COMP 3355
Classical Cryptography

K.P. Chow

University of Hong Kong

Security Services = Basic Computer Security Concepts (CIA)

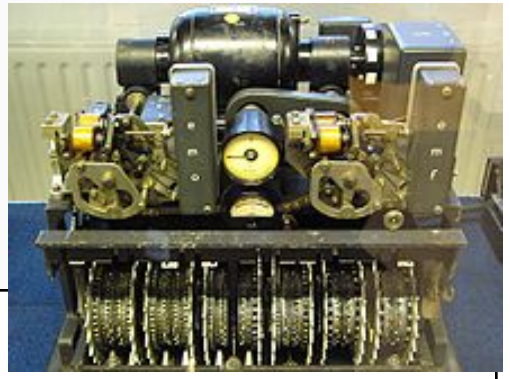
- Confidentiality: prevention of unauthorized disclosure of information
- Integrity: prevention of unauthorized modification of information

Cryptography

- Availability: prevention of unauthorized withholding of information or resources
- Authenticity: able to verify the origin of data
- Accountability: audit information must be selectively kept and protected so that actions affecting security can be traced to the responsible party

Access Control

Types of Encryptions

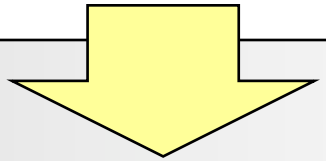


- Symmetric key encryption
  - Encryption and decryption using the same key
  - E.g. AES, 3DES, RC4
- Public key encryption
  - Encryption and decryption using different key
  - E.g. RSA

Let's look at encryption!

- Process of transforming information to make it unreadable to anyone except those possessing the key
- "Encrypted data"

  `fbehu vsdfh fulph lvhdv b`



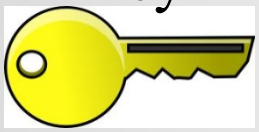
Shift 3 positions

`cyber space crime iseas y`

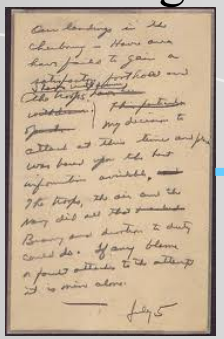
The encryption algorithm is called **Caesar Cipher** and the **key is 3**

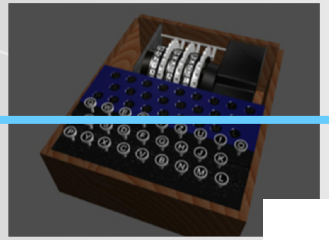
Encryption/ Decryption

Key



Original Message





Encryption Algorithm

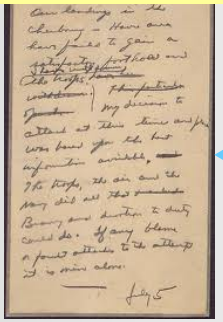
Encrypted Message

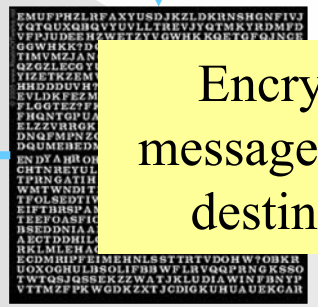
Symmetric Key Cryptographic System

Encrypted message sent over Internet

Decryption Algorithm

Original Message



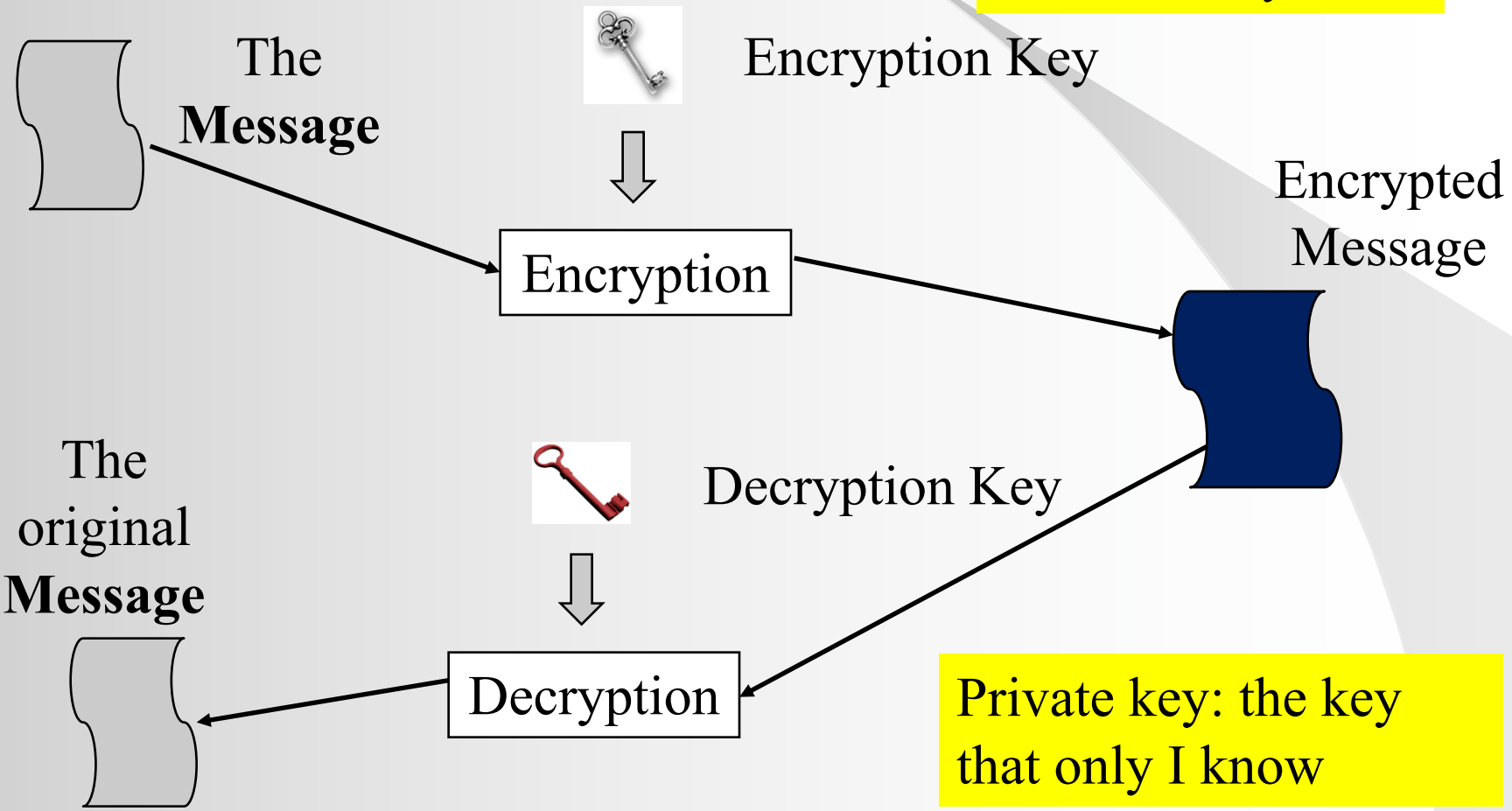


Encrypted message arrives destination

How many keys in symmetric key cryptographic system?

Public Key Encryption: Everyone has 2 keys

Public key: public available key

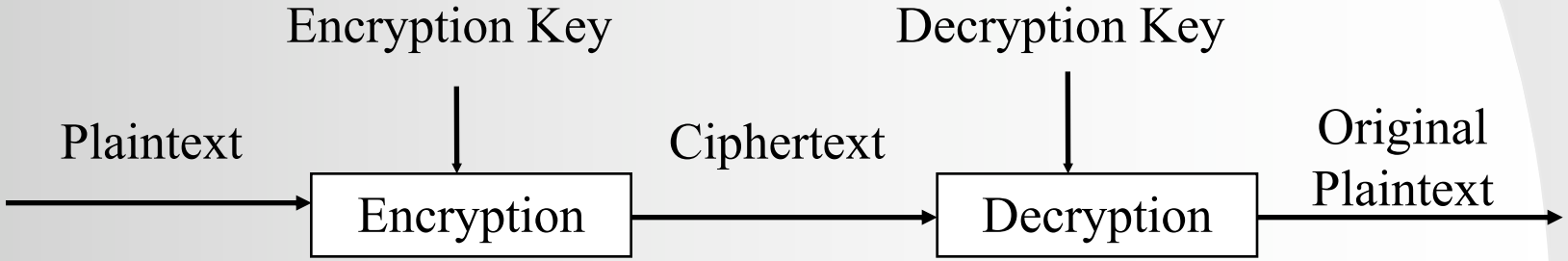


Private key: the key that only I know

Encryption Algorithms

Single Key (symmetric key) Cryptosystem

Two Key (public key) Cryptosystem



Terminology

- Encryption is a process of encoding a message so that the meaning of the message is not obvious
- Decryption is the processing of transforming an encrypted message back into its normal form
- Plaintext: the original form of a message
- Ciphertext: the encrypted form of a message
- Cryptography means hidden meaning, the practice of using encryption to conceal text
- Cryptanalysis studies encryption and encrypted messages, with the goal of finding the hidden meanings of the messages
- Cryptology is the research into and study of encryption and decryption; it includes both cryptography and cryptanalysis

Symmetric key cryptography can protect your data

Symmetric key cryptography is based on simple operations

Representation of Characters

- To simplify the study, we will assume encryption of messages written in standard English alphabet: A … Z
- We shall write plaintext in uppercase letters and ciphertext in lowercase letters
- Most encryption algorithms are mathematical in nature, we shall use the following numeric encoding of each letter:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

R	S	T	U	V	W	X	Y	Z
17	18	19	20	21	22	23	24	25

- Arithmetic on letters: addition and subtraction on letters will be performed on the corresponding code number, e.g. A + 3 = D
- Modular arithmetic: A + B mod n, if the result of A + B > n, it will be reduced by n as many times as necessary to bring it back into the range 0 <= result < n, e.g. Y + 3 = B mod 26

Caesar Cipher (1)

```
wklv phvvdjh lv qrw wrr kdug wr euhdn
```

- The message is enciphered with a 27-symbol alphabet (A-Z) and the blank, the blank is translated to itself
- Attack:
  - 2-letter words: am, is, to be, he, we, …
  - 3-letter words: and, are, you, she, …
  - 3-letter words with repeated character (wrr): see, too, add, odd, off
  - the first 2 characters also form a word: se, to, ad, od, of, try ‘to’ and ‘be’ first

```
wklv phvvdjh lv qrw wrr kdug wr euhdn
T--- ------- -- -OT TOO ---- TO -----   (1)
B--- ------- -- -EB BEE ---- BE -----   (2)
```

Caesar Cipher (2)

- Caesar cipher: each letter is translated to the letter a fixed number of letters after it in the alphabet
- The "real" Caesar cipher by Julius Caesar used a shift of 3:
  - $c = E(p) = p + 3$

```
A B C D E F G H I J K L N . . . W X Y Z
d e f g h i j k l m n o p . . . z a b c
wklv phvvdjh lv qrw wrr kdug wr euhdn
T-IS ------- IS -OT TOO ---- TO -----
3 33         33  33 333      33
```

  - The message is using the "real" Caesar cipher

```
THIS MESSAGE IS NOT TOO HARD TO BREAK
```

Monoalphabetic Substitutions

- Monoaphabetic substitution: the alphabet is scrambled, and each plaintext letter maps to a unique ciphertext letter
- A permutation is a reordering of the elements of a series, e.g.
  - $\pi 1$ = 1, 3, 5, 7, 9, 10, 8, 6, 4, 2
  - $\pi 2$ = 10, 9, 8, 7, 6, 5, 4, 3, 2, 1
- A permutation can be a function, e.g.
  - $\pi 3(i) = 25 - i$, e.g. $\pi 3(A) = z$
  - $\pi 4(i) = (3*i) \bmod 26$, e.g. $\pi 4(A) = a$

  `ABCDEFGHIJKLMNOPQRSTUVWXYZ`

  `adgjmpsvybehknqtwzcfilorux`
- Some permutations cannot be represented as simple equation,

  `ABCDEFGHIJKLMNOPQRSTUVWXYZ`

  `key``abcdfghijlmnopqrstuvwxz`

Cryptanalysis of Monoalphabetic Ciphers (1)

- In English, some letters are used more frequently than others, e.g. letters E, T and A occur far more frequently than J, Q, Z
- Consider the following ciphertext

```
hqfubswlrq lv d phdqv ri dwwdlqlqj vhfxuh frpsxwdwlrq
ryhu lqvhfxuh fkdqqhov
eb xvlqj hqfubswlrq zh glvjxlvh wkh phvvdjh vr wkdw
hyhq li wkh wudqvplvvlrq lv glyhuwhg
wkh phvvdjh zloo qrw eh uhyhdohg
```

Cryptanalysis of Monoalphabetic Ciphers (2)

Letter	Message	
a	0	
b	3	
c	0	
d	11	
e	2	
f	6	
g	4	
h	26	
i	2	
j	5	
k	5	
l	16	
m	0	

Letter	Message	
n	0	
o	4	
p	5	
q	16	
r	9	
s	3	
t	0	
u	8	
v	17	
w	14	
x	5	
y	4	
z	2	
Total	167	

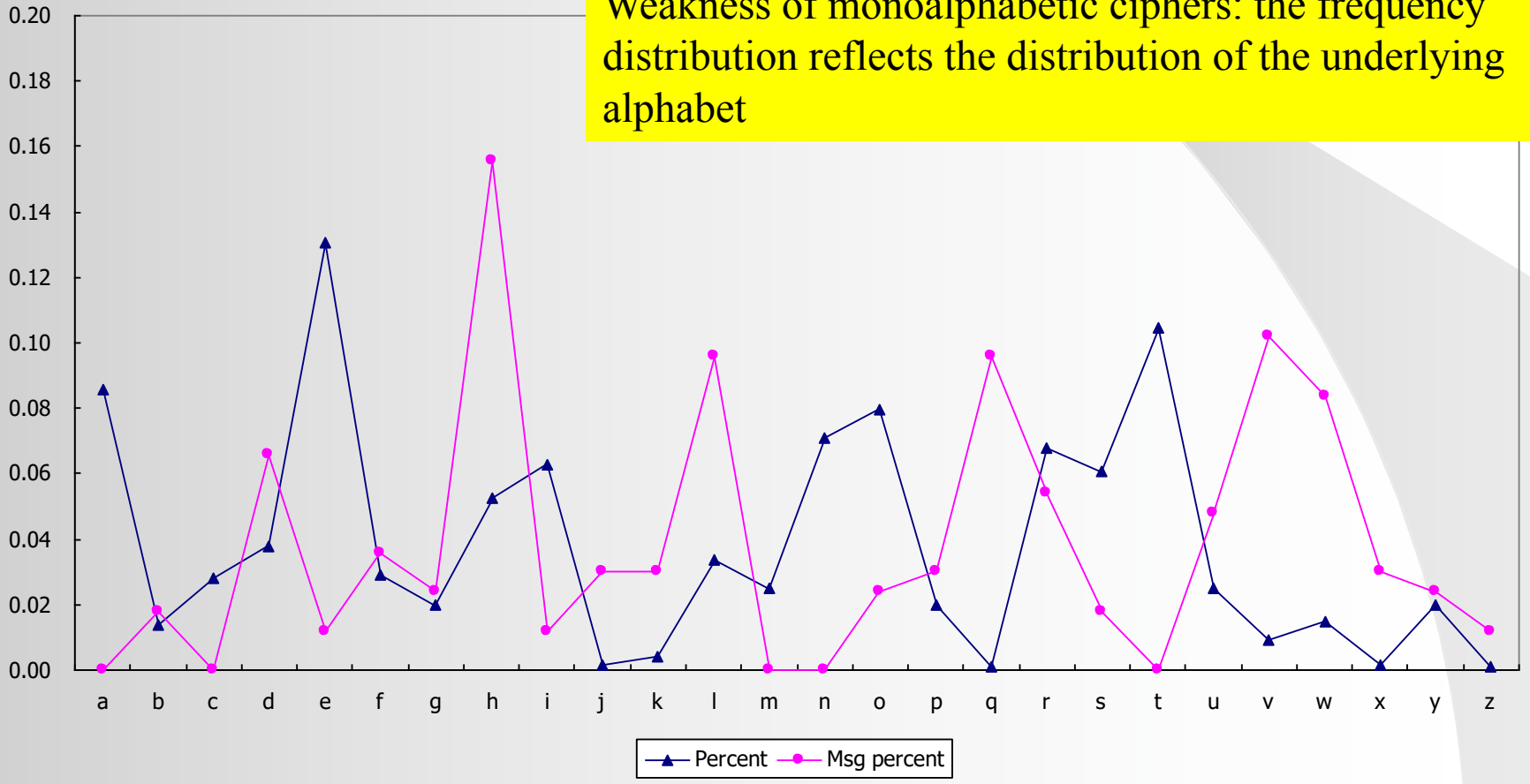
Cryptanalysis of Monoalphabetic Ciphers (3)

Letter	%	Msg	Msg %
a	8.56	0	0.00
b	1.39	3	1.8
c	2.79	0	0
d	3.78	11	6.59
e	13.04	2	1.2
f	2.89	6	3.59
g	1.99	4	2.4
h	5.28	26	15.57
i	6.27	2	1.2
j	0.13	5	2.99
k	0.42	5	2.99
l	3.39	16	9.58
m	2.49	0	0

Letter	%	Msg	Msg %
n	7.07	0	0.00
o	7.97	4	2.40
p	1.99	5	2.99
q	0.12	16	9.58
r	6.77	9	5.39
s	6.07	3	1.80
t	10.45	0	0.00
u	2.49	8	4.79
v	0.92	17	10.18
w	1.49	14	8.38
x	0.17	5	2.99
y	1.99	4	2.40
z	0.08	2	1.20
Total	100%	167	100%

Cryptanalysis of Monoalphabetic Ciphers (4)

- Referring to the frequency distribution, any idea?



Weakness of monoalphabetic ciphers: the frequency distribution reflects the distribution of the underlying alphabet

Polyalphabetic Substitution Ciphers

- Weakness of monoalphabetic ciphers: the frequency distribution reflects the distribution of the underlying alphabet
- A cipher that is more cryptographically secure would display a rather flat distribution, which gives no information to a cryptanalyst

- How to flatten the distribution?
- Combine distributions that are high with ones that are low
  - E.g. T is sometimes enciphered as a and sometimes as b, X is also sometimes enciphered as a and sometimes as b, the high frequency of T mixes with the low frequency of X will produce a more moderate distribution for a and b

Polyalphabetic Substitution Ciphers

- Consider the following 2 tables:
  - Table for **odd** positions (3*a mod 26):

    ABCDEFGHIJKLMNOPQRSTUVWXYZ

    adgjmpsvybehknqtwzcfilorux
  - Table for **even** positions ((5*a)+13 mod 26):

    ABCDEFGHIJKLMNOPQRSTUVWXYZ

    nsxchmrwbglqvafkpuzejotydi
  - E.g.

    TREAT YIMPO SSIBL E

    fumnf dyvtf czysh h
  - S becomes c,z, E becomes m, h
  - Both T's are enciphered as f in the message
- In the above 2 tables, A are enciphered as either 'a' or 'n', both 'a' and 'n' are high frequency letters

Vigenere Cipher

- Extend the number of permutations to 26, i.e. a plaintext letter can be enciphered as any ciphertext letter. Need to maintain 26 mappings! HOW?
- A Vigenere tableau is a collection of 26 permutations: written as a 26x26 matrix, with all 26 letters in each row and each column
- Which permutation to be used to encipher the current character?
  - Use a key (keyword): letters in the key will be used to select a particular permutation (e.g. juliet)
  - E.g. BUT SOFT, WHAT LIGHT THROUGH YONDER WINDOW BREAKS

```
julie tjuli etjul ietju lietj uliet julie tjuli e
BUTSO FTWHA TLIGH TTHRO UGHYO NDERW INDOW BREAK S

koeas ycqsi . . .
```

TABLE 2.5 VIGENÈRE TABLEAU

	0 1 2 0123456789012345678901234 5	
	abcdefghijklmnopqrstuvwxyz	
A	abcdefghijklmnopqrstuvwxyz	0
B	bcdefghijklmnopqrstuvwxyza	1
C	cdefghijklmnopqrstuvwxyzab	2
D	defghijklmnopqrstuvwxyzabc	3
E	efghijklmnopqrstuvwxyzabcd	4
F	fghijklmnopqrstuvwxyzabcde	5
G	ghijklmnopqrstuvwxyzabcdef	6
H	hijklmnopqrstuvwxyzabcdefg	7
I	ijklmnopqrstuvwxyzabcdefgh	8
J	jklmnopqrstuvwxyzabcdefghi	9
K	klmnopqrstuvwxyzabcdefghij	10
L	lmnopqrstuvwxyzabcdefghijk	11
M	mnopqrstuvwxyzabcdefghijkl	12
N	nopqrstuvwxyzabcdefghijklm	13
O	opqrstuvwxyzabcdefghijklmo	14
P	pqrstuvwxyzabcdefghijklmop	15
Q	qrstuvwxyzabcdefghijklmnop	16
R	rstuvwxyzabcdefghijklmnopq	17
S	stuvwxyzabcdefghijklmnopqr	18
T	tuvwxyzabcdefghijklmnopqrs	19
U	uvwxyzabcdefghijklmnopqrst	20
V	vwxyzabcdefghijklmnopqrstu	21
W	wxyzabcdefghijklmnopqrstuv	22
X	xyzabcdefghijklmnopqrstuvw	23
Y	yzabcdefghijklmnopqrstuvwx	24
Z	zabcdefghijklmnopqrstuvwxy	25

julie

BUTSO

kcoeas

	abcdefghijklmnopqrstuvwxyz	
A	abcdefghijklmnopqrstuvwxyz	0
B	bcdefghijklmnopqrstuvwxyza	1
C	cdefghijklmnopqrstuvwxyzab	2
D	defghijklmnopqrstuvwxyzabc	3
E	efghijklmnopqrstuvwxyzabcd	4
F	fghijklmnopqrstuvwxyzabcde	5
G	ghijklmnopqrstuvwxyzabcdef	6
H	hijklmnopqrstuvwxyzabcdefg	7
I	ijklmnopqrstuvwxyzabcdefgh	8
J	jklmnopqrstuvwxyzabcdefghi	9
K	klmnopqrstuvwxyzabcdefghij	10
L	lmnopqrstuvwxyzabcdefghijk	11
M	mnopqrstuvwxyzabcdefghijkl	12
N	nopqrstuvwxyzabcdefghijklm	13
O	opqrstuvwxyzabcdefghijklmo	14
P	pqrstuvwxyzabcdefghijklmop	15
Q	qrstuvwxyzabcdefghijklmnop	16
R	rstuvwxyzabcdefghijklmnopq	17
S	stuvwxyzabcdefghijklmnopqr	18
T	tuvwxyzabcdefghijklmnopqrs	19

Cryptanalysis

- Given a message
  - Is it encrypted?
  - How is it encrypted? Mono- or poly-alphabetic cipher?
  - What is the key?
- Index of coincidence
  - Measure how often character could theoretically appear next to each other, based on the frequency analysis of the text

Index of Coincidence (1)

- A tool to rate how well a particular distribution matches the distribution of letters in English
- Suppose we have a piece of text and is suspected to be encrypted using monalphabetic substitution, the frequencies of ciphertext letters should be the same as the frequencies of the corresponding English letters
- **The index of coincidence is a measure of the variation between frequencies in a distribution**
- Suppose we pick a letter at random from an English text, from Table A
  - The probability that the letter is a is 0.0856 (P(a))
  - The probability that the letter is b is 0.0139 (P(b))
  - …
- Suppose the text is encrypted and the encryption is so good that every letter will appear with equal probability, i.e. 1/26 = 0.0384
- How to measure the nonuniformity of a distribution?

Index of Coincidence (2)

- Measure of roughness (variance): a measure of the size of the peaks and valleys, i.e. the deviation from the uniform distribution (0.0384)
- If the distribution were perfectly flat (each letter with probability 0.0284), the var will be 0
- If the distribution follows Table 1, the var will be 0.0303.

$$
\begin{aligned}
&\text{var} \\
&= \sum_{i=a}^{i=z} \left(P(i) - \tfrac{1}{26} \right)^2 \\
&= \sum_{i=a}^{i=z} \left(P(i)^2 - \tfrac{2}{26} P(i) + \left(\tfrac{1}{26} \right)^2 \right) \\
&= \sum_{i=a}^{i=z} \left(P(i)^2 \right) - \tfrac{2}{26} \sum_{i=a}^{i=z} P(i) + \sum_{i=a}^{i=z} \left(\tfrac{1}{26} \right)^2 \\
&= \sum_{i=a}^{i=z} \left(P(i)^2 \right) - \tfrac{2}{26} * 1 + 26 * \left(\tfrac{1}{26} \right)^2 \\
&= \sum_{i=a}^{i=z} \left(P(i)^2 \right) - 0.0384
\end{aligned}
$$

Index of Coincidence (3)

- Since we don't know how often a particular letter should be without knowing the algorithm that generates the letter, we will approximate the probability from observed frequencies
- In an observed sample of n cipher text, suppose there are Freq(i) instances of character I
- The index of coincidence (CI) is to approximate variance from observed data:

$$IC = \sum_{i=a}^{i=z} \frac{Freq(i) * (Freq(i) - 1)}{n * (n - 1)}$$

Alphabets	IC
1	0.068
2	0.052
3	0.047
4	0.044
5	0.044
10	0.041
large	0.038

- The index of coincidence (IC) ranges from 0.0384 (polyalphabetic substitution with perfectly flat distribution) to 0.068 (monoalphabetic substitution from common English)

Index of Coincidence Example

```
ZHYME ZVELK OJUBW CEYIN CUSML RAVSR YARNH CEARI UJPGP VARDU
QZCGR NNCAW JALUH GJPJR YGEGO FULUS QFFPV EYEDQ GOLKA LVOSJ
TFRTR YEJZS RVNCI HYJNM ZDCRO DKHCR MMLNR FFLFN QGOLK ALVOS
JWMIK QKUBP SAYOJ RRQYI NRNYC YQZSY EDNCA LEILX RCHUG IEBKO
YTHGV VCKHC JEQGO LKALV OSJED WEAKS GJHYC LLFTY IGSVT FVPMZ
NRZOL CYUZS FKOQR YRYAR ZFGKI QKRSV IRCEY USKVT MKHCR MYQIL
ZRCRL GQARZ OLKHY KSNFN RRNCZ TWUOC JNMKC MDEZP IRJEJ W
```

- Number of characters = 346
- Index of coincidence = 0.0434 (not monoalphabetic substitution)
- The no. of alphabet should be >= 3
  - Kasiski method: determine when a pattern of encrypting permutations has repeated to predict the number of alphabets used for substitutions

Cryptanalysis of Polyalphabetic Substitutions

- Information in the frequency distribution of letters are not available: all are flat distributions
  - Looks more secure than monoalphabetic substitutions
- How to break the polyalphabetic substitution?
  1. Determine the number of alphabets employed
  2. Break the ciphertext into pieces that were enciphered with the same alphabet
  3. Solve each piece as a monoalphabetic substitution
- 2 tools are used:
  - Index of coincidence: estimate if polyalphabetic substitution is used or not
  - Kasiski method: determine when a pattern of encrypting permutations has repeated to predict the number of alphabets used for substitutions

Kasiski Method for Repeated Patterns breaking the Vigenère cipher

- Based on the regularity of English: letter groupings and full words are repeated, e.g.
  - English uses endings -th, -ing, -ed, -ion, -tion, -ation, …
  - English uses beginings im-, in- un-, re-, …
  - English uses patterns -eek-, -oot-, -our-, …
  - Words with high frequency: of, and, to, with, are, is, …
- Principle of Kasiski method: if a message is encoded with *n* alphabets in cyclic rotation, and if a particular word or letter group appears *k* times in a plaintext message, it should be encoded approximately *k/n* times from the same alphabet
- E.g. if a keyword is 6 characters long, there are only 6 different ways to position the keyword over the plaintext word, a plaintext word or letter group that appears more than 6 times must be encrypted at least twice by the same position of the keyword and those occurrences will be enciphered identically

Kasiski Method

- Look for repeated fragments in the ciphertext and compile a list of the distances that separate the repetitions. Then, the keyword length is likely to divide many of these distances.
- If a repeated substring in a plaintext is encrypted by the same substring in the keyword, then the ciphertext contains a repeated substring and the distance of the two occurrences is a multiple of the keyword length.
- Not every repeated string in the ciphertext arises in this way; but, the probability of a repetition by chance is noticeably smaller.

Kasiski Method for Repeated Patterns (2)

- Consider the Dickens "It was the best of times …" with keyword dickens:

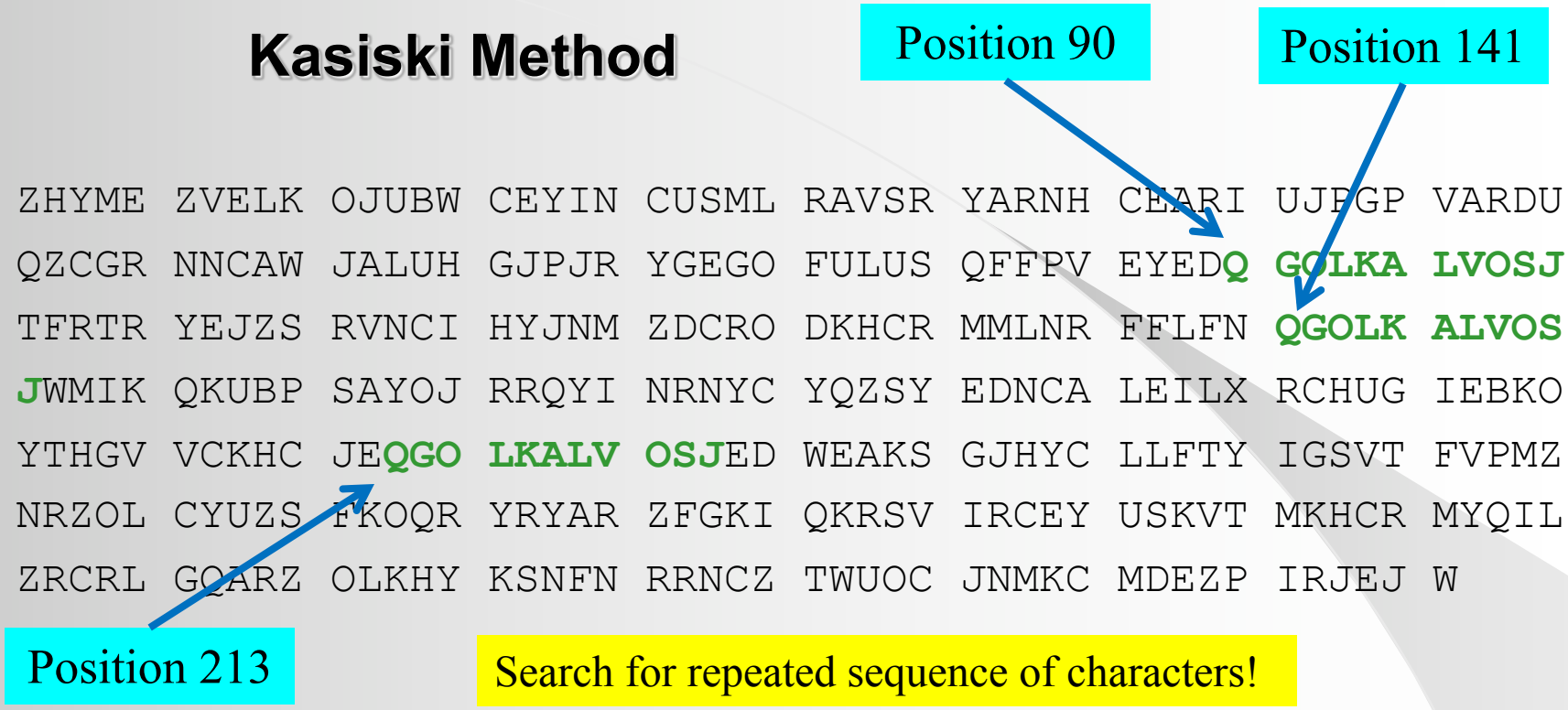
```
dicke nsdic kensd icken sdick ensdi ckens dicke
ITWAS THEBE STOFT IMESI TWAST HEWOR STOFT IMESI
nsdic kensd icken sdick ensdi ckens dicke nsdic
TWAST HEAGE OFWIS DOMIT WASTH EAGEO FFOOL ISHNE
kensd icken sdick ensdi ckens dicke nsdic kensd
SSITW ASTHE EPOCH OFBEL IEFIT WASTH EEPOC HOFIN
```

- Repeated patterns: ITWAST

Starting position	Distance from previous	Factors
20		
83	63 (83-20)	3, 7, 9, 21, 63
104	21 (104-83)	3, 7, 21

Possible key lengths are 3, 7, 21

Kasiski Method



- 3 occurrences of the 11-character sequence (QGOLKALVOSJ)
  - Distance between first 2 sequences (141-90) = 51
  - Distance between second 2 sequences (213-141) = 72
- The common divisor between 51 and 72 is 3
- Estimated key length is 3

Sample Ciphertext 1

```
ZHYME ZVELK OJUBW CEYIN CUSML RAVSR YARNH CEARI UJPGP VARDU
QZCGR NNCAW JALUH GJPJR YGEGO FULUS QFFPV EYEDQ GOLKA LVOSJ
TFRTR YEJZS RVNCI HYJNM ZDCRO DKHCR MMLNR FFLFN QGOLK ALVOS
JWMIK QKUBP SAYOJ RRQYI NRNYC YQZSY EDNCA LEILX RCHUG IEBKO
YTHGV VCKHC JEQGO LKALV OSJED WEAKS GJHYC LLFTY IGSVT FVPMZ
NRZOL CYUZS FKOQR YRYAR ZFGKI QKRSV IRCEY USKVT MKHCR MYQIL
ZRCRL GQARZ OLKHY KSNFN RRNCZ TWUOC JNMKC MDEZP IRJEJ W
```

- Number of characters = 346
- Index of coincidence = 0.0434 (not monoalphabetic substitution)
- By Kasiski method, the estimated key length is 3, divide the ciphertext into 3 sequences:
  - $c_1\ c_4\ c_7\ c_{10} \ldots \rightarrow$ Z M V K …
  - $c_2\ c_5\ c_8\ c_{11} \ldots \rightarrow$ H E E O …
  - $c_3\ c_6\ c_9\ c_{12} \ldots \rightarrow$ Y Z L J …

Sample Ciphertext 1 – IC (2)

```
ZHYME ZVELK OJUBW CEYIN CUSML RAVSR YARNH CEARI UJPGP VARDU
QZCGR NNCAW JALUH GJPJR YGEGO FULUS QFFPV EYEDQ GOLKA LVOSJ
TFRTR YEJZS RVNCI HYJNM ZDCRO DKHCR MMLNR FFLFN QGOLK ALVOS
JWMIK QKUBP SAYOJ RRQYI NRNYC YQZSY EDNCA LEILX RCHUG IEBKO
YTHGV VCKHC JEQGO LKALV OSJED WEAKS GJHYC LLFTY IGSVT FVPMZ
NRZOL CYUZS FKOQR YRYAR ZFGKI QKRSV IRCEY USKVT MKHCR MYQIL
ZRCRL GQARZ OLKHY KSNFN RRNCZ TWUOC JNMKC MDEZP IRJEJ W
```

- For sequence 1
  - Number of characters : 116
  - IC = 0.06747
- For sequence 2
  - Number of characters : 115
  - IC = 0.06499
- For sequence 3
  - Number of characters : 115
  - IC = 0.07597

Analyzing Polyalphabetic Cipher

1. Use the Kasiski method to predict likely “numbers” of enciphering alphabets
2. If no “numbers” emerge fairly regularly, the encryption is probably not a polyalphabetic substitution
3. Compute the index of coincidence to validate the predictions from step 1
4. When steps 1 and 3 indicate a promising value, separate the ciphertext into appropriate subsets and independently compute index of coincidence of each subset

The "Perfect" Substitution Cipher

- The ideal substitution would use many alphabets for an unrecognizable distribution and no apparent pattern for the choice of an alphabet at a particular point
- An infinite non-repeating sequence of alphabets would confuse the Kasiski method
  - A repeated plaintext would not be encrypted the same way twice
  - Suppose some piece of ciphertext was a duplicate of a previous piece, the duplicate would almost certainly be accidental: the distance between 2 duplicated sequences would not denote a period in the encryption pattern
- The index of coincidence would be close to 0.038
- Any possible attack?

What's the problem?

One-Time Pad

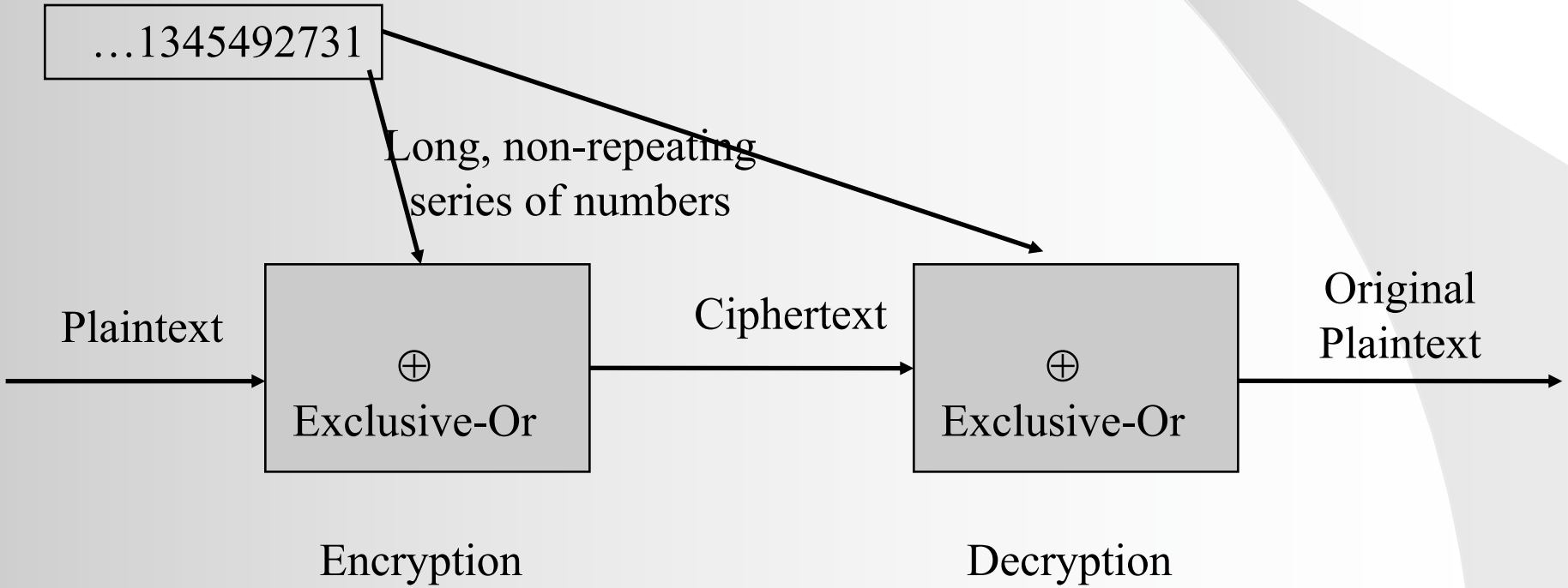
- The encryption method is based on a large nonrepeating set of keys, which is written on sheets of paper, glued together into a pad
- Assuming each sheet of paper contains a key of 20 characters long
- The sender needs to send a message of 300 characters
  1. The sender will tear off the next 15 pages of keys
  2. Write the key one at a time above the letters of the plaintext
  3. Encipher the plaintext with a chart like Vigenere tableau
  4. The sender then destroys the used keys
- The receiver need a pad identical to that of the sender, on receiving a message, the receiver
  1. Takes the appropriate number of keys
  2. Decipher the message as if it were a polyalphabetic substitution with a long key

Problems:

- Absolute synchronization between sender and receiver
- Need an unlimited number of keys

Vernam Cipher (1)

- Type of one-time pad devised by G. Vernam for AT&T
- The encryption involves an arbitrarily long nonrepeating sequence of numbers that are combined with the plaintext



Vernam Cipher (2)

- Plaintext: VERNAMCIPHER
  - plaintext: V E R N A M C I P H E R
  - encoded: 21 4 17 13 0 12 2 8 15 7 4 17
  - random: 76 48 16 82 33 3 58 11 60 5 48 88
  - sum: 97 52 33 95 33 15 60 19 75 12 52 105
  - mod 26: 19 0 7 17 18 15 8 19 23 12 0 1
  - ciphertext: T A H R S P I T X M A B
- Ciphertext: tahrspitxmab
- Any possible attack?
  - The random number generator

Cracking Random Number Generators (1)

- Most common type of random number generator: linear congruential random number generator
  - Seed: the initial value r_0
  - Successive random number r_{i+1}

    $r_{i+1} = (a * r_i + b) \bmod n$
  - The generator produces random integers between 0 and n – 1
- If r_0 and a are relatively prime to n, each number between 0 and n – 1 will be generated before the sequence repeats
- Once repetition begins, the entire sequence repeats in order

Cracking Random Number Generators (2)

- Assuming you have intercepted a message with header MEMO (12 4 12 14) and the first 4 letters in the ciphertext are 15 15 10 13
- You can deduce the first four random numbers as follow:

  r0 = c0 – 12 mod n = 15 – 12 mod n = 3 mod n

  r1 = c1 – 4 mod n = 15 – 4 mod n = 11 mod n

  r2 = c2 – 12 mod n = 10 – 12 mod n = –2 mod n

  r3 = c3 – 14 mod n = 13 – 14 mod n = –1 mod n

$c_i = r_i + p_i \bmod n$

$r_i = c_i - p_i \bmod n$

Transpositions (Permutations)

- A transposition is an encryption in which the letters of the messages are rearranged (also called permutation because it is a rearrangement of the symbols of a message)
- Transpositions try to break established patterns
- **Columnar transposition**: rearrangement of the characters of the plaintext into columns
- E.g. THIS IS A MESSAGE TO SHOW HOW A COLUMN TRANSPOSITION WORKS

```
THISI        tssoh oaniw haaso lrsto imghw
SAMES        utpir seeoa mrook istwc nasns
SAGET
OSHOW
HOWAC
OLUMN
ARTRA
NSPOS
ITION
WORKS
```

Transpositions - Complexity

- Transposition involves no additional work except arranging the letters and reading them off again: the execution time of the algorithm is proportional to the length of the message
- The cipher requires storage for all characters of the message
- Output characters cannot be produced until all characters of the message have been read: not good for long message
- Alternative: permute the characters of the plaintext with a fixed period d, e.g. d=5, and the permutation is (2 4 5 1 3)

```
12345
THISI
SAMES
SAGET
OSHOW
HOWAC
OLUMN
ARTRA
NSPOS
ITION
WORKS
```

```
24513 24513
hsiti aessm aetsg sowoh owchw
...
```

Cryptanalysis of Columnar Transposition

- Based on characteristic patterns of pairs of adjacent letters, called digrams
- Some letter pairs appear frequently, e.g. -re-, -th-, -en-, -ed-, …
- Steps to analyze columnar transposition:
  1. Compute the letter frequencies: all letters appear with their normal frequencies implies that a transposition has been performed
  2. Break the text into columns by compare a block of ciphertext characters against characters successively farther away in the ciphertext (anagramming):
     1. Do common digrams appear?
     2. Do most of the digrams look reasonable?

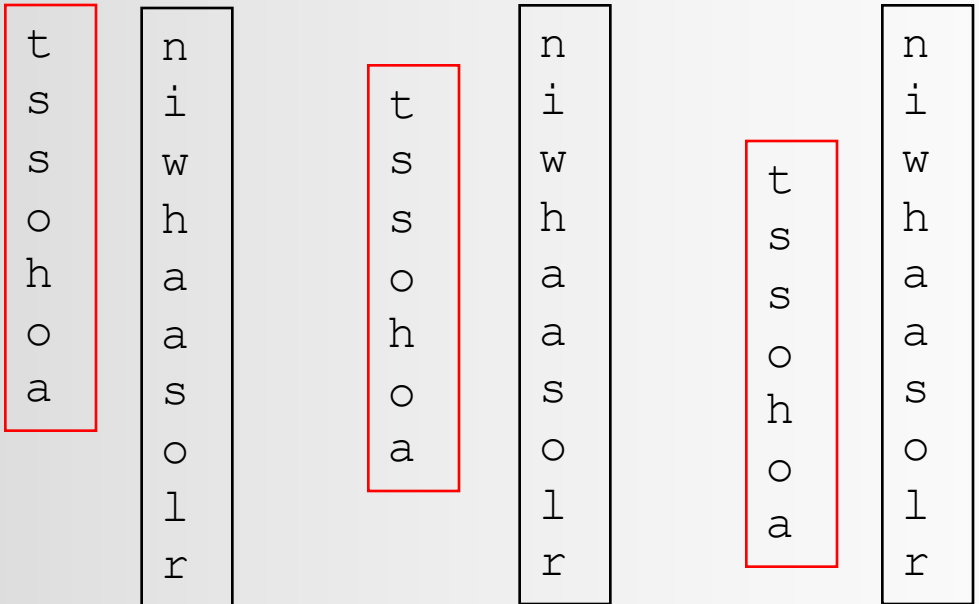


TABLE 2.3.3
Count of 2-Grams

	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	7	125	251	304	13	65	151	13	311	13	67	681	182	1216	5	144	0	764	648	1019	89	137	37	17	202	15
B	114	7	2	1	394	0	0	0	74	7	0	152	6	0	118	0	0	81	28	6	89	2	0	0	143	0
C	319	0	52	1	453	0	0	339	202	0	86	98	4	3	606	0	1	113	23	237	92	0	0	0	25	0
D	158	3	4	33	572	1	20	1	273	5	0	19	27	8	111	0	1	49	75	2	91	15	6	0	40	0
E	492	27	323	890	326	106	93	16	118	4	27	340	253	1029	30	143	25	1436	917	301	36	160	153	113	90	3
F	98	0	0	0	150	108	0	0	188	0	0	35	1	1	326	0	0	142	3	58	54	0	0	0	5	0
G	122	0	0	2	271	0	20	145	95	0	0	23	3	51	129	0	0	150	29	28	58	0	0	0	6	0
H	646	2	5	3	2053	0	0	2	426	0	0	6	6	14	287	0	0	56	10	85	31	0	4	0	15	0
I	236	51	476	285	271	80	174	1	10	0	31	352	184	1550	554	62	5	212	741	704	7	155	0	14	1	49
J	18	0	0	0	26	0	0	0	5	0	0	0	0	0	45	0	0	1	0	0	48	0	0	0	0	0
K	14	1	0	1	187	1	0	7	56	0	4	7	1	20	7	0	0	3	39	1	1	0	0	0	4	0
L	359	5	6	197	513	28	29	0	407	0	21	378	22	1	208	11	0	9	104	68	72	15	3	0	219	0
M	351	65	5	0	573	2	0	0	259	0	0	2	126	8	240	139	0	5	47	1	65	1	0	0	37	0
N	249	2	281	761	549	46	630	6	301	4	30	33	47	88	239	2	3	5	340	743	56	31	8	1	71	2
O	48	57	91	130	21	731	46	14	52	8	44	234	397	1232	125	164	0	861	201	223	533	188	194	7	23	2
P	241	0	1	0	310	0	0	42	75	0	0	144	13	1	268	103	0	409	32	51	81	0	0	0	3	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	73	0	0	0	0	0
R	470	15	79	129	1280	14	80	8	541	0	94	75	139	149	510	25	0	97	300	273	88	65	8	1	140	0
S	200	4	94	9	595	8	0	186	390	0	30	48	37	7	234	128	3	9	277	823	192	0	13	0	27	0
T	381	2	22	1	872	4	1	2161	865	0	0	62	27	9	756	2	0	295	257	131	120	3	54	0	125	3
U	72	87	103	51	91	11	80	2	54	0	3	230	69	318	4	81	0	306	256	263	6	3	0	2	3	1
V	65	0	0	2	522	0	0	0	223	0	0	0	1	0	46	0	0	0	2	0	1	1	0	0	5	0
W	282	1	0	4	239	0	0	175	259	0	0	5	0	44	159	0	0	13	45	2	0	0	0	0	3	0
X	9	0	15	0	17	0	0	1	15	0	0	0	1	0	1	47	0	0	0	23	0	0	0	5	0	0
Y	17	1	3	2	84	0	0	0	20	0	1	5	11	5	64	9	0	9	44	5	4	0	3	0	2	1
Z	18	0	0	0	36	0	0	0	17	0	0	1	0	0	4	0	0	0	0	0	1	0	0	0	0	2

Cryptanalysis of Columnar Transposition (2)

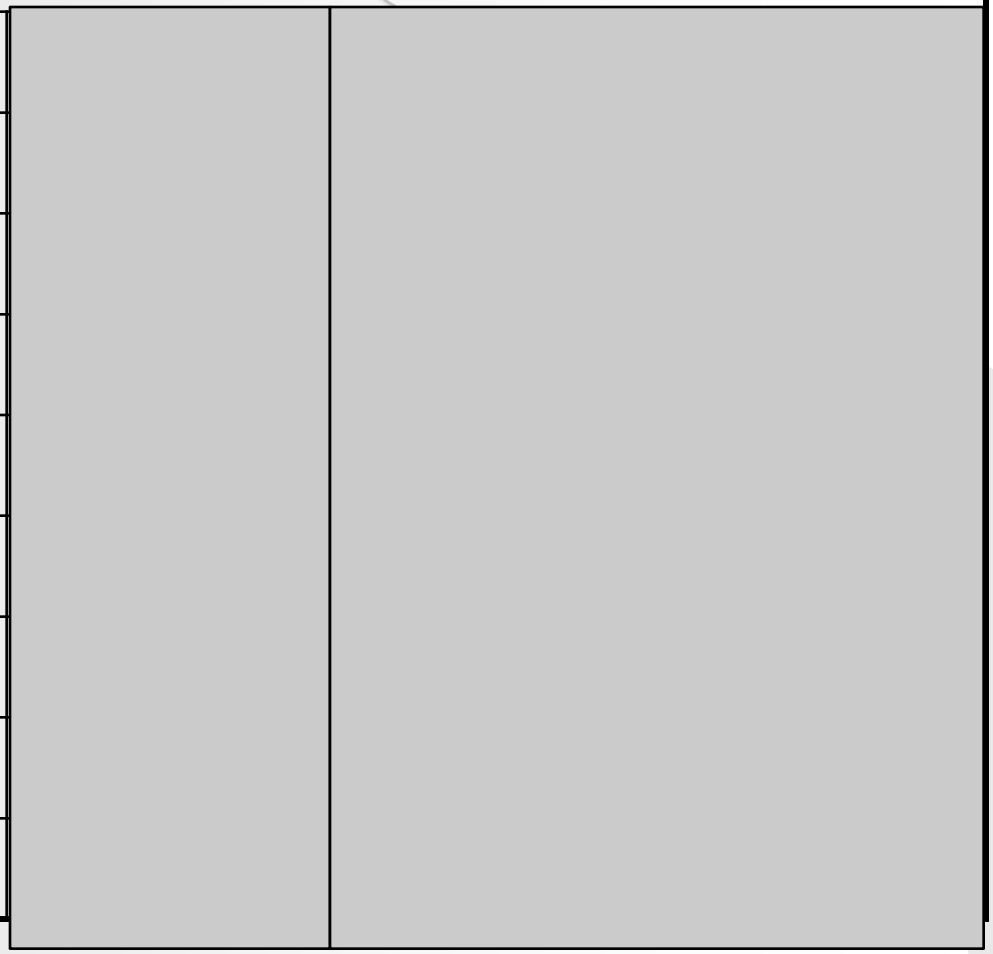
first letter	(second letter, digram, relative freq)
t	n tn 9
s	i si 390
s	w sw 13
o	h oh 52
h	a ha 646
o	a oa 48
a	s as 648
Mean	258
Std. Dev.	297

TABLE 2.3.3
Count of 2-Grams

	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	7	125	251	304	13	65	151	13	311	13	67	681	182	1216	5	144	0	764	648	1019	89	137	37	17	202	15
B	114	7	2	1	394	0	0	0	74	7	0	152	6	0	118	0	0	81	28	6	89	2	0	0	143	0
C	319	0	52	1	453	0	0	339	202	0	86	98	4	3	606	0	1	113	23	237	92	0	0	0	25	0
D	158	3	4	33	572	1	20	1	273	5	0	19	27	8	111	0	1	49	75	2	91	15	6	0	40	0
E	492	27	323	890	326	106	93	16	118	4	27	340	253	1029	30	143	25	1436	917	301	36	160	153	113	90	3
F	98	0	0	0	150	108	0	0	188	0	0	35	1	1	326	0	0	142	3	58	54	0	0	0	5	0
G	122	0	0	2	271	0	20	145	95	0	0	23	3	51	129	0	0	150	29	28	58	0	0	0	6	0
H	646	2	5	3	2053	0	0	2	426	0	0	6	6	14	287	0	0	56	10	85	31	0	4	0	15	0
I	236	51	476	285	271	80	174	1	10	0	31	352	184	1550	554	62	5	212	741	704	7	155	0	14	1	49
J	18	0	0	0	26	0	0	0	5	0	0	0	0	0	45	0	0	1	0	0	48	0	0	0	0	0
K	14	1	0	1	187	1	0	7	56	0	4	7	1	20	7	0	0	3	39	1	1	0	0	0	4	0
L	359	5	6	197	513	28	29	0	407	0	21	378	22	1	208	11	0	9	104	68	72	15	3	0	219	0
M	351	65	5	0	573	2	0	0	259	0	0	2	126	8	240	139	0	5	47	1	65	1	0	0	37	0
N	249	2	281	761	549	46	630	6	301	4	30	33	47	88	239	2	3	5	340	743	56	31	8	1	71	2
O	48	57	91	130	21	731	46	14	52	8	44	234	397	1232	125	164	0	861	201	223	533	188	194	7	23	2
P	241	0	1	0	310	0	0	42	75	0	0	144	13	1	268	103	0	409	32	51	81	0	0	0	3	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	73	0	0	0	0	0
R	470	15	79	129	1280	14	80	8	541	0	94	75	139	149	510	25	0	97	300	273	88	65	8	1	140	0
S	200	4	94	9	595	8	0	186	390	0	30	48	37	7	234	128	3	9	277	823	192	0	13	0	27	0
T	381	2	22	1	872	4	1	2161	865	0	0	62	27	9	756	2	0	295	257	131	120	3	54	0	125	3
U	72	87	103	51	91	11	80	2	54	0	3	230	69	318	4	81	0	306	256	263	6	3	0	2	3	1
V	65	0	0	2	522	0	0	0	223	0	0	0	1	0	46	0	0	0	2	0	1	1	0	0	5	0
W	282	1	0	4	239	0	0	175	259	0	0	5	0	44	159	0	0	13	45	2	0	0	0	0	3	0
X	9	0	15	0	17	0	0	1	15	0	0	0	1	0	1	47	0	0	0	23	0	0	0	5	0	0
Y	17	1	3	2	84	0	0	0	20	0	1	5	11	5	64	9	0	9	44	5	4	0	3	0	2	1
Z	18	0	0	0	36	0	0	0	17	0	0	1	0	0	4	0	0	0	0	0	1	0	0	0	0	2

Cryptanalysis of Columnar Transposition (2)

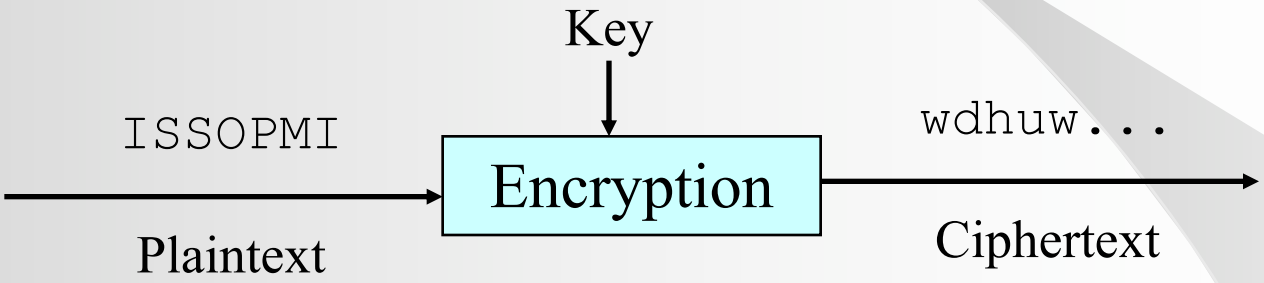
first letter	(second letter, digram, relative freq)
t	n tn 9
s	i si 390
s	w sw 13
o	h oh 52
h	a ha 646
o	a oa 48
a	s as 648
Mean	258
Std. Dev.	297



Most probably match!

Stream and Block Ciphers (1)

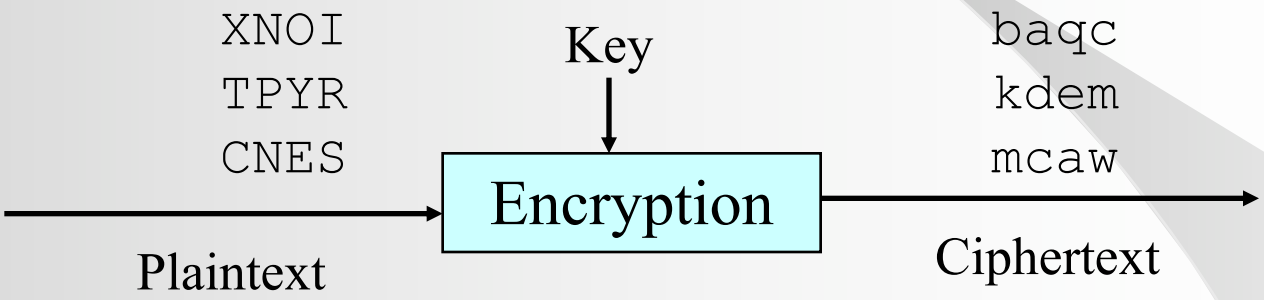
- Stream cipher: convert one symbol of plaintext immediately into a symbol of ciphertext, e.g. the substitution cipher



- Advantages:
  - Speed of transformation
  - Low error propagation: each symbol is separately encoded
- Disadvantages:
  - Low diffusion: all information of a symbol is contained in the symbol, subject to attack like frequency count, digram analysis, IC and Kasiski method
  - Possible for malicious insertions and modifications

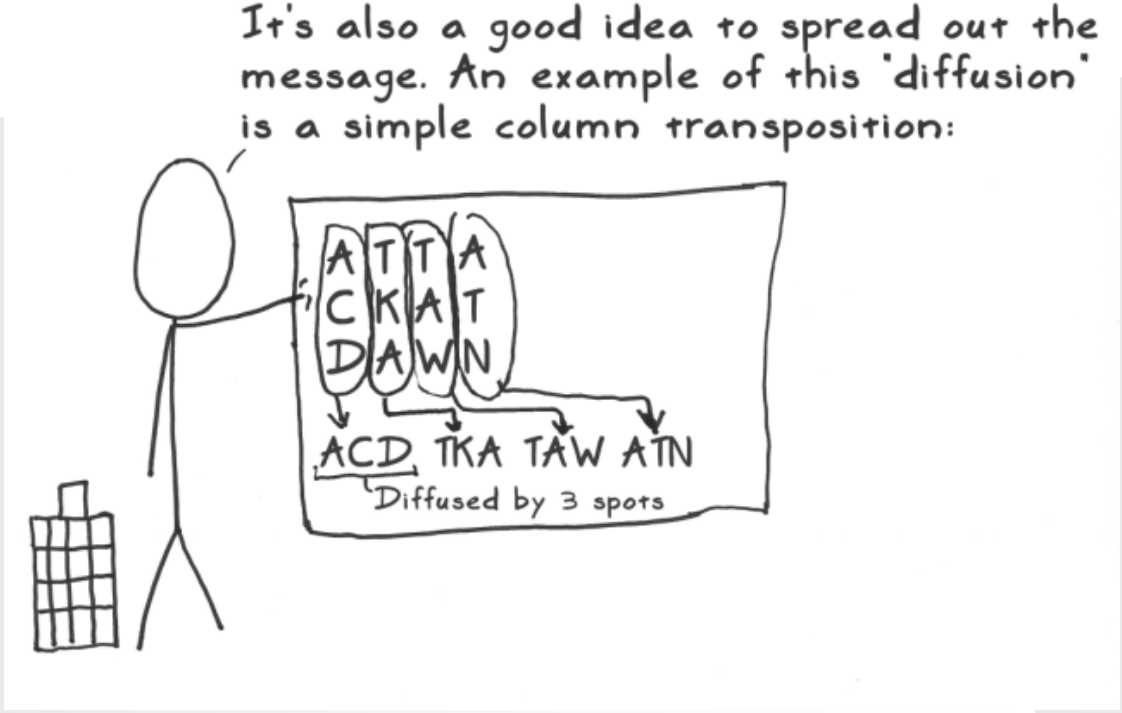
Stream and Block Ciphers (2)

- Block cipher: encrypt a group of plaintext symbol as one block, e.g. the columnar transposition cipher



- Advantages:
  - Diffusion: information of plaintext is diffused into several ciphertext symbols
  - Immunity to insertions: impossible to insert a single symbol in a block
- Disadvantages:
  - Slowness of encryption
  - Error propagation: an error will affect all other characters in the same block

Confusion and Diffusion



Confusion

- Cipher that makes the relationship between the plaintext/key pair and the ciphertext as complex as possible, i.e. difficult to determine how a message and a key were transformed into the corresponding ciphertext
- The attacker should not be able to predict when changing one character in the plaintext will do to the ciphertext $\Rightarrow$ the attacker will take a long time to determine the relationship among plaintext, key, and ciphertext
- Bad confusion: Caesar cipher
- Good confusion: Polyalphabetic substitution with a long key

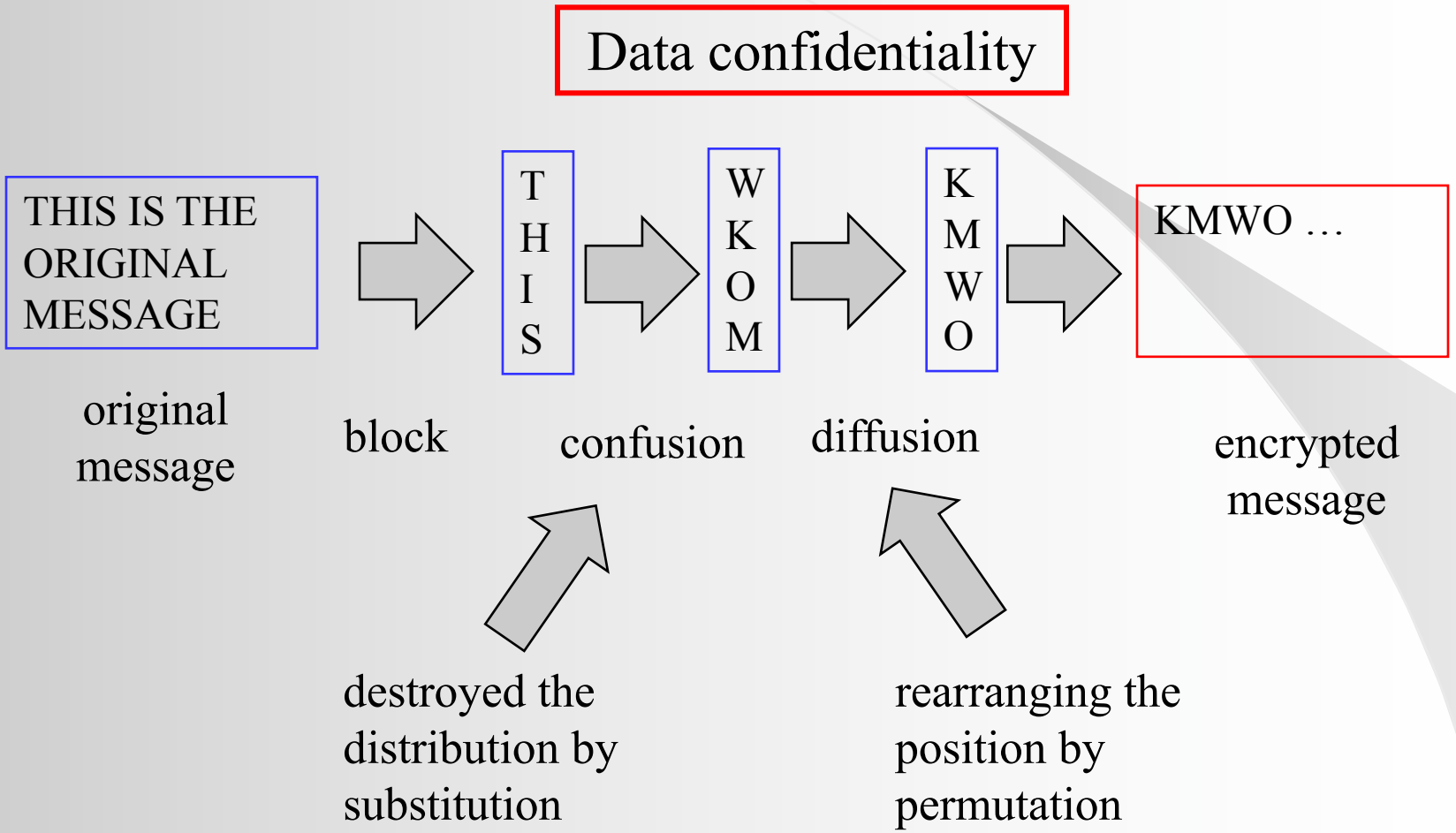
Diffusion

- Cipher that spreads the information from the plaintext over the entire ciphertext
- Changes in the plaintext should affect many parts of the ciphertext
- Good diffusion means that the attacker needs access to many ciphertexts in order to infer the algorithm
- Bad diffusion: the substitution ciphers
- Some diffusion: the transposition ciphers

DES provides good confusion and diffusion

- Substitution provides confusion: makes the output a nonlinear function of the input
- Permutation provides diffusion: spread the dependence of the output from a small number of input bit positions to a larger number

Purpose of cryptography



Rotor Machines

- Rotor machines implement polyalphabetic substitution ciphers with a long period
- A rotor machine consists of a bank of t rotors: the perimeter of each rotor R_i has 26 electrical contacts (one for each letter) on both its front and rear faces
- Each contact on the front face is wired to a contact on the rear face to implement a mapping f_i from plaintext to ciphertext letters (substitution)

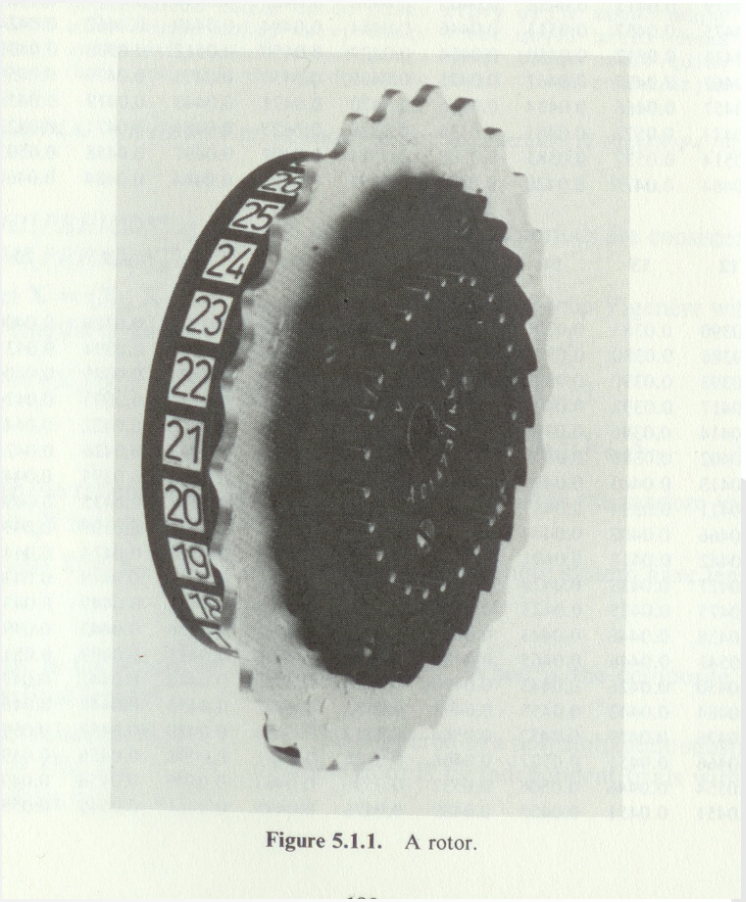
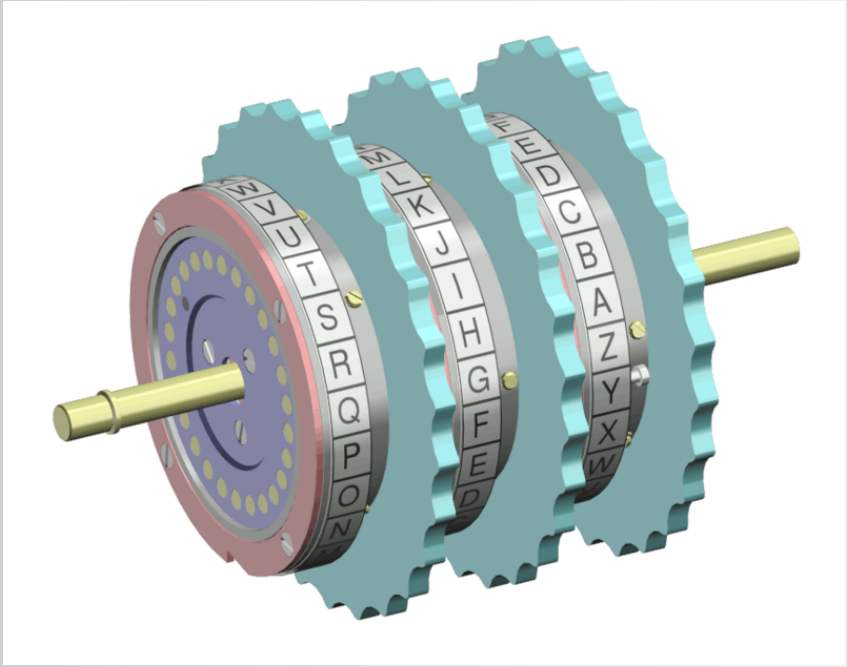
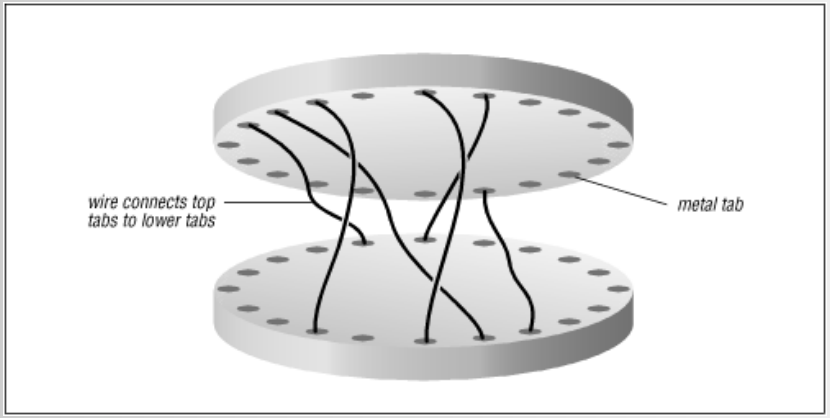


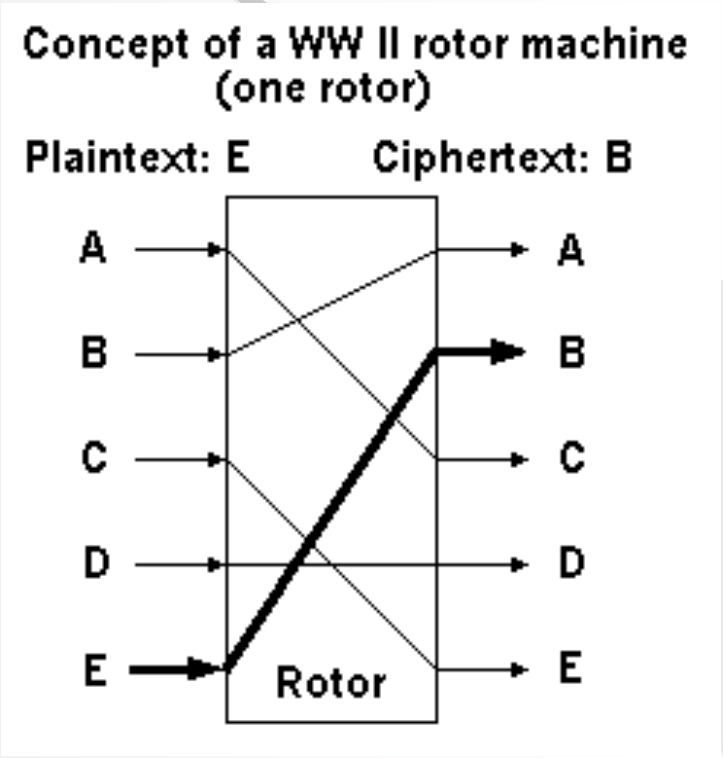
Figure 5.1.1. A rotor.

The Rotor

Rotor Machine (2)

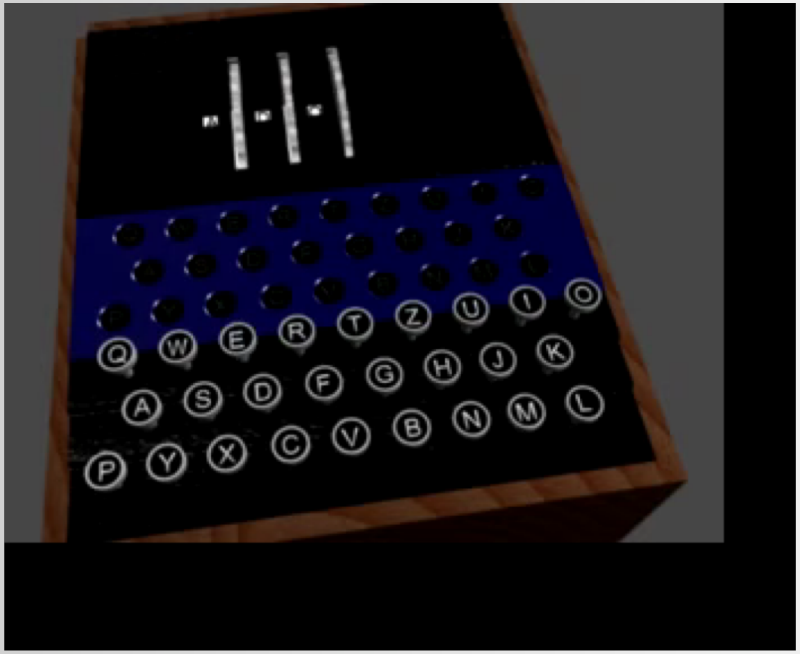




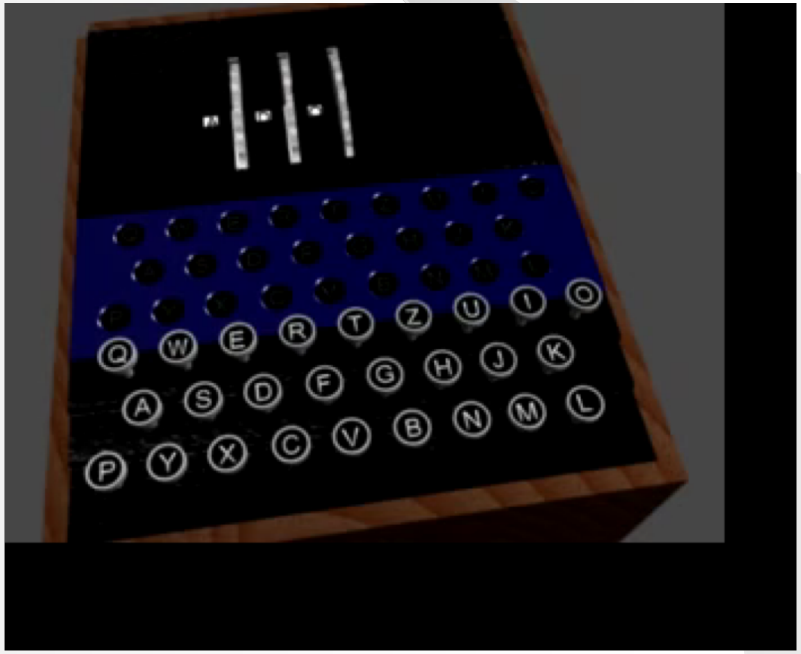


Rotor Machine – Encrypt and Decrypt

Encrypt



Decrypt



Rotor Machines (3)

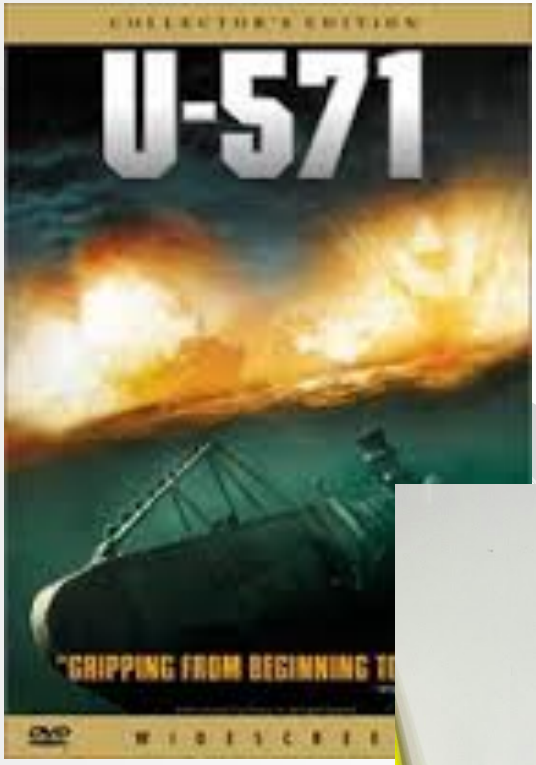
- R_i can rotate into 26 positions: each position alters the mapping
- When R_i is in position j_i, the mapping is

  $F_i(a) = (f_i(a - j_i) \bmod 26 + j_i) \bmod 26$
- A machine with t rotors implements a substitution cipher composed of $F_1, \ldots, F_t$, and the ith letter in the plaintext is enciphered as

  $E(m_i) = F_1 \bigcirc F_2 \bigcirc \ldots \bigcirc F_1(m_i)$
- The wirings plus initial positions of the rotors determine the starting key
- As each plaintext letter is enciphered, one or more of the rotors moves to a new position, changing the key
- A machine with t rotors does not return to its starting position until after 26^t successive encipherments, e.g. a machine with 5 rotors has a period of 11,881,376 letters
- The Enigma machine is a rotor machine
- Unix crypt: one-rotor machine based on German Enigma, but with a 256-element rotor

Movie U-571
The Enigma

- When the German crew abandoned their damaged submarine, a boarding party from Bulldog got on board and recovered a working Enigma machine, its cipher keys, keybooks and other cryptological records.



Enigma Machine

From http://www.dgp.toronto.edu/~lockwood/enigma/enigma.htm

Enigma Machine (2)