CHƯƠNG 2: MÃ CỔ ĐIỂN

2.1. Các khái niệm về mã đối xứng

2.1.1. Mật mã đối xứng

Mật mã đối xứng sử dụng cùng một khóa cho việc mã hóa và giải mã. Có thể nói mã đối xứng là mã một khóa hay mã khóa chia sẻ. Ở đây người gửi và người nhận chia sẻ khóa chung K, mà họ có thể trao đổi bí mật với nhau. Ta xét hai hàm ngược nhau: E là hàm mã hóa biến đổi bản rõ thành bản mã và D là hàm giải mã biến đổi bản mã trở về bản rõ. Giả sử X là văn bản cần mã hóa gọi là bản rõ và Y là dạng văn bản đã được thay đổi qua việc mã hóa gọi là bản mã. Khi đó ta ký hiệu:

$$Y = E_K(X)$$

$$X = D_K(Y)$$

Mọi thuật toán mã cổ điển đều là mã khóa đối xứng, vì ở đó thông tin về khóa được chia sẻ giữa người gửi và người nhận. Mã đối xứng là kiểu duy nhất trước khi phát minh ra khóa mã công khai vào những năm 1970, **mã công khai** còn được gọi là mã không đối xứng. Hiện nay các mã đối xứng và công khai tiếp tục phát triển và hoàn thiện. Mã công khai ra đời hỗ trợ mã đối xứng chứ không thay thế nó, do đó mã đối xứng đến nay vẫn được sử dụng rộng rãi.

Sau đây ta đưa ra định nghĩa một số khái niệm cơ bản về mã hóa.

Bản rõ X là bản tin gốc. Bản rõ có thể được chia nhỏ để có kích thước phù hợp.

Bản mã Y là bản tin gốc đã được mã hoá. Nói chung kích thước bản mã không nhỏ hơn kích thước bản rõ. Nhưng ở đây ta thường xét phương pháp mã hóa mà không làm thay đổi kích thước của bản rõ, tức là chúng có cùng độ dài.

Mã là thuật toán E chuyển bản rõ thành bản mã. Thông thường chúng ta cần thuật toán mã hóa mạnh, cho dù kẻ thù biết được thuật toán, nhưng khi không biết thông tin về khóa, cũng không tìm được bản rõ.

Khóa K là thông tin tham số dùng để mã hoá, chỉ có người gửi và người nhận biết. Khóa là độc lập với bản rõ và có độ dài phù hợp với yêu cầu bảo mật.

Mã hoá là quá trình chuyển bản rõ thành bản mã, thông thường bao gồm việc áp dụng thuật toán mã hóa và một số quá trình xử lý thông tin kèm theo.

Giải mã chuyển bản mã thành bản rõ, đây là quá trình ngược lại của mã hóa.

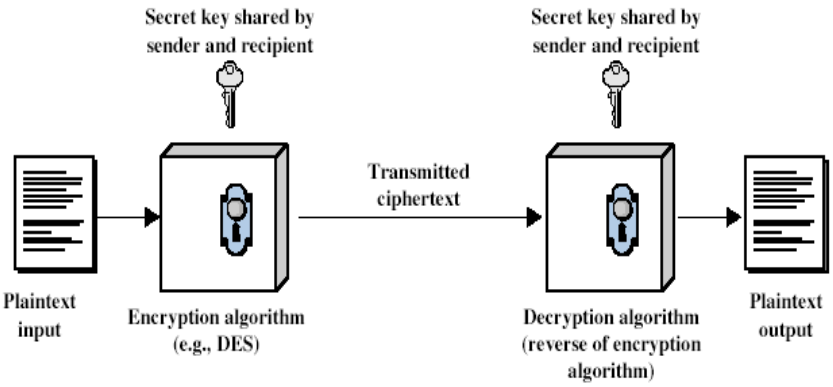
Mật mã là chuyên ngành khoa học của Khoa học máy tính nghiên cứu về các nguyên lý và phương pháp mã hoá. Hiện nay người ta đưa ra nhiều chuẩn an ninh cho các lĩnh vực khác nhau của công nghệ thông tin.

Thám mã nghiên cứu các nguyên lý và phương pháp giải mã mà không biết khóa. Thông thường khi đưa các mã mạnh ra làm chuẩn dùng chung giữa các người sử dụng, các mã đó đã được các kẻ thám mã cũng như những người phát triển mã tìm hiểu nghiên cứu kỹ về các phương pháp giải một phần bản mã với các thông tin không đầy đủ.

Lý thuyết mã bao gồm cả mật mã và thám mã. Nó là một thể thống nhất, để đánh giá một mã mạnh hay không, đều phải xét từ cả hai khía cạnh đó. Các nhà khoa học mong

muốn tìm ra các mô hình mã hóa khái quát cao đáp ứng nhiều chính sách an ninh khác nhau.

Mô hình mã đối xứng



2.1.2. Các yêu cầu

Một mã đối xứng có các đặc trưng trong cách xử lý thông tin của thuật toán mã, giải mã, tác động của khóa vào bản mã, độ dài của khóa. Mối liên hệ giữa bản rõ, khóa và bản mã càng phức tạp càng tốt, nếu tốc độ tính toán là chấp nhận được. Cụ thể hai yêu cầu để sử dụng an toàn mã khóa đối xứng là:

- Thuật toán mã hoá mạnh; có cơ sở toán học vững chắc đảm bảo rằng mặc dù công khai thuật toán, mọi người đều biết, nhưng việc thám mã là rất khó khăn và phức tạp, nếu không biết khóa.
- Khóa mật chỉ có người gửi và người nhận biết; có kênh an toàn để phân phối khóa giữa các người sử dụng chia sẻ khóa. Mối liên hệ giữa khóa và bản mã là không thể nhận biết được.

2.1.3. Hệ mật mã.

Hệ mật mã được đặc trưng bởi các yếu tố sau :

- Kiểu của thao tác mã hoá được sử dụng trên bản rõ:
  - Phép thế - thay thế các ký tự trên bản rõ bằng các ký tự khác trên bản mã.
  - Hoán vị - thay đổi vị trí các ký tự trong bản rõ, tức là thực hiện hoán vị các ký tự của bản rõ.
  - Tích của chúng, tức là kết hợp cả hai kiểu thay thế và hoán vị các ký tự của bản rõ.
- Số khóa được sử dụng khi mã hóa và giải mã: một khóa duy nhất - khóa đối xứng hoặc hai khóa - khóa không đối xứng. Ngoài ra còn xem xét số khóa có thể được dùng có nhiều không. Khóa càng nhiều, thì việc mò tìm khóa càng lâu.

Một đặc trưng của mã nữa là cách mà bản rõ được xử lý, theo:

- Khối - dữ liệu được chia thành từng khối có kích thước xác định và áp dụng thuật toán mã hóa với tham số khóa cho từng khối.
- Dòng - từng đơn vị thông tin đầu vào thường là bit hoặc byte được xử lý liên tục tạo phần tử đầu ra tương ứng.

2.1.4. Tìm duyệt tổng thể (Brute-Force)

Về mặt lý thuyết phương pháp duyệt tổng thể là luôn thực hiện được, do có thể tiến hành thử từng khóa, mà số khóa là hữu hạn. Phần lớn công sức của các tấn công đều tỷ lệ thuận với kích thước khóa. Khóa càng dài thời gian tìm kiếm càng lâu và thường tăng theo hàm mũ. Ta có thể giả thiết là kẻ thám mã có thể dựa vào bối cảnh để biết hoặc nhận biết được bản rõ.

Sau đây là một số thống kê về mối liên hệ giữa độ dài khóa, kích thước không gian khóa, tốc độ xử lý và thời gian tìm duyệt tổng thể. Chúng ta nhận thấy với độ dài khóa từ 128 bit trở lên, thời gian yêu cầu là rất lớn, lên đến hàng tỷ năm, như vậy có thể coi phương pháp duyệt tổng thể là không hiện thực.

Key Size (bits)	Number of Alternative Keys	Time required at 1 encryption/μs	Time required at 10^6 encryptions/μs
32	$2^{32} = 4.3 \times 10^9$	$2^{31}\ \mu s = 35.8$ minutes	2.15 milliseconds
56	$2^{56} = 7.2 \times 10^{16}$	$2^{55}\ \mu s = 1142$ years	10.01 hours
128	$2^{128} = 3.4 \times 10^{38}$	$2^{127}\ \mu s = 5.4 \times 10^{24}$ years	5.4×10^{18} years
168	$2^{168} = 3.7 \times 10^{50}$	$2^{167}\ \mu s = 5.9 \times 10^{36}$ years	5.9×10^{30} years
26 characters (permutation)	$26! = 4 \times 10^{26}$	$2 \times 10^{26}\ \mu s = 6.4 \times 10^{12}$ years	6.4×10^{6} years

2.1.5. Độ an toàn.

Có thể phân loại an toàn thành hai kiểu như sau:

An toàn không điều kiện: Ở đây cho dù máy tính thực hiện được bao nhiêu phép toán trong một giây, mã hoá không thể bị bẻ, vì bản mã không cung cấp đủ thông tin để xác định duy nhất bản rõ. Việc dùng bộ đệm ngẫu nhiên một lần làm khóa để mã dòng cho dữ liệu mà ta sẽ xét cuối bài này được coi là an toàn không điều kiện. Ngoài ra chưa có thuật toán mã hóa nào được coi là an toàn không điều kiện.

An toàn tính toán: Với nguồn lực máy tính có giới hạn và thời gian có hạn (chẳng hạn thời gian tính toán không quá tuổi của vũ trụ) mã hoá coi như không thể bị bẻ. Trong trường hợp này không quan trọng máy tính mạnh như thế nào, có thể coi như mã hóa an toàn về mặt tính toán. ***Nói chung từ nay về sau, một thuật toán mã hóa mà an toàn tính toán, sẽ được coi là an toàn.***

2.2. Mã cổ điển

Mã hoá cổ điển là phương pháp mã hoá đơn giản nhất xuất hiện đầu tiên trong lịch sử mã hoá. Thuật toán đơn giản và dễ hiểu. Những phương pháp mã hoá này là cở sở cho việc nghiên cứu và phát triển thuật toán mã hoá đối xứng được sử dụng ngày nay.

Mọi mã cổ điển đều là mã đối xứng và có hai loại mã cổ điển là mã thay thế và mã hoán vị (hay còn gọi là dịch chuyển):

Mã thay thế là phương pháp mà từng kí tự (nhóm kí tự) trong bản rõ được thay thế bằng một kí tự (một nhóm kí tự) khác để tạo ra bản mã. Bên nhận chỉ cần thay thế ngược lại trên bản mã để có được bản rõ ban đầu.

Mã hoán vị là phương pháp mà các kí tự trong bản rõ vẫn được giữ nguyên, chúng chỉ được sắp xếp lại vị trí để tạo ra bản mã. Tức là các kí tự trong bản rõ hoàn toàn không bị thay đổi bằng kí tự khác mà chỉ đảo chỗ của chúng để tạo thành bản mã.

Trước hết ta xét các mã cổ điển sử dụng phép thay thế các chữ của bản rõ bằng các chữ khác của bảng chữ để tạo thành bản mã.

Ở đây các chữ của bản rõ được thay bằng các chữ hoặc các số hoặc các ký tự khác.

Hoặc nếu xem bản rõ như môt dãy bit, thì phép thế thay các mẫu bit bản rõ bằng các mẫu bit bản mã.

2.2.1. Mã Ceasar

Đây là mã thế được biết sớm nhất, được nghĩ ra bởi Julius Ceasar. Lần đầu tiên được sử dụng trong quân sự. Việc mã hoá được thực hiện đơn giản là thay mỗi chữ trong bản rõ bằng chữ thứ ba tiếp theo trong bảng chữ cái.

Ví dụ. Mã bản rõ: "Meet me after the toga party" bằng bản mã: "`PHHW PH DIWHU WKH WRJD SDUWB`".

Ở đây thay chữ m bằng chữ đứng thứ 3 sau m là p (vì thứ tự từ điển từ m là: m, n, o, p); thay chữ e bằng chữ đứng thứ 3 sau e là h (vì thứ tự từ điển từ e là e, f, g, h). Ta viết các chữ trong bản mã bằng chữ in hoa cho dễ phân biệt bản rõ với bản mã.

Có thể định nghĩa việc mã hoá trên qua ánh xạ trên bảng chữ cái như sau: các chữ ở dòng dưới là mã của các chữ tương ứng ở dòng trên:

```
a b c d e f g h i j k l m n o p q r s t u v w x y z
D E F G H I J K L M N O P Q R S T U V W X Y Z A B C
```

Về toán học, nếu ta gán số thứ tự cho mỗi chữ trong bảng chữ cái, bắt đầu từ thứ tự 0, thì các chữ ở dòng trên có số thứ tự tương ứng là số ở dòng dưới:

```
a b c d e f g h i j k  l  m
0 1 2 3 4 5 6 7 8 9 10 11 12

n  o  p  q  r  s  t  u  v  w  x  y  z
13 14 15 16 17 18 19 20 21 22 23 24 25
```

Khi đó mã Ceasar được định nghĩa qua phép tịnh tiến các chữ như sau:

$$c = E(p) = (p + k) \bmod (26)$$

$$p = D(c) = (c - k) \bmod (26)$$

Ở đây, p là số thứ tự của chữ trong bản rõ và c là số thứ tự của chữ tương ứng của bản mã; k là khóa của mã Ceasar. Khóa k là số bước tịnh tiến các chữ trong bảng chữ. Do đó có 26 khóa khác nhau. Độ dài khóa biểu diễn qua bit ở đây là 5, vì đó là số bit ít nhất cần thiết để biểu diễn 26 giá trị khác nhau.

Thám mã Ceasar là việc làm đơn giản, do số khóa có thể có là rất ít. Chỉ có 26 khóa có thể, vì a chỉ có thể ánh xạ vào một trong số 26 chữ cái của bảng chữ cái tiếng Anh: A, B, C, … Các chữ khác sẽ được xác định bằng số bước tịnh tiến tương ứng của a. Kẻ thám mã có thể thử lần lượt từng khóa một, tức là sử dụng phương pháp tìm duyệt tổng thể. Vì số khóa ít nên việc tìm duyệt là khả thi. Cho trước bản mã, thử 26 cách dịch chuyển khác nhau, ta sẽ đoán nhận thông qua nội dung các bản rõ nhận được.

Ví dụ. Bẻ bản mã: "GCUA VQ DTGCM" bằng cách thử các phép tịnh tiến khác nhau của bảng chữ, ta chọn được bước tịnh tiến thích hợp là 24 và cho bản rõ là "easy to break".

2.2.2. Các bảng mã chữ đơn

Bây giờ ta khắc phục nhược điểm của mã Ceasar bằng cách mã hoá các chữ không chỉ là dịch chuyển bảng chữ, mà có thể tạo ra các bước nhảy khác nhau cho các chữ. Trong một mã mỗi chữ của bản rõ được ánh xạ đến một chữ khác nhau của bản mã. Do đó mỗi cách mã như vậy sẽ tương ứng với một hoán vị của bảng chữ và hoán vị đó chính là khoá của mã đã cho. Như vậy độ dài khoá ở đây là 26 và số khoá có thể có là 26!. Số khoá như vậy là rất lớn.

Ví dụ. Ta có bản mã tương ứng với bản rõ trong mã bảng chữ đơn như sau:

```
Plain:  abcdefghijklmnopqrstuvwxyz
Cipher: DKVQFIBJWPESCXHTMYAUOLRGZN
Plaintext:  ifwewishtoreplaceletters
Ciphertext: WIRFRWAJUHYFTSDVFSFUUFYA
```

- Tính an toàn của mã trên bảng chữ đơn. Tổng cộng có 26! xấp xỉ khoảng 4×10^{26} khoá. Với khá nhiều khoá như vậy nhiều người nghĩ là mã trên bảng chữ đơn sẽ an toàn. Nhưng không phải như vậy. Vấn đề ở đây là do các đặc trưng về ngôn ngữ. Tuy có số lượng khoá lớn, nhưng do các đặc trưng về tần suất xuất hiện của các chữ trong bản rõ và các chữ tương ứng trong bản mã là như nhau, nên kẻ thám mã có thể đoán được ánh xạ của một số chữ và từ đó mò tìm ra chữ mã cho các chữ khác. Ta sẽ xét khía cạnh này cụ thể trong mục sau.

- Tính dư thừa của ngôn ngữ và thám mã. Ngôn ngữ của loài người là dư thừa. Có một số chữ hoặc các cặp chữ hoặc bộ ba chữ được dùng thường xuyên hơn các bộ chữ cùng độ dài khác. Chẳng hạn như các bộ chữ sau đây trong tiếng Anh "th lrd s m shphrd shll nt wnt". Tóm lại trong nhiều ngôn ngữ các chữ không được sử dụng thường xuyên như nhau. Trong tiếng Anh chữ E được sử dụng nhiều nhất; sau đó đến các chữ T, R, N, I, O, A, S. Một số chữ rất ít dùng như: Z, J, K, Q, X. Bằng phương pháp thống kê, ta có thể xây dựng các bảng các tần suất các chữ đơn, cặp chữ, bộ ba chữ.

- Sử dụng bảng tần suất vào việc thám mã

Điều quan trọng là mã thế trên bảng chữ đơn không làm thay đổi tần suất tương đối của các chữ, có nghĩa là ta vẫn có bảng tần suất trên nhưng đối với bảng chữ mã tương ứng. Điều đó được phát hiện bởi các nhà khoa học Ai cập từ thế kỷ thứ 9. Do đó có cách thám mã trên bảng chữ đơn như sau:

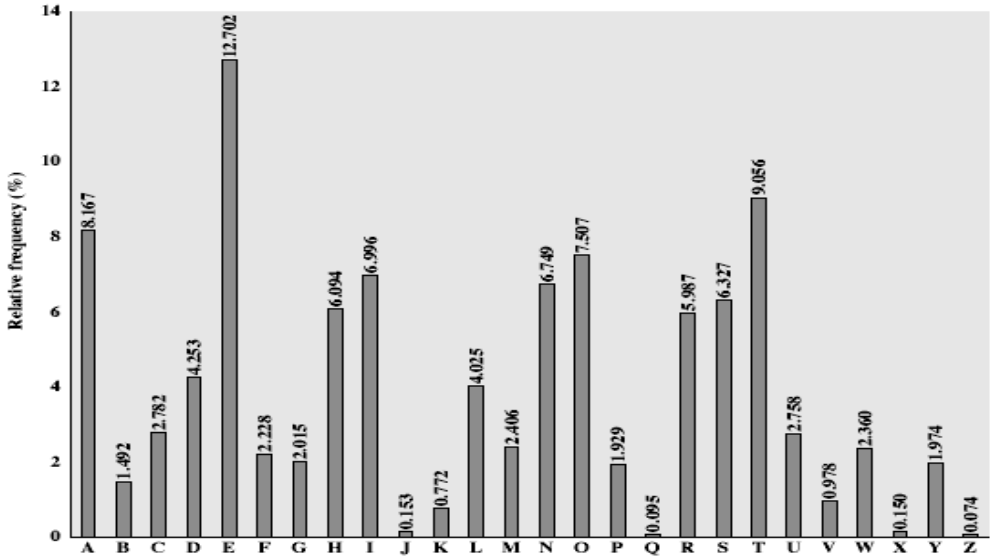
- Tính toán tần suất của các chữ trong bản mã

- So sánh với các giá trị đã biết

- Tìm kiếm các chữ đơn hay dùng A-I-E, bộ đôi NO và bộ ba RST; và các bộ ít dùng JK, X-Z..

- Trên bảng chữ đơn cần xác định các chữ dùng các bảng bộ đôi và bộ ba trợ giúp.

Bảng tần suất chữ cái tiếng Anh:



Ví dụ. Thám mã bản mã trên bảng chữ đơn, cho bản mã:

UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZ

VUEPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSXEPYEP

OPDZSZUFPOUDTMOHMQ

- Tính tần suất các chữ
- Đoán P và Z là e và t.
- Khi đó ZW là th và ZWP là the.
- Suy luận tiếp tục ta có bản rõ:

it was disclosed yesterday that several informal but

direct contacts have been made with political

representatives in moscow

2.2.3. Mã Playfair

Như chúng ta đã thấy không phải số khoá lớn trong mã bảng chữ đơn đảm bảo an toàn mã. Một trong các hướng khắc phục là mã bộ các chữ, tức là mỗi chữ sẽ được mã bằng một số chữ khác nhau tùy thuộc vào các chữ mà nó đứng cạnh. Playfair là một trong các mã như vậy, được sáng tạo bởi Charles Wheastone vào năm 1854 và mang tên người bạn là Baron Playfair. Ở đây mỗi chữ có thể được mã bằng một trong 7 chữ khác nhau tùy vào chữ cặp đôi cùng nó trong bản rõ.

Ma trận khoá Playfair. Cho trước một từ làm khoá, với điều kiện trong từ khoá đó không có chữ cái nào bị lặp. Ta lập ma trận Playfair là ma trận cỡ 5 x 5 dựa trên từ khoá đã cho và gồm các chữ trên bảng chữ cái, được sắp xếp theo thứ tự như sau:

- Trước hết viết các chữ của từ khoá vào các hàng của ma trận bắt từ hàng thứ nhất.

- Nếu ma trận còn trống, viết các chữ khác trên bảng chữ cái chưa được sử dụng vào các ô còn lại. Có thể viết theo một trình tự qui ước trước, chẳng hạn từ đầu bảng chữ cái cho đến cuối.

- Vì có 26 chữ cái tiếng Anh, nên thiếu một ô. Thông thường ta dồn hai chữ nào đó vào một ô chung, chẳng hạn I và J.

- Giả sử sử dụng từ khoá MORNACHY. Lập ma trận khoá Playfair tương ứng như sau:

```
MONAR
CHYBD
EFGIK
LPQST
UVWXZ
```

Mã hoá và giải mã: bản rõ được mã hoá 2 chữ cùng một lúc theo qui tắc như sau:

- Chia bản rõ thành từng cặp chữ. Nếu một cặp nào đó có hai chữ như nhau, thì ta chèn thêm một chữ lọc chẳng hạn X. Ví dụ, trước khi mã **"balloon"** biến đổi thành **"ba lx lo on".**

- Nếu cả hai chữ trong cặp đều rơi vào cùng một hàng, thì mã mỗi chữ bằng chữ ở phía bên phải nó trong cùng hàng của ma trận khóa (cuộn vòng quanh từ cuối về đầu), chẳng hạn "**ar"** biến đổi thành "RM"

- Nếu cả hai chữ trong cặp đều rơi vào cùng một cột, thì mã mỗi chữ bằng chữ ở phía bên dưới nó trong cùng cột của ma trận khóa (cuộn vòng quanh từ cuối về đầu), chẳng hạn "**mu"** biến đổi thành "CM"

- Trong các trường hợp khác, mỗi chữ trong cặp được mã bởi chữ cùng hàng với nó và cùng cột với chữ cùng cặp với nó trong ma trận khóa. Chẳng hạn, **"hs"** mã thành **"BP",** và **"ea"** mã thành **"IM"** hoặc **"JM"** (tuỳ theo sở thích)

An toàn của mã Playfair:

- An toàn được nâng cao so hơn với bảng đơn, vì ta có tổng cộng 26 x 26 = 676 cặp. Mỗi chữ có thể được mã bằng 7 chữ khác nhau, nên tần suất các chữ trên bản mã khác tần suất của các chữ cái trên văn bản tiếng Anh nói chung.

- Muốn sử dụng thống kê tần suất, cần phải có bảng tần suất của 676 cặp để thám mã (so với 26 của mã bảng đơn). Như vậy phải xem xét nhiều trường hợp hơn và tương ứng sẽ có thể có nhiều bản mã hơn cần lựa chọn. Do đó khó thám mã hơn mã trên bảng chữ đơn.

- Mã Playfair được sử dụng rộng rãi nhiều năm trong giới quân sự Mỹ và Anh trong chiến tranh thế giới thứ 1. Nó có thể bị bẻ khoá nếu cho trước vài trăm chữ, vì bản mã vẫn còn chứa nhiều cấu trúc của bản rõ.

2.2.4. Các mã đa bảng

Một hướng khác làm tăng độ an toàn cho mã trên bảng chữ là sử dụng nhiều bảng chữ để mã. Ta sẽ gọi chúng là các mã thế đa bảng. Ở đây mỗi chữ có thể được mã bằng bất kỳ chữ nào trong bản mã tùy thuộc vào ngữ cảnh khi mã hoá. Làm như vậy để trải bằng tần suất các chữ xuất hiện trong bản mã. Do đó làm mất bớt cấu trúc của bản rõ được thể hiện trên bản mã và làm cho thám mã đa bảng khó hơn. Ta sử dụng từ khoá để chỉ rõ chọn bảng nào được dùng cho từng chữ trong bản tin. Sử dụng lần lượt các bảng theo từ khóa đó và lặp lại từ đầu sau khi kết thúc từ khoá. Độ dài khoá là chu kỳ lặp của các bảng chữ. Độ dài càng lớn và nhiều chữ khác nhau được sử dụng trong từ khoá thì càng khó thám mã.

Mã Vigenere

Mã thế đa bảng đơn giản nhất là mã Vigenere. Thực chất quá trình mã hoá Vigenere là việc tiến hành đồng thời dùng nhiều mã Ceasar cùng một lúc trên bản rõ với nhiều khoá khác nhau. Khoá cho mỗi chữ dùng để mã phụ thuộc vào vị trí của chữ đó trong bản rõ và được lấy trong từ khoá theo thứ tự tương ứng.

Giả sử khoá là một chữ có độ dài d được viết dạng $K = K_1K_2\ldots K_d$, trong đó K_i nhận giá trị nguyên từ 0 đến 25. Khi đó ta chia bản rõ thành các khối gồm d chữ. Mỗi chữ thứ i trong khối chỉ định dùng bảng chữ thứ i với tịnh tiến là K_i giống như trong mã Ceasar. Trên thực tế khi mã ta có thể sử dụng lần lượt các bảng chữ và lặp lại từ đầu sau d chữ của bản rõ. Vì có nhiều bảng chữ khac nhau, nên cùng một chữ ở các vị trí khác nhau sẽ có các bước nhảy khác nhau, làm cho tần suất các chữ trong bản mã dãn tương đối đều.

Giải mã đơn giản là quá trình làm ngược lại. Nghĩa là dùng bản mã và từ khoá với các bảng chữ tương ứng, nhưng với mỗi chữ sử dụng bước nhảy lui lại về đầu.

Ví dụ: Để sử dụng mã Vigenere với từ khóa và bản rõ cho trước ta có thể làm như sau:

- Viết bản rõ ra

- Viết từ khoá lặp nhiều lần phía trên tương ứng của nó

- Sử dụng mỗi chữ của từ khoá như khoá của mã Ceasar

- Mã chữ tương ứng của bản rõ với bước nhảy tương ứng.

- Chẳng hạn sử dụng từ khoá deceptive

```
key:        deceptivedeceptivedeceptive
    plaintext: wearediscoveredsaveyourself
    ciphertext:ZICVTWQNGRZGVTWAVZHCQYGL
```

Để mã chữ w đầu tiên ta tìm chữ đầu của khóa là d, như vậy w sẽ được mã trên bảng chữ tịnh tiến 3 (tức là a tịnh tiến vào d). Do đó chữ đầu w được mã bởi chữ Z. Chữ thứ hai trong từ khóa là e, có nghĩa là chữ thứ hai trong bản rõ sẽ được tịnh tiến 4 (từ a tịnh

tiến đến e). Như vậy thứ hai trong bản rõ e sẽ được mã bởi chữ I. Tương tự như vậy cho đến hết bản rõ.

Trên thực tế để hỗ trợ mã Vigenere, người ta đã tạo ra trang Saint – Cyr để trợ giúp cho việc mã và giải mã thủ công. Đó là một bảng cỡ 26 x 26 có tên tương ứng là các chữ cái trong bảng chữ tiếng Anh. Hàng thứ i là tịnh tiến i chữ của bảng chứ cái. Khi đó chữ ở cột đầu tiên chính là khoá của bảng chữ ở cùng hàng. Do đó chữ mã của một chữ trong bản rõ nằm trên cùng cột với chữ đó và nằm trên hàng tương ứng với chữ khoá.

```
   ABCDEFGHIJKLMNOPQRSTUVWXYZ

A  ABCDEFGHIJKLMNOPQRSTUVWXYZ
B  BCDEFGHIJKLMNOPQRSTUVWXYZA
C  CDEFGHIJKLMNOPQRSTUVWXYZAB
D  DEFGHIJKLMNOPQRSTUVWXYZABC
E  EFGHIJKLMNOPQRSTUVWXYZABCD
F  FGHIJKLMNOPQRSTUVWXYZABCDE
G  GHIJKLMNOPQRSTUVWXYZABCDEF
H  HIJKLMNOPQRSTUVWXYZABCDEFG
I  IJKLMNOPQRSTUVWXYZABCDEFGH
J  JKLMNOPQRSTUVWXYZABCDEFGHI
K  KLMNOPQRSTUVWXYZABCDEFGHIJ
L  LMNOPQRSTUVWXYZABCDEFGHIJK
M  MNOPQRSTUVWXYZABCDEFGHIJKL
N  NOPQRSTUVWXYZABCDEFGHIJKLM
O  OPQRSTUVWXYZABCDEFGHIJKLMN
P  PQRSTUVWXYZABCDEFGHIJKLMNO
Q  QRSTUVWXYZABCDEFGHIJKLMNOP
R  RSTUVWXYZABCDEFGHIJKLMNOPQ
S  STUVWXYZABCDEFGHIJKLMNOPQR
T  TUVWXYZABCDEFGHIJKLMNOPQRS
U  UVWXYZABCDEFGHIJKLMNOPQRST
V  VWXYZABCDEFGHIJKLMNOPQRSTU
W  WXYZABCDEFGHIJKLMNOPQRSTUV
X  XYZABCDEFGHIJKLMNOPQRSTUVW
Y  YZABCDEFGHIJKLMNOPQRSTUVWX
Z  ZABCDEFGHIJKLMNOPQRSTUVWXY
```

Bảng Saint Cyr

An toàn của mã Vigenere. Như vậy có chữ mã khác nhau cho cùng một chữ của bản rõ. Suy ra tần suất của các chữ bị là phẳng, nghĩa là tần suất xuất hiện các chữ trên bản mã tương đối đều nhau**.** Tuy nhiên chưa mất hoàn toàn**,** do độ dài của khoá có hạn, nên có thể tạo nên chu kỳ vòng lặp. Kẻ thám mã bắt đầu từ tần suất của chữ để xem có phải đây là mã đơn bảng chữ hay không. Giả sử đây là mã đa bảng chữ, sau đó xác định số bảng chữ trong từ khoá và lần tìm từng chữ. Như vậy cần tăng độ dài từ khoá để tăng số bảng chữ dùng khi mã để "là" tần suất của các chữ.

2.2.5. Mã khóa tự động

Lý tưởng nhất là ta có khoá dài như bản tin. Do đó Vigenere đề xuất khoá tự động sinh cho bằng độ dài bản tin như sau: từ khoá được nối tiếp bằng chính bản rõ để tạo thành khoá. Sau đó dùng mã Vigenere để mã bản rõ đã cho. Khi đó biết từ khoá có thể khôi phục được một số chữ ban đầu của bản rõ. Sau đó tiếp tục sử dụng chúng để giải mã cho văn bản còn lại. Sự cải tiến này làm mất khái niệm chu kỳ, gây khó khăn cho việc thám mã, nhưng vẫn còn đặc trưng tần suất để tấn công.

Ví dụ. Cho từ khoá **deceptive.** Ta viết bản rõ nối tiếp vào từ khoá tạo thành từ khoá mới có độ dài bằng độ dài bản rõ.

key: deceptivewearediscoveredsav

plaintext: wearediscoveredsaveyourself

ciphertext:ZICVTWQNGKZEIIGASXSTSLVVWLA

2.2.6. Bộ đệm một lần

Nếu khóa thực sự ngẫu nhiên được dùng và có độ dài bằng bản rõ thì ta nói đó là bộ đệm một lần. Việc mã hóa (giải mã) được thực hiện bằng phép toán XOR từng bit giữa các bit có vị trí tương ứng ở bản rõ (bản mã) và khóa.

Vì khóa chỉ được dùng một lần và ngẫu nhiên, nên mã hoá sẽ an toàn. Mã sẽ không bẻ được, vì bản mã không có liên quan thống kê gì với bản rõ, do bộ đệm được sinh ngẫu nhiên. Có thể nói mã bộ đệm một lần là an toàn tuyệt đối, vì với bản rõ bất kỳ và bản mã bất kỳ, luôn tồn tại một khóa để ánh xạ bản rõ đó sang bản mã đã cho. Về mặt lý thuyết, xác suất để mọi mẩu tin (có cùng độ dài với bản rõ) trên bảng chữ mã là mã của một bản rõ cho trước là như nhau. Khóa chỉ sử dụng một lần, nên các lần mã là độc lập với nhau.

Vấn đề khó khăn của mã bộ đệm một lần là việc sinh ngẫu nhiên khóa và phân phối khóa an toàn. Do đó bộ đệm một lần ít được sử dụng và chỉ dùng trong trường hợp đòi hỏi bảo mật rất cao.

2.2.7. Các mã hoán vị cổ điển

Trong các mục trước chúng ta đã xét một số mã thay thế, ở đó các chữ của bản rõ được thay thế bằng các chữ khác của bản mã. Bây giờ chúng ta xét đến loại mã khác, mã hoán vị, các chữ trong bản rõ không được thay thế bằng các chữ khác mà chỉ thay đổi vị trí, tức là việc mã hoá chỉ dịch chuyển vị trí tương đối giữa các chữ trong bản rõ. Như vậy, nó giấu bản rõ bằng cách thay đổi thứ tự các chữ, nó không thay đổi các chữ thực tế được dùng. Do đó bản mã có cùng phân bố tần suất xuất hiện các chữ như bản gốc. Tính chất này tạo điều kiện để thám mã có thể phát hiện được.

Mã Rail Fence

Đây là mã hoán vị đơn giản. Viết các chữ của bản rõ theo đường chéo trên một số dòng. Sau đó đọc các chữ theo theo từng dòng sẽ nhận được bản mã. Số dòng chính là

khoá của mã. Vì khi biết số dòng ta sẽ tính được số chữ trên mỗi dòng và lại viết bản mã theo các dòng sau đó lấy bản rõ bằng cách viết lại theo các cột.

Ví dụ. Viết bản tin "meet me after the toga party" lần lượt trên hai dòng như sau

m e m a t r h t g p r y

e t e f e t e o a a t

Sau đó ghép các chữ ở dòng thứ nhất với các chữ ở dòng thứ hai cho bản mã:

MEMATRHTGPRYETEFETEOAAT

Mã dịch chuyển dòng

Giả sử lấy một số cột xác định và chọn một hoán vị chỉ số của các cột đó làm khóa. Viết các chữ của bản rõ lần lượt theo các dòng với số cột xác định. Sau đó đọc lại chúng theo các cột với thứ tự chỉ số ở dòng khóa để nhận được bản mã. Quá trình giải mã được thực hiện ngược lại.

Ví dụ:

Khóa: 4 3 1 2 5 6 7

Bản rõ: a t t a c k p

o s t p o n e

d u n t i l t

w o a m x y z

Ta đọc theo thứ tự các cột từ 1 đến 7 để nhận được bản mã:

Ciphertext: TTNAAPTMTSUOAODWCOIXKNLYPETZ

Rõ ràng trong mã trên mỗi khóa là một hoán vị của 7, nên số khóa khác nhau có thể có là 7! =4032. Chúng ta cần 12 bit để biểu diễn không gian khóa đó (vì $11 < \log_2 4032 \leq 12$). Hay nói độ dài khóa biểu diễn dạng bit là 12.

2.2.8. Mã tích

Mã dùng hoán vị hoặc dịch chuyển không an toàn vì các đặc trưng tần suất của ngôn ngữ không thay đổi. Mã cổ điển chỉ sử dụng một trong hai phương pháp thay thế hoặc hoán vị. Có thể sử dụng một số mã liên tiếp nhau sẽ làm cho mã khó hơn. Do đó người ta nghĩ đến việc kết hợp cả hai phương pháp này trong cùng một mã và có thể sử dụng đan xen hoặc lặp nhiều vòng. Đôi khi ta tưởng lặp nhiều lần cùng một loại mã sẽ tạo nên mã phức tạp hơn, nhưng trên thực tế trong một số trường hợp về bản chất chúng cũng tương đương với một lần mã cùng loại nào đó như: tích của hai phép thế sẽ là một phép thế; tích của hai phép hoán vị sẽ là một phép hoán vị. Nhưng nếu hai loại mã đó khác nhau thì sẽ tạo nên mã mới phức tạp hơn, chính vì vậy phép thế được nối tiếp bằng phép dịch chuyển sẽ tạo nên mã mới khó hơn rất nhiều. Đây chính là chiếc cầu nối từ mã cổ điển sang mã hiện đại.

Điểm yếu của mã cổ điển

- Phương pháp mã hoá cổ điển có thể dễ dàng bị giải mã bằng cách đoán chữ dựa trên phương pháp thống kê tần xuất xuất hiện các chữ cái trên mã và so sánh với bảng thống kê quan sát của bản rõ.

- Để dùng được mã hoá cổ điển thì bên mã hoá và bên giải mã phải thống nhất với nhau về cơ chế mã hoá cũng như giải mã. Nếu không thì hai bên sẽ không thể làm việc được với nhau.

2.3. Cấu trúc mã khối Fiestel

Các nguyên lý mã khối:

- Hầu hết các mã khối đối xứng dựa trên cấu trúc mã Fiestel, do nhà bác học Fiestel đề xuất năm 1973. Đây là điều cần thiết, vì cần phải có khả năng giải mã các bản mã một cách có hiệu quả.
- Mã khối được coi giống như phép thế cực lớn. Cần bảng có 2^{64} đầu vào cho mã khối 64 bit, bảng như vậy là rất lớn. Do đó có thể thay thế bằng cách tạo các khối nhỏ hơn.
- Sử dụng ý tưởng dùng mã tích. Ở đây sẽ kết hợp giữa mã thay thế và mã hoán vị, đồng thời sử dụng nhiều vòng lặp như vậy.

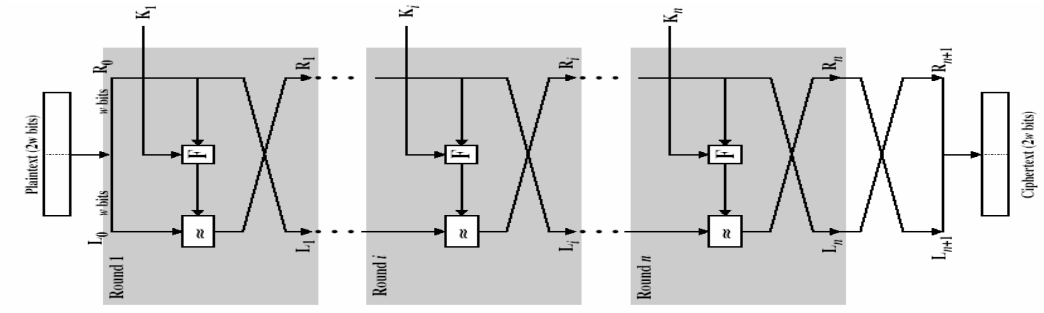
Rối loạn và khuếch tán

- Một tính chất quan trọng của mã tốt là mã cần phải che giấu hoàn toàn các tính chất thống kê của bản tin gốc. Như ta đã thấy mã bộ đệm một lần có thể làm được điều đó, do tính ngẫu nhiên của khóa đệm và độ dài bằng bản tin của nó.
- Shannon nghiên cứu và đề xuất phương pháp thực tế hơn là kết hợp các thành phần khác nhau của bản rõ để xử lý qua nhiều lần và nhận được bản mã.
- **Khuếch tán** là làm tan biến cấu trúc thống kê của bản rõ trên bản mã. Điều đó đạt được nếu mỗi bit của bản rõ tác động đến giá trị của rất nhiều bit trên bản mã hay mỗi bit của bản mã chịu tác động của nhiều bit bản rõ.
- **Rối loạn** là làm cho quan hệ giữa bản mã và khóa càng phức tạp càng tốt. Bản mã có tính rối loạn cao sẽ làm cho việc tìm mò khóa trở nên rất khó khăn, ngay cả khi kẻ tấn công có các đặc trưng thống kê của bản mã và biết cách khóa tác động đến bản mã.

Cấu trúc mã Fiestel

- Horst Fiestel sáng tạo nên mã Fiestel dựa trên mã tích nghịch đảo được, tức là kết hợp mã thế với mã hoán vị và qui trình giải mã là giống với mã hoá, chỉ cần thay đổi vai trò khối bản mã với khối bản rõ và thứ tự các khóa con được dùng. Từ khóa chính sinh ra cho mỗi vòng lặp một khóa con.
- Chia khối đầu vào thành hai nửa bằng nhau:
    - Thực hiện phép thế trên nửa trái. Sử dụng hàm vòng trên nửa phải và khóa con, rồi tác động đến nửa trái.
    - Sau đó hoán vị các nửa, nửa phải chưa được xử lý.
    - Xử lý vòng tiếp theo.

Đây là một thể hiện của mã thế kết hợp với hoán vị của Shannon. Ta xem xét cụ thể cấu trúc mã Fiestel gồm n vòng:



Nguyên tắc thiết kế mã khối Fiestel:

- Tăng kích thước khối sẽ làm tăng độ an toàn nhưng làm giảm tốc độ mã
- Tăng kích thước khóa sẽ làm tăng độ an toàn – tìm khóa khó hơn, nhưng làm chậm mã.
- Tăng số vòng làm tăng độ an toàn nhưng làm chậm mã.
- Phát sinh khóa con càng phức tạp làm cho việc thám mã khó hơn nhưng làm chậm mã.
- Hàm vòng càng phức tạp làm cho việc thám mã khó hơn nhưng làm chậm mã.
- Phần mềm mã hoá/giải mã nhanh và khó thám mã là tiêu chí hay được đề cập đến đối với ứng dụng và kiểm nghiệm thực tế.

TÓM LƯỢC CUỐI BÀI

- Khái niệm mã đối xứng
- Cấu trúc mã khối Fiestel

CÂU HỎI TRẮC NGHIỆM CUỐI BÀI

Câu 1: Mục nào không phải là thành phần của Khóa đối xứng

A. Một khóa chia sẻ người gửi và người nhận
B. Hai thuật toán mã hóa và giải mã
C. Bản rõ và bản mã
D. Thuật toán nén văn bản

Câu 2: Nói chung coi độ khó thám mã không phụ thuộc vào

A. Độ phức tạp của thuật toán mã hóa
B. Độ lớn của không gian khóa
C. Che giấu thuật toán mã hóa
D. Che giấu khóa mật

Câu 3: Thao tác xử lý dữ liệu sau nào không dùng trong mã đối xứng:

A. Dùng phép thế xâu ký tự bản rõ bằng xâu ký tự bản mã
B. Dùng phép hoán vị đảo chỗ các ký tự bản rõ tạo ra bản mã
C. Kết hợp cả hai phép toán trên và có thể xử lý nhiều vòng
D. Che giấu dữ liệu trong môi trường khác

ĐÁP ÁN CÂU HỎI TRẮC NGHIỆM

Câu 1: D, Khóa, thuật toán mã hóa, giải mã, bản rõ bản mã đều là các thành phần của mã đối xứng, nén không dùng đến.
Câu 2: C, luôn coi thuật toán mã hoá là mọi người đều biết
Câu 3: D, Che giấu dữ liệu không phải là thao tác mã hóa, mà là giấu sự tồn tại của dữ liệu mật trong môi trường nào đó

THUẬT NGỮ TRONG BÀI

- Bản rõ là bản tin gốc.
- Bản mã là bản tin gốc đã được mã hoá.
- Mã là thuật toán chuyển bản rõ thành bản mã
- Khoá là thông tin dùng để mã hoá, chỉ có người gửi và người nhận biết.
- Mã hoá chuyển bản rõ thành bản mã
- Giải mã chuyển bản mã thành bản rõ.
- Mật mã nghiên cứu các nguyên lý và phương pháp mã hoá.
- Thám mã nghiên cứu các nguyên lý và phương pháp giải mã mà không biết khoá.
- Lý thuyết mã bao gồm cả mật mã và thám mã
- Mã đối xứng là mã ở đó hai người nhận và gửi chia sẻ chung một khóa.

CÂU HỎI THƯỜNG GẶP

1. Nêu sự khác biệt giữa mã thế và mã hoán vị
2. Thế nào là mã đối xứng mạnh. Nó cần có các tính chất gì?
3. Mô tả kiến trúc mã đối xứng Fiestel

TRẢ LỜI CÂU HỎI THƯỜNG GẶP

1. Mã thế là thay mỗi ký tự bản rõ bằng 1 xâu bản mã; mã hoán vị đảo thứ tự các ký tự trong bản rõ để tạo nên bản mã.
2. Mã đối xứng mạnh cần có các tính chất sau:
   - Kích thước khối dữ liệu mã và khóa tương đối lớn: cân bằng với tốc độ
   - Thuật toán mã hóa mạnh: lặp nhiều vòng, mỗi vòng kết hợp hoán vị với thế; thuật toán sinh khóa con phức tạp cho mỗi vòng. Bản mã có tính chất khoếch tán và tác dụng đồng loạt để khó thám mã
3. Kiến trúc mã khối Fiestel
   - Lặp nhiều vòng, sinh khóa con riêng cho từng vòng
   - Quá trinh giải mã ngược lại với quá trình mã hóa
   - Cân đối việc tăng kích thước khối, khóa và số vòng để đảm bảo an ninh với tốc độ thực hiện

CÂU HỎI TỰ LUẬN

Câu 1. Các khái niệm cơ bản của mã đối xứng là gì?

Câu 2. Hai thuật toán cơ bản nào được dùng trong mã hóa? Các thuật toán như thế nào là mạnh?

Câu 3 Hai kiểu thao tác cơ bản nào được dùng trong các thuật toán mã hoá?

Câu 4 Có bao nhiêu khóa cần cho hai người để trao đổi thông qua mã đối xứng? Không gian khóa như thế nào là tốt? Làm sao để gây khó khăn cho việc thám mã dò tìm khóa?

Câu 5 Sự khác biệt giữa mã khối và mã dòng?

BÀI TẬP TRẮC NGHIỆM

1. **Mã Ceasar là mã**
   a) phép thế trên nhiều bảng chữ
   b) phép thế trên một bảng chữ
   c) phép dịch chuyển, tức là đảo chữ trên bản rõ để nhận được bản mã
   d) phép thế kết hợp với dịch chuyển
2. **Mã Vigenere là mã**
   a) phép thế trên một bảng chữ
   b) phép thế trên nhiều bảng chữ
   c) phép dịch chuyển, tức là đảo chữ trên bản rõ để nhận được bản mã
   d) phép thế kết hợp với dịch chuyển
3. **Cho qui tắc má hóa:**

a b c d e f g h i j k l m n o p q r s t u v w x y z

E F G H I J K L M N O P Q R S T U V W X Y Z A B C D

3.1 Tìm bản rõ của QE LSE:

a. HOA MA
b. MA DOI
c. MA HOA
d. MA HAO

3.2 Tìm bản mã của KNOWLEDGE:

a. RASAPIHKI
b. ORPIASHKI
c. ORSAPIHKI
d. SAPIKHJOA

4. Trong tiếng Việt, giả sử tần suất xuất hiện của 1 số chữ cái là:

p(g)=p(n)=3.5%; p(t)=p(h)=0.9%; p(a)=p(i)=1.8%

4.1 Tìm bản rõ (có thể nhất) của bản mã "THATH":

a. THANH
b. QUAUG
c. NGING
d. NGANG

4.2 Tìm bản rõ (có thể nhất) của bản mã "MT TEMT":

a. NA NHAN
b. HAN HAN
c. HA NHAN
d. AN NHAN

5. Chẳng hạn sử dụng từ MORNACHY cho mã Playfair, ta có bàng mã:

MON AR

C HY BD

E FG I K

L PQ ST

UVWXZ

Chữ lọc là X.

5.1 Tìm bản mã của từ: ”CALE”

a. AEUL
b. AMLU
c. AMUE
d. AMUL

5.2 Tìm bản mã của từ: “BOO”

a. HAVR
b. HRAV
c. HAVA
d. FAVA

6. Dùng mã Vignerer, với khóa KEY trong bảng chữ cái tiếng Anh, tìm bản mã của từ: “ENCRIPE”

a. OSABMDO

b. ORADMDO

c. ORABMDO

d. ORBCNCO

7. Dùng mã dịch theo hàng với khóa 425316, mã hóa bản rõ sau:

“Moi nguoi dan la mot bong hoa dep”

a. GNBAOIMGPNATOMOANEIDOHULOD
b. GNBAOIMANEIDOHULODGPNATOMO
c. GNBAOIMGPOANEIDOHULODNATOM
d. OMOANEIDOGNBAOIMGPNATHULOD

8. Điểm khác nhau giữa mã khối và mã dòng:

a. Mã khối xử lý văn bản lần lượt từng cặp block. Mã dòng xử lý bản theo từng cặp bít hoặc byte.

b. Mã khối chỉ xử lý từng khối trong khi mã dòng xử lý đồng thời nhiều dòng

c. Mã dòng chỉ dùng cho các loại mã cổ điển, mã khối dùng cho mọi kiểu mã.

d. Mã khối xử lý văn bản lần lượt từng block. Mã dòng xử lý bản theo từng bít hoặc byte.