ECE 646 - Lecture 6

Historical Ciphers

Why (not) to study historical ciphers?

AGAINST

Not similar to modern ciphers

Long abandoned

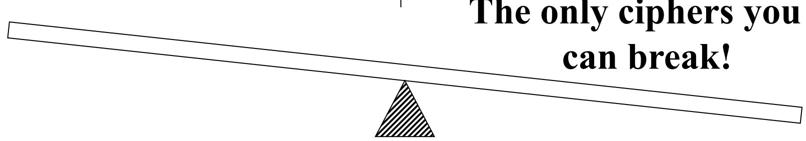
FOR

Basic components became a part of modern ciphers

Under special circumstances modern ciphers reduce to historical ciphers

Influence on world events

The only ciphers you can break!



Secret Writing

Steganography
(hidden messages)

Cryptography
(encrypted messages)

Substitution Transformations

Codes
(replace words)

Substitution Ciphers
(replace letters)

Transposition Ciphers
(change the order of letters)

Selected world events affected by cryptology

1586 - trial of Mary Queen of Scots - substitution cipher

1917 - Zimmermann telegram, America enters World War I

1939-1945 Battle of England, Battle of Atlantic, D-day -
ENIGMA machine cipher

1944 – world's first computer, Colossus -
German Lorenz machine cipher

1950s – operation Venona – breaking ciphers of soviet spies
stealing secrets of the U.S. atomic bomb
– one-time pad

Ciphers used predominantly in the given period(1)

Cryptography			Cryptanalysis	
100 B.C.		**Shift ciphers**		
			Frequency analysis al-Kindi, Baghdad	IX c.
		Monoalphabetic substitution cipher		
1586	Invention of the Vigenère Cipher		Black chambers	XVIII c.
		Homophonic ciphers		
		Vigenère cipher (Simple polyalphabetic substitution ciphers)	Kasiski's method	1863
1919	Invention of rotor machines		Index of coincidence William Friedman	1918
		Electromechanical machine ciphers (Complex polyalphabetic substitution ciphers)		
1926	Vernam cipher (one-time pad)			

Ciphers used predominantly in the given period(2)

Cryptography			Cryptanalysis	
			Reconstructing ENIGMA Rejewski, Poland	1932
1949	Shennon's theory of secret systems		Polish cryptological bombs, and perforated sheets	1939
		one-time pad Stream Ciphers S-P networks	British cryptological bombs, Bletchley Park, UK	1945
			Breaking Japanese "Purple" cipher	
1977	Publication of DES			1977
		DES		
		Triple DES		1990
			DES crackers	
2001		**AES**		2001

Substitution Ciphers (1)

1. Monalphabetic (simple) substitution cipher

$$M = \quad m_1 \quad m_2 \quad m_3 \quad m_4 \quad .\,.\,.\,. \quad m_N$$
$$C = \quad f(m_1) \quad f(m_2) \quad f(m_3) \quad f(m_4) \quad .\,.\,.\,. \quad f(m_N)$$

Generally f is a random permutation, e.g.,

$$f = \begin{pmatrix} a & b & c & d & e & f & g & h & i & j & k & l & m & n & o & p & q & r & s & t & u & v & w & x & y & z \\ s & l & t & a & v & m & c & e & r & u & b & q & p & d & f & k & h & w & y & g & x & z & j & n & i & o \end{pmatrix}$$

Key = f

Number of keys = 26! $\approx 4 \cdot 10^{26}$

Monalphabetic substitution ciphers Simplifications (1)

A. Caesar Cipher

$$c_i = f(m_i) = m_i + 3 \bmod 26$$

$$m_i = f^{-1}(c_i) = c_i - 3 \bmod 26$$

No key

B. Shift Cipher

$$c_i = f(m_i) = m_i + k \bmod 26$$

$$m_i = f^{-1}(c_i) = c_i - k \bmod 26$$

Key = k

Number of keys = 26

Coding characters into numbers

A	⇔	0	N	⇔	13
B	⇔	1	O	⇔	14
C	⇔	2	P	⇔	15
D	⇔	3	Q	⇔	16
E	⇔	4	R	⇔	17
F	⇔	5	S	⇔	18
G	⇔	6	T	⇔	19
H	⇔	7	U	⇔	20
I	⇔	8	V	⇔	21
J	⇔	9	W	⇔	22
K	⇔	10	X	⇔	23
L	⇔	11	Y	⇔	24
M	⇔	12	Z	⇔	25

Caesar Cipher: Example

Plaintext: I CAME I SAW I C O N Q U E R E D

8 2 0 12 4 8 18 0 22 8 2 14 13 16 20 4 17 4 3

---

11 5 3 15 7 11 21 3 25 11 5 17 16 19 23 7 20 7 6

Ciphertext: L F D P H L V D Z L F R Q T X H U H G

Monalphabetic substitution ciphers
Simplifications (2)

C. Affine Cipher

$$c_i = f(m_i) = k_1 \cdot m_i + k_2 \bmod 26$$
$$\gcd(k_1, 26) = 1$$

$$m_i = f^{-1}(c_i) = k_1^{-1} \cdot (c_i - k_2) \bmod 26$$

Key = (k_1, k_2)
Number of keys = 12·26 = 312

Most frequent single letters

Average frequency in a random string of letters:

$$\frac{1}{26} = 0.038 = 3.8\%$$

Average frequency in a long English text:

E	—	13%
T, N, R, I, O, A, S	—	6%-9%
D, H, L	—	3.5%-4.5%
C, F, P, U, M, Y, G, W, V	—	1.5%-3%
B, X, K, Q, J, Z	—	< 1%

Most frequent digrams, and trigrams

Digrams:

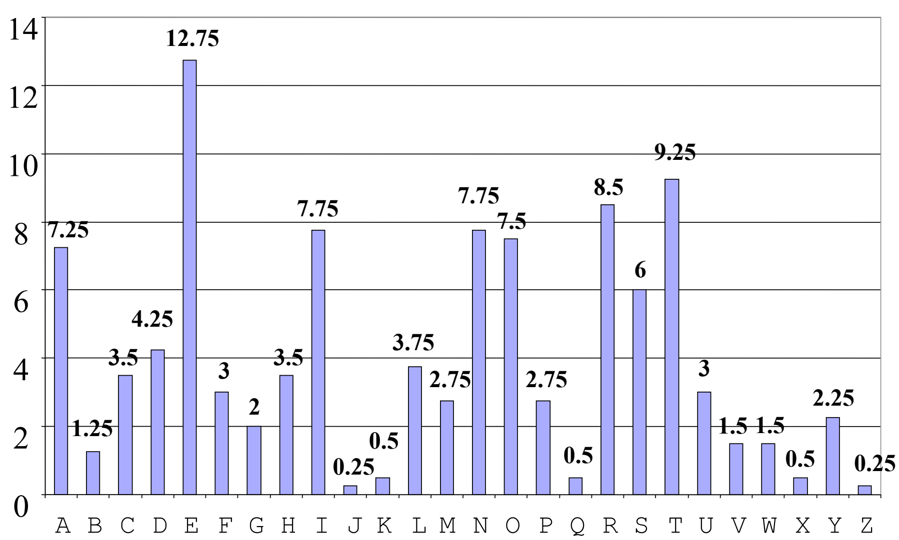
TH, HE, IN, ER, RE, AN, ON, EN, AT

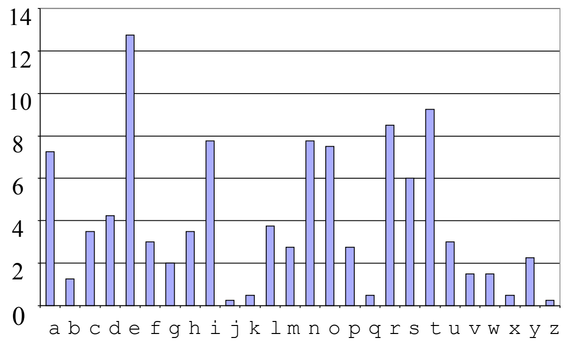
Trigrams:

THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR, DTH

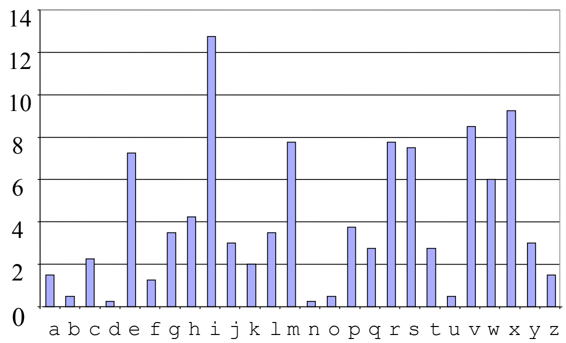
Relative frequency of letters in a long English text

by Stallings

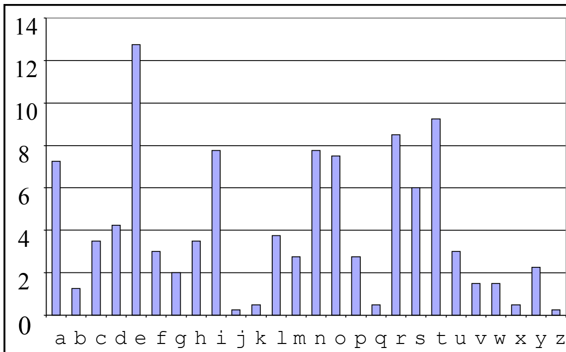




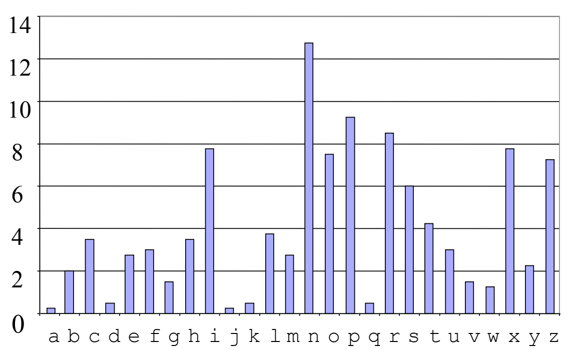
Character frequency
in a long English
plaintext



Character frequency
in the corresponding
ciphertext
for a shift cipher

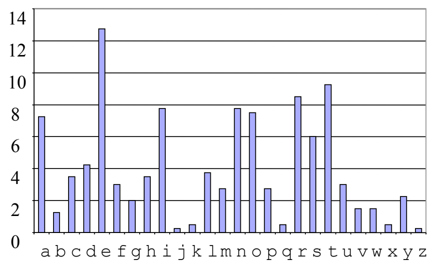


Character frequency
in a long English
plaintext

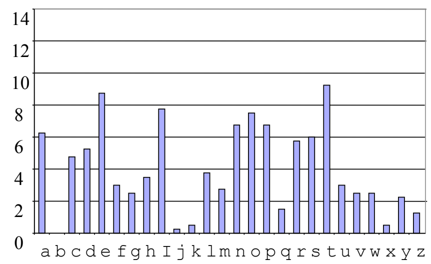


Character frequency
in the corresponding
ciphertext
for a general
monoalphabetic
substitution cipher

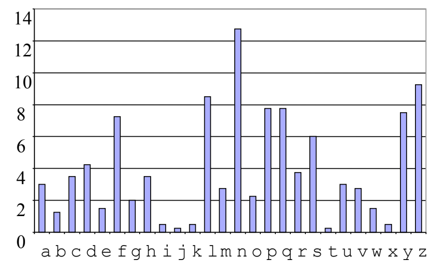
Frequency analysis attack: relevant frequencies



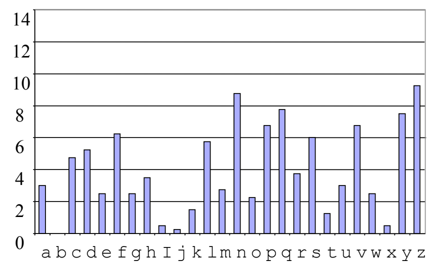
Long English text T



Short English message M



Ciphertext of the long English text T



Ciphertext of the short English message M

Frequency analysis attack (1)

Step 1: Establishing the relative frequency of letters in the ciphertext

Ciphertext:

```
FMXVE DKAPH FERBN DKRXR SREFM ORUDS
DKDVS HVUFE DKAPR KDLYE VLRHH RH
```

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

R	**- 8**
D	**- 7**
E, H, K	**- 5**

Frequency analysis attack (2)

Step 2: Assuming the relative frequency of letters in the corresponding message, and deriving the corresponding equations

Assumption: Most frequent letters in the message: E and T

Corresponding equations:

E → R	f(E) = R
T → D	f(T) = D
4 → 17	f(4) = 17
19 → 3	f(19) = 3

Frequency analysis attack (3)

Step 3: Verifying the assumption for the case of <u>affine cipher</u>

$$f(4) = 17$$
$$f(19) = 3$$

⇩

$$4 \cdot k_1 + k_2 \equiv 17 \pmod{26}$$
$$19 \cdot k_1 + k_2 \equiv 3 \pmod{26}$$

⇩

$$15 \cdot k_1 \equiv -14 \pmod{26}$$

⇩

$$15 \cdot k_1 \equiv 12 \pmod{26}$$

Substitution Ciphers (2)

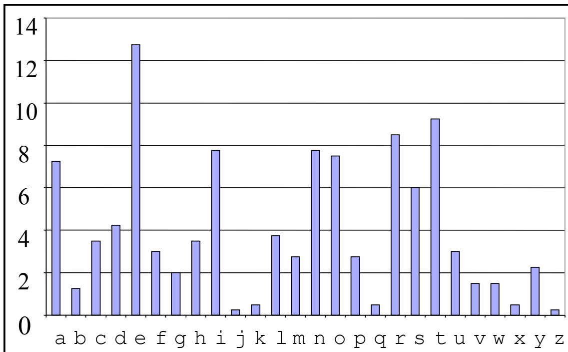
2. Polyalphabetic substitution cipher

$$
\begin{array}{llll}
M = & m_1 & m_2 & \dots & m_d \\
 & m_{d+1} & m_{d+2} & \dots & m_{2d} \\
 & m_{2d+1} & m_{2d+2} & \dots & m_{3d} \\
 & & \dots\dots & & \\
C = & f_1(m_1) & f_2(m_2) & \dots & f_d(m_d) \\
 & f_1(m_{d+1}) & f_2(m_{d+2}) & \dots & f_d(m_{2d}) \\
 & f_1(m_{2d+1}) & f_2(m_{2d+2}) & \dots & f_d(m_{3d}) \\
 & & & \dots\dots &
\end{array}
$$

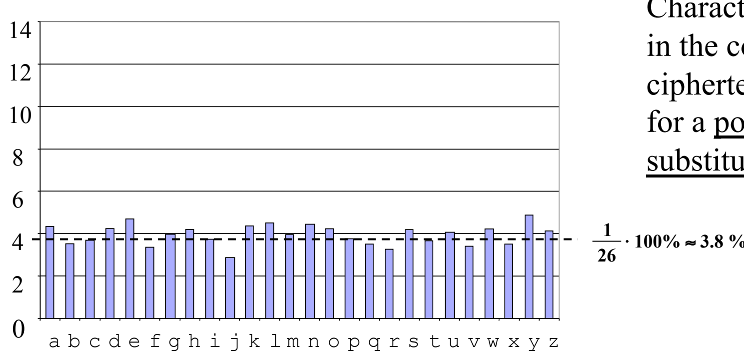
d is a **period** of the cipher

$$\text{Key} = d, f_1, f_2, \dots, f_d$$

Number of keys for a given period d = $(26!)^d \approx (4 \cdot 10^{26})^d$



Character frequency in a <u>long</u> English plaintext



Character frequency in the corresponding ciphertext for a <u>polyalphabetic substitution cipher</u>

Polyalphabetic substitution ciphers Simplifications (1)

A. Vigenère cipher: polyalphabetic shift cipher

Invented in 1568

$$c_i = f_{i \bmod d}(m_i) = m_i + k_{i \bmod d} \bmod 26$$

$$m_i = f^{-1}_{i \bmod d}(m_i) = m_i - k_{i \bmod d} \bmod 26$$

$$\text{Key} = k_0, k_1, \ldots, k_{d-1}$$

Number of keys for a given period $d = (26)^d$

Vigenère Square

plaintext:		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
	3	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
		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	a
		c	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
		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
		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
		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	e
		g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f
		h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g
		i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h
		j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i
		k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j
		l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k
		m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l
Key = “nsa”	**1**	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m
		o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n
		p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
		q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
		r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
	2	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r
		t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s
		u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
		v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u
		w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v
		x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
		y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x
		z	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

Vigenère Cipher - Example

Plaintext: TO BE OR NOT TO BE

Key: NSA

Encryption:

T O B
E O R
N O T
T O B
E

G G B
R G R
A G T
G G B
R

Ciphertext: GGBRGRAGTGGBR

Determining the period of the polyalphabetic cipher
Kasiski's method

Ciphertext: G G B R G R A G T G G B R

Distance = 9

Period d is a divisor of the distance between identical blocks of the ciphertext

In our example: d = 3 or 9

Index of coincidence method (1)

n_i - number of occurances of the letter *i* in the ciphertext

i = a .. z

N - length of the ciphertext

p_i = frequency of the letter *i* for a long ciphertext

$$p_i = \lim_{N \to \infty} \frac{n_i}{N}$$

$$\sum_{i=a}^{z} p_i = 1$$

Index of coincidence method (2)

Measure of roughness:

$$M.R. = \sum_{i=a}^{z} \left(p_i - \frac{1}{26} \right)^2 = \sum_{i=a}^{z} p_i^2 - \frac{1}{26}$$

M.R.	**0.028**	**0.014**	**0.006**	**0.003**
period	**1**	**2**	**5**	**10**

Index of coincidence method (3)

Index of coincidence

The approximation of $\sum_{i=a}^{z} p_i^2$

Definition:

Probability that two random elements of the ciphertext are identical

Formula:

$$I.C. = \sum_{i=a}^{z} \frac{\binom{n_i}{2}}{\binom{N}{2}} = \frac{\sum_{i=a}^{z} (n_i - 1) \cdot n_i}{(N - 1) \cdot N}$$

Index of coincidence method (4)

Measure of roughness

$$M.R. = I.C. - \frac{1}{26} = \frac{\sum_{i=a}^{z} (n_i - 1) \cdot n_i}{(N - 1) \cdot N} - \frac{1}{26}$$

M.R.	**0.028**	**0.014**	**0.006**	**0.003**
period	**1**	**2**	**5**	**10**

Polyalphabetic substitution ciphers Simplifications (2)

B. Rotor machines used before and during the WWII

Country	Machine	Period
Germany:	Enigma	$d=26\cdot25\cdot26 = 16{,}900$
U.S.A.:	M-325, Hagelin M-209	
Japan:	"Purple"	
UK:	Typex	$d=26\cdot(26-k)\cdot26$, $k=5, 7, 9$
Poland:	Lacida	$d=24\cdot31\cdot35 = 26{,}040$

Substitution Ciphers (3)

3. Running-key cipher

$$
\begin{array}{llllll}
M = & m_1 & m_2 & m_3 & m_4 & \ldots\, m_N \\
K = & k_1 & k_2 & k_3 & k_4 & \ldots\, k_N
\end{array}
$$

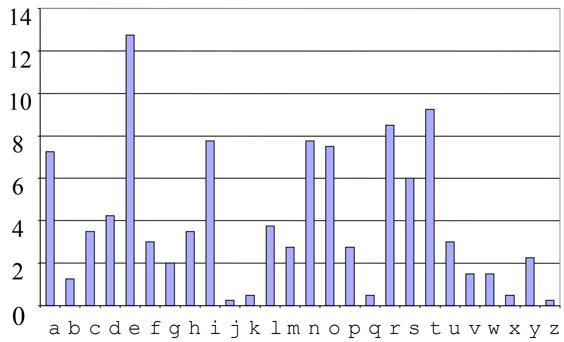
K is a fragment of a book

$$
\begin{array}{llllll}
C = & c_1 & c_2 & c_3 & c_4 & \ldots\, c_N
\end{array}
$$

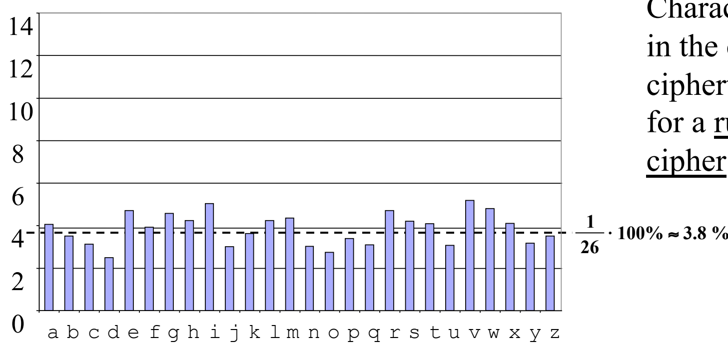
$$c_i = m_i + k_i \bmod 26$$

$$m_i = c_i - k_i \bmod 26$$

Key: book (title, edition), position in the book (page, row)



Character frequency in a long English plaintext



Character frequency in the corresponding ciphertext for a running-key cipher

Substitution Ciphers (4)

4. Polygram substitution cipher

$$
\begin{array}{lccccccc}
M = & m_1 & m_2 & \ldots & m_d & - & M_1 \\
 & m_{d+1} & m_{d+2} & \ldots & m_{2d} & - & M_2 \\
 & m_{2d+1} & m_{2d+2} & \ldots & m_{3d} & - & M_3 \\
 & & & \ldots\ldots & & & \\
C = & c_1 & c_2 & \ldots & c_d & - & C_1 \\
 & c_{d+1} & c_{d+2} & \ldots & c_{2d} & - & C_2 \\
 & c_{2d+1} & c_{2d+2} & \ldots & c_{3d} & - & C_3 \\
 & & & \ldots\ldots & & &
\end{array}
$$

d is the length of a message block

$C_i = f(M_i) \qquad M_i = f^{-1}(C_i)$

Key = d, f

Number of keys for a given block length d = $(26^d)!$

Playfair Cipher

1854

Key:

PLAYFAIR IS A DIGRAM CIPHER

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

Convention 1

message	P O	L A	N D
ciphertext	A K	A Y	Q R

Convention 2

message	P O	L A	N D
ciphertext	K A	R S	R Q

Hill Cipher

1929

Ciphering:

$$C_{[1xd]} = M_{[1xd]} \cdot K_{[dxd]}$$

$$(c_1, c_2, \ldots, c_d) = (m_1, m_2, \ldots, m_d) \begin{pmatrix} k_{11}, k_{12}, \ldots, k_{1d} \\ \\ k_{d1}, k_{d2}, \ldots, k_{dd} \end{pmatrix}$$

ciphertext block = *message block* · *key matrix*

Hill Cipher

Deciphering:

$$M_{[1xd]} = C_{[1xd]} \cdot K^{-1}_{[dxd]}$$

message block = ciphertext block · inverse key matrix

where

$$K_{[dxd]} \cdot K^{-1}_{[dxd]} = \begin{pmatrix} 1, 0, \ldots, 0, 0 \\ 0, 1, \ldots, 0, 0 \\ \ldots\ldots\ldots\ldots \\ 0, 0, \ldots, 1, 0 \\ 0, 0, \ldots, 0, 1 \end{pmatrix}$$

key matrix · inverse key matrix = identity matrix

Hill Cipher - Known Plaintext Attack (1)

Known:

$C_1 = (c_{11}, c_{12}, \ldots, c_{1d})$ $M_1 = (m_{11}, m_{12}, \ldots, m_{1d})$
$C_2 = (c_{21}, c_{22}, \ldots, c_{2d})$ $M_2 = (m_{21}, m_{22}, \ldots, m_{2d})$
……………………………………………………
$C_d = (c_{d1}, c_{d2}, \ldots, c_{dd})$ $M_d = (m_{d1}, m_{d2}, \ldots, m_{dd})$

We know that:

$(c_{11}, c_{12}, \ldots, c_{1d}) = (m_{11}, m_{12}, \ldots, m_{1d}) \cdot K_{[dxd]}$

$(c_{21}, c_{22}, \ldots, c_{2d}) = (m_{21}, m_{22}, \ldots, m_{2d}) \cdot K_{[dxd]}$

……………………………………………………

$(c_{d1}, c_{d2}, \ldots, c_{dd}) = (m_{d1}, m_{d2}, \ldots, m_{dd}) \cdot K_{[dxd]}$

Hill Cipher - Known Plaintext Attack (2)

$$\begin{pmatrix} c_{11}, c_{12}, \dots, c_{1d} \\ c_{21}, c_{22}, \dots, c_{2d} \\ \dots\dots\dots\dots\dots\dots \\ c_{d1}, c_{d2}, \dots, c_{dd} \end{pmatrix} = \begin{pmatrix} m_{11}, m_{12}, \dots, m_{1d} \\ m_{21}, m_{22}, \dots, m_{2d} \\ \dots\dots\dots\dots\dots\dots\dots \\ m_{d1}, m_{d2}, \dots, m_{dd} \end{pmatrix} \begin{pmatrix} k_{11}, k_{12}, \dots, k_{1d} \\ k_{21}, k_{22}, \dots, k_{2d} \\ \\ k_{d1}, k_{d2}, \dots, k_{dd} \end{pmatrix}$$

$$C_{[dxd]} = M_{[dxd]} \cdot K_{[dxd]}$$

$$K_{[dxd]} = M^{-1}{}_{[dxd]} \cdot C_{[dxd]}$$

Substitution Ciphers (5)

4. Homophonic substitution cipher

$\mathcal{M} = \{ A, B, C, \dots, Z \}$
$\mathcal{C} = \{ 0, 1, 2, 3, \dots, 99 \}$

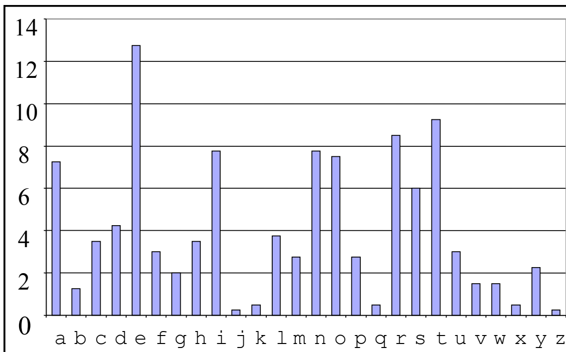
$c_i = f(m_i, \text{random number})$
$m_i = f^{-1}(c_i)$

f: E → 17, 19, 27, 48, 64
A → 8, 20, 25, 49
U → 45, 68, 91
……
X → 33

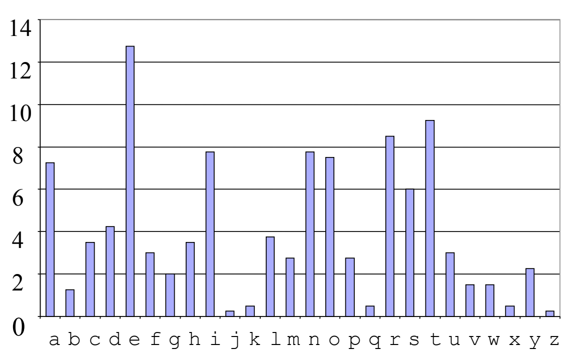
Transposition ciphers

$$M = m_1 \quad m_2 \quad m_3 \quad m_4 \quad .... \quad m_N$$
$$C = m_{f(1)} \quad m_{f(2)} \quad m_{f(3)} \quad m_{f(4)} \quad .... \quad m_{f(N)}$$

Letters of the plaintext are rearranged without changing them



Character frequency in a long English plaintext



Character frequency in the corresponding ciphertext for a transposition cipher

Transposition cipher
Example

Plaintext: CRYPTANALYST

Key: KRIS

Encryption:

2	3	1	4
K	R	I	S
C	R	Y	P
T	A	N	A
L	Y	S	T

Ciphertext: YNSCTLRAYPAT

One-time Pad
Vernam Cipher

Gilbert Vernam, AT&T
Major Joseph Mauborgne

1926

$$c_i = m_i \oplus k_i$$

m_i	0111011010100101011010 1
k_i	1101110111011010111011 0
c_i	1010101101111111100001 1

All bits of the key must be chosen at random and never reused

One-time Pad
Equivalent version

$c_i = m_i + k_i \bmod 26$

m_i	`TO BE OR NOT TO BE`
k_i	`AX TC VI URD WM OF`
c_i	`TL UG JZ HFW PK PJ`

All letters of the key must be chosen at random and never reused

Perfect Cipher

Claude Shannon
Communication Theory of Secrecy Systems, 1948

$$\forall_{\substack{m \in M \\ c \in C}} \quad P(M=m \mid C=c) = P(M = m)$$

The cryptanalyst can guess a message with the same probability without knowing a ciphertext as with the knowledge of the ciphertext

Is substitution cipher a perfect cipher?

C = *XRZ*

P(M=*ADD* | C=*XRZ*) = 0

P(M=*ADD*) $\neq$ 0

Is one-time pad a perfect cipher?

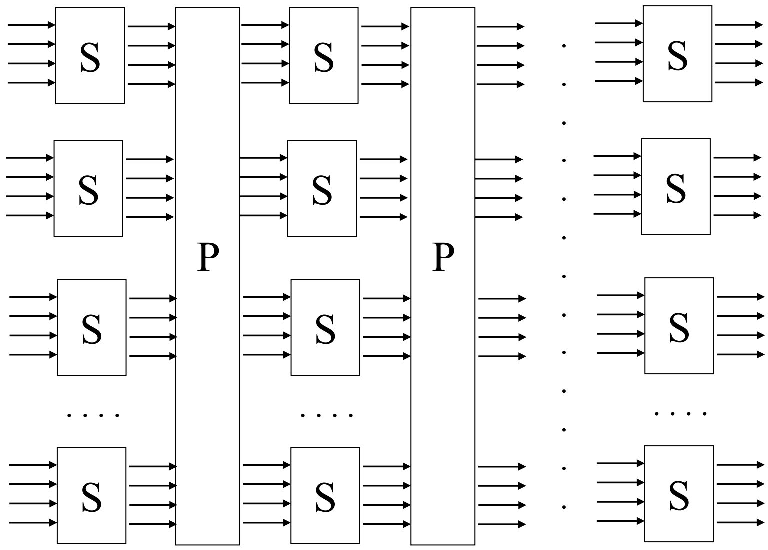
C = *XRZ*

P(M=*ADD* | C=*XRZ*) $\neq$ 0

P(M=*ADD*) $\neq$ 0

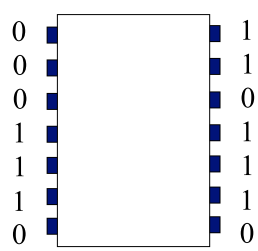
M might be equal to
CAT, PET, SET, ADD, BBC, AAA, HOT, HIS, HER, BET, WAS, NOW, etc.

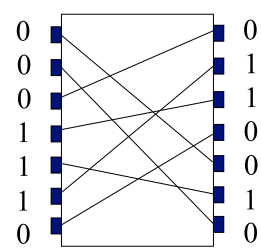
S-P Networks



Basic operations of S-P networks

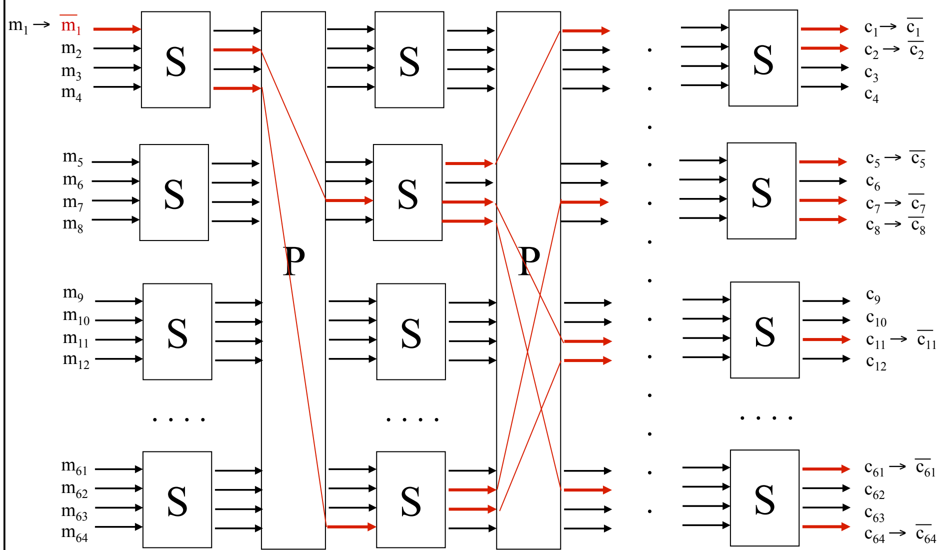
Substitution **Permutation**





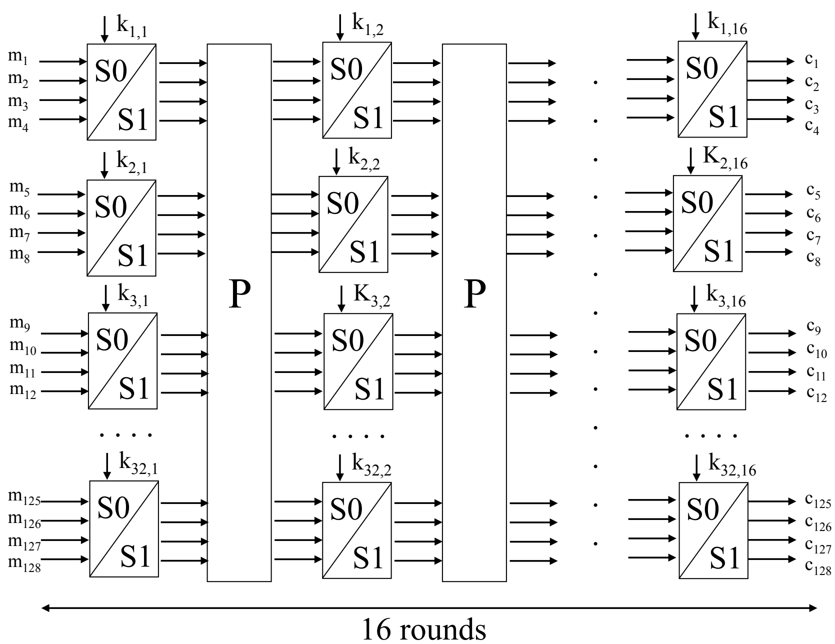
S-box **P-box**

Avalanche effect



LUCIFER

Horst Feistel, Walt Tuchman
IBM



LUCIFER- external look