CLASSICAL ENCRYPTION TECHNIQUES

Plaintext- original message

Ciphertext – coded message

Enciphering, encryption – process of converting from plaintext to ciphertext

Deciphering, decryption – restoring the plaintext from the ciphertext

Cryptography – area of study schemes for enciphering

Cryptographic system, cipher – scheme of enciphering

Cryptanalysis – techniques for deciphering a message without knowledge of the enciphering details

Cryptology – areas of cryptography and cryptanalysis

OUTLINE

1. SYMMETRIC CIPHER MODEL
2. SUBSTITUTION TECHNIQUES
3. TRANSPOSITION TECHNIQUES
4. ROTOR MACHINES
5. STEGANOGRAPHY

SYMMETRIC CIPHER MODEL

Symmetric (conventional) encryption scheme has the following ingredients

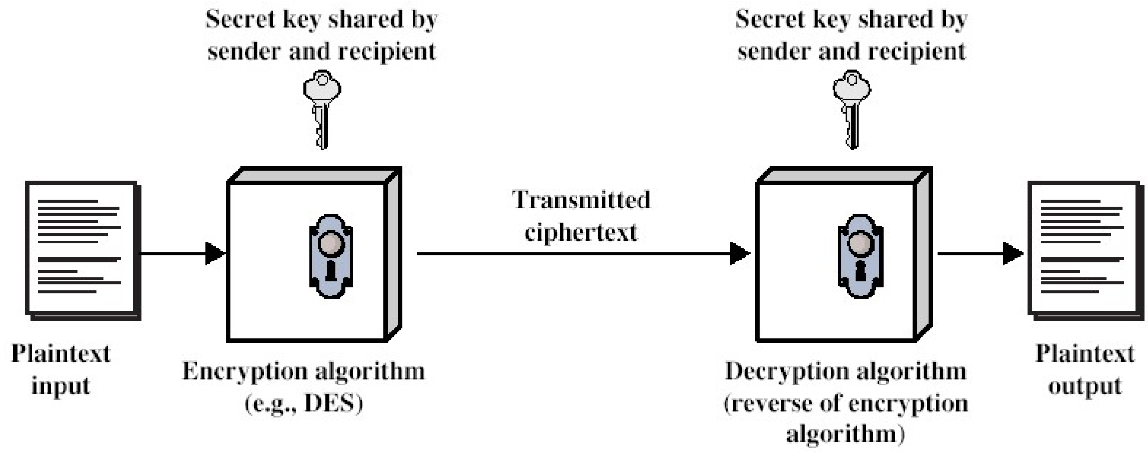


Figure 2.1 Simplified Model of Symmetric Encryption

There are 2 requirements for secure use of conventional encryption:

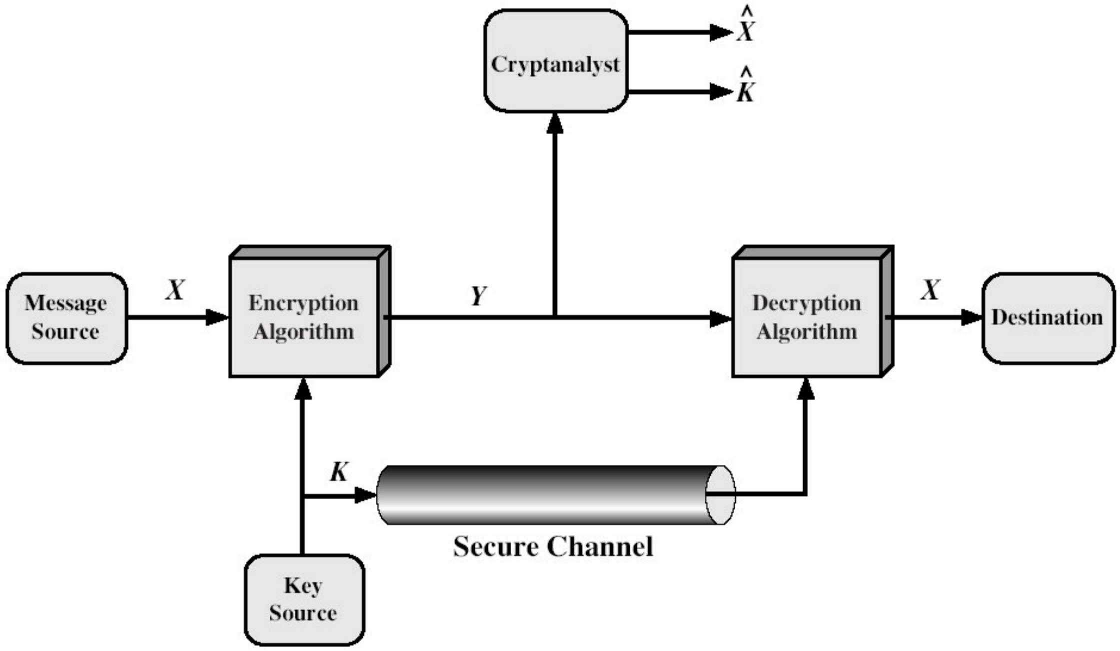
1. We need a strong encryption algorithm – the opponent should be unable to decrypt ciphertext or to discover the key even if s/he is in the possession of a number of ciphertexts together with the plaintext that produced each ciphertext

2. Sender and receiver must have obtained copies of the secret key in a secure fashion and must keep the key secure. If someone can discover the key and knows the algorithm, all communication using this key is readable

We assume that it is impractical to decrypt a message on the basis of the ciphertext plus knowledge of the encryption/decryption algorithm, i.e. we do not need to keep the algorithm secret; we need to keep only the key secret.

Let's consider essential elements of a symmetric encryption scheme:

SYMMETRIC CIPHER MODEL (CONT 1)



We can write:

$Y = E_K(X)$

$X = D_K(Y)$

Opponent knows Y, E, D. He may be interested to recover X or/and K. Knowledge of K gives him opportunity to read future messages.

CRYPTOGRAPHY

Cryptographic systems are characterized by

1. The type of operations used for transforming plaintext to ciphertext (substitution, transposition). Fundamental requirement – no information be lost
2. The number of keys used (1 key – symmetric, single-key, secret-key; 2 keys – asymmetric, two-key, public-key)
3. The way in which the plaintext is processed (block cipher, stream cipher). Stream cipher may be viewed as a block cipher with block size equal to 1 element.

CRYPTANALYSIS

There are two general approaches to attacking a conventional encryption scheme:

1. **Cryptanalysis:** attempts to use characteristics of the plaintext or even some plaintext-ciphertext pairs to deduce a specific plaintext or key being used
2. **Brute-force attack:** every possible key is tried until an intelligible translation into plaintext is obtained. On average, half of all possible keys should be tried to achieve success.

CRYPTANALYSIS (CONT 1)

Unconditionally secure encryption scheme – ciphertext generated by the scheme does not contain enough information to determine uniquely the corresponding plaintext, no matter how much ciphertext is available. Excepting a scheme known as one-time pad, there is no encryption algorithm that is unconditionally secure. Therefore, encryption algorithm should meet one or both of the following criteria:

- The cost of breaking the cipher exceeds the value of the encrypted information
- The time required to break the cipher exceeds the useful lifetime of the information

Such algorithm is called **computationally secure**. Table below shows how much time is involved for various key sizes. The 56-bit key size is used with the DES (Data Encryption Standard), 168-bit – for triple DES, 128-bit – for AES (Advanced Encryption Standard). Results are also shown for substitution codes that use 26-character key, in which all possible permutations of the 26 characters serve as keys. It is assumed that it take 1 μs to perform a single decryption or encryption (in last column – 10^6 decryptions per 1 μs)

Table 2.2 Average Time Required for Exhaustive Key Search

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} 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

CRYPTANALYSIS (CONT 2)

All forms of cryptanalysis for symmetric encryption try to exploit the fact that traces of structure or pattern in the plaintext may survive encryption and be discernible in the ciphertext. Cryptanalysis for public-key schemes tries to use mathematical properties of pair of keys to deduce one from the other.

SUBSTITUTION TECHNIQUE

A substitution technique is one in which the letters of plaintext are replaced by other letters or by numbers. If the plaintext is viewed as a sequence of bits, then substitution involves replacing plaintext bit patterns with ciphertext bit patterns

CAESAR CIPHER

It was used by Julius Caesar. The Caesar cipher involves replacing each letter of the alphabet with the letter standing three places further down the alphabet

For example

Plain: meet me after the toga party

Cipher: PHHW PH DIWHU WKH WRJD SDUWB

Transformation is made using the following mapping:

Plain: 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

Cipher: 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

Let us assign a numerical equivalent to each letter from 0 to 25. Then the algorithm may be expressed as follows. For each plaintext letter p, substitute the ciphertext letter C:

C=E(p)=(p+3) mod 26

A shift may be of any amount, so that general Caesar algorithm is

CAESAR CIPHER (CONT 1)

$C=E(p)=(p+k) \bmod 26$,

where k takes on a value in the range 1 to 25. The decryption algorithm is simply

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

If it is known that a given ciphertext is a Caesar cipher, then a brute-force cryptanalysis is easily performed: simply try all possible 25 keys.

Three important characteristics of this problem enable us to use brute-force cryptanalysis:

1. The encryption and decryption algorithms are known
2. There are only 25 keys to try
3. The language of the plaintext is known and easily recognizable

In most networking situations algorithms are assumed to be known. Brute-force analysis is impractical when algorithm employs large size of keys. The 3^{rd} characteristic is also significant. If the language of the plaintext is not known, then the plaintext output may not be recognizable.

CAESAR CIPHER (CONT 2)

KEY	PHHW PH DIWHU WKH WRJD SDUWB
1	oggv og chvgt vjg vqic rctva
2	nffu nf bgufs uif uphb qbsuz
3	meet me after the toga party
4	ldds ld zesdq sgd snfz ozqsx
5	kccr kc ydrcp rfc rmey nyprw
6	jbbq jb xcqbo qeb qldx mxoqv
7	iaap ia wbpan pda pkcw lwnpu
8	hzzo hz vaozm ocz ojbv kvmot
9	gyyn gy uznyl nby niau julns
10	fxxm fx tymxk max mhzt itkmr
11	ewwl ew sxlwj lzw lgys hsjlq
12	dvvk dv rwkvi kyv kfxr grikp
13	cuuj cu qvjuh jxu jewq fqhjo
14	btti bt puitg iwt idvp epgin
15	assh as othsf hvs hcuo dofhm
16	zrrg zr nsgre gur gbtn cnegl
17	yqqf yq mrfqd ftq fasm bmdfk
18	xppe xp lqepc esp ezrl alcej
19	wood wo kpdob dro dyqk zkbdi
20	vnnc vn jocna cqn cxpj yjach
21	ummb um inbmz bpm bwoi xizbg
22	tlla tl hmaly aol avnh whyaf
23	skkz sk glzkx znk zumg vgxze
24	rjjy rj fkyjw ymj ytlf ufwyd
25	qiix qi ejxiv xli xske tevxc

Figure 2.3 Brute-Force Cryptanalysis of Caesar Cipher

CAESAR CIPHER (CONT 3)

Furthermore, if the input is compressed in some manner, again recognition is difficult. Below is example of compression by ZIP:

```
¯+Wµ"— Ω—O)≤4{∞‡¸ ë~Ω%ràu·¯Í ◊¯Z-
Ú≠2Ò#Åæ∂ œ«q7¸Ωn·®3N◊Ú Œz'Y-ƒ∞Í[±Û_ èΩ,<NO¬±«˘xã  Åä£èü3Å
x}ö§kºÂ
_yÍ ˆΔÉ] ˛  J⁄˚iTê&ı'c<uΩ-
 ÄD(G WÄC~y_ïõÄW PÔı«ÎÜ†ç], ¡˘Ì^üÑ
π˘≈˘L˘9Ogflo¯&Œ≤¬≤ØÔ§˝: ˘Œ!SGqèvo˜ úError!
```

Figure 2.4 Sample of Compressed Text

If this file is then encrypted with a simple substitution cipher (expanded to include more than just 26 characters), then the plaintext may not be recognized

MONOALPHABETIC CIPHERS

With only 25 keys Caesar cipher is far from secure. A dramatic increase in the key space may be achieved by allowing an arbitrary substitution. If instead of

Plain: *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*

Cipher: 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

the cipher line can be any permutation of the 26 alphabetic symbols, then there are 26! or greater than $4*10^{26}$ possible keys. There is however another line of attack. If the cryptanalyst knows the nature of the plaintext (e.g., non-compressed English text), then the analyst can exploit the regularities of the language.

MONOALPHABETIC CIPHERS (CONT 1)

Let's consider example of ciphertext:

UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZ
VUEPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSX
EPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMOHMQ

As a first step, relative frequency of the letters can be determined and compared to a standard frequency distribution for English:

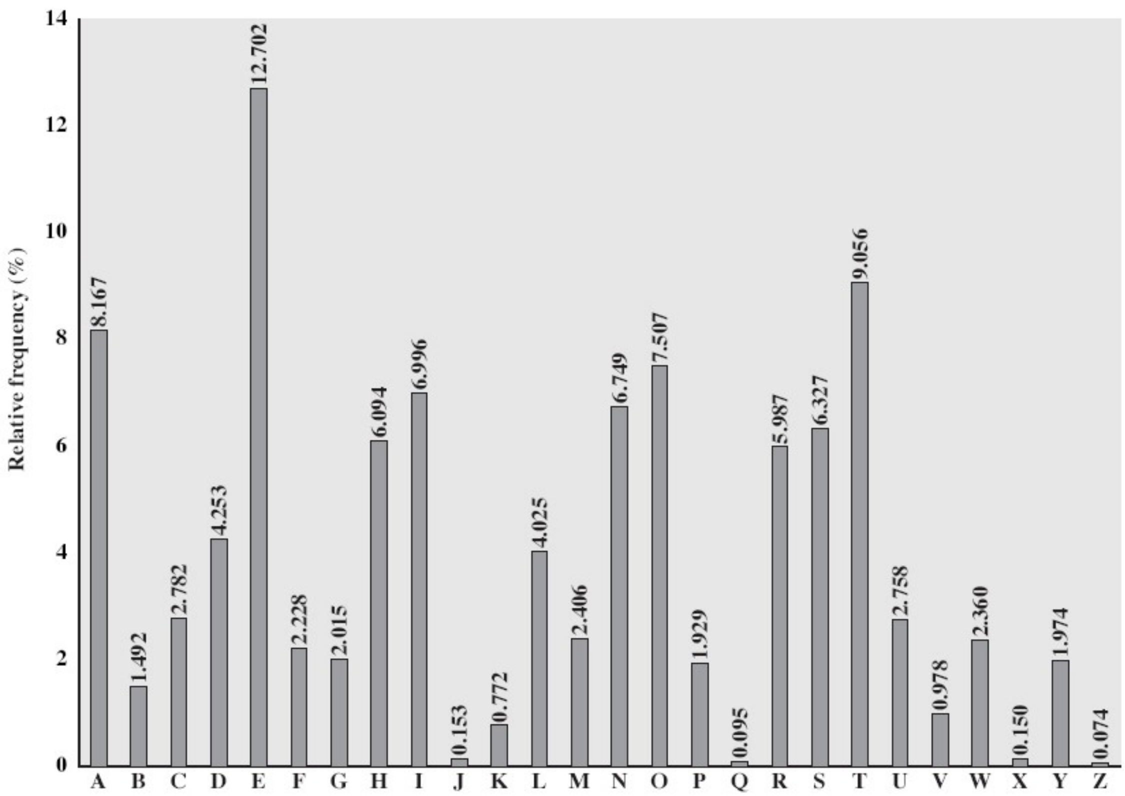


Figure 2.5 Relative Frequency of Letters in English Text

The relative frequencies of the letters in the ciphertext (in percentages):

MONOALPHABETIC CIPHERS (CONT 2)

P 13.33	H 5.83	F 3.33	B 1.67	C 0.00
Z 11.67	D 5.00	W 3.33	G 1.67	K 0.00
S 8.33	E 5.00	Q 2.50	Y 1.67	L 0.00
U 8.33	V 4.17	T 2.50	I 0.83	N 0.00
O 7.50	X 4.17	A 1.67	J 0.83	R 0.00
M 6.67				

Comparing this with Fig.2.5, it seems likely that cipher letters P and Z are the equivalents of plain letters e and t, but it is not certain which is which. The letters S,U,O,M, and H are all of the relatively high frequency and probably correspond to plain letters from the set {a,h,i,n,o,r,s}. The letters with the lowest frequencies (A,B,G,Y,I,J) are likely included in the set {b,j,k,q,v,x,z}. Now we could make some tentative assignments and start to fill plaintext to see if it looks like a reasonable "skeleton" of a message.

Another way, to consider frequency of two-letter combinations, is known as digrams. The most common digram is th. In our ciphertext, the most common digram is ZW, which appears 3 times. So, we make correspondence: Z – t, W – h. Then, P is equated with e. Now notice that sequence ZWP appears in the ciphertext, and we can translate it as "the". Next, notice ZWSZ in the first line. If they form a complete word, it will be th_t. If so, S equates with a. So far, then, we have

MONOALPHABETIC CIPHERS (CONT 3)

```
UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZ
 t a        e  e te  a that e e a       a
VUEPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSX
   e t   ta t ha e ee  a e  th    t  a
EPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMOHMQ
 e  e e tat  e   the  et
```

Continued analysis of frequencies plus trial and error may lead us to the solution:

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

Two principal methods are used in substitution ciphers to lessen the extent to which the structure of the plaintext survives in the ciphertext: One approach is to encrypt multiple letters of the plaintext (Playfair Cipher, Hill Cipher), and the other is to use multiple cipher alphabets (Polyalphabetic Ciphers)

PLAYFAIR CIPHER

The best-known multiple-letter encryption cipher is the Playfair (invented in 1854 by Sir Charles Wheatstone, but it bears the name of his friend Baron Playfair of St. Andrews, who championed the cipher at the British foreign office), which treats digrams in the plaintext as single units and translates these units into ciphertext digrams.

PLAYFAIR CIPHER (CONT 1)

The Playfair algorithm is based on the use of a 5x5 matrix of letters constructed using a keyword. In the case of keyword *monarchy*, matrix is as follows:

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

The matrix is constructed by filling in the letters of the keyword (minus duplicates) from left to right and from top to bottom, and then filling in the remainder of the matrix with the remaining letters in alphabetic order. The letters I and J count as one letter. Plaintext is encrypted two letters at a time, according to the following rules:

1. Repeating plaintext letters that would fall in the same pair are separated with a filler letter, such as x, so that *balloon* will be treated as *ba lx lo on*

2. Plaintext letters that would fall in the same row of matrix are each replaced with the letter to the right, with the first element of the row circularly following the last. For example, *ar* is encrypted as *RM*.

3. Plaintext letters that fall in the same column are each replaced by the letter beneath, with the top element of the row circularly following the last. For example, *mu* is encrypted as *CM*.

PLAYFAIR CIPHER (CONT 2)

4. Otherwise, each plaintext letter is replaced by the letter that lies in its own row and the column occupied by the other plaintext letter. Thus, *hs* becomes *BP*, and *ea* becomes *IM* (or *JM*, as the encipherer wishes).

As far as number of digrams is 26x26=676 is significantly greater than number of letters, frequency analysis becomes much more difficult. For these reasons, Playfair cipher was for a long time considered unbreakable. It was used as standard field system by the British Army in World War I and still enjoyed considerable use by U.S.Army and other Allied forces during World War II.

Despite this level of confidence in its security, the Playfair cipher is relatively easy to break because it still leaves much of the structure of the plaintext language intact. A few hundred letters of ciphertext are generally sufficient.

HILL CIPHER

It was developed by the mathematician Lester Hill in 1929. The encryption algorithm takes m successive plaintext letters and substitutes for them m ciphertext letters. The substitution is determined by m linear equations in which each character is assigned a numerical value:

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

For m=3, the system can be described as follows:

$C1=(k11p1+k12p2+k13p3) \bmod 26$

$C2=(k21p1+k22p2+k23p3) \bmod 26$

$C3=(k31p1+k32p2+k33p3) \bmod 26$

HILL CIPHER (CONT 1)

This can be expressed in terms of column vectors and matrices:

C=KP mod 26,

where C and P are column vectors of length 3, representing the plaintext and ciphertext, and K is 3x3 matrix, representing the encryption key.

Operations are performed mod 26.

For example, consider the plaintext “paymoremoney”, and use the encryption key

K=

17	17	5
21	18	21
2	2	19

The first 3 letters of the plaintext are represented by the vector (15 0 24). Then K(15 0 24) = (375 819 486) mod 26 = (11 13 18) = LNS. Continuing in this fashion, the ciphertext for the entire plaintext is LNSHDLEWMTRW.

Decryption requires using the inverse of the matrix K. The inverse K^{-1} of a matrix K is defined by K K^{-1} = K^{-1} K=I, where I is the unit matrix (1-s on the diagonal, other elements – zeroes). The inverse of the matrix does not always exist, but when it does, it satisfies the preceding equation. In this case, the inverse is

HILL CIPHER (CONT 2)

K^{-1}=

4	9	15
15	17	6
24	0	17

This is demonstrated as follows:

$K K^{-1}$ =

443	442	442
858	495	780
494	52	365

And after taking mod 26 of the elements above, unit matrix is obtained.

In general terms, the Hill system can be expressed as follows:

$C=E_K(P)=KP \bmod 26$

$P= D_K(C)=K^{-1}C \bmod 26 = K^{-1}KP = P$

As with Playfair, the strength of the Hill cipher is that it completely hides single-letter frequencies.

Although the Hill cipher is strong against a ciphertext-only attack (opponent has only ciphertext), it is easily broken with a known plaintext attack (opponent has pairs plaintext – ciphertext). For an m*m Hill cipher, suppose we have m plaintext-ciphertext pairs, each of length m. We label the pairs $Pj=(p_{1j}, p_{2j},\ldots, p_{mj})$ and $Cj=(c_{1j}, c_{2j},\ldots, c_{mj})$ such that $Cj=KPj$ for $1<=j<=m$ and for some unknown key matrix K. Now define two m*m matrices $X=(p_{ij})$ and $Y=(c_{ij})$.

HILL CIPHER (CONT 3)

Then we can form matrix equation Y=KX. If X has an inverse, then we can determine $K=YX^{-1}$. If X is not invertible, then a new version of X can be formed until an invertible X is obtained.

Suppose that the plaintext "friday" is encrypted using a 2*2 Hill cipher to yield the ciphertext PQCFKU. Thus, we know that

K(5 17) = (15 16);

K(8 3) = (2 5);

K(0 24) = (10 20).

Using the first 2 plaintext-ciphertext pairs, we have

$$\begin{pmatrix} 15 & 2 \\ 16 & 5 \end{pmatrix} = K \begin{pmatrix} 5 & 8 \\ 17 & 3 \end{pmatrix} \bmod 26$$

The inverse of X can be computed:

$$\begin{pmatrix} 5 & 8 \\ 17 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} 9 & 2 \\ 1 & 15 \end{pmatrix}$$

$$K = \begin{pmatrix} 15 & 2 \\ 16 & 5 \end{pmatrix} \begin{pmatrix} 9 & 2 \\ 1 & 15 \end{pmatrix} = \begin{pmatrix} 137 & 60 \\ 149 & 107 \end{pmatrix} \bmod 26 = \begin{pmatrix} 7 & 8 \\ 19 & 3 \end{pmatrix}$$

Let's check now that this key matrix produces required transformation:

$$\begin{pmatrix} 7 & 8 \\ 19 & 3 \end{pmatrix} \begin{pmatrix} 5 \\ 17 \end{pmatrix} = \begin{pmatrix} 35+136 \\ 95+51 \end{pmatrix} = \begin{pmatrix} 171 \\ 146 \end{pmatrix} \bmod 26 = \begin{pmatrix} 15 \\ 16 \end{pmatrix}$$

$$\begin{pmatrix} 7 & 8 \\ 19 & 3 \end{pmatrix} \begin{pmatrix} 8 \\ 3 \end{pmatrix} = \begin{pmatrix} 56+24 \\ 152+9 \end{pmatrix} = \begin{pmatrix} 80 \\ 161 \end{pmatrix} \bmod 26 = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 7 & 8 \\ 19 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 24 \end{pmatrix} = \begin{pmatrix} 192 \\ 72 \end{pmatrix} \bmod 26 = \begin{pmatrix} 10 \\ 20 \end{pmatrix}$$