Cryptography and Network Security
Chapter 2
Mã cổ điển

Fourth Edition

by William Stallings

Lecture slides by Lawrie Brown

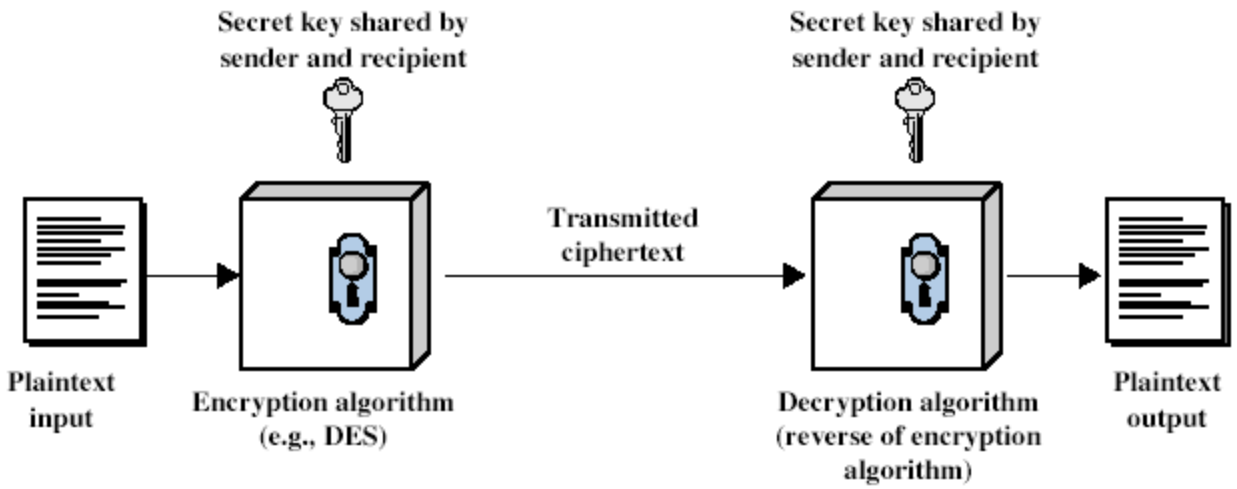
Mã đối xứng

- Hay là mã một khoá – khoá mã hoá = khoá giải mã – khoá khoá thỏa thuận
- Người gửi và người nhận chia sẻ khoá chung
- Mọi thuật toán mã cổ điển đều là mã đối xứng
- Là kiểu duy nhất trước khi phát minh ra khoá mã công khai vào những năm 1970
- Và đến nay vẫn được sử dụng rộng

Các khái niệm cơ bản

- **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.**

Mô hình mã đối xứng
Symmetric Cipher Model



Các yêu cầu

- Hai yêu cầu để sử dụng an toàn mã khoá đối xứng là
  - thuật toán mã hoá mạnh
  - khoá mật chỉ có người gửi và người nhận biết
- Về mặt toán học:

  $Y = E_K(X)$

  $X = D_K(Y)$
- Giả thiết rằng thuật toán mã hoá mọi người đều biết
- Có kênh an toàn để phân phối

Mật mã

- Hệ mật mã được đặc trưng bởi:
  - kiểu của các thao tác mã hoá được sử dụng:
    - phép thế,
    - thay đổi vị trí (hoán vị) hay
    - tích của chúng.
  - Số khoá được sử dụng:
    - Khoá duy nhất - khoá riêng hoặc
    - hai khoá hay khoá công khai
  - Cách mà bản rõ được xử lý:
    - khối hay
    - dòng bít

Thám mã
Cryptanalysis

- Mục đích là tìm khoá chứ không phải một bản tin cụ thể
- Có các cách chung như:
  - tấn công để thám mã và
  - tìm duyệt toàn bộ

Các kiểu tấn công thám mã

- **Chỉ dùng bản mã: biết thuật toán và bản mã, dùng phương pháp thống kê, xác định bản rõ.**
- **Biết bản rõ: biết được bản mã/bản rõ để tấn công mã**
- **Bản rõ được chọn: chọn bản rõ và nhận được bản mã đẻ tấn công mã.**
- **Bản mã được chọn: chọn bản mã và nhận được bản rõ để tấn công mã.**
- **Bản tin được chọn: chọn được bản rõ hoặc mã và mã hoặc giải mã để tấn công mã.**

Các khái niệm tiếp theo

- **An toàn không điều kiện: không quan trọng máy tính mạnh như thế nào, mã hoá không thể bị bẻ vì bản mã không cung cấp đủ thông tin để xác định duy nhất bản rõ.**
- **An toàn tính toán: với nguồn lực máy tính giới hạn (chẳng hạn thời gian tính toán không quá tuổi của vũ trụ) mã hoá không thể bị bẻ.**

Tìm duyệt - Brute Force Search

- Luôn có thể thử từng khoá
- Phần lớn công sức của các tấn công đều tỷ lệ thuận với kích thước khoá.
- Giả thiết là biết hoặc nhận biết được bản rõ.

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

Các mã thế cổ điển
Classical Substitution Ciphers

- Ở đây các chữ của bản rõ được thay bằng các chữ khác hoặc các số hoặc các ký hiệu.
- Hoặc nếu xem bản rõ như môt dãy bít, thì phép thế thay các mẫu bít bản rõ bằng các mẫu bít bản mã

Mã Ceasar

- Mã thế được biết sớm nhất
- Được sáng tạo bởi Julius Ceasar
- Đầu tiên được sử dụng trong quân sự
- Thay mỗi chữ bằng chữ thứ ba tiếp theo
- Ví dụ:

```
    meet me after the toga
party
 PHHW PH DIWHU WKH WRJD SDUWB
```

Caesar Cipher

- Có thể định nghĩ qua phép dịch chuyển

```
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 gán số cho mỗi chữ

```
   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
 0 1  2  3  4  5  6  7 8 9  10 11 12 13 14 15  16 17 18
   19 20 21 22 23 24 25
```

- thì mã Ceasar được định nghĩa như sau

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

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

Thám mã Ceasar

- Chỉ có 26 khoá có thể:
  - A ánh xạ vào A, B, C, …, Z
- Có thể thử lần lượt
- Sử dụng tìm duyệt
- Cho bản mã, hãy thử mọi cách dịch chuyển các chữ
- Sẽ đoán nhận thông qua nội dung các bản rõ nhận được
- Ví dụ: bẻ bản mã "GCUA VQ DTGCM" cho "easy to break".

Các mã bảng chữ đơn

- Không chỉ là dịch chuyển bảng chữ
- Có thể tạo các bước nhảy các chữ tùy ý
- Mỗi chữ của bản rõ được ánh xạ đến một chữ ngẫu nhiên khác nhau của bản mã.
- Như vậy độ dài khoá là 26
- Ví dụ:
- Plain: abcdefghijklmnopqrstuvwxyz
- Cipher: DKVQFIBJWPESCXHTMYAUOLRGZN
- Plaintext: ifwewishtoreplaceletters
- Ciphertext: WIRFRWAJUHYFTSDVFSFUUFYA

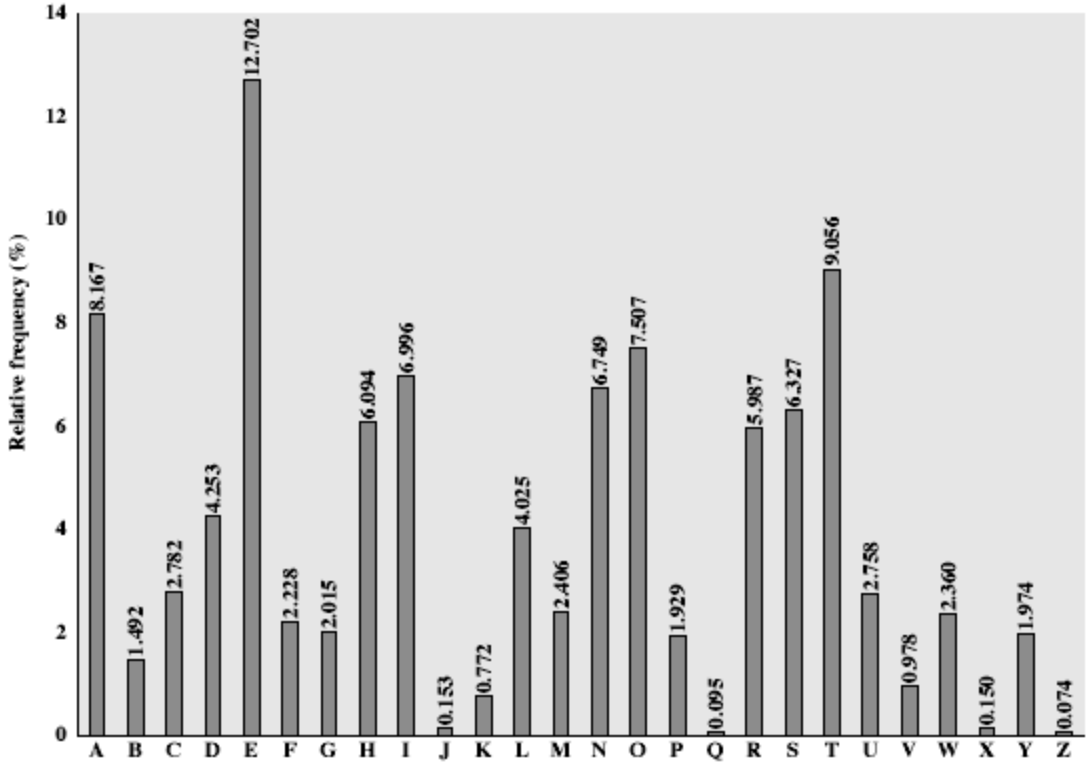
Tính an toàn của mã trên bảng chữ đơn

- Tổng cộng có 26! = 4 x 10^{26} khoá
- Với khá nhiều khoá như vậy nhiều người nghĩ là an toàn
- Nhưng không phải như vậy: Sai!!!
- Vấn đề ở đây là do các đặc trưng về ngôn ngữ.

Tính dư thừa của ngôn ngữ và thám mã

- Ngôn ngữ của loài người là dư thừa
- Như "th lrd s m shphrd shll nt wnt"
- Các chữ không được sử dụng thường xuyên như nhau
- Trong tiếng Anh chữ e được sử dụng nhiều nhất
- Sau đó đến T,R,N,I,O,A,S
- Một số chữ rất ít dùng như: Z,J,K,Q,X
- Có bảng các tần suất các chữ đơn, cặp chữ, bộ ba chữ.

Bảng tần suất chữ cái tiếng Anh
English Letter Frequencies



Sử dụng 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.
- Được phát hiện bởi các nhà khoa học Ai cập từ thế kỷ thứ 9.
- 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.

Ví dụ thám mã

- Cho bản mã:

  UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZVUEP
  HZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSXEPYEPOP
  DZSZUFPOMBZWPFUPZHMDJUDTMOHMQ

  – 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:

    it was disclosed yesterday that several informal but direct contacts have been made with political representatives of the viet cong in moscow

Mã Playfair

- Không phải số khoá lớn trong mã bảng chữ đơn đảm bảo an toàn mã.
- Một hướng khắc phục là **mã bộ các chữ**.
- Playfair là một trong các mã như vậy
- Được sáng tạo bởi Charles Wheastone 1854 và mang tên người bạn là Baron Playfair

Ma trận khoá Playfair

- Là ma trận gồm các chữ 5 x 5 dựa trên một từ khoá.
- Viết các chữ của từ khoá vào ma trận
- nếu còn trống, viết các chữ khác vào các ô còn lại
- Chẳng hạn sử dụng từ MORNACHY
- VD: MU -> CM, HD -> DT, O mã bằng cùng hàng, cột

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

Mã hoá và giải mã

- bản rõ được mã hoá 2 chữ cùng một lúc
  - Nếu một cặp nào đó là chữ lặp, thì chèn thêm một từ lọc chẳng hạn X. Ví dụ, trước khi mã **“balloon”** biến đổi thành **“ba lx lo on”**
  - Nếu cả hai chữ đều rơi vào cùng một hàng, thì mã mỗi chữ bằng chữ ở phía bên phải nó (cuộn vòng quanh từ cuối về đầu), chẳng hạn “**ar”** biến đổi thành “RM”
  - Nếu cả hai chữ đều rơi vào cùng một cột, thì mã mỗi chữ bằng chữ ở phía bên dưới nó (cuộn vòng quanh từ cuối về đầu), chẳng hạn “**mu”** biến đổi thành “CM”
  - Trong các trường hợp khác, mỗi chữ được mã bởi chữ cùng hàng với nó và cùng cột với chữ cùng cặp với nó. 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 với bảng đơn
- Vì có 26 x 26 = 676 cặp
- Cần phải có bảng tần suất của 676 cặp để thám mã (so với 26 của mã bảng đơn)
- Và tương ứng sẽ nhiều bản mã hơn.
- Được sử dụng rộng rãi trong 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 còn chứa nhiều cấu trúc của bản rõ.

Các mã đa bảng

- **Các mã thế đa bảng**
- Một hướng khác tăng độ an toàn là sử dụng bảng chữ lặp để mã.
- Làm cho thám mã đa bảng khó hơn và trải bằng tần suất các chữ.
- Sử dụng khoá để chọn bảng nào được dùng cho từng chữ trong bản tin
- Sử dụng lần lượt các bảng
- Lặp lại từ đầu sau khi kết thúc từ khoá.

Mã Vigenere

- Mã thế đa bảng đơn giản nhất là mã Vigenere
- Hiệu quả như dùng nhiều mã Ceasar cùng một lúc
- Khoá là một chữ có độ dài K = K1K2…Kd
- Chữ thứ i chỉ định dùng bảng chữ thứ i với tịnh tiến là Ki
- Sử dụng lần lượt các bảng chữ.
- Lặp lại từ đầu sau d chữ của bản tin
- Giải mã đơn giản là làm việc ngược lại.

Ví dụ dùng mã Vigenère

- Viết bản rõ ra
- Viết từ khoá lặp nhiều lần trên nó
- Sử dụng mỗi chữ của từ khoá như khoá của mã Ceasar
- Mã chữ tương ứng của bản rõ.
- Sử dụng từ khoá deceptive `key:`
- `abcdefghijklmnopqrstuvwxyz`

```
Khóa: deceptivedeceptivedeceptive
   Plaintext va Ciphertext
     wearediscoveredsaveyourself
     ZICVTWQNGRZGVTWAVZHCQYGLMGJ
```

Hỗ trợ

- Hỗ trợ đơn giản có thể giúp cho việc mã và giải mã
- Trang Saint – Cyr là sự trợ giúp bằng tay
  - trang với bản chữ cái lặp
  - Gióng dòng chữ cái “A” với chữ khoá chẳng hạn “C”
  - Sau đó đọc ra từng chữ bản mã dựa vào bản rõ và Trang hỗ trợ
- Có thể uốn vòng quanh thành đĩa mã
- Hoặc tạo thành Bảng Vigenere

An toàn của mã Vigenere

- Có chữ mã khác nhau cho cùng chữ của bản rõ
- Suy ra tần suất của các chữ bị là phẳng
- Tuy nhiên chưa mất hoàn toàn
- Bắt đầu từ tần suất của chữ: xem có phải là đơn bảng chữ hay không
- Sau đó xác định số bảng chữ và tìm từng cái
- Tăng dần số bảng chữ cần dùng để “là” tần suất

Phương pháp thám mã Kasiski

- Phương pháp phát triển bởi Babbage và Kasiski
- Các phép lặp trong bản mã cho phép xác định chu kỳ
- Tìm bản rõ như nhau và chu kỳ chính xác
- Kết quả sẽ như nhau ở đầu ra
- Không phải khi nào cũng tìm được
- May mắn sẽ tìm được chu kỳ
- Tấn công từng bảng chữ đơn với cùng kỹ thuật như trước

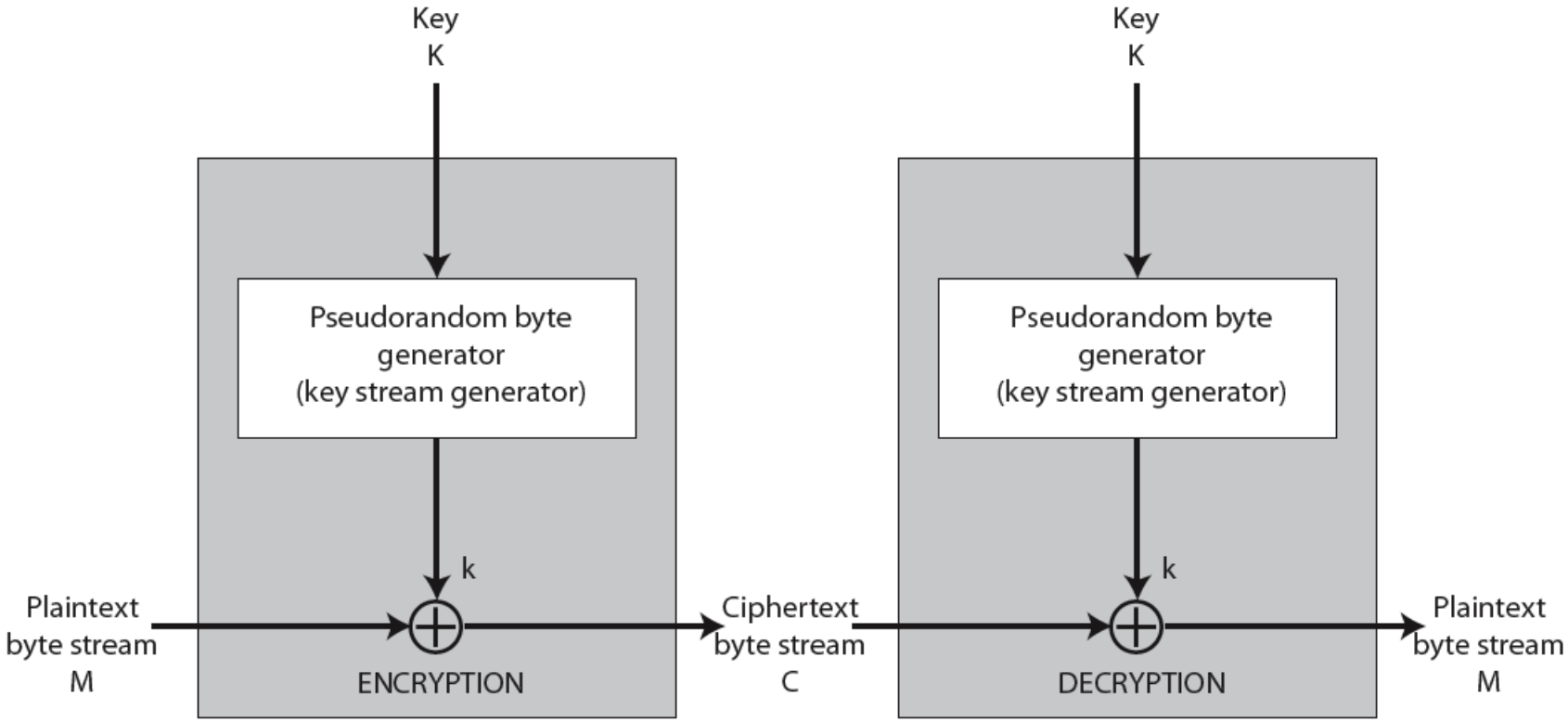
Mã khoá tự động-Autokey Cipher

- Lý tưởng có khoá dài như bản tin
- Vigenere đề xuất khoá tự động
- Với từ khoá được nối vào đầu bản tin tạo thành khoá
- Biết từ khoá có thể khôi phục được một số chữ ban đầu
- Tiếp tục sử dụng chúng cho văn bản còn lại
- Nhưng vẫn còn đặc trưng tần suất để tấn công
- Ví dụ: cho từ khoá **deceptive**
- Bảng chữ: abcdefghijklmnopqrstuvwxyz
- key: deceptivewearediscoveredsav
- plaintext: wearediscoveredsaveyourself
- ciphertext: ZICVTWQNGKZEIIGASXSTSLVVWLA

Bộ đệm một lần (One-Time Pad)

- Nếu khoá thực sự ngẫu nhiên được dùng, thì mã hoá sẽ an toàn
- Gọi là bộ đệm một lần
- sẽ không bẻ được vì bản mã không có liên quan thống kê gì với bản rõ.
- Vì với bản rõ bất kỳ và bản mã bất kỳ, luôn tồn tại một khoá để ánh xạ bản rõ đó sang bản mã đã cho.
- Có thể sử dụng khoá một lần
- Vấn đề là sinh và phân phối an toàn khoá.

Sinh số giả ngẫu nhiên
Cấu trúc mã dòng



Các mã hoán vị đổi chỗ

- Xét mã hoán vị (các chữ trong bản rõ) hay đổi chỗ (dịch chuyển) cổ điển
- Nó giấu bản rõ bằng cách thay đổi thứ tự các chữ
- Không thay đổi các chữ thực tế được dùng
- Có thể nhận biết được vì có cùng phân bố tần suất như bản gốc.

Mã Rail Fence

- Viết các chữ của bản tin theo đường chéo trên một số dòng
- Sau đó đọc theo dòng nhận được bản mã
- Ví dụ: cho bản rõ

  “meet me after the toga party”

sao cho các chữ kề nhau ở các hang khác nhau:

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

e t e f e t e o a a t

nhận bản mã – viết lại theo hàng:

MEMATRHTGPRYETEFETEOAAT

Mã dịch chuyển dòng

- Sơ đồ phức tạp hơn
- Viết các chữ của bản tin theo các dòng trên số cột xác định
- Sau đó thay đổi thứ tự các cột theo một khoá trước khi đọc lại chúng theo dòng
- Ví dụ:

Key:	4	3	1	2	5	6	7
Plaintext:	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

Ciphertext: TTNAAPTMTSUOAODWCOIXKNLYPETZ

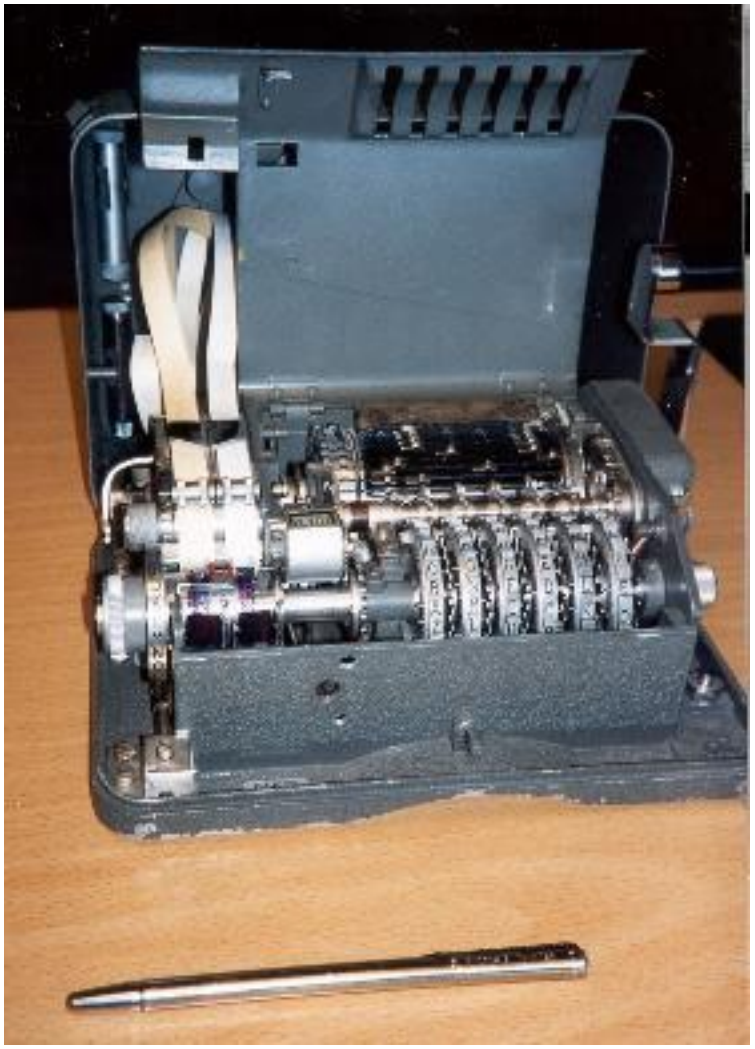
Các mã tích

- Mã dùng hoán vị hoặc dịch chuyển không an toàn vì các đặc trưng của ngôn ngữ
- Vì vậy sử dụng một số mã liên tiếp sẽ làm cho mã khó hơn, nhưng
  - Tích hai hoán vị sẽ tạo nên hoán vị phức tạp hơn
  - Tich hai phép dịch chuyển tạo nên dịch chuyển phức tạp hơn
  - Phép thế được nối tiếp bằng phép dịch chuyển tạo nên mã mới khó hơn rất nhiều
- Đây là chiếc cầu từ cổ điển sang hiện đại

Máy quay

- Trước khi có mã hiện đại, máy quay là mã tích thông dụng nhất
- Được sử dụng rộng rãi trong chiến tranh thế giới thứ hai: Đức, đồng minh và Nhật
- Tạo nên mã thế rất đa dạng và phức tạp
- Sử dụng một số lõi hình trụ, mỗi lõi ứng với một phép thế, khi quay sẽ thay đổi sau khi mỗi chữ được mã.
- Với 3 hình trụ có 26 x 26 x 26 = 17576 bảng chữ

Hagelin Rotor Machine



Giấu tin - Steganography

- Là lựa chọn dùng kết hợp hoặc đồng thời với mã
- Dấu sự tồn tại của bản tin
  - Trong bản tin dài chỉ dùng một tập con các chữ/từ được đánh giấu bằng cách nào đó
  - Sử dụng mực không nhìn thấy
  - Dấu trong các file âm thanh hoặc hình ảnh
- Có nhược điểm: chỉ giấu được lượng thông tin nhỏ các bít.

Kết luận - Summary

- Đã xét:
  - Các thuật ngữ và kỹ thuật mã cổ điển
  - Các mã thế đơn bảng chữ
  - Thám mã sử dụng tần suất của các chữ
  - Mã Playfair
  - Mã thế đa bảng chữ
  - Mã hoán vị (đổi chỗ)
  - Tích các mã và máy quay
  - Giấu tin