Cryptography and Network Security

Sixth Edition

by William Stallings

Chapter 2

Classical Encryption Techniques

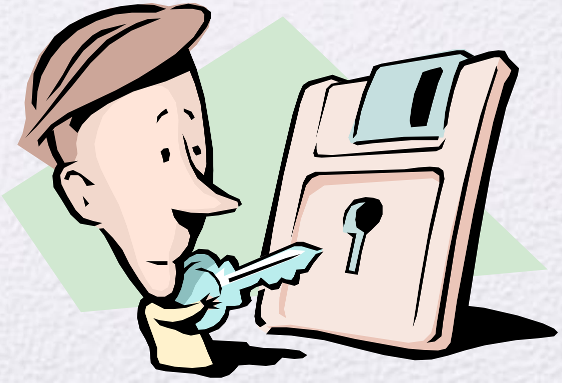
"I am fairly familiar with all the forms of secret writings, and am myself the author of a trifling monograph upon the subject, in which I analyze one hundred and sixty separate ciphers," said Holmes.

—The Adventure of the Dancing Men,

Sir Arthur Conan Doyle

Symmetric Encryption

- Also referred to as conventional encryption or single-key encryption
- Was the only type of encryption in use prior to the development of public-key encryption in the 1970s
- Remains by far the most widely used of the two types of encryption



Basic Terminology

- Plaintext
  - The original message
- Ciphertext
  - The coded message
- Enciphering or encryption
  - Process of converting from plaintext to ciphertext
- Deciphering or decryption
  - Restoring the plaintext from the ciphertext
- Cryptography
  - Study of encryption
- Cryptographic system or cipher
  - Schemes used for encryption
- Cryptanalysis
  - Techniques used for deciphering a message without any knowledge of the enciphering details
- Cryptology
  - Areas of cryptography and cryptanalysis together

Simplified Model of Symmetric Encryption

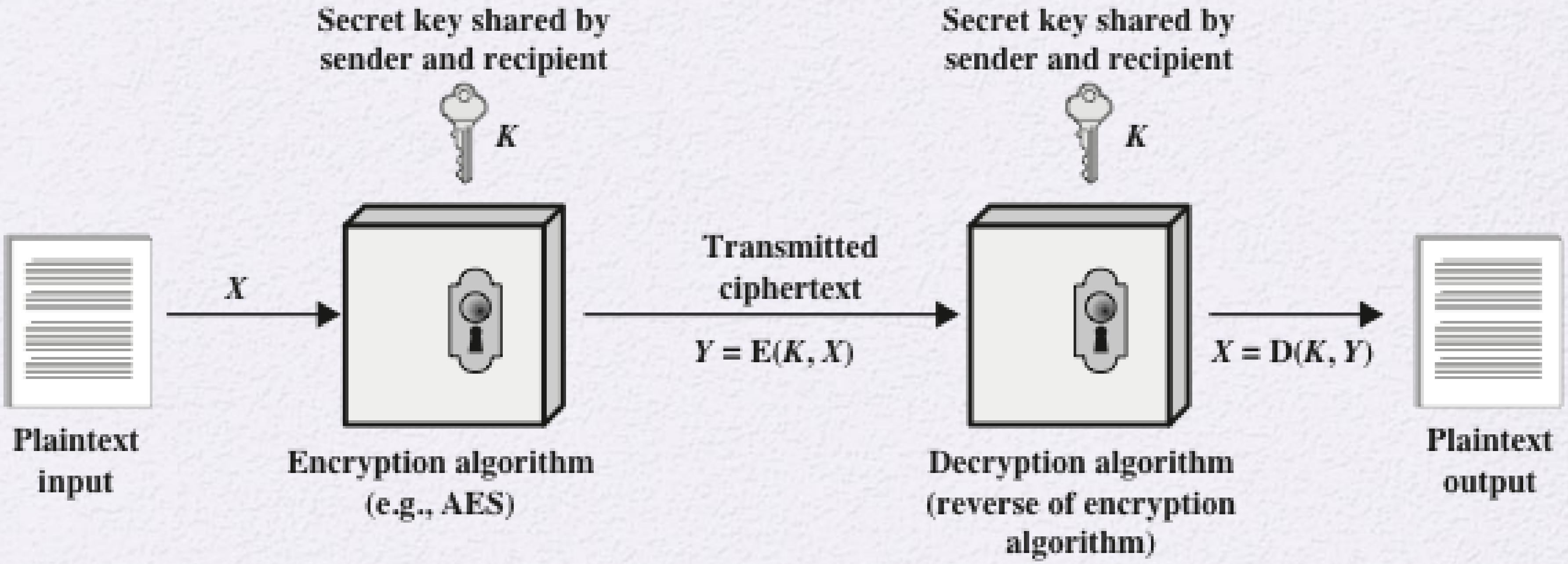


Figure 2.1 Simplified Model of Symmetric Encryption

A symmetric encryption scheme has five ingredients (Figure 2.1):

- Plaintext:
- Encryption algorithm
- Secret key
- Ciphertext
- Decryption algorithm

There are two requirements for secure use of conventional encryption:

- We need a strong encryption algorithm. At a minimum, we would like the algorithm to be such that an opponent who knows the algorithm and has access to one or more ciphertexts would be unable to decipher the ciphertext or figure out the key
- 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. In other words, we do not need to keep the algorithm secret; we need to keep only the key secret. This feature of symmetric encryption is what makes it feasible for widespread use

Model of Symmetric Cryptosystem

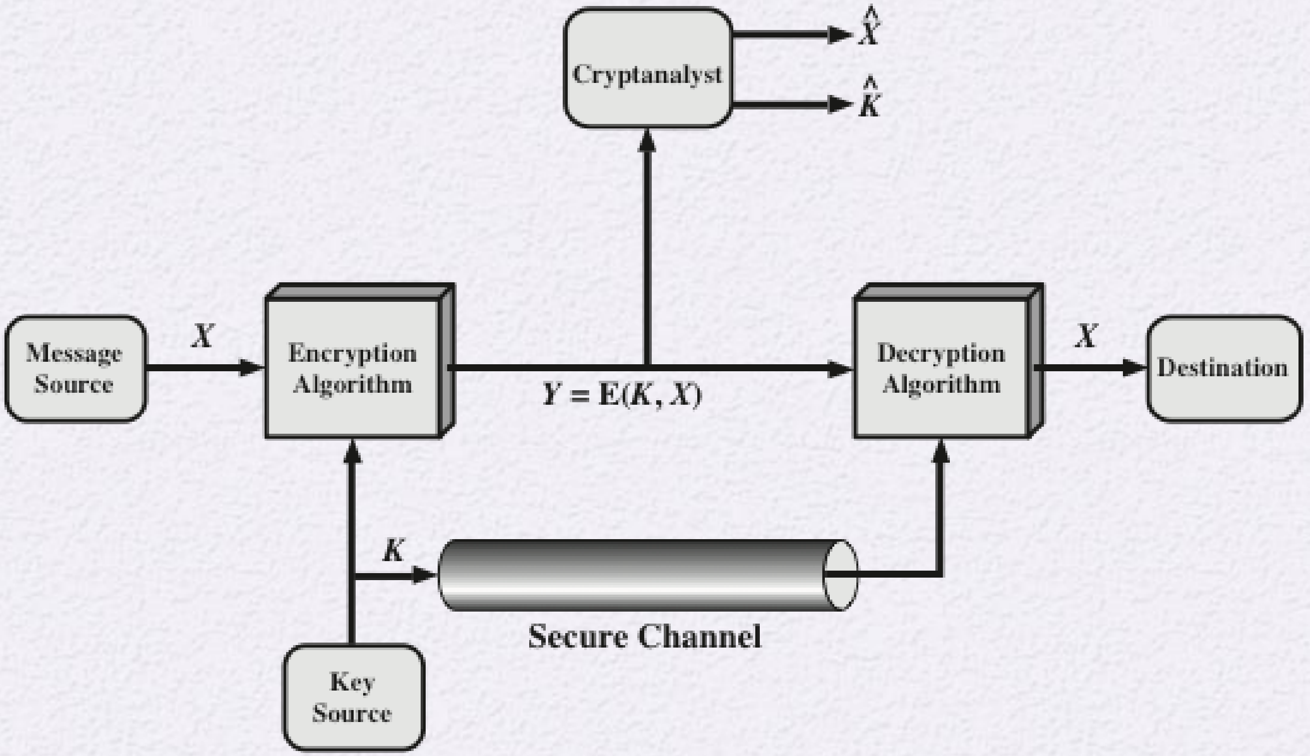


Figure 2.2 Model of Symmetric Cryptosystem

Cryptographic Systems

- Characterized along three independent dimensions:

The type of operations used for transforming plaintext to ciphertext	The number of keys used	The way in which the plaintext is processed
Substitution	Symmetric, single-key, secret-key, conventional encryption	Block cipher
Transposition	Asymmetric, two-key, or public-key encryption	Stream cipher

Cryptanalysis and Brute-Force Attack

Cryptanalysis

- Attack relies on the nature of the algorithm plus some knowledge of the general characteristics of the plaintext
- Attack exploits the characteristics of the algorithm to attempt to deduce a specific plaintext or to deduce the key being used

Brute-force attack

- Attacker tries every possible key on a piece of ciphertext until an intelligible translation into plaintext is obtained
- On average, half of all possible keys must be tried to achieve success

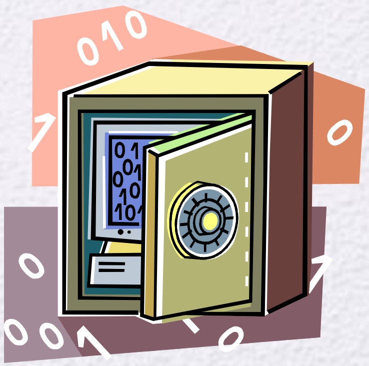
Table 2.1

Types of Attacks on Encrypted Messages

Type of Attack	Known to Cryptanalyst
Ciphertext Only	• Encryption algorithm • Ciphertext
Known Plaintext	• Encryption algorithm • Ciphertext • One or more plaintext-ciphertext pairs formed with the secret key
Chosen Plaintext	• Encryption algorithm • Ciphertext • Plaintext message chosen by cryptanalyst, together with its corresponding ciphertext generated with the secret key
Chosen Ciphertext	• Encryption algorithm • Ciphertext • Ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key
Chosen Text	• Encryption algorithm • Ciphertext • Plaintext message chosen by cryptanalyst, together with its corresponding ciphertext generated with the secret key • Ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key

Encryption Scheme Security

- Unconditionally secure
  - No matter how much time an opponent has, it is impossible for him or her to decrypt the ciphertext simply because the required information is not there
- Computationally secure
  - 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



Brute-Force Attack

Involves trying every possible key until an intelligible translation of the ciphertext into plaintext is obtained

On average, half of all possible keys must be tried to achieve success

To supplement the brute-force approach, some degree of knowledge about the expected plaintext is needed, and some means of automatically distinguishing plaintext from garble is also needed

Substitution Technique

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

Caesar Cipher

- Simplest and earliest known use of a substitution cipher
- Used by Julius Caesar
- Involves replacing each letter of the alphabet with the letter standing three places further down the alphabet
- Alphabet is wrapped around so that the letter following Z is A

plain: meet me after the toga party

cipher: PHHW PH DIWHU WKH WRJD SDUWB

Caesar Cipher Algorithm

- Can define transformation as:

  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

  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

- Mathematically give each letter a number

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

- Algorithm can be expressed as:

$$c = E(3, p) = (p + 3) \bmod (26)$$

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

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

- Where k takes on a value in the range 1 to 25; the decryption algorithm is simply:

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

Brute-Force Cryptanalysis of Caesar Cipher

(This chart can be found on page 35 in the textbook)

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

Sample of Compressed Text

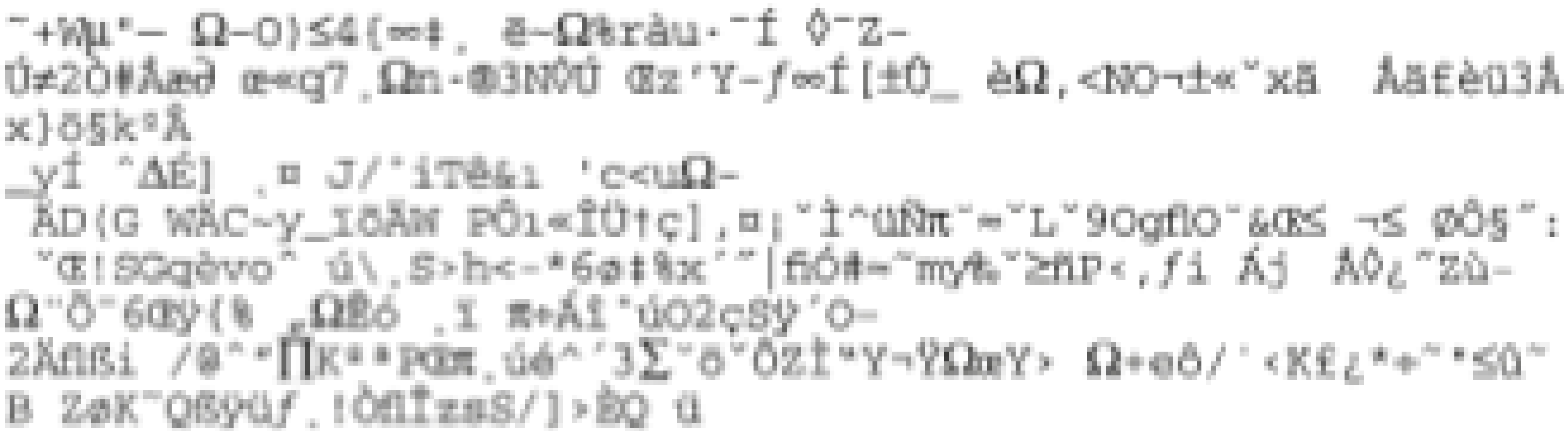


Figure 2.4 Sample of Compressed Text

Monoalphabetic Cipher

- Permutation
  - Of a finite set of elements *S* is an ordered sequence of all the elements of *S* , with each element appearing exactly once
  - For example, if S = {a, b, c}, there are six permutations of S: abc, acb, bac, bca, cab, cba
- If the "cipher" line can be any permutation of the 26 alphabetic characters, then there are 26! or greater than 4 x 10^{26} possible keys
  - This is 10 orders of magnitude greater than the key space for DES
  - Approach is referred to as a *monoalphabetic substitution cipher* because a single cipher alphabet is used per message

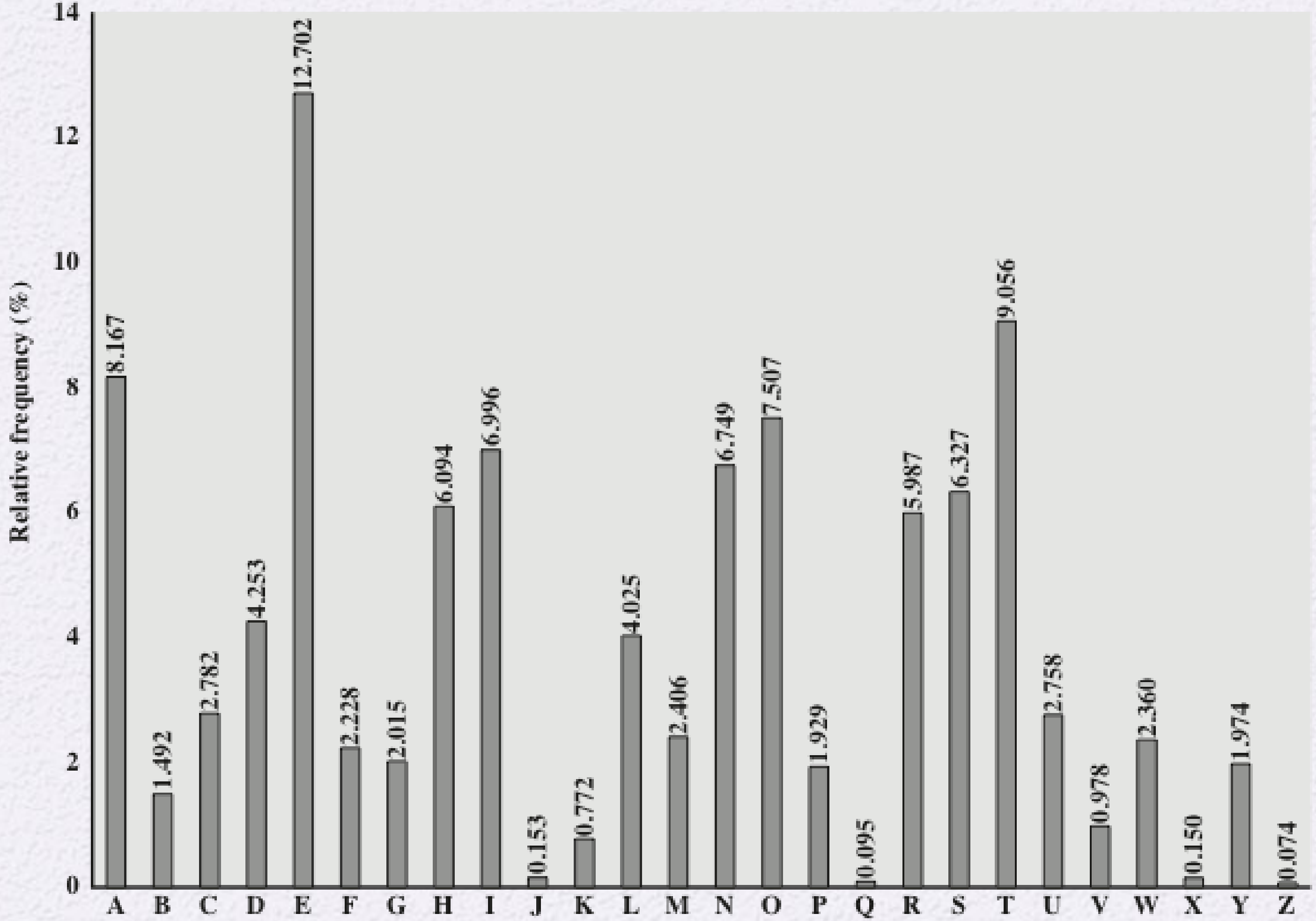
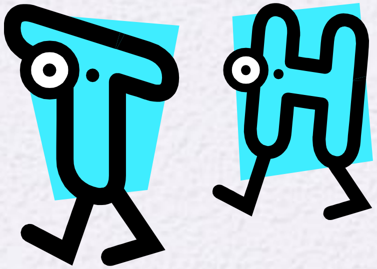


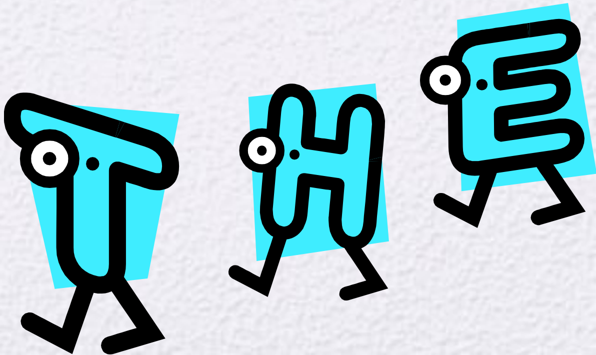
Figure 2.5 Relative Frequency of Letters in English Text

Monoalphabetic Ciphers

- Easy to break because they reflect the frequency data of the original alphabet
- Countermeasure is to provide multiple substitutes (homophones) for a single letter
- Digram
  - Two-letter combination
  - Most common is *th*



- Trigram
  - Three-letter combination
  - Most frequent is *the*



- UzqSovUoHxmoPvgPozPevSgzWSzoPfPeSxUDBmeTSxaIz vUePHzHmDzSHzoWSfPaPPDTSvPqUzWymxUzUHSx ePyePoPDzSzUfPomBzWPfUPzHmDJUDTmoHmq
- As a first step, the relative frequency of the letters can be determined and compared to a standard frequency distribution for English, such as is shown in Figure 2.5 (based on [LEWA00]). If the message were long enough, this technique alone might be sufficient, but because this is a relatively short message, we cannot expect an exact match. In any case, the relative frequencies of the letters in the ciphertext (in percentages) are as follows:

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 breakdown with Figure 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 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 (namely, A, B, G, Y, I, J) are likely included in the set {b, j, k, q, v, x, z}.
- There are a number of ways to proceed at this point. We could make some tentative assignments and start to fill in the plaintext to see if it looks like a reasonable "skeleton" of a message. A more systematic approach is to look for other regularities. For example, certain words may be known to be in the text. Or we could look for repeating sequences of cipher letters and try to deduce their plaintext equivalents.

- A powerful tool is to look at the frequency of two-letter combinations, known as digrams. A table similar to Figure 2.5 could be drawn up showing the relative frequency of digrams. The most common such digram is th. In our ciphertext, the most common digram is ZW, which appears three times. So we make the correspondence of Z with t and W with h. Then, by our earlier hypothesis, we can equate P with e. Now notice that the sequence ZWP appears in the ciphertext, and we can translate that sequence as "the." This is the most frequent trigram (three-letter combination) in English, which seems to indicate that we are on the right track.

- Next, notice the sequence ZWSZ in the first line. We do not know that these four letters form a complete word, but if they do, it is of the form th_t. If so, S equates with a.

- So far, then, we have

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 t

- Only four letters have been identified, but already we have quite a bit of the message. Continued analysis of frequencies plus trial and error should easily yield a solution from this point. The complete plaintext, with spaces added between words, follows:
- it was disclosed yesterday that several informal but direct contacts have been made with political representatives of the viet cong in moscow
- Monoalphabetic ciphers are easy to break because they reflect the frequency data of the original alphabet. A countermeasure is to provide multiple substitutes

Playfair Cipher

- Best-known multiple-letter encryption cipher
- Treats digrams in the plaintext as single units and translates these units into ciphertext digrams
- Based on the use of a 5 x 5 matrix of letters constructed using a keyword
- Invented by British scientist Sir Charles Wheatstone in 1854
- Used as the standard field system by the British Army in World War I and the U.S. Army and other Allied forces during World War II

Playfair Key Matrix

- Fill in letters of keyword (minus duplicates) from left to right and from top to bottom, then fill in the remainder of the matrix with the remaining letters in alphabetic order
- Using the keyword MONARCHY:

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

- In this case, the keyword is monarchy. 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 are in the same pair are separated with a filler letter, such as x, so that balloon would be treated as ba lx lo on.

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

- 3. Two plaintext letters that fall in the same column are each replaced by the letter beneath, with the top element of the column circularly following the last. For example, mu is encrypted as CM.
- 4. Otherwise, each plaintext letter in a pair 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).

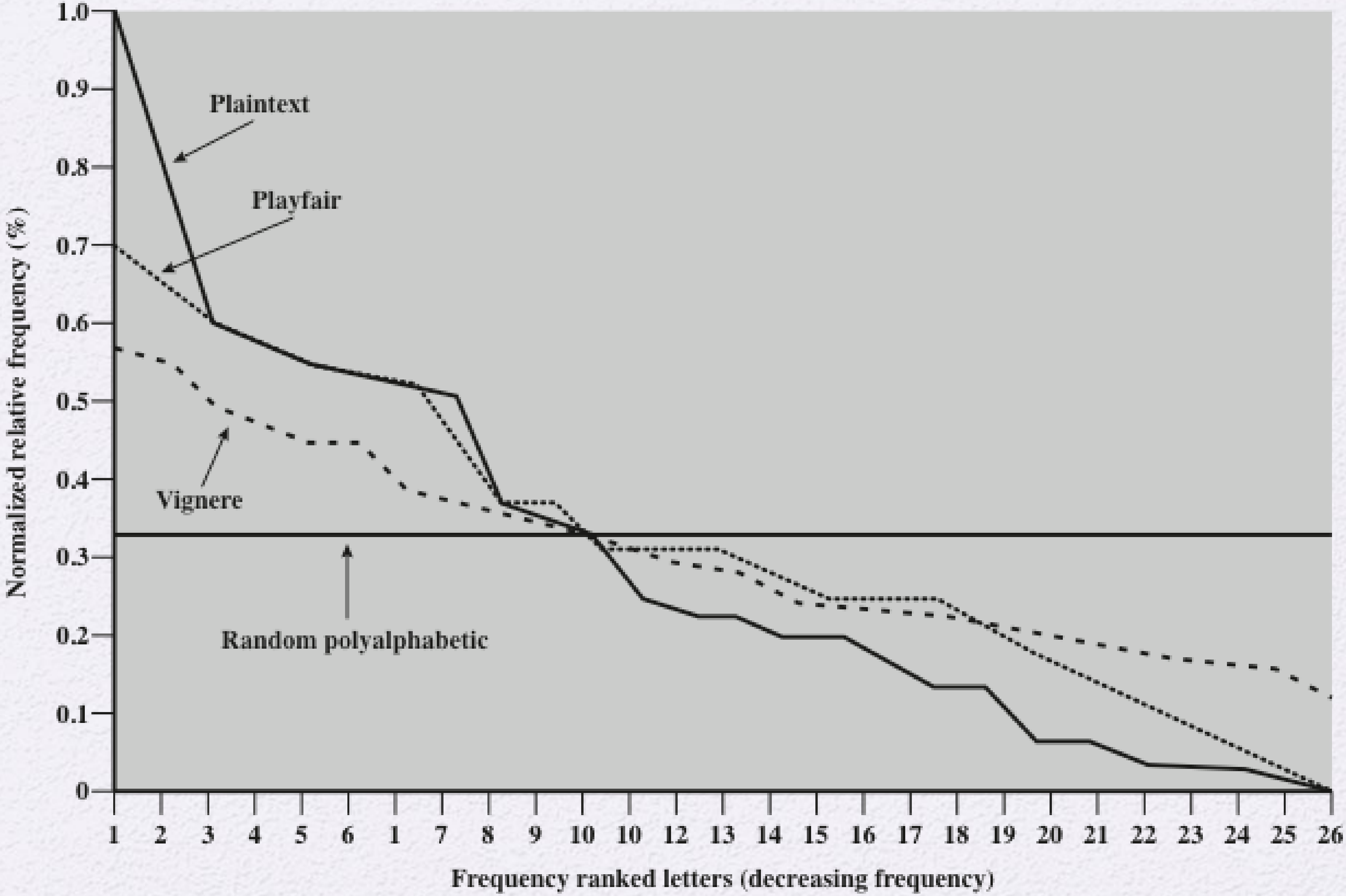


Figure 2.6 Relative Frequency of Occurrence of Letters

Hill Cipher

- Developed by the mathematician Lester Hill in 1929
- Strength is that it completely hides single-letter frequencies
  - The use of a larger matrix hides more frequency information
  - A 3 x 3 Hill cipher hides not only single-letter but also two-letter frequency information
- Strong against a ciphertext-only attack but easily broken with a known plaintext attack

Hill Cipher

- Matrix Multiplication:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \times \begin{pmatrix} E & F \\ G & H \end{pmatrix} = \begin{pmatrix} AE+BG & AF+BH \\ CE+DG & CF+DH \end{pmatrix}$$

- Modulus Theorem

THEOREM:
For an integer a and modulus m, let

$$R = \text{remainder of } \frac{|a|}{m}$$

Then the residue r of a modulus m is given by:

$$r = \begin{cases} R & if \quad a \geq 0 \\ m - R & if \quad a < 0 \quad and \quad R \neq 0 \\ 0 & if \quad a < 0 \quad and \quad R = 0 \end{cases}$$

Hill Cipher

- Modular Inverse:
- Matrix Inverse: $AA^{-1} = I$
- Modular Inverse for Mod m: $(a \cdot a^{-1})$ Mod $m = 1$
- For Modular Inverses, *a* and *m* must NOT have any prime factors in common
- By using Extended Euclidean algorithm, we can get the inverse
- Letter to Number Substitution

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

Hill Cipher

- p : plaintext , c = ciphertext , k = encryption key
- Encryption: c = (k * p) mod 26
- Decryption: p = (k^{-1} * c) mod 26
- k^{-1} = (inv(det(k)) * adj(k)) mod 26 , Where:
- Inv = inverse , det = determinant , adj = adjoint
- $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ det = a*d – c*b adj = $\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Polyalphabetic Ciphers

- Polyalphabetic substitution cipher
  - Improves on the simple monoalphabetic technique by using different monoalphabetic substitutions as one proceeds through the plaintext message

All these techniques have the following features in common:

- A set of related monoalphabetic substitution rules is used
- A key determines which particular rule is chosen for a given transformation

Vigenère Cipher

- Best known and one of the simplest polyalphabetic substitution ciphers
- In this scheme the set of related monoalphabetic substitution rules consists of the 26 Caesar ciphers with shifts of 0 through 25
- Each cipher is denoted by a key letter which is the ciphertext letter that substitutes for the plaintext letter a

Example of Vigenère Cipher

- To encrypt a message, a key is needed that is as long as the message
- Usually, the key is a repeating keyword
- For example, if the keyword is *deceptive*, the message "we are discovered save yourself" is encrypted as:

key: deceptivedeceptivedeceptive
plaintext: wearediscoveredsaveyourself
ciphertext: ZICVTWQNGRZGVTWAVZHCQYGLMGJ

Vigenère Autokey System

- A keyword is concatenated with the plaintext itself to provide a running key
- Example:

key: deceptivewearediscoveredsav

plaintext: wearediscoveredsaveyourself

ciphertext: ZICVTWQNGKZEIIGASXSTSLVVWLA

- Even this scheme is vulnerable to cryptanalysis
  - Because the key and the plaintext share the same frequency distribution of letters, a statistical technique can be applied

Vernam Cipher

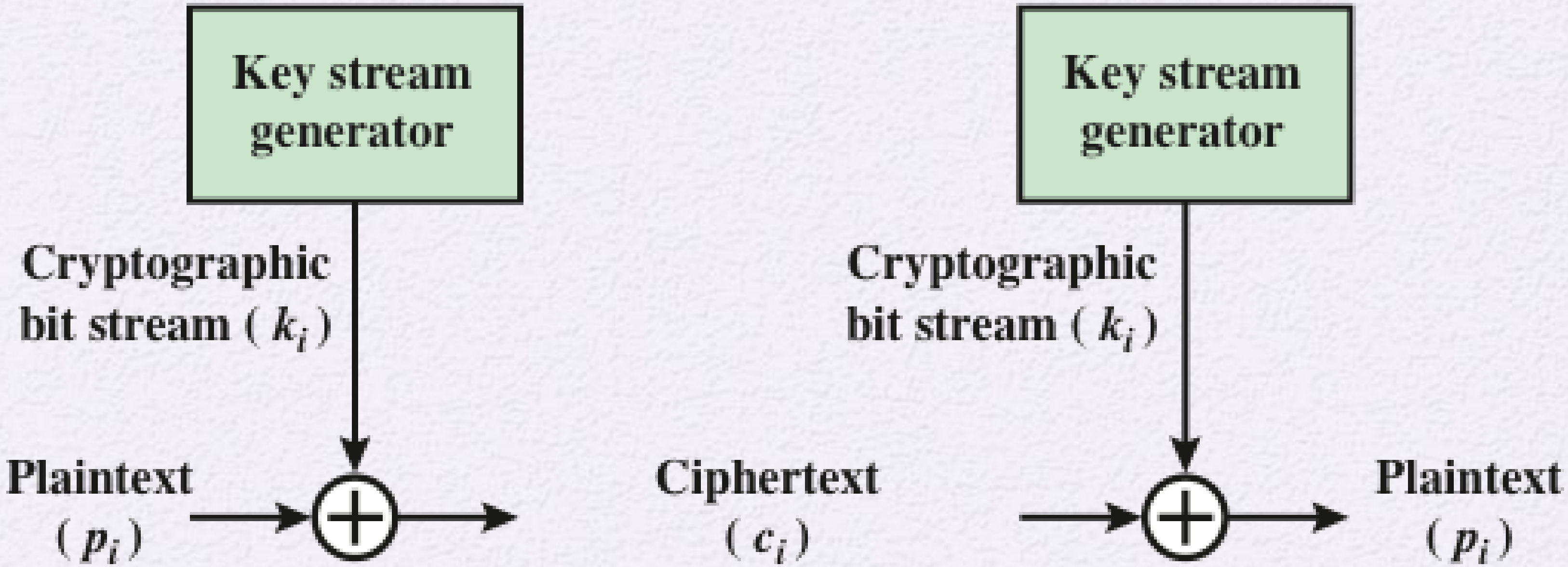


Figure 2.7 Vernam Cipher

One-Time Pad

- Improvement to Vernam cipher proposed by an Army Signal Corp officer, Joseph Mauborgne
- Use a random key that is as long as the message so that the key need not be repeated
- Key is used to encrypt and decrypt a single message and then is discarded
- Each new message requires a new key of the same length as the new message
- Scheme is unbreakable
  - Produces random output that bears no statistical relationship to the plaintext
  - Because the ciphertext contains no information whatsoever about the plaintext, there is simply no way to break the code

Difficulties

- The one-time pad offers complete security but, in practice, has two fundamental difficulties:
  - There is the practical problem of making large quantities of random keys
    - Any heavily used system might require millions of random characters on a regular basis
  - Mammoth key distribution problem
    - For every message to be sent, a key of equal length is needed by both sender and receiver
- Because of these difficulties, the one-time pad is of limited utility
  - Useful primarily for low-bandwidth channels requiring very high security
- The one-time pad is the only cryptosystem that exhibits *perfect secrecy* (see Appendix F)

Rail Fence Cipher

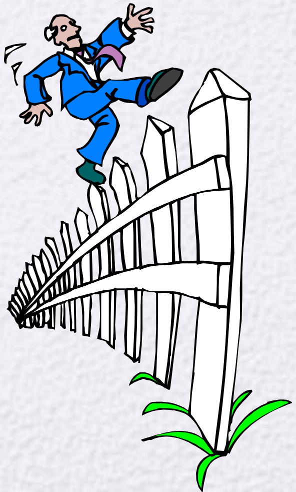
- Simplest transposition cipher
- Plaintext is written down as a sequence of diagonals and then read off as a sequence of rows
- To encipher the message "meet me after the toga party" with a rail fence of depth 2, we would write:

  m e m a t r h t g p r y

  e t e f e t e o a a t

  Encrypted message is:

  MEMATRHTGPRYETEFETEOAAT



Rail fence cipher - example

- Plaintext: transposition
- Number of "rails" = 3

```
T           S           I           N
   R    N     P     S     T     O
      A           O           I
```

- Cipher text: TSIN RNPSTO AOI

Row Transposition Cipher

- Is a more complex transposition
- Write the message in a rectangle, row by row, and read the message off, column by column, but permute the order of the columns
  - The order of the columns then becomes the key to the algorithm

Key:	4 3 1 2 5 6 7
Plaintext:	a t t a c k p
	o s t p o n e
	d u n t i l t
	w o a m x y z
Ciphertext:	TTNAAPTMTSUOAODWCOIXKNLYPETZ

- Key: 4 3 1 2 5 6 7
- Input: t t n a a p t
- m t s u o a o
- d w c o i x k
- n l y p e t z
- output: nScyaUoPTTWlTmDnaoIePaxTToKz

- To visualize the result of this double transposition, designate the letters in the original plaintext message by the numbers designating their position. Thus, with 28 letters in the message, the original sequence of letters is
- 01 02 03 04 05 06 07

  08 09 10 11 12 13 14

  15 16 17 18 19 20 21

  22 23 24 25 26 27 28

- After the first transposition, we have
- 03 10 17 24 04 11 18

  25 02 09 16 23 01 08

  15 22 05 12 19 26 06

  13 20 27 07 14 21 28

- which has a somewhat regular structure. But after the second transposition, we have
- 17 09 05 27 24 16 12

  07 10 02 22 20 03 25

  15 13 04 23 19 14 11

  01 26 21 18 08 06 28

This is a much less structured permutation and is much more difficult to cryptanalyze.

Rotor Machines

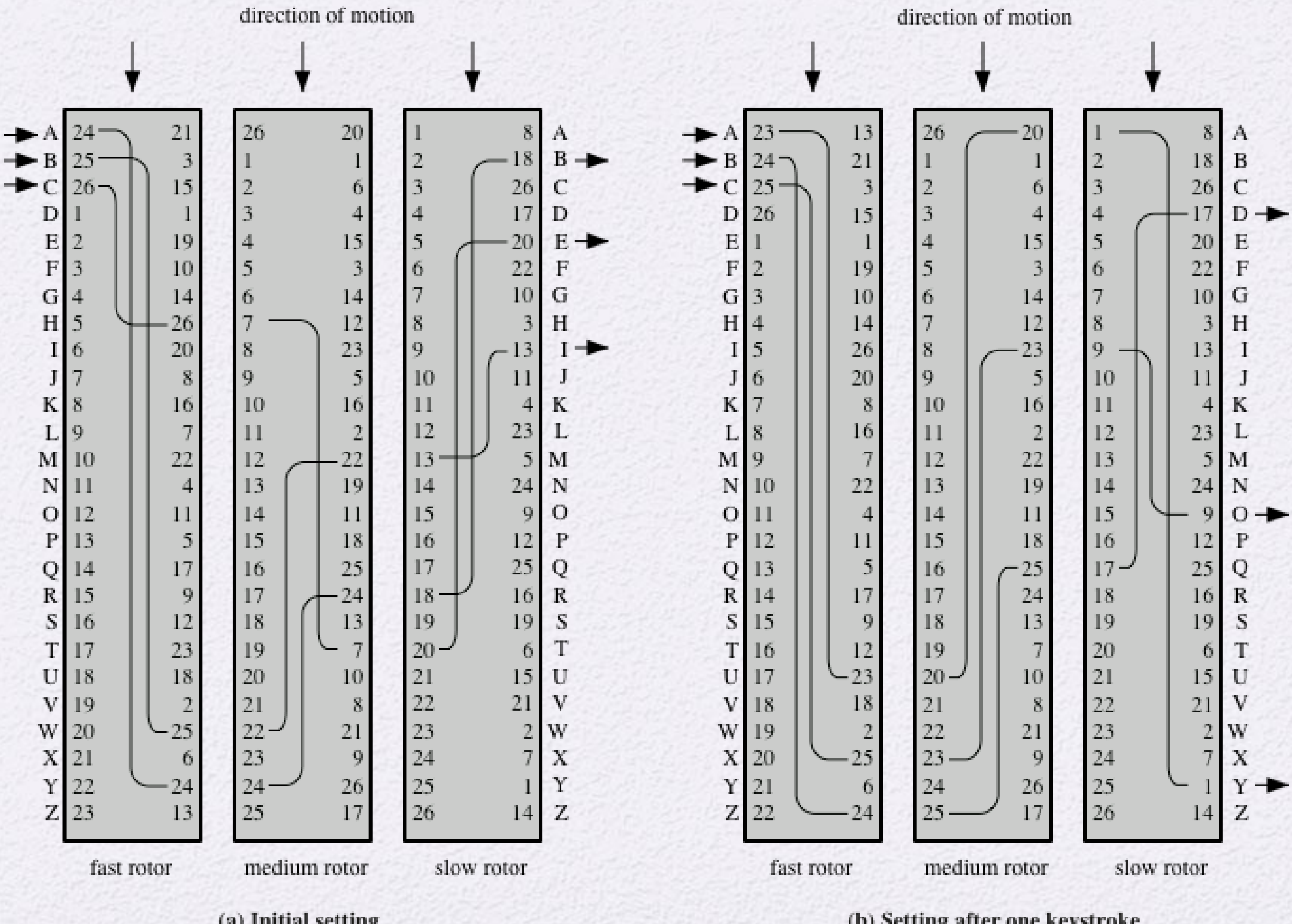


Figure 2.8 Three-Rotor Machine With Wiring Represented by Numbered Contacts

Steganography

3rd March

Dear George,

Greetings to all at Oxford. Many thanks for your
letter and for the Summer examination package.
All Entry Forms and Fees Forms should be ready
for final despatch to the Syndicate by Friday
20th or at the very latest, I'm told, by the 21st.
Admin has improved here, though there's room
for improvement still; just give us all two or three
more years and we'll really show you! Please
don't let these wretched 16+ proposals destroy
your basic O and A pattern. Certainly this
sort of change, if implemented immediately,
would bring chaos.

Sincerely yours.

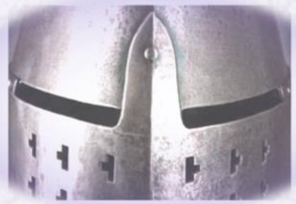
Figure 2.9 A Puzzle for Inspector Morse
(from *The Silent World of Nicholas Quinn*, by Colin Dexter)

Other Steganography Techniques

- Character marking
  - Selected letters of printed or typewritten text are over-written in pencil
  - The marks are ordinarily not visible unless the paper is held at an angle to bright light
- Invisible ink
  - A number of substances can be used for writing but leave no visible trace until heat or some chemical is applied to the paper
- Pin punctures
  - Small pin punctures on selected letters are ordinarily not visible unless the paper is held up in front of a light
- Typewriter correction ribbon
  - Used between lines typed with a black ribbon, the results of typing with the correction tape are visible only under a strong light

Summary

- Symmetric Cipher Model
  - Cryptography
  - Cryptanalysis and Brute-Force Attack
- Transposition techniques
- Rotor machines



- Substitution techniques
  - Caesar cipher
  - Monoalphabetic ciphers
  - Playfair cipher
  - Hill cipher
  - Polyalphabetic ciphers
  - One-time pad
- Steganography