# TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI TP. HÒ CHÍ MINH



# KHOA CÔNG NGHỆ THÔNG TIN AN TOÀN THÔNG TININFORMATION SECURITY

CHƯƠNG 3 KỸ THUẬT MÃ HOÁ

MÃ HOÁ CỔ ĐIỂN

Giảng viên: TS. Trần Thế Vinh

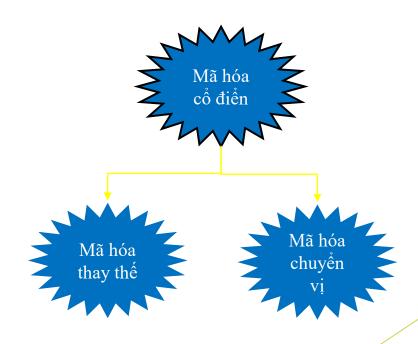


#### Hệ thống mật mã cổ điển:

- Tất cả các hệ thống này đều dựa trên sơ đồ mã hóa đối xứng.
- Dịch vụ bảo mật duy nhất mà các hệ thống này cung cấp là tính bảo mật thông tin.
- Không giống như các hệ thống hiện đại (coi dữ liệu là hệ nhị phân (binary)), các hệ thống cổ điển hoạt động trên bảng chữ cái như một thành phần cơ bản.

#### Có 2 loại mật mã cổ điển:

- Mật mã thay thế (Substitution Cipher)
- Mật mã chuyển vị (Transposition Cipher)



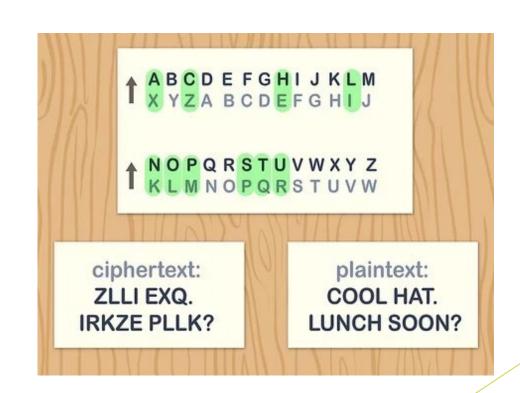


Mật mã thay thế (Substitution Cipher):

Trong một mật mã thay thế, bất kỳ ký tự nào của bản rõ từ tập hợp các ký tự cố định đã cho sẽ được thay thế bằng một số ký tự khác từ cùng một tập hợp (tùy thuộc vào khóa để có sự thay thế khác nhau). Ví dụ: với độ dịch chuyển là 2 thì A sẽ được thay thế bằng C, B trở thành D, .v..v.

Mã mã thay thế được chia thành:

- Mật mã đơn chữ cái
- Mật mã đa chữ cái



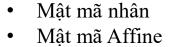


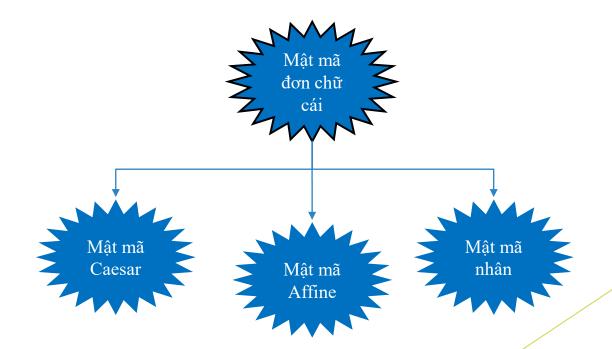
#### Mật mã đơn chữ cái:

Trong mật mã đơn chữ cái, mỗi ký hiệu trong bản rõ được ánh xạ tới một ký hiệu trong bản mã. Cho dù một ký hiệu xuất hiện bao nhiêu lần trong bản rõ thì nó sẽ tương ứng với cùng một ký hiệu trong bản mã nghĩa là độ dài của bản rõ và bản mã là như nhau.

Phân loại của mật mã đơn chữ cái:

Mật mã Shift (hay còn gọi là mật mã Caesar, mật mã cộng)







#### Mật mã đa chữ cái:

Trong mật mã đa chữ cái, mọi ký hiệu trong bản rõ được ánh xạ tới một ký hiệu trong bản mã khác bất kể sự xuất hiện của nó. Mỗi lần xuất hiện khác nhau của một ký tự có ánh xạ khác nhau tới bản mã.

#### Phân loại của mật mã đa chữ cái:

- Mật mã Auto-key.
- Mật mã Vigenere
- Mật mã Playfair
- Mật mã Hill
- One Time Pad
- Mật mã Rotor



#### Mật mã chuyển vị:

Mật mã chuyển vị không xử lý việc thay thế ký hiệu này bằng một ký hiệu khác. Nó tập trung vào việc thay đổi vị trí của các ký hiệu trong bản rõ. Một ký hiệu ở vị trí đầu tiên trong bản rõ có thể xuất hiện ở vị trí thứ 7 trong bản mã.

#### Mật mã chuyển vị bao gồm:

- Mật mã chuyển vị cột.
- Mật mã Rail-Fence.



# Mật mã thay thế - Mật mã Caesar

Một trong những mật mã được ghi chép đầu tiên là mật mã của Caesar, được sử dụng bởi một chỉ huy nổi tiếng trong thư từ của ông ấy. Trong mật mã này, mỗi chữ cái được thay thế bằng một chữ cái nằm bên phải k ký tự trong bảng chữ cái theo modulo bằng số chữ cái trong bảng chữ cái:

$$C_k(j) = (j+k) \pmod{n},$$

**Trong đó**: j – số thứ tự của chữ cái trong bảng chữ cái,  $C_k(j)$  – Số thứ tự của chữ cái thay thế n - sức mạnh của bảng chữ cái đầu vào (số lượng chữ cái trong bảng chữ cái được sử dụng).



# Mật mã thay thế - Mật mã Caesar

Do đó, khóa mã hóa ở đây là số k, xác định kích thước của độ lệch. Rõ ràng, tra cứu nghịch đảo là:

$$C_k^{-1}(j) = C_{n-k} = (j + n - k) \pmod{n}$$

Nếu cần thiết, bảng chữ cái có thể được mở rộng với dấu chấm câu, chữ in hoa, số, để mật mã có thể xử lý tất cả các ký tự của văn bản nguồn. Tổng số khóa hợp lệ bằng n và một trong các khóa chuyển đổi văn bản thành chính nó.



# Mật mã thay thế - Mật mã Caesar

Trong loại mật mã Caesar này, khóa được cho bởi số k $(0 \le k \le n-1)$  và một từ khóa hoặc câu ngắn.

Bảng chữ cái được viết ra, và bên dưới nó, bắt đầu từ vị trí k là từ khoá. Các chữ cái còn lại được viết theo thứ tự bảng chữ cái sau từ khóa. Kết quả là, chúng ta nhận được tra cứu cho từng chữ cái. Không bắt buộc tất cả các chữ cái của từ khóa phải khác nhau, chỉ cần viết ra từ khóa mà không lặp lại các chữ cái giống nhau.

Từ khoá: "yes", k =2			
Bảng gốc	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		
Bảng mới	x, z, y, e, s, a, b, c, d, f, g, h, i, j, k, l, m, n, o, p, q, r, t, u, v, w		
Văn bản nguồn	data		
Văn bản mã hoá	expx		

Số lượng chìa khóa trong hệ thống của Caesar với từ khóa là n!. Để giải mã, cần sử dụng khóa mã hóa đã biết để xác định sự tương ứng của bản gốc và thay thế bảng chữ cái và thực hiện thay thế ngược lại.



# Mật mã thay thế - Hệ thống mật mã Affine

Mật mã Affine là một loại mật mã thay thế một bảng chữ cái, trong đó mỗi chữ cái trong bảng chữ cái được ánh xạ tới số tương đương của bảng chữ cái đó. Nó được mã hóa bằng một hàm toán học đơn giản và được chuyển đổi trở lại thành một chữ cái.

Trong mật mã Affine, các chữ cái của bảng chữ cái có kích thước m trước tiên được ánh xạ tới các số nguyên từ 0 đến m-1.

Key cho mật mã Affine bao gồm 2 số (a và b). 1 trong 2 số (a hoặc b) phải là nguyên tố cùng nhau với m ((a,m)=1)

Công thức về mặt toán học có dạng:

$$E(x) = (ax + b) \bmod m$$

Trong đó:

- m là độ dài của bảng chữ cái.
- a và b: khoá của mật mã.
- $\bullet \quad (a,m) = 1$

Khi giải mã, chúng ta thực hiện các chức năng ngược lại (hoặc nghịch đảo) trên bản mã để lấy ra bản rõ.

$$D(x) = a^{-1}(x - b) \bmod m$$

Trong đó:

•  $1 = a^{-1}a \mod m$ 

# AND TONG NGHE THONGTH

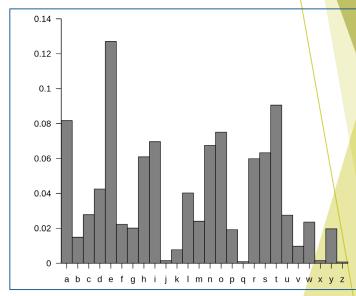
# Mật mã thay thế - Hệ thống mật mã Affine

Mật mã và hệ thống mật mã Affine của Caesar thuộc về lớp các hệ thống mật mã đơn chữ cái, nghĩa là, đối với khóa đã chọn, một số chữ cái của bản rõ ban đầu sẽ luôn được thay thế bằng cùng một chữ cái trong bản mã.

Do đó, các mật mã này có thể được tiết lộ bằng cách phân tích tần số mật mã. Phân tích tần số sử dụng thuộc tính của văn bản được mã hóa mà tần suất xuất hiện của các ký tự trong đó trùng với tần suất xuất hiện của các ký tự tương ứng trong văn bản thuần túy.

Nếu chúng ta xem xét rằng tần suất xuất hiện của các ký tự khác nhau trong các văn bản của ngôn ngữ tương ứng được phân phối không đồng đều (ví dụ: tần suất tương đối của sự xuất hiện của chữ "A" trong văn bản bằng tiếng Anh là 0.0817 và các chữ cái "V" là 0, 0098), sau đó, tính tần suất xuất hiện tương đối của các chữ cái trong bản mã, chúng ta có thể giả định rằng ký hiệu thường được tìm thấy trong bản mã tương ứng với ký hiệu thường thấy nhất trong các văn bản bằng ngôn ngữ tương ứng, và do đó tìm khóa k cho mật mã của Caesar.

Để tiết lộ các tham số a và hệ thống mật mã Affine, bạn cần tìm sự tương ứng của hai chữ cái - một chữ cái thường được tìm thấy trong bản mã và chữ cái thứ hai về tần xuất. Hiệu quả của việc phân tích tần số mật mã của Caesar với một từ khóa phụ thuộc phần lớn vào độ dài của cụm từ khóa được sử dụng.





# Mật mã thay thế - Mật mã Vigenere

Vào thế kỷ XVI, nhà ngoại giao Pháp Blaise de Vigenère đã đề xuất sửa đổi mật mã thay thế, sau này được gọi theo tên của ông. Trong mật mã này, khóa được đưa ra bởi một cụm từ của các chữ cái d. Cụm từ khóa được ký bằng một sự lặp lại bên dưới tin nhắn. Chữ cái của bản mã phải được tìm thấy tại giao điểm của cột, được xác định bằng chữ cái của bản rõ, và dòng được xác định bằng chữ cái của khóa:

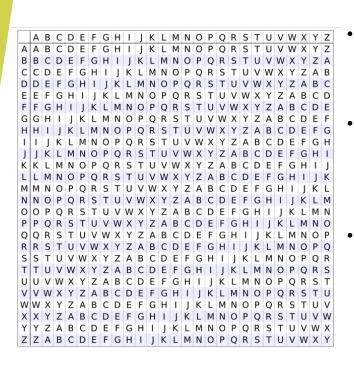
$$Vig_d(mi) = (m_i + k_{i \bmod d}) (mod n),$$

Trong đó:  $m_{i,} k_{i}$ ,  $Vig_d(m_i)$  - số thứ tự trong bảng chữ cái của các ký tự tiếp theo của bản rõ, khóa và bản mã, tương ứng. Chuyển đổi ngược lại như sau:

$$Vig_d^{-1}(mi) = (mi - k_{i \mod d} + n) (mod n).$$

#### Mật mã thay thế - Mật mã Vigenere

Bản rõ: HELLO Khóa: KEYKE Bản mã: RIJVS



Sự khác biệt cơ bản giữa mật mã này và tất cả các mật mã trước đó là nó thuộc về lớp thuật toán đa chữ cái - vì dễ thấy, các ký tự khác nhau của bản rõ có thể tương ứng với cùng một chữ cái của bản mã, tùy thuộc vào ký tự chính mà chúng được ngụy trang.

Điều này làm cho việc đếm tần suất xuất hiện của các ký tự trong bản mã trở nên vô nghĩa. Để phân tích mật mã của mật mã Vigenere, bạn có thể sử dụng phương pháp Kaziski. Vào giữa thế kỷ XIX, nhà toán học người Đức Kaziski đã đề xuất xác định độ dài của cụm mật khẩu bằng khoảng cách giữa các đoan giống nhau của bản mã

Giả sử hai đoạn mật mã giống hệt nhau đã được tìm thấy, khoảng cách giữa chúng là 20 ký tự. Điều này có thể có nghĩa là hai đoạn văn bản thuần túy giống hệt nhau đã được mã hóa với cùng một vị trí khóa. Điều này cho thấy rằng cụm mật khẩu dài 4, 5, 10 hoặc 20 ký tự. Bằng cách học (hoặc đoán) độ dài của cụm mật khẩu l, có thể thực hiện phân tích mật mã tần số của bản mã để lấy mẫu từng ký tự l của bản mã.





# Mật mã thay thế - Mật mã Vigenere

Để tăng độ dài của cụm mật khẩu và độ phức tạp của phân tích mật mã, bạn có thể sử dụng thành phần mật mã Vigenere, là mã hóa nhiều Vigenere với các cụm mật khẩu khác nhau. Nó có một phương trình

$$Vig^*(m_i) = (m_i + k_{i \mod dk} + l_{i \mod dl} + \dots + s_{i \mod ds}) \pmod n,$$

Trong đó  $k_{i,}$   $l_{i,}$  ...,  $s_i$  - ký tự của các cụm mật khẩu khác nhau. Để tăng khả năng chống giải mã, các cụm mật khẩu phải có các chu kỳ đơn giản khác nhau dk, dl, ..., ds.

Điều này sẽ khiến một nhà phân tích mật mã gặp khó khăn hơn nhiều trong việc tìm ra giá trị của chu kỳ lặp lại của một cụm mật khẩu thông thường, sẽ bằng với tích của độ dài của tất cả các cụm mật khẩu.



#### HÌNH VUÔNG POLYBIUS

Một sửa đổi khác của sự thay thế một chữ cái là hình vuông Polybius, trong đó ký hiệu bảng chữ cái được thay thế bằng một cặp số hoặc ký hiệu theo một quy tắc nhất định. Hãy xem xét một hình vuông, thường được gọi là bảng Polybius. Một chữ cái được tham chiếu bởi tọa độ của nó theo hàng và cột. Chữ cái đầu tiên trong hàng đầu tiên được mã là "11", chữ cái thứ tư trên hàng thứ hai sẽ được viết là "24",...

	1	2	3	4	5
1	A	В	C	D	E
2	F	G	Н	I,J	K
3	L	M	N	O	P
4	Q	R	S	T	U
5	V	W	X	Y	Z

#### Ví dụ minh họa:

Câu mẫu: TRUONG GIAO THONG VAN TAI TPHCM

Mật mã: 444245343322 22241134 4423343322 511133 441124 4435231332

#### HÌNH VUÔNG POLYBIUS



Bảng chữ cái được viết trong một hình chữ nhật như vậy, và sơ đồ ghi được giữ bí mật và tạo thành khóa mã hóa. Để có được các ma trận gần với hình vuông (6x6, 5x7, 6x5), có thể bao gồm các ký tự đặc biệt,

#### Phá mã:

Đối với hình vuông của Polybius, cũng có thể áp dụng phương pháp phân tích mật mã tần số. Những kết hợp chữ cái (bigram) được tìm thấy trong bản mã thường xuyên nhất sẽ tương ứng với các chữ cái được sử dụng phổ biến nhất trong bảng chữ cái. Thay thế chúng trong bản mã, các chữ cái còn lại có thể được thay thế bằng ngữ nghĩa của văn bản kết quả. Tổng số khóa (các tùy chọn khác nhau để đặt bảng chữ cái trong ma trận) là n!.

#### MÃ HOÁ BIGRAM

Vào đầu thế kỷ XVI, Johann Trisemus, một tu viện trưởng người Đức, đề nghị mã hóa hai ký tự cùng một lúc. Mật mã sử dụng một nguyên tắc tương tự được gọi là bigram. Thông thường, những mật mã như vậy sử dụng các bảng tương tự như hình vuông Polybius, chứa đầy các ký hiệu của bảng chữ cái được sử dụng. Mật mã bigram nổi tiếng nhất được gọi là Playfair. Nó đã được Vương quốc Anh sử dụng trong Chiến tranh thế giới thứ nhất. Văn bản mở được chia thành các cặp ký tự (bigrams) và văn bản mã hóa được xây dựng từ đó theo ba quy tắc rất đơn giản sau:





#### MÃ HOÁ BIGRAM

- 1. Nếu cả hai chữ cái bigram của văn bản gốc thuộc cùng một cột của bảng, thì các chữ cái của mật mã được coi là các chữ cái nằm bên dưới chúng. Do đó, ký tự RF cung cấp ký tự mật mã FO. Nếu chữ cái của bản rõ ở hàng dưới cùng, thì chữ cái tương ứng từ hàng trên cùng được lấy làm mật mã, ví dụ bigram SY cho ký tự mật mã YB. (Biểu đồ của một chữ cái hoặc một cặp chữ cái giống hệt nhau cũng tuân theo quy tắc này và văn bản GG cung cấp cho mật mã EE).
- 2. Nếu cả hai chữ cái bigram của văn bản nguồn thuộc cùng một hàng của bảng thì các chữ cái ở bên phải chúng được coi là chữ cái của mật mã. Do đó, biểu đồ GR cung cấp văn bản mật mã RA. Nếu ký tự của bản rõ nằm ở cột ngoài cùng bên phải, thì một ký tự từ cột ngoài cùng bên trái của cùng một hàng được lấy cho mật mã và biểu đồ AM cung cấp cho mật mã MB.
- 3. Nếu cả hai chữ cái bigram của văn bản mở nằm ở các hàng và cột khác nhau, thì hai chữ cái như vậy được lấy thay thế, để cả bốn chữ cái đó đại diện cho một hình chữ nhật. Ví dụ: bigram EW được mã hóa là HU. Việc điền vào ô vuông với bảng chữ cái có thể là ngẫu nhiên, hoặc nó có thể được xác định bởi một cụm từ khóa nhất định, tất cả các ký hiệu trong đó (nhưng không lặp lại) được viết ở đầu ma trận và sau đó là các chữ cái khác của bảng chữ cái được viết ra theo thứ tự.



Key: BIGRAM Bån rõ: HELLO-Bån mã:JFTTQ.





Năm 1854, Charles Wheatstone, người Anh, đã phát triển một mã hóa mới với bigram, được gọi là hình vuông kép. Mã hóa ở đây tương tự như mật mã Playfair, nhưng bigram được mã hóa bằng cách sử dụng hai bảng được điền ngẫu nhiên bằng các bảng chữ cái. Đối với một cặp ký tự từ tin nhắn ban đầu, một hình chữ nhật được tạo thành hai bảng theo quy tắc

Chữ cái đầu tiên trong bảng bên trái là một góc, chữ cái thứ hai ở bên phải là một góc khác. Các chữ cái của mã hoá bigram được lấy từ hai đỉnh còn lại của hình chữ nhật.

Nếu cả hai chữ cái nằm trong cùng một hàng, thì các chữ cái mật mã được lấy từ cùng một hàng, nhưng ở cột tiếp theo của bảng (đối với cột cuối cùng sẽ là từ cột đầu tiên). Một ví dụ về mã hóa bigram hai hình vuông được hiển thị:

В	I	G	R	A	M
C	D	E	F	H	J
K	L	N	0	P	Q
S	T	U	V	W	X
Y	Z		•	,	-

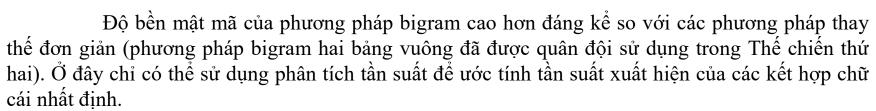


Key: BIGRAM và WELCOME

Bản rõ: **UKRAINA.**Bản mã: TNWFCLC,



#### MÃ HOÁ BIGRAM VỚI HAI BẢNG VUÔNG



Và mặc dù phân tích cho thấy rằng một số bigram được tìm thấy trong các văn bản của một ngôn ngữ nhất định thường xuyên hơn những ngôn ngữ khác, nhưng số lượng lớn các bigram thì không cho phép chúng ta xác định sự tương ứng giữa bigrams của bản rõ và bản mã.

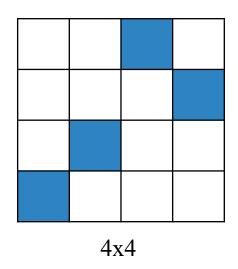
Nhà phân tích mật mã chỉ phải duyệt qua tất cả các tùy chọn có thể để sắp xếp các ký hiệu trong bảng (với n! Tùy chọn khác nhau). Hãy xem xét các thuật toán thuộc loại hoán vị.

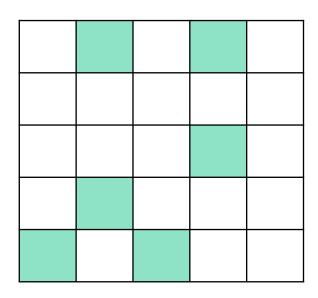




#### HÌNH VUÔNG CARDANO

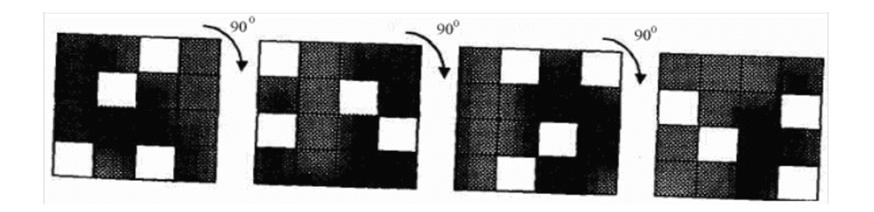
Vào thế kỷ 16, nhà toán học và triết học người Ý J. Cardano đã đề xuất một loại mật mã mới dựa trên một hoán vị rất đơn giản và đồng thời đáng tin cậy của các chữ cái trong bản rõ. Để mã hóa, ông ấy đề xuất sử dụng một hình vuông với một số lỗ hở trên đó. Các lỗ được cắt theo cách sao cho khi xoay hình vuông 90, 180 và 270 độ, tất cả các vị trí của hình vuông ban đầu lần lượt xuất hiện trong các khe và tại một thời điểm. Mật mã được đặt tên là hình vuông Cardano, một ví dụ về hình vuông 4x4 và 5x5 được hiển thị trong Hình.





#### HÌNH VUÔNG CARDANO

Khi mã hóa, đầu tiên hình vuông được đặt trên một tờ giấy ở vị trí ban đầu và phần đầu tiên của thông báo được viết vào các khe, sau đó hình vuông được quay 90° và phần thứ hai được viết vào các khe, v.v. Sau khi tất cả các ô của hình vuông được lấp đầy, bản mã được tạo ra.

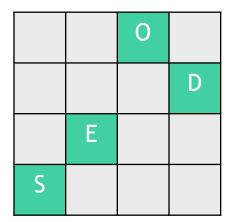




#### HÌNH VUÔNG CARDANO

Thông tin: **ODESSA** – **UKRAINA** 

#### Ma trận 4x4



S		0	
	Α		D
	Е		
S		-	

S		0	
	A	U	D
K	E		
S	R	-	

S	Α	0	
1	A	U	D
K	E	N	
S	R	-	Α

Bån mã: SAO IAUDKEN SR-A



#### HÌNH VUÔNG CARDANO

Để giải mã thông điệp, bạn cần có một bản sao chính xác của hình vuông được sử dụng để mã hóa (vị trí của các khe trên hình vuông là khóa).

Số lượng các tùy chọn khác nhau cho vị trí của các khe trong hình vuông NxN là

 $4^{\left(\frac{N^2}{4}\right)}$ , đối với hình vuông 6x6 cho 262.144 tùy chọn (tương đương với khóa 18 bit). Tuy nhiên, mật mã này (cũng như tất cả các hoán vị) bị suy yếu trong thực tế, trong quá trình phân tích mật mã, có thể sử dụng các đặc điểm của ngữ âm của ngôn ngữ quốc gia bản địa (các tổ hợp ký hiệu phổ biến nhất hoặc không được chấp nhận cho ngôn ngữ này, độ dài trung bình của từ, v..v..





#### MẬT MÃ HOÁN VỊ VỚI MỘT TỪ KHOÁ

Các chữ cái của từ khóa không lặp lại được viết ở hàng đầu tiên của bảng, do đó xác định số cột của từ khóa đó. Các chữ cái của tin nhắn được ghi lại trong bảng theo từng dòng. Bảng được hình thành theo cách này được sắp xếp theo cột, tiêu chí sắp xếp là thứ tự truyền ký tự của dòng đầu tiên trong bảng chữ cái. Sau khi phân loại, văn bản được mã hóa được viết lại theo từng cột.

Key: CALL

Thông tin: ODESSA – UKRAINA

C	A	L	L
0	D	E	S
S	A		-
	U	K	R
A	I	N	A

L	L	A	C
E	S	D	0
	-	A	S
K	R	U	
N	A	I	A

Mã hoá: E KNS-RADAUIOS A