Classic Encryption Techniques

Tsay, Yih-Kuen

Dept. of Information Management

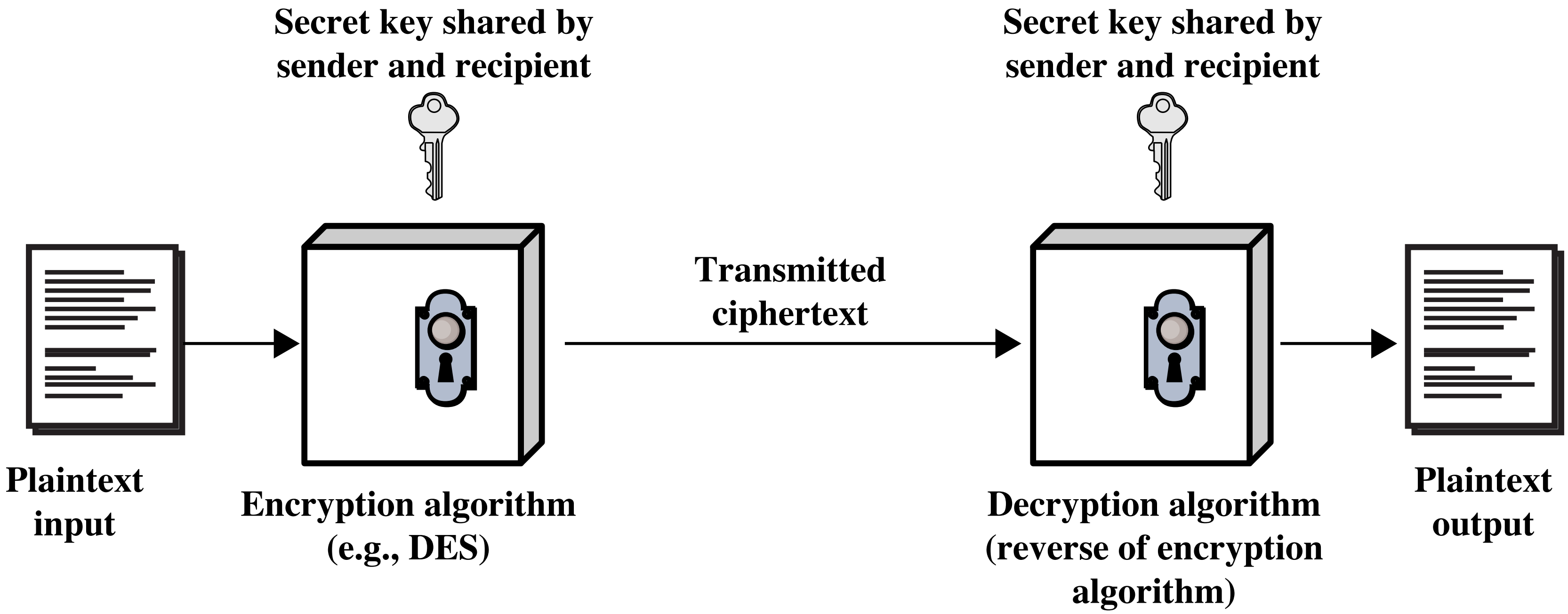
National Taiwan University

Symmetric Encryption/Ciphers

- Also known as
  - conventional,
  - single-key, or
  - secret-key

  encryption
- Encryption and decryption performed with the same key
- Most widely used type of ciphers

Simplified Model of Symmetric Encryption



Source: Figure 2.1, Stallings 2006

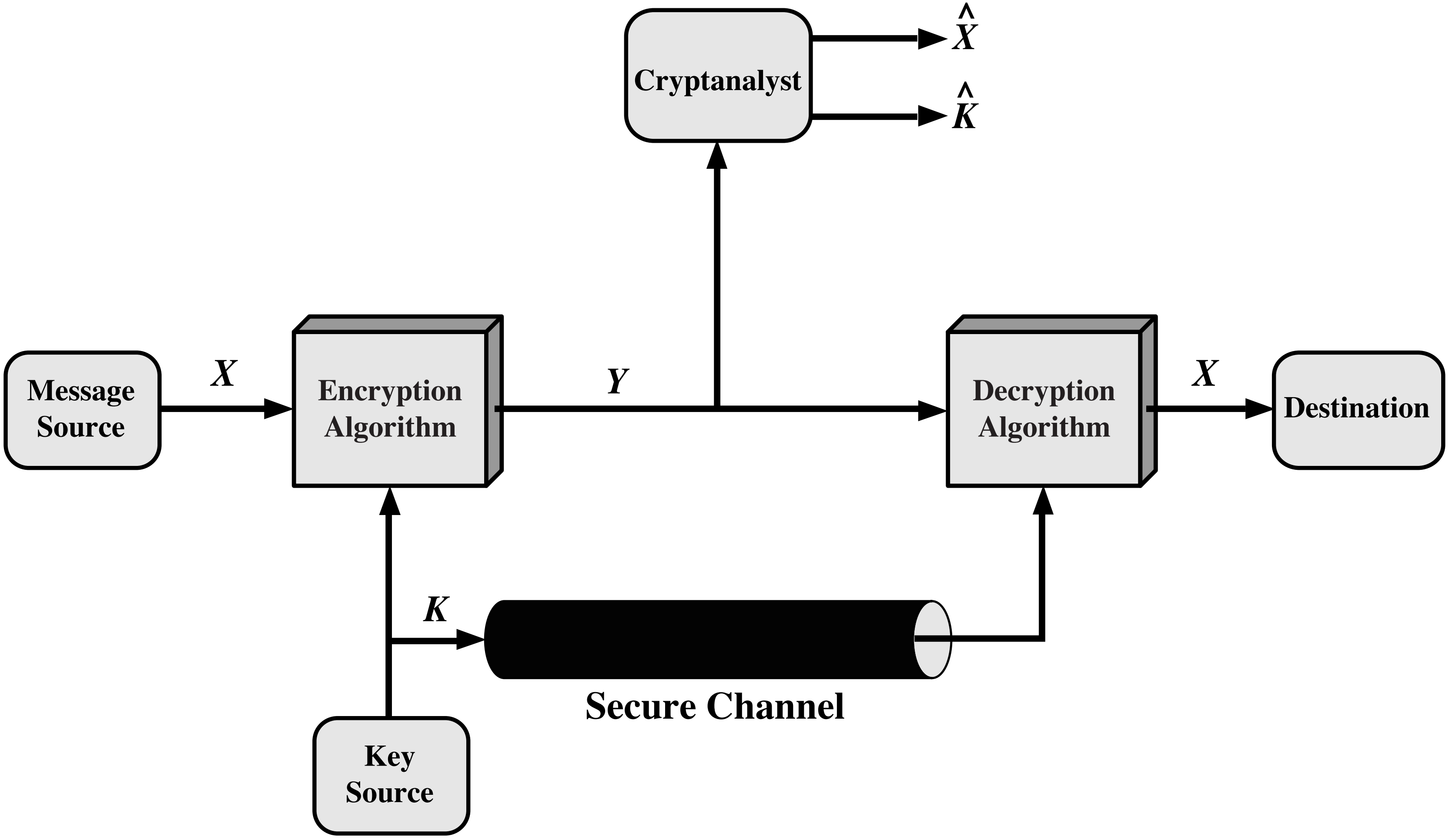
Symmetric Encryption in Essence

- Setting:
  - X: the plaintext
  - Y: the ciphertext
  - E: the encryption algorithm
  - D: the decryption algorithm
  - K: the secret key
- $Y = E(K, X)$ or $Y = E_K(X)$
- $X = D(K, Y)$ or $X = D_K(Y)$
- E_K and D_K are the inverse function of each other!

Security of Secret-Key Encryption

- Encryption algorithm must be strong enough: impossible to decrypt a message based on the ciphertext alone
- Depends on the secrecy of the key, not the secrecy of the algorithm
- Do not need to keep the algorithm secret; only need to keep the key secret
- Feasible for wide-spread use

Model of Conventional Cryptosystem



$$Y = E(K, X);\ X = D(K, Y)$$

Source: Figure 2.2, Stallings 2006

Dimensions of Cryptographic Systems

- The type of operations used for the security-related transformation:
  - substitution and/or
  - transposition (permutation)
- The number of keys used:
  - one key (symmetric encryption) or
  - two keys (asymmetric encryption)
- The way in which the plaintext is processed:
  - block cipher or
  - stream cipher

Cryptanalysis

Cryptanalysis is the process of attempting to discover plaintext or key or both.

- Ciphertext only: all that is available is the ciphertext.
  - the brute-force approach
  - statistical approaches (must first have some general idea about the type of plaintext)
- Known plaintext: feasible if certain plaintext patterns are known to appear in a message.
- Chosen plaintext: feasible if the analyst is able to insert chosen messages into the system.
- Chosen ciphertext
- Chosen text

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 •Purported 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 •Purported ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key

Source: Table 2.1, Stallings 2006

Strength of Encryption Schemes

- Unconditionally secure: unbreakable no matter how much ciphertext is available
- Computationally secure:
  - The cost exceeds the value of the encrypted information
  - The time required exceeds the useful lifetime of the information

Exhaustive Key Search

Key size (bits)	Number of alternative keys	Time required at 1 decryption/μs	Time required at 10^6 decryptions/μs
32	$2^{32} = 4.3 \times 10^9$	$2^{31}\ \mu s = 35.8$ minutes	2.15 milliseconds
56	$2^{56} = 7.2 \times 10^{16}$	$2^{55}\ \mu s = 1142$ years	10.01 hours
128	$2^{128} = 3.4 \times 10^{38}$	$2^{127}\ \mu s = 5.4 \times 10^{24}$ years	5.4×10^{18} years
168	$2^{168} = 3.7 \times 10^{50}$	$2^{167}\ \mu s = 5.9 \times 10^{36}$ years	5.9×10^{30} years
26 characters (permutation)	$26! = 4 \times 10^{26}$	$2 \times 10^{26}\ \mu s = 6.4 \times 10^{12}$ years	6.4×10^6 years

Source: Table 2.2, Stallings 2006

Substitution Techniques

A *substitution technique* is one in which the letters of plaintext are replaced by other letters or by numbers or symbols.

- Caesar Cipher
- Monoalphabetic Ciphers
- Playfair Cipher
- Hill Cipher
- Polyalphabetic Ciphers

The Caesar Cipher

- Each letter replaced with the letter standing three places further down the alphabet

plain:	`abcdefghijklmnopqrstuvwxyz`
cipher:	`DEFGHIJKLMNOPQRSTUVWXYZABC`
plain:	`meet me after the toga party`
cipher:	`PHHW PH DIWHU WKH WRJD SDUWB`

- The shift or key (which is 3) may be generalized to get General Caesar cipher:

$C = E_k(p) = (p + k) \bmod 26$, where $1 \leq k \leq 25$

Decryption: $p = D_k(C) = (C - k) \bmod 26$

Cryptanalysis of Caesar Cipher

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

Source: Figure 2.3, Stallings 2006

Breaking General Caesar Ciphers

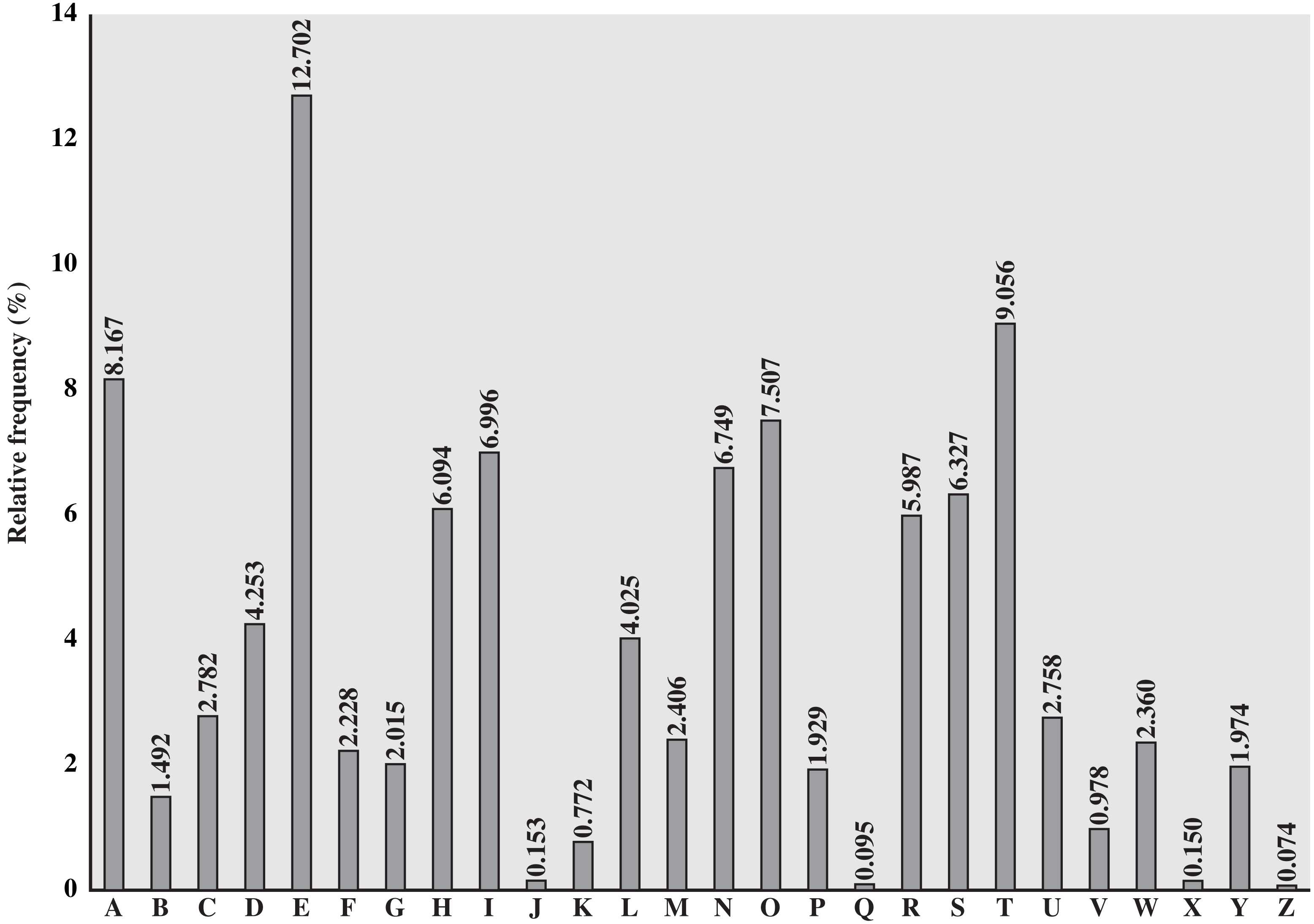
Three characteristics of general Caesar ciphers enable us to use a brute-force cryptanalysis:

- Encryption and decryption algorithms known
- Only 25 keys to try
- Language of the plaintext known and easily recognizable

Mono-alphabetic Ciphers

- Substitution represented by an arbitrary permutation of the alphabet
- $26!$ possible permutations (or keys) for English
- If language of the plaintext is known, regularities of the language may be exploited

Relative Frequency of English Letters



Source: Figure 2.5, Stallings 2006

Breaking a Mono-alphabetic Cipher

UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZ
VUEPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSX
EPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMOHMQ

1. Examine the relative frequency.

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								

Guess: P $\rightarrow$ e and Z $\rightarrow$ t (or the other way),
{S,U,O,M,H} $\rightarrow$ {r,n,i,o,a,s}, {A,B,G,Y,I,J} $\rightarrow$ {w,v,b,k,x,q,j,z}.

Breaking a Mono-alphabetic Cipher (cont.)

2. Look for other regularities, particularly the frequency of two-letter combinations (digrams).
   Guess: ZW → th, Z → t, P → e.
3. ZWSZ → th_t,
   Guess: S → a.

```
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
```

Improving Mono-alphabetic Ciphers

- Easy to break, because they reflect the frequency data of the original alphabet
- A countermeasure: provide multiple substitutes (homophones) for a single letter
- Still, multi-letter patterns survive in the ciphertext
- Two better approaches for improvement:
  - Encrypt multiple letters of plaintext: Playfair Cipher
  - Use multiple cipher alphabets: Hill Cipher

The Playfair Cipher

- Treats digrams in the plaintext as single units.
- Based on the use of a 5×5 matrix of letters constructed using a keyword.
- For example,

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

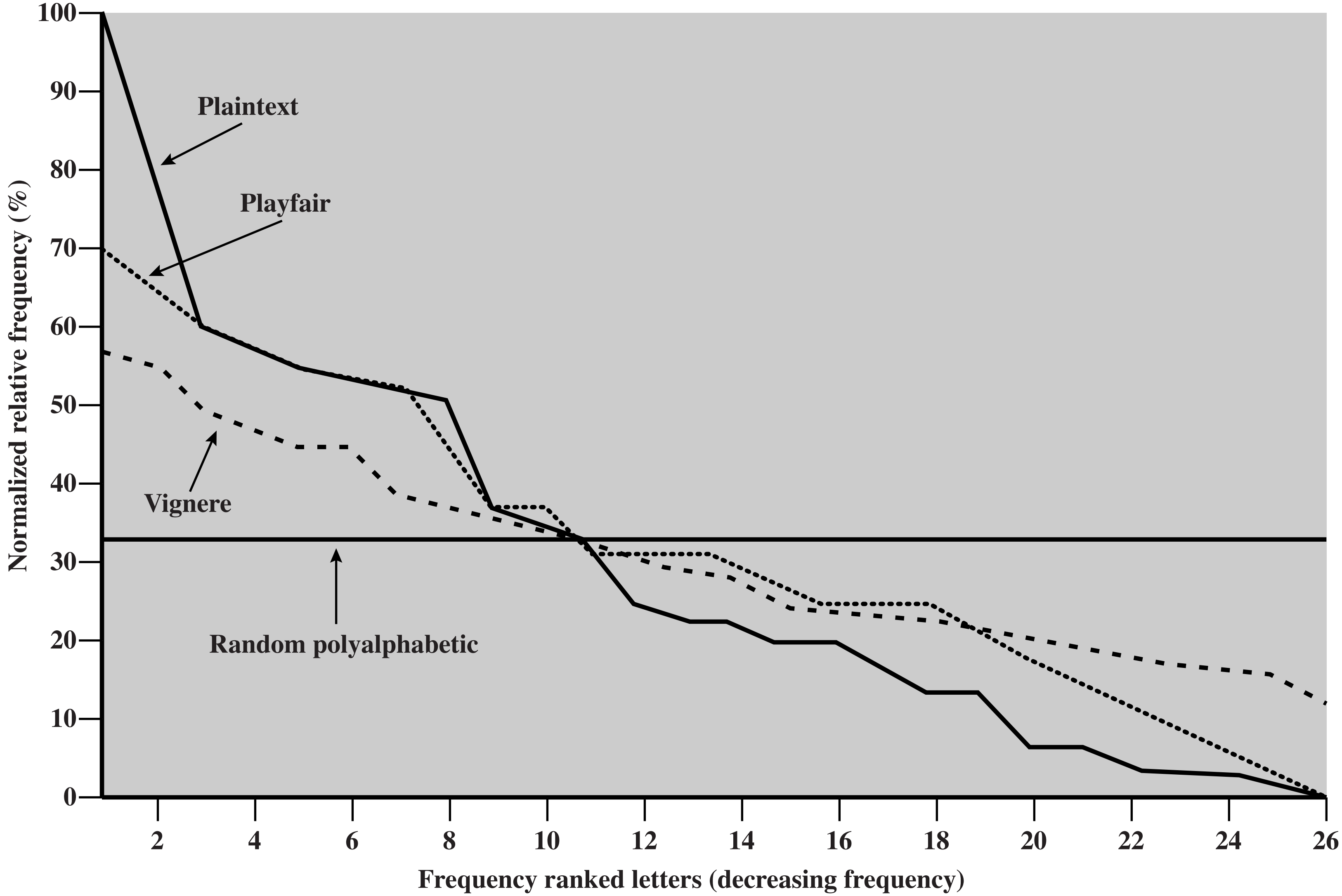
The Playfair Cipher (cont.)

Encryption rules by example:

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

1. balloon (the plaintext) → ba lx lo on (repeating letters in the same pair separated by filler x)
2. ON → NA (ON on the same row)
3. BA → IB (BA on the same column)
4. LX → SU, LO → PM

Relative Frequency of Letter Occurrences



Source: Figure 2.6, Stallings 2006

The Hill Cipher

- m (successive) plaintext letters $\longrightarrow$ m ciphertext letters
- Substitution determined by m linear equations, with $a = 0, b = 1, \ldots z = 25$
- For $m = 3$,

$$C_1 = (k_{11}p_1 + k_{12}p_2 + k_{13}p_3) \bmod 26$$
$$C_2 = (k_{21}p_1 + k_{22}p_2 + k_{23}p_3) \bmod 26$$
$$C_3 = (k_{31}p_1 + k_{32}p_2 + k_{33}p_3) \bmod 26$$

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \pmod{26}$$

The Hill Cipher (cont.)

- **P**,**C**: column vectors of length m, representing the plaintext and ciphertext
- **K**: invertible $m \times m$ matrix, representing the encryption key

$$\mathbf{C} = E_{\mathbf{K}}(\mathbf{P}) = \mathbf{KP}$$

$$\mathbf{P} = D_{\mathbf{K}}(\mathbf{C}) = \mathbf{K}^{-1}\mathbf{C} = \mathbf{K}^{-1}\mathbf{KP} = \mathbf{P}$$

- Strong against a ciphertext-only attacks, but easily broken with a known plaintext attack

Breaking the Hill Cipher

Given: $\mathbf{K}\begin{pmatrix} 5 \\ 17 \end{pmatrix} = \begin{pmatrix} 15 \\ 16 \end{pmatrix}$, $\mathbf{K}\begin{pmatrix} 8 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$,
$\mathbf{K}\begin{pmatrix} 0 \\ 24 \end{pmatrix} = \begin{pmatrix} 10 \\ 20 \end{pmatrix}$

From the first two pairs: $\mathbf{K}\begin{pmatrix} 5 & 8 \\ 17 & 3 \end{pmatrix} = \begin{pmatrix} 15 & 2 \\ 16 & 5 \end{pmatrix}$

Calculating the needed inverse: $\begin{pmatrix} 5 & 8 \\ 17 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} 9 & 2 \\ 1 & 15 \end{pmatrix}$

Breaking the Hill Cipher (cont.)

Calculating the key: $\mathbf{K} = \begin{pmatrix} 15 & 2 \\ 16 & 5 \end{pmatrix} \begin{pmatrix} 9 & 2 \\ 1 & 15 \end{pmatrix} = \begin{pmatrix} 7 & 8 \\ 19 & 3 \end{pmatrix}$

Checking the third pair: $\begin{pmatrix} 7 & 8 \\ 19 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 24 \end{pmatrix} = \begin{pmatrix} 10 \\ 20 \end{pmatrix}$

Calculating the Inverse of a Matrix

Let A be an invertible matrix (with a nonzero determinant). Its inverse A^{-1} can be computed as follows:

$$[A^{-1}]_{ij} = (-1)^{i+j} \times D_{ji} \times \det{}^{-1}(A)$$

where D_{ji} is the subdeterminant obtained by deleting the j-th row and the i-th column of A.

$$\det{}^{-1}\begin{pmatrix} 5 & 8 \\ 17 & 3 \end{pmatrix} = (-121)^{-1} = 9^{-1} = 3 \pmod{26}$$

$$\begin{pmatrix} 5 & 8 \\ 17 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} 3 \times 3 & -8 \times 3 \\ -17 \times 3 & 5 \times 3 \end{pmatrix} = \begin{pmatrix} 9 & 2 \\ 1 & 15 \end{pmatrix} \pmod{26}$$

Poly-alphabetic Ciphers

- To improve on simple monoalphabetic ciphers, juggle different monoalphabetic substitutions
- This is called *polyalphabetic* cipher
- Common features:
  - A set of related monoalphabetic substitution rules
  - A key determines which particular rule is chosen

The Vigenère Cipher

- Best-known polyalphabetic cipher
- Monoalphabetic substitution rules consist of the 26 general Caesar ciphers
- Each cipher is denoted by a key letter, which is the ciphertext letter that substitutes for letter ‘a’

key:	`deceptivedeceptivedeceptive`
plain:	`wearediscoveredsaveyourself`
cipher:	`ZICVTWQNGRZGVTWAVZHCQYGLMGJ`

- Multiple ciphertext letters for each plaintext letter

The Modern Vigenère Tableau

Key \ 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
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

Source: Table 2.3, Stallings 2006

The Vernam Cipher

- The encryption scheme is expressed as

$$C_i = p_i \oplus k_i$$

where p_i = i-th binary digit of plaintext,
k_i = i-th binary digit of key, and
C_i = i-th binary digit of ciphertext

- The one-time pad scheme uses a random key for the Vernam cipher; in principle, unbreakable
- Rarely used due to key management problems

One-Time Pad Is Unbreakable

Assume a 27×27 Vigenère substitution cipher.

cipher: `ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS`

key: *pxlmvmsydofuyrvzwc tnlebnecvgdupahfzzlmnyih*

plain: `mr mustard with the candlestick in the hall`

cipher: `ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS`

key: *mfugpmiydgaxgoufhklllmhsqdqogtewbqfgyovuhwt*

plain: `miss scarlet with the knife in the library`

Cannot conclude one of the two keys is more likely than the other.

Transposition Techniques

Transposition ciphers perform some sort of permutation on the plaintext letters.

- The rail fence technique
- Columnar transpositions
- Multiple-stage transpositions

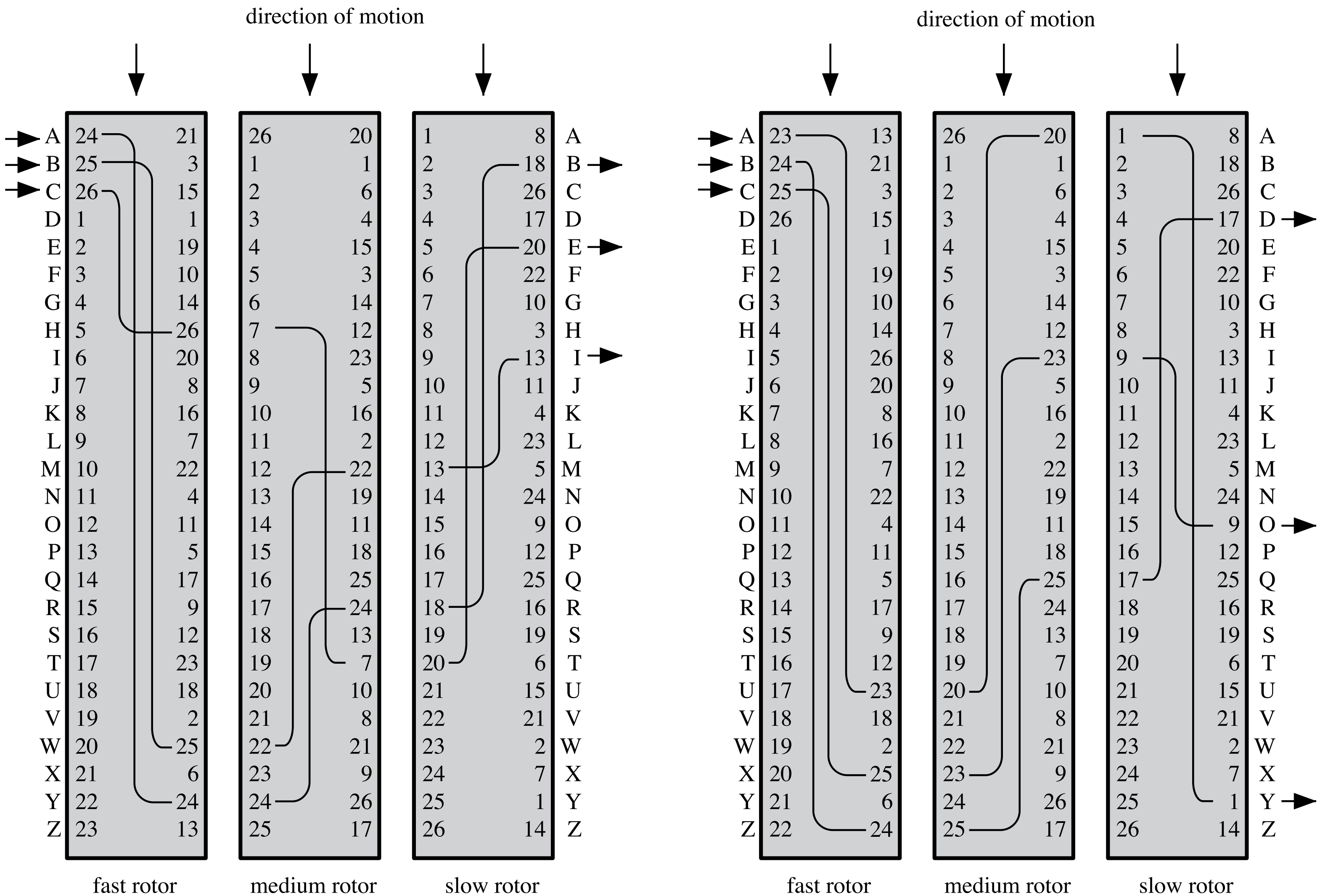
Columnar Transpositions

- Write the message in a rectanlge, row by row, and read the message off, column by column, but permute the order of the columns
- For example,

key:	4	3	1	2	5	6	7
plain:	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

cipher: TTNAAPTMTSUOAODWCOIXKNLYPETZ

A Three-Rotor Machine



(a) Initial setting **(b) Setting after one keystroke**

Source: Figure 2.7, Stallings 2006

Rotor Machines

- A rotor machine consists of a set of cylinders that rotate like an odometer.
- A cylinder has 26 input pins, each connecting to a unique output pin.
- A rotating cylinder defines a poly-alphabetic substitution algorithm with a period of 26.
- A three-rotor machine has a period of $26 \times 26 \times 26 = 17,576$; four-rotor $456,976$; five-rotor $11,881,376$.

Steganography

The methods of steganography conceal the existence of the message (whereas the methods of cryptography render the message unintelligible to outsiders).

- Character marking
- Invisible ink
- Pin punctures
- Typewriter correction ribbon

A Puzzle

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,

Source: Figure 2.8, Stallings 2006