Introduction to Information Security

Lecture 2: Classical Ciphers

2007. 6.

Prof. Byoungcheon Lee
sultan (at) joongbu . ac . kr

Information and Communications University

Contents

1. History of cryptographic research
2. Substitution ciphers
   - ✓ Caesar ciphers
   - ✓ Affine ciphers
   - ✓ Monoalphabetic substitution cipher
   - ✓ Homophonic substitution cipher
   - ✓ Polyalphabetic substitution cipher
   - ✓ Vigenere cipher
   - ✓ Hill cipher
   - ✓ One-time pad
3. Transposition ciphers
   - ✓ Transposition cipher
   - ✓ scytale cipher
4. Product ciphers

1. History of Cryptologic Research

History of Cryptologic Research

1900BC : Non-standard hieroglyphics

1500BC : Mesopotamian pottery glazes

50BC : Caesar cipher

1518 : Trithemius’ cipher book

1558 : Keys invented

1583 : Vigenere’s book

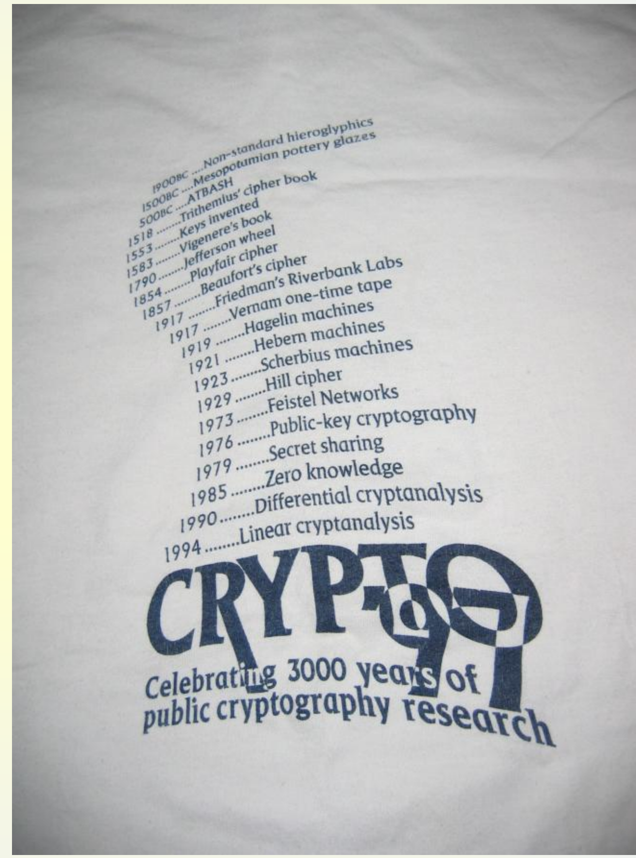
1790 : Jefferson wheel

1854 : Playfair cipher

1857 : Beaufort’s cipher

1917 : Friedman’s Riverbank Labs

1917 : Vernam one-time pads



History of Cryptologic Research

1919 : Hegelin machines

1921 : Hebern machines

1929 : Hill cipher

1973 : Feistel networks

1976 : Public key cryptography

1979 : Secret sharing

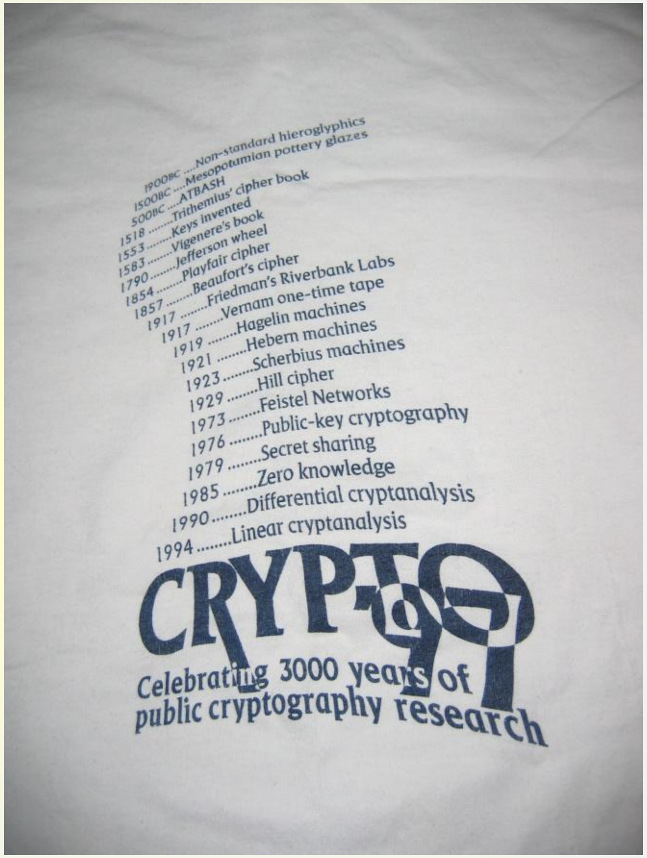
1985 : Zero knowledge

1990 : Differential cryptanalysis

1994 : Linear cryptanalysis

1997 : Triple-DES

1998 ~ 2001 : AES



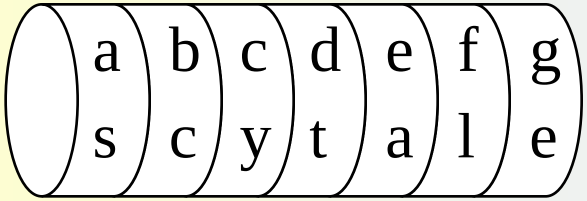
History of Cryptologic Research

	Period	Features	Examples
Manual Crypto	ancient ~ 1920	Substitution Transposition	Scytale, Caesar, Vigenere, Beaufort (USA)
Machine Crypto	1920 ~ 1950	Using complex machine	Enigma (Germany in 2nd WW) M-209 (USA in 2nd WW)
Modern Crypto Computer Crypto	1950 ~ current	Using computer Shannon's theory	DES, SEED, AES RSA, ECC, KCDSA

Using Cryptologic Technology

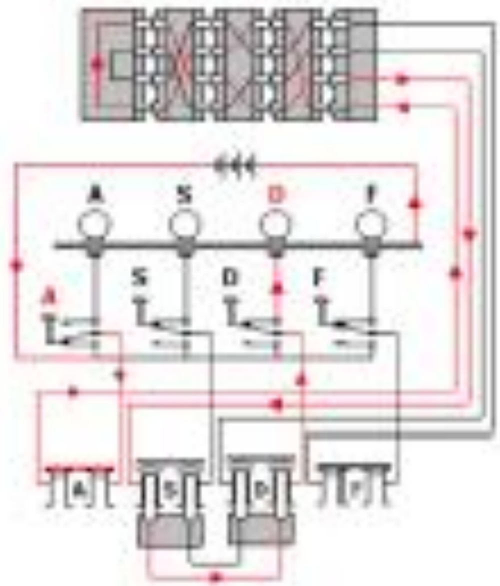
- Before modern crypto : limited usage
  - National security, diplomatic, war
  - Used by limited people
  - Researched by limited people
- Current crypto : widely open, standardized, commerce
  - Internet, e-commerce
  - Anybody is using
  - Research and development by anyone

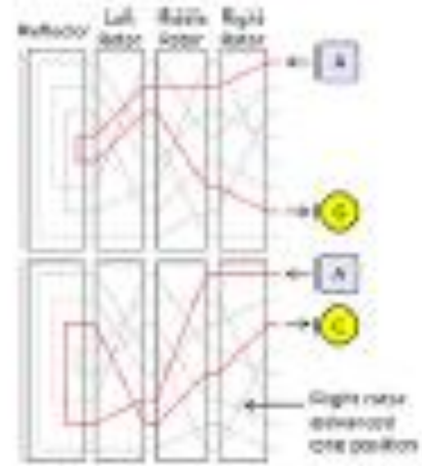
Scytale



as bc cy dt ea fl ge

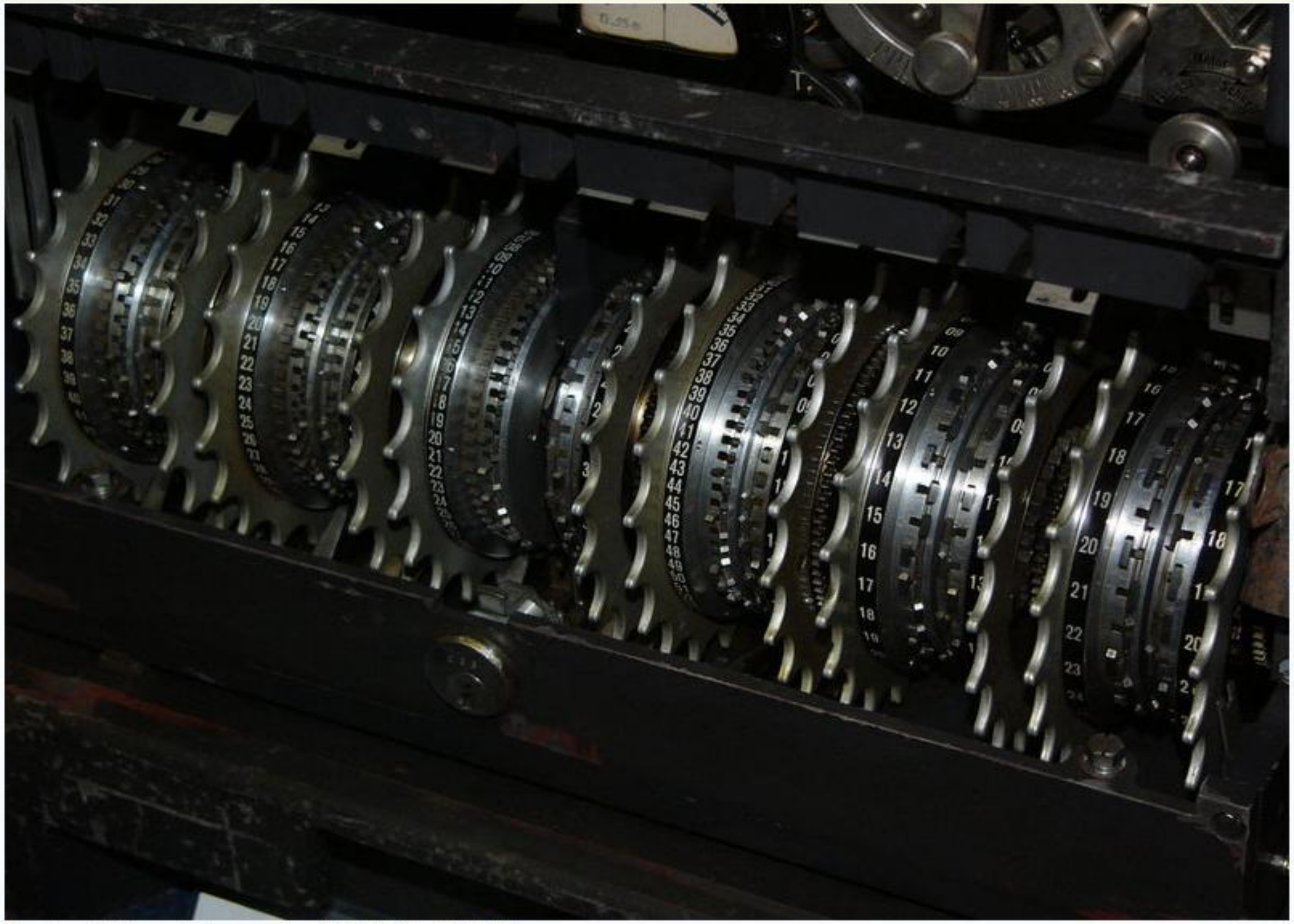
Enigma





Article at
http://en.wikipedia.org/wiki/Enigma_machine

Lorenz SZ42 Cipher Machine



Classical Encryption Techniques

- Basic building blocks of all encryption techniques
  - Substitution: replacement
  - Transposition: relocation
- Substitution ciphers
  - Caesar cipher
  - Monoalphabetic ciphers
  - Playfair cipher
  - Hill cipher
  - Polyalphabetic ciphers: Vigenere cipher
  - Vernam cipher/One-time pad: perfect cipher
- Transposition techniques
  - Rotor machines: Enigma, Purple

2. Substitution Cipher

Caesar ciphers
Affine ciphers
Hill cipher
Monoalphabetic substitution cipher
Homophonic substitution cipher
Polyalphabetic substitution cipher
Vigenere cipher
One-time pad

Caesar Ciphers

Julius Caesar, the Roman emperor
Also known as shift cipher

Mathematically assign numbers to each alphabet

a	b	c	d	e	f	g	h	i	j	k	...	z
0	1	2	3	4	5	6	7	8	9	10	…	25

Caesar cipher :

$C = E_K(M) = M + K \mod 26$
$K = 3$

$M = D_K(C) = C - K \mod 26$
$K = 3$

Caesar Ciphers

Define transformation as:

a	b	c	d	e	f	g	h	i	j	k	...	z
D	E	F	G	H	I	J	K	L	M	N	...	C

Encryption example

i	**n**	**f**	**o**	**r**	**m**	**a**	**t**	**i**	**o**	**n**
L	**Q**	**I**	**R**	**U**	**P**	**D**	**W**	**L**	**R**	**Q**

Weakness

- **Key space is too short – only 26 possible keys**
- **Brute force search**

Example: Break ciphertext “L ORYH LFX"

Affine Ciphers

Generalization of Caesar cipher

Encryption

$$C = E_K(M) = K_1 M + K_2 \bmod 26$$

$$\gcd(K_1, 26) = 1$$

Decryption

$$M = D_K(C) = (C - K_2) K_1^{-1} \bmod 26$$

Example: decrypt the following ciphertext

WZDUY ZZYQB OTHTX ZDNZD KWQHI BYQBP WZDUY ZXZDSS

How? Using English character frequency analysis…

English Character Frequencies

Letter	Frequency(%)	Letter	Frequency(%)	Letter	Frequency(%)
e	12.7	d	4.3	p	1.9
t	9.1	l	4.0	b	1.5
a	8.2	c	2.8	v	1.0
o	7.5	u	2.8	k	0.8
i	7.0	m	2.4	j	0.2
n	6.7	w	2.3	x	0.1
s	6.3	f	2.2	q	0.1
h	6.1	g	2.0	z	0.1
r	6.0	y	2.0		

(1) Pr(e)=0.12, (2) Pr(t,a,o,i,n,s,h,r) = 0.06 ~0.09
(3) Pr(d,l)=0.04 (4) Pr(c,u,m,w,f,g,y,p,b)= 0.015~0.023
(5) Pr(v,k,j,x,q,z) <=0.01

Affine Ciphers

Z occurs 8 times → E,T,A,O,I ???
D occurs 5 times → E,T,A,O,I ???
Y occurs 4 times → E,T,A,O,I ???
W,Q,B occur 3 times → E,T,A,O,I ???

Z → E, D → T :
try to solve

$$25 = 4K_1 + K_2 \bmod 26$$

$$3 = 19K_1 + K_2 \bmod 26$$

$$K_1 = 2, K_2 = 17$$ ← ***reject***

Try possible solutions until you get meaningful plaintext

Exercise: try yourself

Hill Cipher

$e_K(x) : (y_1,y_2,\ldots,y_m) = (x_1,x_2,\ldots,x_m)\ K$

where K is m x m matrix and gcd(det K, 26) =1

$d_K(y) = y\ K^{-1}$

(Ex) $K = \begin{vmatrix} 11 & 8 \\ 3 & 7 \end{vmatrix}$ $K^{-1} = \begin{vmatrix} 7 & 18 \\ 23 & 11 \end{vmatrix}$

x : july, (j,u)= (9,20), (l,y) = (11,24)

(9,20) K = (3,4) = (D,E),

(11,24) K = (11,22) = (L,W)

Monoalphabetic Substitution Ciphers

Example : 1-1 Substitution rule

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

Example : Encryption

i	n	f	o	r	m	a	t	i	o	n
P	R	N	X	Y	C	E	D	P	X	R

Key space : 26!

Cryptanalysis: Using English character frequency analysis…

Homophonic Substitution Ciphers

Letters which occur frequently may be mapped into more than one letter in the ciphertext to flatten the frequency distribution.

Alphabet is mapped into the numbers 0 to 99
For example,
E(12.7%) → 17, 19, 23, 47, 64
A(8.2%) → 8, 20, 25, 49
R(6.0%) → 1, 29, 65
T(9.1%) → 16, 31, 85, 87

Polyalphabetic Substitution Ciphers

Hide the frequency distribution by making multiple substitutions.
Apply *d* different permutations.

$$m = m_1, m_2, \ldots, m_d, m_{d+1}, m_{d+2}, \ldots, m_{2d}, \ldots$$

$$E_K(m) = \pi_1(m_1), \pi_2(m_2), \ldots, \pi_d(m_d), \pi_1(m_{d+1}), \pi_2(m_{d+2}), \ldots, \pi_d(m_{2d}), \ldots$$

- **Vigenere cipher**
- **Beauford cipher**

Polyalphabetic Substitution Ciphers

Vigenère Ciphers

- **Multiple caesar cipher**

$$k = (k_1, k_2, \ldots, k_d), |k| = 26^d$$

$$c = E_k(m_1, m_2, \ldots, m_d) = (c_1, c_2, \ldots, c_d) = m_i + k_i \bmod 26 \ \text{ for } i = 1, \ldots, d$$

$$m = D_k(c_1, c_2, \ldots, c_d) = (m_1, m_2, \ldots, m_d) = c_i - k_i \bmod 26 \ \text{ for } i = 1, \ldots, d$$

Beauford ciphers (used in US civil war)

$$k = (k_1, k_2, \ldots, k_d), |k| = 26^d$$

$$c = E_k(m_1, m_2, \ldots, m_d) = (c_1, c_2, \ldots, c_d) = k_i - m_i \bmod 26 \ \text{ for } i = 1, \ldots, d$$

$$m = D_k(c_1, c_2, \ldots, c_d) = (m_1, m_2, \ldots, m_d) = k_i - c_i \bmod 26 \ \text{ for } i = 1, \ldots, d$$

Vigenère Ciphers

Look-up table for Vigenère Ciphers

키워드 \ 평문	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
A	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	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	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	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	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	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	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	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	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	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	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	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	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
N	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	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	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	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	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
S	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	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	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	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	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	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	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	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 Ciphers

Plaintext	t h i s c r y p t o s y s t e m i s n o t s e c u r e
Keyword	S E C U R I T Y S E C U R I T Y S E C U R I T Y S E C
Ciphertext	L L K M T Z R N L S U S J B X K A W P I K A X A M V G

Polyalphabetic Substitution Ciphers

Cryptanalysis of polyalphabetic substitution ciphers

1. **Determine the period**
2. **Determine each substitution keys**

How to determine the period?

1. **Kasiski method : use repetitions in the ciphertext**
2. **Index of coincidence by Friedman: compute the index of coincidence and estimate the period**

Refer to
http://www.rhodes.edu/mathcs/faculty/barr/Math103CUSummer04/FriedmanKasiski.pdf

Kasiski Method

Example: Vigenère Ciphers

```
key:        deceptivedeceptivedeceptive
plaintext:  wearediscoveredsaveyourself
ciphertext: ZICVTWQNGRZGVTWAVZHCQYGLMGJ
```

Method developed by Kasiski

- **Letter groups in ciphertext are repeated because repeated letter groups in the plaintext line up with the keyword.**
- **If letter groups repeated in ciphertext, then keyword length may be a divisor of their separations.**

- in this example “VTW” is repeated in 9 letters apart
- suggests size of d is 3 or 9

Index of Coincidence

The **index of coincidence** for a (cipher)text is the probability that two letters selected from it are identical. It is denoted *I.*

If the text has n_0 A's, n_1 B's, n_2 C's, ..., n_{25} Z's, and

$$n = n_0 + n_1 + n_2 + \cdots + n_{25}$$

is the number of letters in the text, then

$$I = \frac{n_0(n_0 - 1) + n_1(n_1 - 1) + \cdots + n_{25}(n_{25} - 1)}{n(n-1)}.$$

Index of Coincidence

For a typical English document, I=0.0656

letter	count	letter	count
A	141	N	119
B	36	O	132
C	36	P	28
D	103	Q	1
E	188	R	95
F	37	S	64
G	34	T	182
H	102	U	59
I	123	V	13
J	4	W	55
K	18	X	3
L	56	Y	2
M	27	Z	0

$$
\begin{aligned}
&\text{Prob}\binom{\text{identical}}{\text{pair}} \\
&= \frac{C(141,2)+C(36,2)+\cdots+C(23,2)+C(0,2)}{C(1679,2)} \\
&= \frac{141(141-1)+36(36-1)+\cdots+0(0-1)}{1679(1679-1)} \\
&= \frac{184838}{2817362} \approx 0.0656
\end{aligned}
$$

Index of Coincidence

For a randomized (ideally encrypted) document, I=0.0384615

Example What is the index of coincidence for a collection of 2600 letters consisting of 100 A's, 100 B's, 100 C's, ..., 100 Z's?

Answer:

$$\text{Prob}\binom{\text{identical}}{\text{pair}}$$

$$= \frac{C(100,2) + C(100,2) + \cdots + C(100,2) + C(100,2)}{C(2600,2)}$$

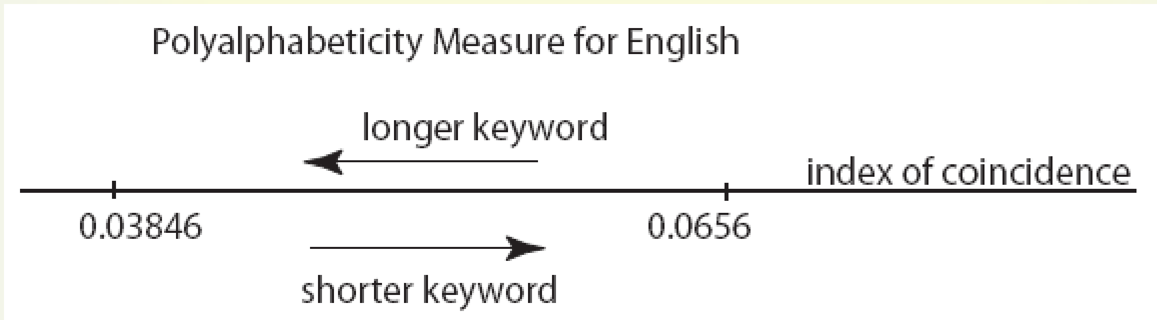
$$= \frac{26 \cdot 100 \cdot 99}{2600 \cdot 2599}$$

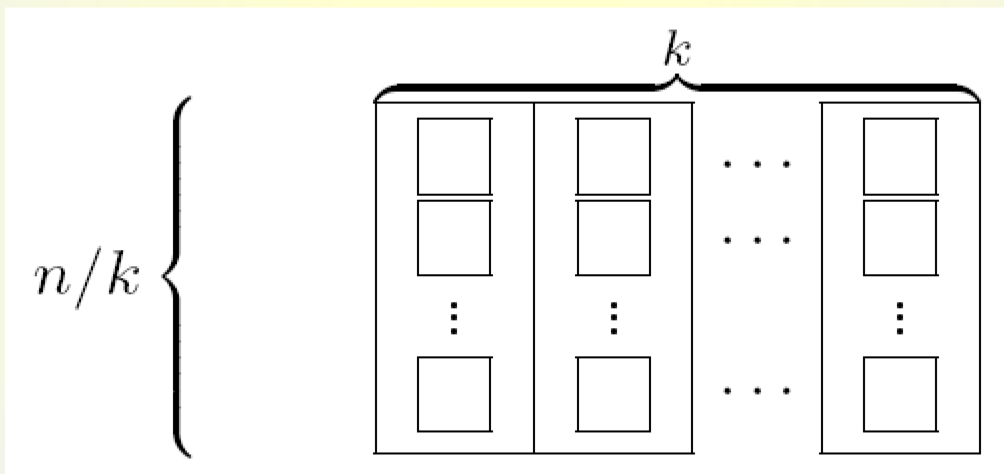
$$\approx \frac{1}{26}$$

$$= 0.0384615$$

Index of Coincidence

We can estimate the keyword length using the index of coincidence.





k : Estimated keyword length

Index of Coincidence

We can estimate the keyword length using the index of coincidence.

$$I \approx \frac{S+D}{C(n,2)}$$

$$= \frac{k \cdot \frac{\frac{n}{k}(\frac{n}{k}-1)}{2\cdot 1} \cdot 0.065 + \frac{k(k-1)}{2\cdot 1}(\frac{n}{k})^2 \cdot 0.03846}{\frac{n(n-1)}{2\cdot 1}}$$

$$= \frac{(n-k)0.065 + n(k-1)0.03846}{k(n-1)}$$

Solve for k:

$$\boxed{k \approx \frac{0.0265n}{(0.065 - I) + n(I - 0.0385)}}$$

Index of Coincidence

Example: Estimate the keyword length of the following distribution in ciphertext

letter	count	letter	count
A	60	N	28
B	50	O	83
C	42	P	44
D	64	Q	69
E	51	R	13
F	63	S	29
G	19	T	66
H	48	U	87
I	56	V	63
J	67	W	19
K	23	X	43
L	45	Y	39
M	44	Z	67

Solution There are $n = 1282$ letters.

$$\begin{aligned} I &= \frac{60 \cdot 59 + 50 \cdot 49 + \cdots + 67 \cdot 66}{1282 \cdot 1281} \\ &= \frac{35761}{821121} \\ &= 0.04355 \end{aligned}$$

$$\begin{aligned} k &\approx \frac{0.0265 \cdot 1282}{(0.065 - 0.04355) + 1282(0.04355 - 0.03846)} \\ &= 5.197 \end{aligned}$$

Estimated keyword length is 5

One-time Pad (Vernam cipher)

- **Use a random key as long as the message size and use the key only once**
- **Unbreakable**
  - **Since ciphertext bears no statistical relationship to the plaintext**
  - **Since for any plaintext & any ciphertext there exists a key mapping one to other**
- **Have the problem of safe distribution of key**

Ex) Binary alphabet

P :	o	n	e	t	i
P’:	01101111	01101110	01100101	01110100	01101001
K :	01011100	01010001	11100000	01101001	01111010
C :	00110011	00111111	10000101	00011101	00010011

Perfect Cipher : $p(x|y) = p(x)$ for all $x \in P$, $y \in C$

Impossible COA

3. Transposition Ciphers

Transposition cipher
Scytale cipher
Rotor machines

Transposition Ciphers

- Rearrange characters of plaintext to produce ciphertext
- Frequency distribution of the characters is not changed by encryption

- Example:

Encryption permutation

1	2	3	4	5	6
3	5	1	6	4	2

Decryption permutation

1	2	3	4	5	6
3	6	1	5	2	4

plaintext	i n f o r m a t i o n s e c u r i t y x y z a b
ciphertext	F R I M O N I N A S O T U I E T R C Y A Y B Z X

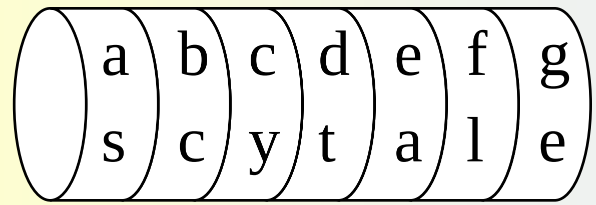
Transposition Ciphers

- **Cryptanalysis :**
  - **Period d is guessed by trying possible periods**
  - **A knowledge of the most frequent pairs and triples in a language is used with anagramming.**
  - **Use language characteristics**
- **Frequent pairs on a relative scale to 10**
  - **TH : 10.00, HE : 9.50, IN : 7.17, ER : 6.65, RE : 5.92**
- **Frequent triples on a relative scale to 10**
  - **THE : 10.00, AND : 2.81, TIO : 2.24, ATI : 1.67**

Exercise: decrypt the following ciphertext

LDWEOHETTHSESTRUHTELOBSEDEFEIVNT

Scytale Cipher



as bc cy dt ea fl ge

4. Product Ciphers

ADFGVX
Shannon
SP Network

ADFGVX

- **Product of substitution and permutation**

Substitution table

	A	D	F	G	V	X
A	f	x	a	9	u	1
D	n	g	0	l	d	o
F	5	b	k	2	h	z
G	m	j	s	y	t	v
V	7	4	3	e	8	i
X	c	w	q	6	r	p

c → XA

Substitution result

c	o	n	v	e	n	t	i	o	n	a	l
X	D	D	G	V	D	G	V	D	D	A	D
A	X	A	X	G	A	V	X	X	A	F	G

c	r	y	p	t	o	g	r	a	p	h	y
X	X	G	X	G	D	D	X	A	X	F	G
A	V	G	X	V	X	D	V	F	X	V	G

ADFGVX

Permutation table

C	I	P	H	E	R
1	**4**	**5**	**3**	**2**	**6**
X	A	D	X	D	A
G	X	V	G	D	A
G	V	V	X	D	X
D	A	A	F	D	G
X	A	X	V	G	G
X	X	G	V	D	X
D	D	X	V	A	F
X	X	F	V	G	G

Keyword (row C I P H E R)
permutation (row 1 4 5 3 2 6)

Ciphertext

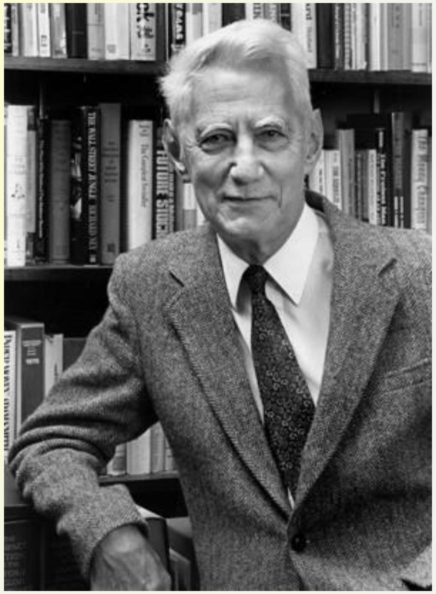
XGGDXXDX
DDDDGDAG
XGXFVVVV
AXVAAXDX
DVVAXGXF
AAXGFXFG

Shannon's Proposal

- **C. Shannon, "*Communication Theory for Secrecy Systems",* 1949**
  - **Compose different kind of simple and insecure ciphers to create complex and secure cryptosystems → called "product cipher"**
  - **Incorporate confusion and diffusion**
  - **Substitution-Permutation Network**

http://www.bell-labs.com/news/2001/february/26/1.html

http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html



Claude Shannon

Confusion and Diffusion

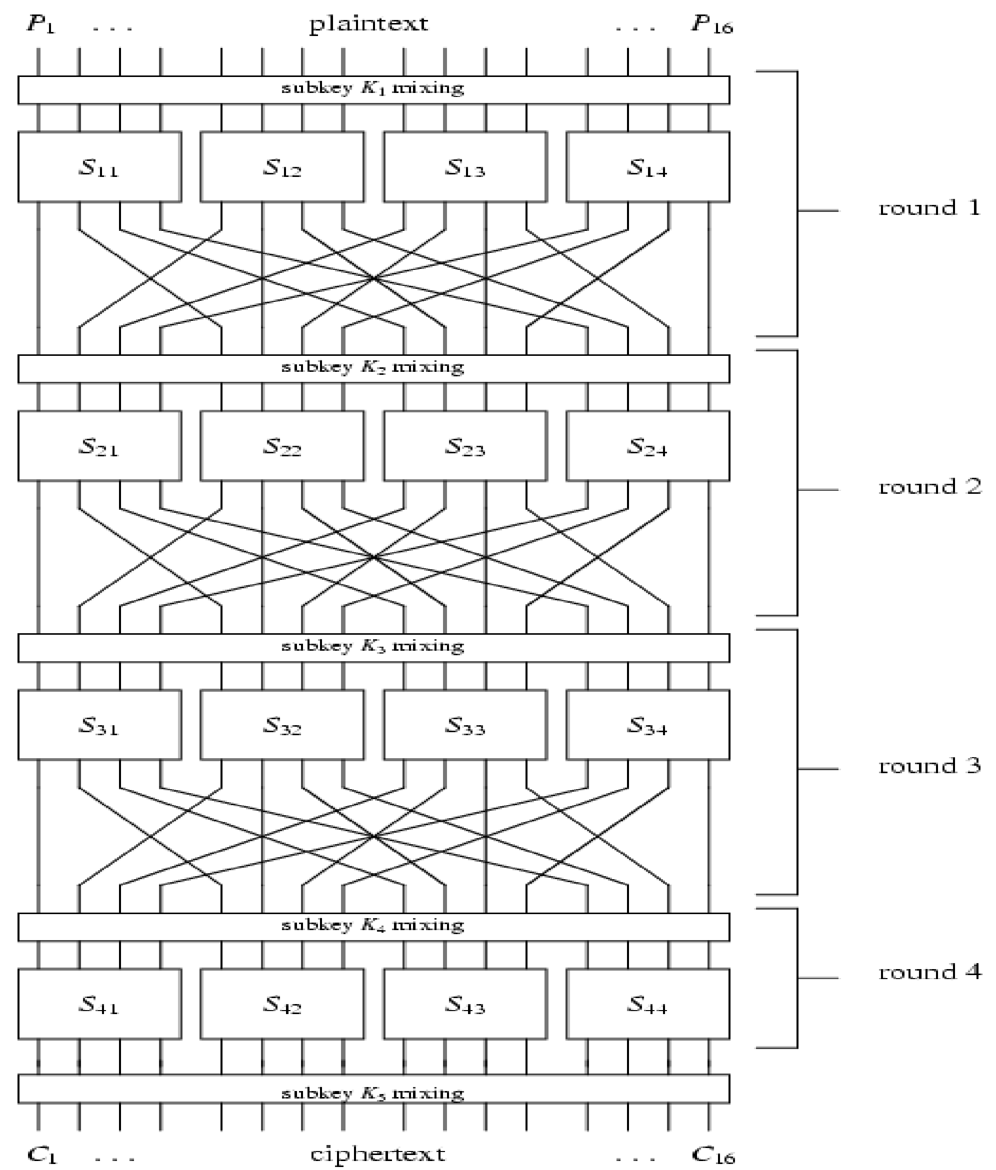
- **Confusion (substitution) :**
  - **The ciphertext statistics should depend on the plaintext statistics in a manner too complicated to be exploited by the enemy cryptanalyst**
  - **Makes relationship between ciphertext and key as complex as possible**

- **Diffusion (permutation) :**
  - **Each digit of the plaintext should influence many digits of the ciphertext, and/or**
  - **Each digit of the secret key should influence many digits of the the ciphertext.**
  - **Dissipates statistical structure of plaintext over bulk of ciphertext**

SP Network

- Substitution-Permutation network
  - Substitution (S-box) : secret key is used
  - Permutation (P-box) : no secret key, fixed topology
- Provide confusion and diffusion
- S-P networks are expected to have
  - Avalanche property: a single input bit change should force the complementation of approximately half of the output bits
  - Completeness property: each output bit should be a complex function of every input bits
- Theoretical basis of modern block ciphers

SP Network



P_1 . . . plaintext . . . P_{16}
subkey K_1 mixing
S_{11}
S_{12}
S_{13}
S_{14}
round 1
subkey K_2 mixing
S_{21}
S_{22}
S_{23}
S_{24}
round 2
subkey K_3 mixing
S_{31}
S_{32}
S_{33}
S_{34}
round 3
subkey K_4 mixing
S_{41}
S_{42}
S_{43}
S_{44}
round 4
subkey K_5 mixing
C_1 . . . ciphertext . . . C_{16}

Kerckhoff's Principle

- **Auguste Kerckhoff, 1883**
  - **A cryptosystem should be secure even if everything about the system, except the key, is public knowledge.**
  - **Eric Raymond extends this principle in support of open source software, saying "Any security software design that doesn't assume the enemy possesses the source code is already untrustworthy; therefore, never trust closed source".**
  - **The majority of civilian cryptography makes use of publicly-known algorithms. By contrast, ciphers used to protect classified government or military information are often kept secret**

Homework #2

1. **Design and implement a C program for encryption, decryption, and cryptanalysis of the affine cipher. For the cryptanalysis your program must not use the enumeration of all possible keys but should use the frequency of characters to make optimal guesses about the key.**

2. **Decryption of Vigenère Ciphers. Solve the problem 9 in page 61 of the textbook.**