã học là một ngành khoa học có một lịch sử khoảng 4000 năm
- Các phương pháp mã hóa đơn giản đầu tiên mà loài người đã sử dụng là của người Ba Tư cổ và người Do Thái cổ.
- Lịch sử mật mã học => hai thời kỳ như sau:
  - Thời kỳ tiền khoa học: Từ trước công nguyên cho tới năm 1949 : Mang tính nghệ thuật
  - Lịch sử của mật mã học hiện đại được đánh dấu vào năm 1949 khi Claude Shannon đưa ra lý thuyết thông tin.
  - Đầu những năm 1970 là sự phát triển của các thuật toán mã hóa khối đầu tiên: Lucipher và DES

- Vào cuối những năm 1970 phát triển các thuật toán khóa công khai sau khi Whitfield Diffie và Martin Hellman công bố bài báo “New Directions in Cryptography” làm nền tảng cho sự ra đời của các hệ mã khóa công khai và các hệ chữ ký số.
- Các hệ mã khối vẫn tiếp tục được phát triển thay thế cho DES vào cuối thế kỷ 20 như IDEA, AES hoặc 3DES (một cải tiến của DES).
- Các hàm băm MD5 (một hàm băm thuộc họ MD do Ron Rivest phát triển) và SHA1 .
- MD5 và SHA1 đã bị hack, các nhà mật mã học đã khuyến cáo sử dụng các hàm băm mạnh hơn (như SHA-256, SHA-512) trong các ứng dụng.

2.3.Phân loại các thuật toán mật mã

- Các thuật toán mã hóa khóa bí mật (hệ mã mật khóa bí mật hay khóa đối xứng SKC (Symmetric Key Cryptosytems), ví dụ : Caesar, DES, AES ...
- Các thuật toán mã hóa khóa công khai (các hệ mã khóa công khai PKC)(Public Key Cryptosystems). Còn gọi là các hệ mã khóa bất đối xứng (Asymmetric Key Cryptosytems). Khóa sử dụng cho các thuật toán này là 2 khóa : Public Key và Private key
- Các thuật toán tạo chữ ký số (Digital Signature Algorithms) : RSA, ElGammma...
- Các hàm băm (Hash functions).

Phân loại theo cách sử lý Input/Ouput

- Các thuật toán mã hóa khối (chẳng hạn như DES, AES …) xử lý bản rõ được chia thành các khối có độ dài giống nhau M_i .
- Các thuật toán mã hóa dòng (RC4 …) coi bản rõ là một luồng bit, byte liên tục.

2.4. Ứng dụng của mật mã học

- Bảo mật (Confidentiality) truyền thông hoặc giao dịch hoặc các thông điệp trên một hệ thống máy tính (các file, các dữ liệu trong một cơ sở dữ liệu …).
- Xác thực (Authentication): đảm bảo nguồn gốc của một thông điệp, người dùng.
- Toàn vẹn (Integrity): đảm bảo dữ liệu không bị thay đổi bất hợp pháp trên mạng truyền thông cũng như khi lưu trữ.
- Dịch vụ không thể chối từ (Non-Repudiation):Không thể phủ nhận việc tham gia vào một giao dịch hợp lệ.
- Ngoài ra còn các dịch vụ quan trọng khác như chữ ký điện tử, dịch vụ chứng thực danh tính (CA)

2.5. Cơ sở toán học của mật mã

- Khái niệm cơ bản về lý thuyết thông tin Entropy,
- Tốc độ của ngôn ngữ (Rate of Language)
- Độ phức tạp của thuật toán,
- Độ an toàn của thuật toán,
- Kiến thức toán học: đồng dư số học (modulo), số nguyên tố, định lý phần dư trung hoa, định lý Fermat . . . và các thuật toán kiểm tra số nguyên tố

Những vấn đề chính

- Lý thuyết thông tin
- Lý thuyết độ phức tạp (tham khảo tài liệu)
- Độ an toàn của thuật toán (tham khảo tài liệu)
- Lý thuyết số học.

2.5.1 . Lý thuyết thông tin

2.5.1.1 . ENTROPY : Đơn vị đo lượng thông tin

Khối lượng thông tin trong một thông báo là số bít nhỏ nhất cần thiết để mã hoá tất cả những ý nghĩa có thể của thông báo đó.

- Ví dụ, trường "NGAY" trong tuần chứa không quá 3 bít thông tin, bởi vậy thông tin ngày có thể mã hoá với 3 bít dữ liệu.
- Trường GIOI_TINH được thể hiên bởi 1 bít thông tin "0" và "1"

- Khối lượng thông tin trong một thông báo M đo bởi Entropy của thông tin đó, ký hiệu là H(M).
- Entropy của thông báo “GIOI_TINH” 1 bít, ký hiệu H(gioi_tinh) = 1. (n=2)
- Entropy của thông báo “NGAY” trong tuần là 3 . (n=8)

Trong trường hợp tổng quát, Entropy của một thông báo là $\log_2 n$, với n là số khả năng có thể (ý nghĩa) của thông báo.

$$H(M) = \log_2 n$$

2.5.1.2.Tốc độ của ngôn ngữ. (Rate of Language)

- Tốc độ thực tế (actual rate) của ngôn ngữ là:

$$r = H(M)/N$$

• N là độ dài của thông báo M . Tốc độ của tiếng Anh bình thường là 0.28 do đó mỗi chữ cái tiếng Anh có 1.3 bit có nghĩa.

- Tốc độ tuyệt đối (absolute rate) là số bits lớn nhất cần thiết để mã hóa các ký tự của một ngôn ngữ . Nếu có L ký tự trong một ngôn ngữ, thì tốc độ tuyệt đối là :

$$R = \log_2 L$$

- Đây là số Entropy lớn nhất của mỗi ký tự đơn lẻ. Đối với tiếng Anh gồm 26 chữ cái, tốc độ tuyệt đối là $\log_2 26 = 4.7$bits/chữ cái(letter).
- Độ dư thừa của ngôn ngữ (Redundancy) tự nhiên.
- Độ dư thừa (Redundancy) của một ngôn ngữ ký hiệu là D :

$$D = R - r.$$

- Đối với tiếng Anh:

  D = 1 - 0.28 =0.72 letters/letter

  D = 4.7 – 1.3 = 3.4 bits/letter

Như vậy mỗi chữ cái có 1.3 bit nghĩa và 3.4 bit dư thừa (xấp xỉ 72%).

2.5.2. Lý thuyết số học

2.5.2.1. Phép toán Modulo

- Các phép toán modulo , bao gồm các phép giao hoán, kết hợp và phân phối.

  (a+b) mod n = ((a mod n) + (b mod n)) mod n

  (a- b) mod n = ((a mod n) - (b mod n)) mod n

  (axb) mod n = ((a mod n) x (b mod n)) mod n

  (ax(b + c)) mod n = (((a x b) mod n) + ((a x c) mod n)) mod n
- Các phép tính trong các hệ mã mật hầu hết đều liên quan đến một phép toán modulo .

2.5.2.2. Số nguyên tố

- $a \in Z, b \in N^*; q \in Z$ và $r \in N$ sao cho $a=bq+r$, $0 \le r \ge b$; q được ký hiệu là a/b (thương số), r – số dư của a%b hay a modulo b
- Một số nguyên dương $c \in Z$ gọi là ƯSC của a,b nếu $c \mid a$ và $c \mid b$; ƯSC $gcd \in Z$ của $a,b \in Z$ được gọi là ƯSCLN , gcd = gcd(a,b) hay $gcd=a \wedge b$ nếu $c \mid a, c \mid b \Rightarrow c \mid gcd$
- $lcm \in Z$ gọi là BSC của a,b nếu $a \mid lcm$ và $b \mid lcm$; $lcm \in N$ là BSCNN của a,b nếu $a \mid c$, $b \mid c \Rightarrow gcd \mid c$;

  Ký hiệu lcm=lcm(a,b) hay $lcm=a \vee b$.

- Định nghĩa
  Với $a \geq 2$ gọi là một SNT nếu nó chia hết cho 1 và a.
  Tập hợp các SNT ký hiệu là : p{2,3,5,7,11,13,..,}

- Định nghĩa
  $a,b \in Z$ gọi là nguyên tố cùng nhau ($a \perp b$) nếu a và b chỉ có một ƯSC duy nhất là 1, ($a \wedge b=1$)

Một số khái niệm

- Tập nguyên Z{0,±1,±2... ±n}
- Vành (A,+,*)
- Nhóm (G)
- Trường (F,+,*,a-1)
- Phép đồng dư

- Phép đồng dư :

$x \equiv y \pmod m$; $x<m$; $x,y \in [0\text{-}n]$

Hay : $x = y+km \Rightarrow x-y = km$

- ✔ x chia cho m có số dư r
  y chia cho m có số dư r
- ✔ x-y $\Rightarrow$ bội số của m ; m là số chia của x-y

Ta goi x là thặng dư của y theo modulo m ; x là đồng dư của y

- Phương trình Diophante (pt bất định)

$ax^n+by^n = c^n \exists x,y \{ Z \}$ → nghiệm của pt

- Vành Z_N (vành đồng dư modul0 N)

Tập các số nguyên $Z_N = \{0, 1, \ldots, N-1\}$ trong đó N là một số tự nhiên dương với hai phép toán cộng (+) và nhân (.) tạo thành một vành đồng dư modulo N (hay còn gọi là tập thặng dư đầy đủ theo modulo N):

  - Phép cộng:

$\forall a, b \in Z_N : a+b = (a+b) \bmod N.$

  - Phép nhân:

$\forall a, b \in Z_N : a \cdot b = (a * b) \bmod N.$

2.5.2.3. Nghịch đảo modulo

- Trên trường số thực R, số nghịch đảo của 5 là 1/5, bởi vì 5 x 1/5=1.
- Trên vành số nguyên Z_N khái niệm về số nghịch đảo của một số như sau:
- Giả sử $a \in Z_N$ và $\exists\ b \in Z_N$ sao cho $a.b \equiv 1 \bmod N$. Khi đó b là duy nhất và được gọi là nghịch đảo của a trên trường Z_N và ký hiệu là $a^{-1} = b$.
- Việc tìm phần tử nghịch đảo của một số $a \in Z_N$ thực chất là tìm hai số b và k sao cho: $a.b = k.N + 1$ trong đó $b, k \in Z_N$. Hay viết gọn lại là:

$$a^{-1} \equiv b \pmod N$$

- Định lý về sự tồn tại của phần tử nghịch đảo:

  Nếu gcd(a, N) = 1 thì tồn tại duy nhất 1 số $b \in Z_N$ là phần tử nghịch đảo của a, nghĩa là thỏa mãn a.b = (a*b) mod N = 1.

Lúc này phương trình đồng dư có dạng :

$$a*b - 1 = kN \quad ; \text{ trong đó } k \in Z_N$$

2.5.2.3. Hàm Phi_Ơle

- Với mỗi số nguyên N , giá trị của hàm phi Ơle của N là tổng số tất cả các số nguyên $\in Z_N$ và nguyên tố cùng nhau với N .
- Nếu P là một số nguyên tố thì giá trị hàm phi Ơle của P: *Φ(P)* = P – 1 hoặc nếu N = p*q trong đó p và q là hai số nguyên tố thì *Φ (N)* = (p-1)*(q-1).
- Tổng quát :

$$\phi(N) = (p_1 - 1)p_1^{\alpha_1 - 1}(p_2 - 1)p_2^{\alpha_2 - 1}...(p_k - 1)p_k^{\alpha_k - 1}$$

- Đinh lý Ơle phát biểu như sau:

  $\forall$ a $\in$ $Z^*_N = Z_N - \{0\}$ và (a, N) = 1 ta có $a^{\emptyset(N)} \equiv 1 mod(N)$. Có nghĩa $a^{\emptyset(N)}$ chính là giá trị nghịch đảo của a trên Z_N.
- Đinh lý Fermat nhỏ (Trường hợp riêng của định lý Ơle): Nếu P là một số nguyên tố thì

  $\forall$ a $\in$ Z^*_P ta có $a^{P-1} \equiv 1(modP)$.
- Đây là một trong những định lý đẹp nhất của số học.

- Với mỗi số nguyên N vành Z^*_N gồm các phần tử thuộc Z_N và nguyên tố cùng nhau với N, hay nói cách khác: $Z^*_N = \{x: x \in Z_N, (x, N) = 1\} = \{x: x \in Z_N, x^{\varphi(N)} = 1\}$.
- Với mỗi phần tử $a \in Z_N$, bậc t của a (ký hiệu là ord (a)) là số nhỏ nhất sao cho : $a^t = 1$. Theo định lý Ơle ta suy ra $\varphi(N)$ chia hết cho t.
- Ví dụ: N=21 ta có bảng sau

$a \in Z^*_{21}$	1	2	4	5	8	10	11	13	16	17	19	20
Ord(a)	1	6	3	6	2	6	6	2	3	6	6	2

- Nếu bậc của a $\in Z^*_N$ bằng $\varphi(N)$ thì a được gọi là phần tử sinh hay phần tử nguyên thủy của tập Z^*_N và nếu tập Z^*_N chỉ có một phần tử sinh thì nó được gọi là một cyclic.

Ví dụ : N=3 , a=2

$\varphi(N)$ =(N-1) =2 ; (Nϵ P)

Ord(a) = t=2 vì $a^t \bmod N = 2^2 \bmod 3 = 1$

a = $\varphi(N)$ =2 vậy 2 là phần tử nguyên thủy của $Z^*_{(2)}$

2.5.3. Một số thuật giải trên trường modulo

2.5.3.1. Thuật giải Euclic tính gcd của hai số nguyên dương

Input : a,b $\in$N,a>b$\geq$1
Output gcd(a,b)
while b>0 do
r=a%b;a=b;b=r
Return(a)

2.5.3.2. Beazout algorithm:

Tính d=gcd(a,b)và x,y : ax+by=d

Input: a,b nguyên , không âm :a $\geq$b
Output: d=gcd(a,b); x,y:ax+by=d;

1) If b=0 then d=a; x=1;y=0.
2) x_2=1; x_1= 0; y_2=0; y_1=1.
3) while(b>0)do
 a)q=a/b; r=a-q*b ; x=x_2-q*x_1 ; y=y_2-q*y_1;
 b).a=b ; b=r ; x_2=x_1; x_1=x ; y_2=y_1; y_1=y;
4) d=a; x=x_2; y=y_2.
5) Return(d,x,y).

2.5.3.3. Phép lũy thừa modulo

- Định nghĩa
  Cho $x \in Z_m$, và $p \in N^*$; $p = \sum_{0 \le i \le l} p_i^{2i}$; Phép toán $x^p \bmod m$ được gọi là phép lũy thừa modulo.
- Ta có : $x^p = x^{p^0} \times x^{p^2} \times x^{p^4} \times \ldots \times x^{p^l}$
- Thuật giải :
  Input : x $\in$ Zm,
  Output : x^p mod m

  (1) y = 1. Nếu p = 0, Return y.
  (2) A = x. nếu P_0 = 1, thì y = x.
  (3) Cho i chạy từ 1 đến l, Do:
  a. A =A^2 mod m ;
  b. Nếu p_i = 1 thì y = (A*y) mod m.
  (4) Return y.

2.3.5.4. Thuật giải tính modulo nghịch đảo

Input : $a \in Z_N$

Output :tìm $x \equiv a^{-1} (\text{mod} n)$ nếu tồn tại

i) Dùng giải thuật Beazout tính

$x,y \in Z$: $ax+ny=d$ với $gcd=gcd(a,n)$.

ii) If $gcd > 1$,

$a^{-1}(\text{mod } n)$ not exist.

iii) If $gcd = 1$,

Return $x(\text{mod } n)$.

2.5.3.5. Thuật toán lũy thừa nhanh

Input: a, m, N.
Output: a^m mod N.
Begin :
Phân tích m thành dạng nhị phân m = b_k ,b_{k-1}…b_0.
j = 0, kq = a;
while (k>=j)
{
    if (bj==1)
    kq = (kq * a) mod N;
    a = (a * a) mod N;
    j = j + 1;
}
return kq;
end

2.4.3.6.Thuật giải Euclic nhị phân

- Input x,y>0
- Output gcd (x,y)

a. g=1
b. While x,y even ,Do
    i. x=x/2
    ii. y=y/2
    iii. g=2g
c. While(x>0),Do
    i. While x even Do x=x/2.
    ii. While y even Do y=y/2.
    iii. t=|x-y|/2.
    iv. If x≥y Then x=t,else y=t.
d. g=gy.
e. Return g.

- Yêu cầu : nắm vững lý thuyết
- Làm các bài tập trong giờ thực hành (8 tiết học)
- Tham khảo các code trong phần bài tập

HẾT CHƯƠNG 2