Chương 3 giới thiệu các khái niệm cơ bản về mật mã, hệ mã hóa, các phương pháp mã hóa. Phần tiếp theo của chương trình bày một số giải thuật cơ bản của mã hóa khóa đối xứng (DES, 3-DES và AES), mã hóa khóa bất đối xứng (RSA), các hàm băm (MD5 và SHA1), chữ ký số, chứng chỉ số và PKI. Phần cuối của chương đề cập vấn đề quản lý và phân phối khóa, và một số giao thức đảm bảo an toàn thông tin dựa trên mã hóa.

1. Khái quát về mã hóa thông tin và ứng dụng

1.1 Các khái niệm

Mật mã

Theo từ điển Webster's Revised Unabridged Dictionary: "cryptography is the act or art of writing secret characters", hay *mật mã (cryptography) là một hành động hoặc nghệ thuật viết các ký tự bí mật*. Còn theo từ điển Free Online Dictionary of Computing: "cryptography is encoding data so that it can only be decoded by specific individuals", có nghĩa là *mật mã là việc mã hóa dữ liệu mà nó chỉ có thể được giải mã bởi một số người chỉ định*.

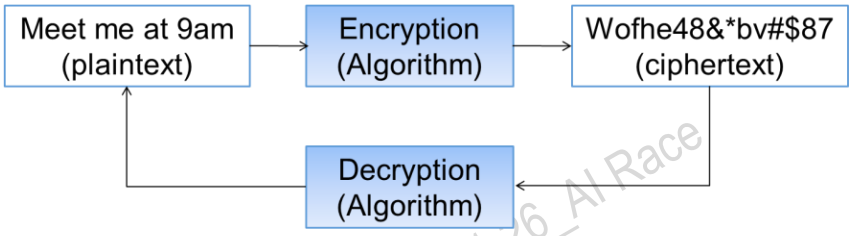
Bản rõ, Bản mã, Mã hóa và Giải mã

Bản rõ (Plaintext), hay thông tin chưa mã hóa (Unencrypted information) là thông tin ở dạng có thể hiểu được.

Bản mã (Ciphertext), hay thông tin đã được mã hóa (Encrypted information) là thông tin ở dạng đã bị xáo trộn.

Mã hóa (Encryption) là hành động xáo trộn (scrambling) bản rõ để chuyển thành bản mã.

Giải mã (Decryption) là hành động giải xáo trộn (unscrambling) bản mã để chuyển thành bản rõ.



Hình 3.1. Các khâu Mã hóa (Encryption) và Giải mã (Decryption) của một hệ mã hóa

Hình 3.1 minh họa các khâu của một hệ mã hóa, trong đó khâu mã hóa thực hiện ở phía người gửi: chuyển bản rõ thành bản mã và khâu giải mã được thực hiện ở phía người nhận: chuyển bản mã thành bản rõ.

Giải thuật mã hóa & giải mã, Bộ mã hóa, Khóa/Chìa, Không gian khóa

Giải thuật mã hóa (Encryption algorithm) là giải thuật dùng để mã hóa thông tin và giải thuật giải mã (Decryption algorithm) dùng để giải mã thông tin.

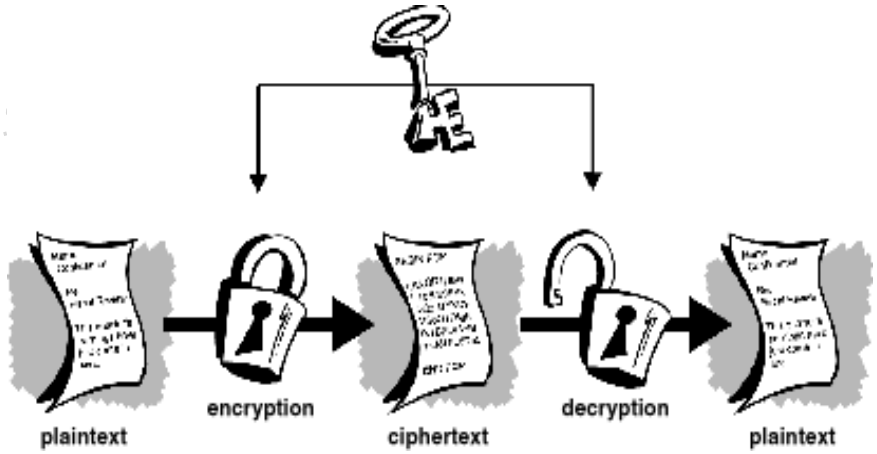
Một bộ mã hóa (Cipher) gồm một giải thuật để mã hóa và một giải thuật để giải mã thông tin.

Khóa/Chìa (Key) là một chuỗi được sử dụng trong giải thuật mã hóa và giải mã.

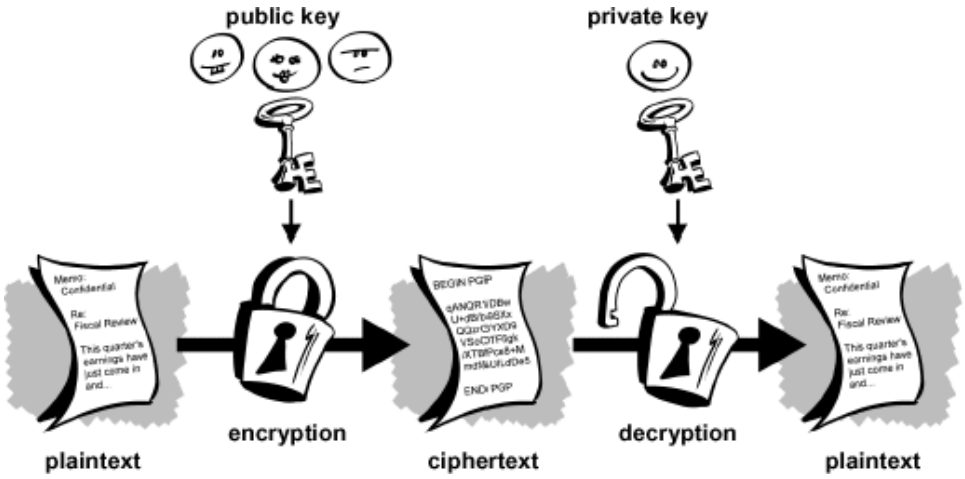
Không gian khóa (Keyspace) là tổng số khóa có thể có của một hệ mã hóa. Ví dụ, nếu sử dụng khóa kích thước 64 bit thì không gian khóa là 2^{64}.

Mã hóa khóa đối xứng, Mã hóa khóa bất đối xứng, Hàm băm, Thám mã

Mã hóa khóa đối xứng (Symmetric key cryptography) là dạng mã hóa trong đó một khóa được sử dụng cho cả khâu mã hóa và khâu giải mã. Do khóa sử dụng chung cần phải được giữ bí mật nên mã hóa khóa đối xứng còn được gọi là mã hóa khóa bí mật (Secret key cryptography). Hình 3.2 minh họa hoạt động của một hệ mã hóa khóa đối xứng, trong đó một khóa bí mật duy nhất được sử dụng cho cả hai khâu mã hóa và giải mã một thông điệp.



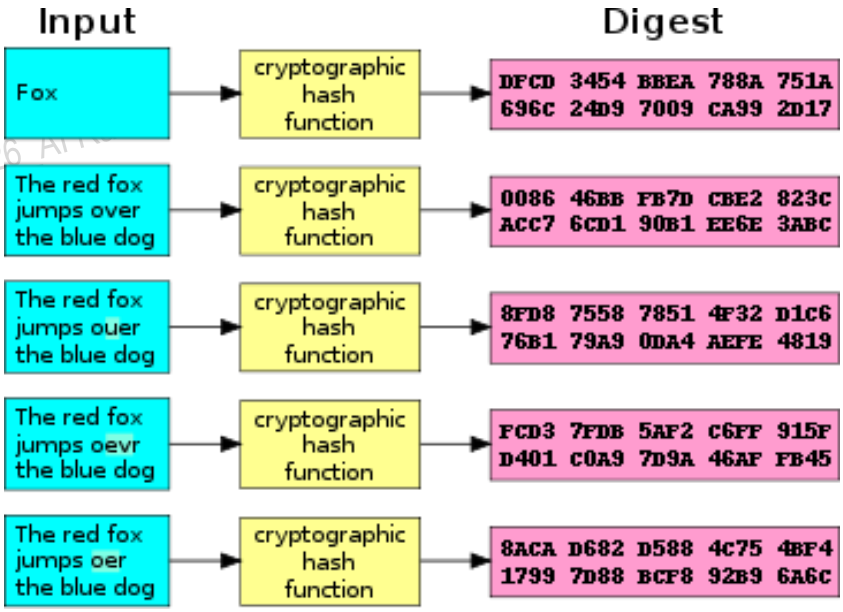
Hình 3.2. Mã hóa khóa đối xứng sử dụng chung 1 khóa bí mật



Hình 3.3. Mã hóa khóa bất đối xứng sử dụng một cặp khóa

Mã hóa khóa bất đối xứng (Asymmetric key cryptography) là dạng mã hóa trong đó một cặp khóa được sử dụng: khóa công khai (public key) dùng để mã hóa, khóa riêng (private key) dùng để giải mã. Chỉ có khóa riêng cần phải giữ bí mật, còn khóa công khai có thể phổ biến rộng rãi. Do khóa để mã hóa có thể công khai nên đôi khi mã hóa khóa bất đối xứng còn được gọi là mã hóa khóa công khai (Public key cryptography). Hình 3.3 minh họa hoạt động của một hệ mã hóa khóa bất đối xứng, trong đó một khóa công khai (public key) được sử dụng cho khâu mã hóa và khóa riêng (private key) cho khâu giải mã thông điệp.

Hàm băm (Hash function) là một ánh xạ chuyển các dữ liệu có kích thước thay đổi về dữ liệu có kích thước cố định. Hình 3.4 minh họa đầu vào (Input) và đầu ra (Digest) của hàm băm. Trong các loại hàm băm, hàm băm 1 chiều (One-way hash function) là hàm băm, trong đó việc thực hiện mã hóa tương đối đơn giản, còn việc giải mã thường có độ phức tạp rất lớn, hoặc không khả thi về mặt tính toán.



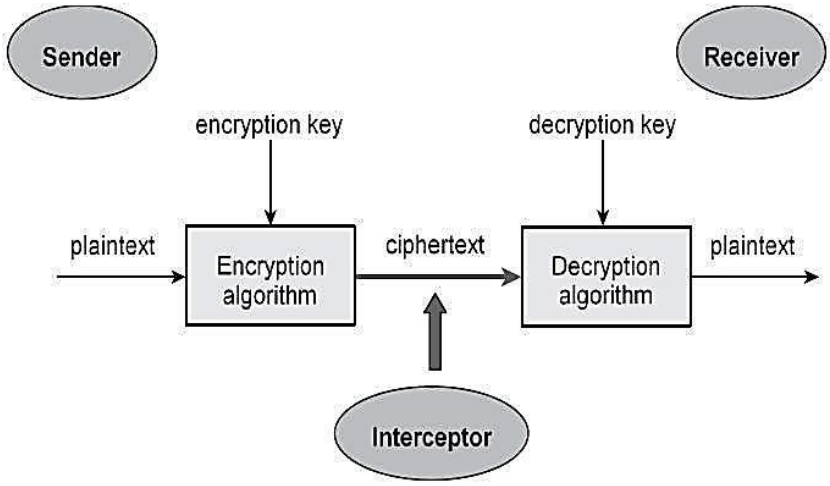
Hình 3.4. Minh họa đầu vào (Input) và đầu ra (Digest) của hàm băm

Thám mã hay phá mã (Cryptanalysis) là quá trình giải mã thông điệp đã bị mã hóa mà không cần có trước thông tin về giải thuật mã hóa và khóa mã.

1.2 Các thành phần của một hệ mã hóa

Một hệ mã hóa hay hệ mật mã (Cryptosystem) là một bản cài đặt của các kỹ thuật mật mã và các thành phần có liên quan để cung cấp dịch vụ bảo mật thông tin. Hình 3.5 nêu các thành phần của một hệ mã hóa đơn giản dùng để đảm bảo tính bí mật của thông tin từ người gửi (Sender) truyền đến người nhận (Receiver) mà không bị một bên thứ ba nghe lén (Interceptor). Các thành phần của một hệ mã

hóa đơn giản gồm bản rõ (plaintext), giải thuật mã hóa (Encryption Algorithm), bản mã (ciphertext), giải thuật giải mã (Decryption Algorithm), khóa mã hóa (encryption key) và khóa giải mã (decryption key). Một thành phần quan trọng khác của một hệ mã hóa là không gian khóa (Keyspace) - là tập hợp tất cả các khóa có thể có. Ví dụ, nếu chọn kích thước khóa là 64 bit thì không gian khóa sẽ là 2^{64}. Nhìn chung, hệ mã hóa có độ an toàn càng cao nếu không gian khóa lựa chọn càng lớn.



Hình 3.5. Các thành phần của một hệ mã hóa đơn giản

1.3 Lịch sử mã hóa

Có thể nói mã hóa hay mật mã là con đẻ của toán học nên sự phát triển của mật mã đi liền với sự phát triển của toán học. Tuy nhiên, do nhiều giải thuật mật mã đòi hỏi khối lượng tính toán lớn nên mật mã chỉ thực sự phát triển mạnh cùng với sự ra đời và phát triển của máy tính điện tử. Sau đây là một số mốc trong sự phát triển của mật mã và ứng dụng mật mã:

- Các kỹ thuật mã hoá thô sơ đã được người cổ Ai cập sử dụng cách đây 4000 năm.
- Người cổ Hy lạp, Ấn độ cũng đã sử dụng mã hoá cách đây hàng ngàn năm.
- Các kỹ thuật mã hoá chỉ thực sự phát triển mạnh từ thế kỷ 1800 nhờ công cụ toán học, và phát triển vượt bậc trong thế kỷ 20 nhờ sự phát triển của máy tính và ngành công nghệ thông tin.
- Trong chiến tranh thế giới thứ I và II, các kỹ thuật mã hóa được sử dụng rộng rãi trong liên lạc quân sự sử dụng sóng vô tuyến. Quân đội các nước đã sử dụng các công cụ phá mã, thám mã để giải mã các thông điệp của quân địch.
- Năm 1976 chuẩn mã hóa DES (Data Encryption Standard) được Cơ

quan mật vụ Mỹ (NSA – National Security Agency) thừa nhận và sử dụng rộng rãi.

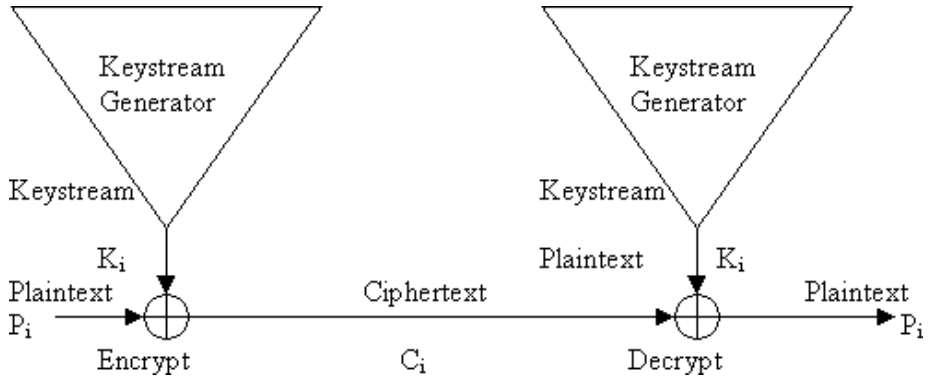
- Năm 1976, hai nhà khoa học Whitman Diffie và Martin Hellman đã đưa ra khái niệm mã hóa khóa bất đối xứng (Asymmetric key cryptography), hay mã hóa khóa công khai (Public key cryptography) đưa đến những thay đổi lớn trong kỹ thuật mật mã. Theo đó, các hệ mã hóa khóa công khai bắt đầu được sử dụng rộng rãi nhờ khả năng hỗ trợ trao đổi khóa dễ dàng hơn và do các hệ mã hóa khóa bí mật gặp khó khăn trong quản lý và trao đổi khóa, đặc biệt khi số lượng người dùng lớn.
- Năm 1977, ba nhà khoa học Ronald Rivest, Adi Shamir, và Leonard Adleman giới thiệu giải thuật mã hóa khóa công khai RSA. Từ đó, RSA trở thành giải thuật mã

hóa khóa công khai được sử dụng rộng rãi nhất do RSA có thể vừa được sử dụng để mã hóa thông tin và sử dụng trong chữ ký số.

- Năm 1991, phiên bản đầu tiên của PGP (Pretty Good Privacy) ra đời.
- Năm 2000, chuẩn mã hóa AES (Advanced Encryption Standard) được thừa nhận và ứng dụng rộng rãi.

1.4 Mã hóa dòng và mã hóa khối

1.4.1 Mã hóa dòng

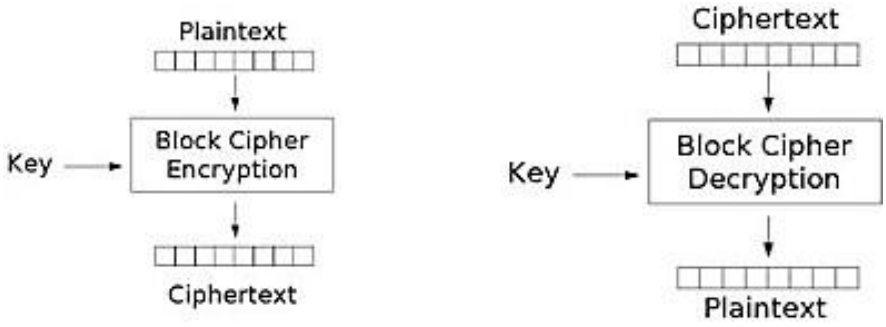


Hình 3.6. Mã hóa dòng (Stream cipher)

Mã hóa dòng (Stream cipher) là kiểu mã hóa mà từng bit, hoặc ký tự của bản rõ được kết hợp với từng bit, hoặc ký tự tương ứng của khóa để tạo thành bản mã. Hình 3.6 biểu diễn quá trình mã hóa (Encrypt) và giải mã (Decrypt) trong mã hóa dòng. Theo đó, ở bên gửi các bit P_i của bản rõ (plaintext) được liên tục đưa vào kết hợp với bit tương ứng K_i của khóa để tạo thành bit mã C_i; Ở bên nhận, bit mã C_i được kết hợp với bit khóa C_i để khôi phục bit rõ P_i. Một bộ sinh dòng khóa

(Keystream Generator) được sử dụng để liên tục sinh các bit khóa K_i từ khóa gốc K. Các giải thuật mã hóa dòng tiêu biểu như A5, hoặc RC4 được sử dụng rộng rãi trong viễn thông.

1.4.2 Mã hóa khối



Hình 3.7. Mã hóa khối (Block cipher)

Mã hóa khối (Block cipher) là kiểu mã hóa mà dữ liệu được chia ra thành từng khối có kích thước cố định để mã hóa và giải mã. Hình 3.7 biểu diễn quá trình mã hóa và giải mã trong mã hóa khối. Theo đó, ở bên gửi bản rõ (Plaintext) được chia thành các khối

(block) có kích thước cố định, sau đó từng khối được mã hóa để chuyển thành khối mã. Các khối mã được ghép lại thành bản mã (Ciphertext). Ở bên nhận, bản mã lại được chia thành các khối và từng lại được giải mã để chuyển thành khối rõ. Cuối cùng ghép các khối rõ để có bản rõ hoàn chỉnh. Các giải thuật mã hóa khối tiêu biểu như DES, 3-DES, IDEA, AES được sử dụng rất rộng rãi trong mã hóa dữ liệu với kích thước khối 64, hoặc 128 bit.

1.5 Ứng dụng của mã hóa

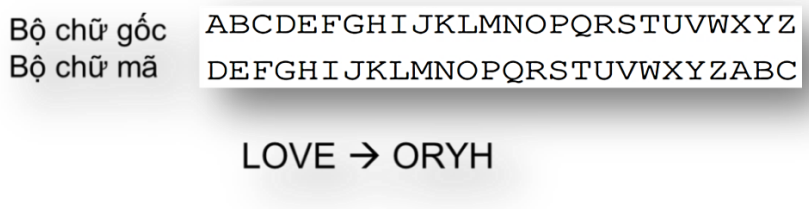
Mã hoá thông tin có thể được sử dụng để đảm bảo an toàn thông tin với các thuộc tính: bí mật (confidentiality), toàn vẹn (integrity), xác thực (authentication), không thể chối bỏ (non-repudiation). Cụ thể, các kỹ thuật mã hóa được ứng dụng rộng rãi trong các hệ thống, công cụ và dịch vụ bảo mật như:

- Dịch vụ xác thực (Kerberos, SSO, RADIUS,…)
- Điều khiển truy nhập
- Các công cụ cho đảm bảo an toàn cho truyền thông không dây
- Các nền tảng bảo mật như PKI, PGP
- Các giao thức bảo mật như SSL/TLS, SSH, SET, IPSec
- Các hệ thống bảo mật kênh truyền, như VPN.

2. Các phương pháp mã hóa

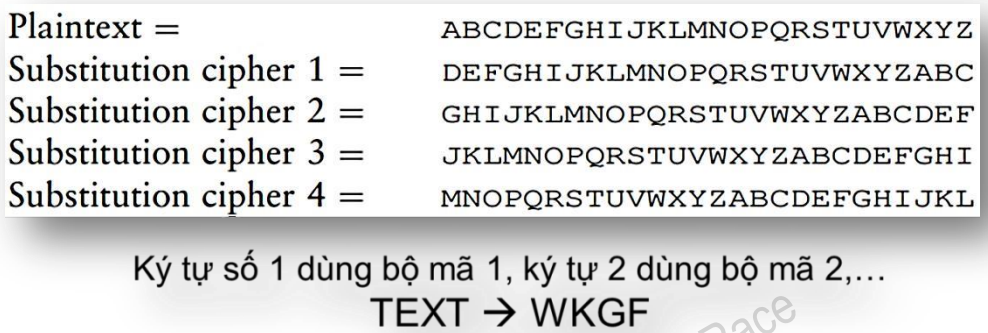
2.1 Phương pháp thay thế

Phương pháp thay thế (Substitution) là phương pháp thay thế một giá trị này bằng một giá trị khác, như thay một ký tự bằng một ký tự khác, hoặc thay một bit bằng một bit khác. Hình 3.8 biểu diễn bộ chữ gốc, bộ chữ mã và ví dụ mã hóa sử dụng hệ mã hóa nổi tiếng thời La Mã là Caesar cipher. Nguyên tắc của Caesar cipher là dịch 3 chữ trong bộ ký tự tiếng Anh sang bên phải (A→D, B→E, C→F,….). Bản rõ "LOVE" được mã hóa thành "ORYH".



Hình 3.8. Mã hóa bằng hệ mã hóa Caesar cipher

Để tăng độ an toàn của phương pháp thay thế, người ta có thể sử dụng nhiều bộ chữ mã, như minh họa trên Hình 3.9 với 4 bộ chữ mã (Substitution cipher), với nguyên tắc thay thế: ký tự số 1 ở bản rõ thay thế sử dụng bộ chữ mã số 1, ký tự số 2 sử dụng bộ chữ mã số 2,…, ký tự số 5 sử dụng bộ chữ mã số 1, ký tự số 6 sử dụng bộ chữ mã số 2,… Nếu các bộ chữ mã được sắp đặt ngẫu nhiên thì một ký tự xuất hiện ở các vị trí khác nhau trong bản rõ sẽ được chuyển đổi thành các ký tự khác nhau trong bản mã. Điều này giúp tăng độ an toàn do làm tăng độ khó trong việc phân tích đoán bản rõ từ bản mã.

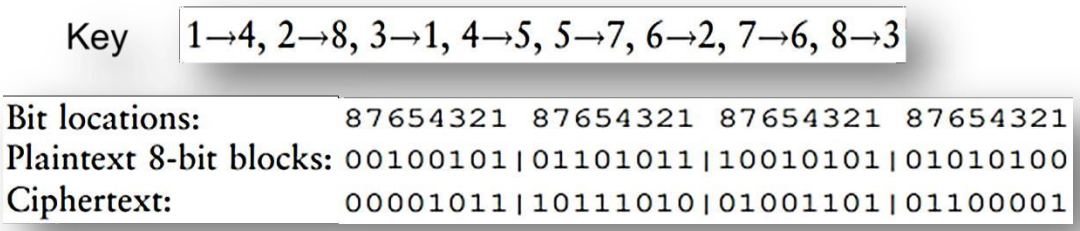


Hình 3.9. Phương pháp thay thế với 4 bộ chữ mã

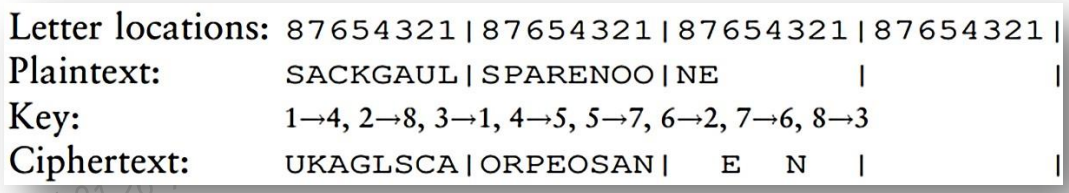
2.2 Phương pháp hoán vị

Phương pháp hoán vị, hoặc đổi chỗ (permutation) thực hiện sắp xếp lại các giá trị trong một khối bản rõ để tạo bản mã. Thao tác hoán vị có thể thực hiện với từng bit hoặc từng byte (ký tự). Hình 3.10 minh họa ví dụ mã hóa bằng phương pháp hoán vị thực hiện đổi chỗ các bit, trong đó việc đổi chỗ được thực hiện theo

khóa (Key) trong khối 8 bit, tính từ bên phải. Hình 3.11 minh họa ví dụ mã hóa bằng phương pháp hoán vị thực hiện đổi chỗ các ký tự, trong đó việc đổi chỗ được thực hiện theo khóa trong khối 8 ký tự, tính từ bên phải. Với bản rõ "SACKGAULSPARENOONE" ta có 3 khối, 2 khối đầu đủ 8 ký tự, còn khối cuối chỉ có 2 ký tự "NE" nên phải chèn thêm dấu trắng cho đủ khối 8 ký tự.



Hình 3.10. Phương pháp hoán vị thực hiện đổi chỗ các bit



Hình 3.11. Phương pháp hoán vị thực hiện đổi chỗ các ký tự

2.3 Phương pháp XOR

Phương pháp mã hóa XOR sử dụng phép toán logic XOR để tạo bản mã, trong đó từng bit của bản rõ được XOR với bit tương ứng của khóa. Để giải mã, ta thực hiện XOR từng bit của bản mã với bit tương ứng của khóa. Hình 3.12 minh họa quá trình mã hóa

bản rõ "CAT" với khóa "VVV". Theo đó, các ký tự của bản rõ và khóa được chuyển thành mã ASCII và biểu diễn dưới dạng nhị phân. Sau đó, thực hiện phép toán XOR trên các bit tương ứng của bản rõ và khóa để tạo bản mã (Cipher).

Text Value	Binary Value
CAT as bits	0 1 0 0 0 0 1 1 0 1 0 0 0 0 0 1 0 1 0 1 0 1 0 0
VVV as key	0 1 0 1 0 1 1 0 0 1 0 1 0 1 1 0 0 1 0 1 0 1 1 0
Cipher	0 0 0 1 0 1 0 1 0 0 0 1 0 1 1 1 0 0 0 0 0 0 1 0

Hình 3.12. Mã hóa bằng phương pháp XOR

2.4 Phương pháp Vernam

Phương pháp Vernam sử dụng một tập ký tự để nối vào các ký tự của bản rõ để tạo bản mã. Tập ký tự này được gọi là *one-time pad* và mỗi ký tự trong tập chỉ dùng 1 lần trong một tiến trình mã hóa. Với bộ chữ tiếng Anh có 26 chữ, mã hóa bằng phương pháp Vernam được thực hiện như sau:

- Các ký tự của bản rõ và các ký tự của tập nối thêm (*one-time pad*) được chuyển thành số trong khoảng 1-26;
- Cộng giá trị của ký tự trong bản rõ với giá trị tương ứng trong tập nối thêm;
- Nếu giá trị cộng lớn hơn 26 thì đem trừ cho 26 (đây chính là phép modulo – chia lấy phần dư).
- Chuyển giá trị số thành ký tự mã.

Plaintext:	S	A	C	K	G	A	U	L	S	P	A	R	E	N	O	O	N	E
Plaintext value:	19	01	03	11	07	01	21	12	19	16	01	18	05	14	15	15	14	05
One-time pad text:	F	P	Q	R	N	S	B	I	E	H	T	Z	L	A	C	D	G	J
One time pad value:	06	16	17	18	14	19	02	09	05	08	20	26	12	01	03	04	07	10
Sum of plaintext and pad:	25	17	20	29	21	20	23	21	24	24	21	44	17	15	18	19	21	15
After modulo Subtraction:				03								18						
Ciphertext:	Y	Q	T	C	U	T	W	U	X	X	U	R	Q	O	R	S	U	O

Hình 3.13. Mã hóa bằng phương pháp Vernam

Hình 3.13 minh họa mã hóa bản rõ "SACKGAULSPARENOONE" bằng phương pháp Vernam với tập nối thêm "FPQRNSBIEHTZLACDGJ".

2.5 Phương pháp sách hoặc khóa chạy

Phương pháp sách, hoặc khóa chạy thực hiện việc mã hóa và giải mã sử dụng các khóa mã chứa trong các cuốn sách. Hiện nay phương pháp này thường được dùng trong các bộ phim trinh thám do tính chất kỳ bí của nó. Ví dụ như, với bản mã "259,19,8; 22,3,8; 375,7,4; 394,17,2" và cuốn sách được dùng chứa khóa là "A Fire Up on the Deep", ta có thể giải mã như sau:

- Trang 259, dòng 19, từ thứ 8 là *sack*
- Trang 22, dòng 3, từ thứ 8 là *island*
- Trang 375, dòng 7, từ thứ 4 là *sharp*
- Trang 394, dòng 17, từ thứ 2 là *path*

Bản rõ tương ứng của bản mã "259,19,8;22,3,8;375,7,4;394,17,2" là "sack island sharp path".

2.6 Phương pháp hàm băm

Các hàm băm (Hash functions) là các giải thuật để tạo các bản tóm tắt (digest) của thông điệp, thường được sử dụng để nhận dạng và đảm bảo tính toàn vẹn của thông điệp. Độ dài của thông điệp đầu vào là bất kỳ, nhưng đầu ra hàm băm

thường có độ dài cố định. Chi tiết về các hàm băm được ở mục 3.3.3. Các hàm băm thông dụng gồm:

- Các hàm băm MD2, MD4, MD5 với độ dài chuỗi đầu ra là 128 bit;
- Hàm băm MD6 cho chuỗi đầu ra có độ dài trong khoảng 0 đến 512 bit;
- Các hàm băm SHA0, SHA1 với độ dài chuỗi đầu ra là 160 bit;
- Các hàm băm SHA2, gồm SHA256, SHA384, SHA512 cho phép một số lựa chọn chuỗi đầu ra tương ứng 256, 384 và 512 bit;
- Hàm băm SHA3 cho chuỗi đầu ra có độ dài trong khoảng 0 đến 512 bit;
- Hàm băm CRC32 với chuỗi đầu ra 32 bit sử dụng trong kiểm tra dư thừa mạch vòng.