Chapter 2

Classical Encryption Techniques

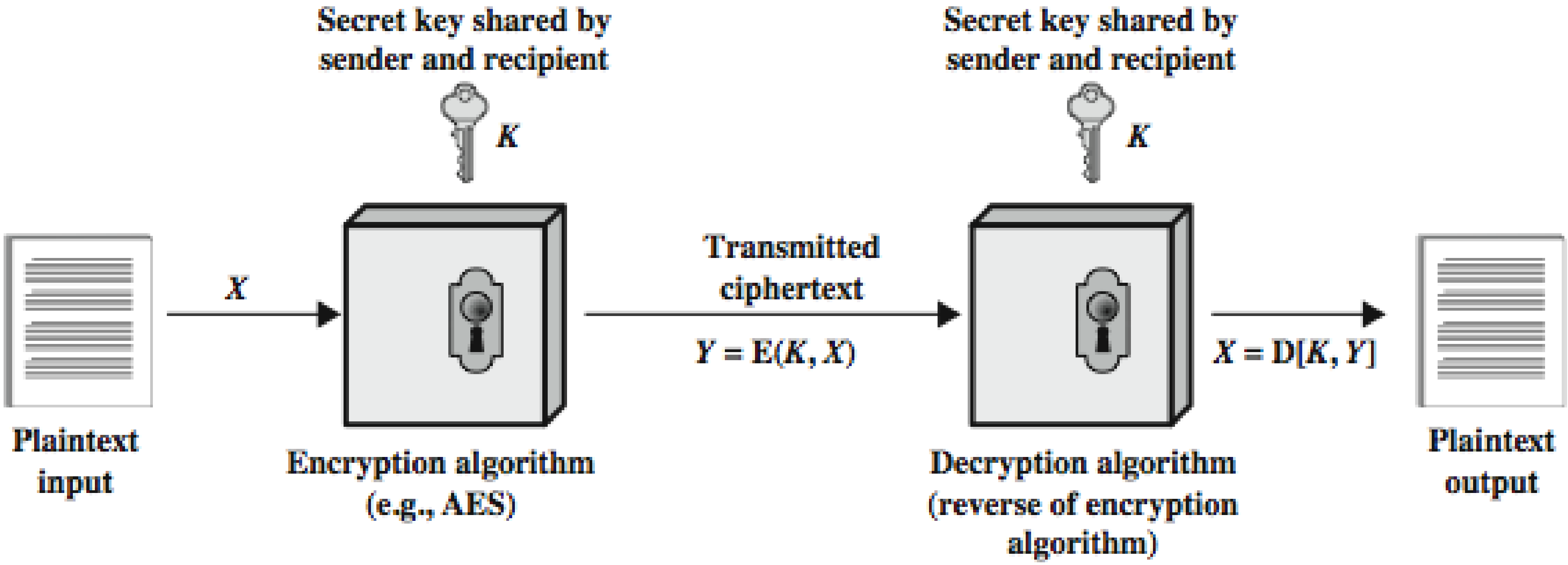
Dr. Shin-Ming Cheng

NTUST CONNECTIVITY LAB

Symmetric Encryption

- conventional/private-key/single-key
- sender and recipient share a common key
- all classical encryption algorithms are private-key
- was only type prior to invention of public-key in 1970's and by far most widely used

Symmetric Cipher Model



Some Basic Terminology

- **Plaintext**: original message
- **Ciphertext**: coded message
- **Cipher**: algorithm for transforming plaintext to ciphertext
- **Key**: Infor used in cipher known only to sender/receiver
- **encipher (encrypt)**: converting plaintext to ciphertext
- **decipher (decrypt)**: recovering ciphertext from plaintext

Some Basic Terminology

- **Cryptography:** study of encryption principles/methods
- **Cryptanalysis (codebreaking):** study of principles/methods of deciphering ciphertext *without* knowing key
- **Cryptology:** field of both cryptography and cryptanalysis

Requirements

- two requirements for secure use of symmetric encryption:
  - a strong encryption algorithm
  - a secret key known only to sender / receiver
- mathematically have:

$$Y = \underline{\mathrm{E}}(\mathrm{K}, X)$$

$$X = \underline{\mathrm{D}}(\mathrm{K}, Y)$$

- assume encryption algorithm is known
- implies a secure channel to distribute key

Cryptography 密碼學

- can characterize cryptographic system by:
  - type of encryption operations used
    - substitution
    - transposition
    - product
  - number of keys used
    - single-key or private
    - two-key or public
  - way in which plaintext is processed
    - block
    - stream

Cryptanalysis 密碼分析

- objective to recover key not just message
- cryptanalytic attack
- brute-force attack
- All future and past messages encrypted with that key are compromised.

Cryptanalysis

- cryptanalytic attack
  - relies on the nature of the algorithm plus perhaps some knowledge of the general characteristics of the plaintext or even some sample plaintext-ciphertext pairs
  - exploits the characteristics of the algorithm to attempt to deduce a specific plaintext or to deduce the key being used.
- brute-force attack
  - try 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.

Cryptanalytic Attacks

- ciphertext only
  - only know algorithm & ciphertext, is statistical, know or can identify plaintext
- known plaintext
  - know/suspect plaintext & ciphertext
- chosen plaintext
  - select plaintext and obtain ciphertext
- chosen ciphertext
  - select ciphertext and obtain plaintext
- chosen text
  - select plaintext or ciphertext to en/decrypt

More Definitions

- Unconditional security
  - no matter how much computer power or time is available, the cipher cannot be broken
    - the ciphertext provides insufficient information to uniquely determine the corresponding plaintext
- Computational security
  - given limited computing resources (e.g., time needed for calculations is greater than age of universe), the cipher cannot be broken

Brute Force Search

- always possible to simply try every key
- most basic attack, proportional to key size
- assume either <u>know/recognise plaintext</u>

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} µs = 35.8 minutes	2.15 milliseconds
56	$2^{56} = 7.2 \times 10^{16}$	2^{55} µs = 1142 years	10.01 hours
128	$2^{128} = 3.4 \times 10^{38}$	2^{127} µs = 5.4×10^{24} years	5.4×10^{18} years
168	$2^{168} = 3.7 \times 10^{50}$	2^{167} µs = 5.9×10^{36} years	5.9×10^{30} years
26 characters (permutation)	$26! = 4 \times 10^{26}$	2×10^{26} µs = 6.4×10^{12} years	6.4×10^6 years

Symmetric key Cryptography

- Classical Cryptography
  - Shift Cipher
  - Substitution Cipher
  - Vigenère Cipher
- Modern Cryptography
  - Data Encryption Standard (DES)
  - Triple DES (3DES)
  - Advanced Encryption Standard (AES)

Classical Substitution Ciphers

- where letters of plaintext are replaced by other letters or by numbers or symbols
- or if plaintext is viewed as a sequence of bits, then substitution involves replacing plaintext bit patterns with ciphertext bit patterns

Caesar Cipher: Shift Cipher

- earliest known substitution cipher
- by Julius Caesar (100 BC – 44 BC)
- first attested use in military affairs
- replaces each letter by 3rd letter on
- example:

```
meet me after the toga party
PHHW PH DIWHU WKH WRJD SDUWB
```

Caesar Cipher

› 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

› then have Caesar cipher as:

$c = \mathrm{E}(\mathrm{k}, p) = (p + k) \bmod (26)$

$p = \mathrm{D}(\mathrm{k}, \mathrm{c}) = (\mathrm{c} - k) \bmod (26)$

Cryptanalysis of Caesar Cipher

- only have 26 possible ciphers
  - `A` maps to `A,B,..Z`
- could simply try each in turn
- a **brute force search**
- given ciphertext, just try all shifts of letters
- do need to recognize when have plaintext
- eg. break ciphertext "`GCUA VQ DTGCM`"

Shift Cipher

› English Letter Frequencies

› The most common bigrams are in decreasing order
  – TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA

› The most common trigrams are in decreasing order
  – THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR

› Useful cryptanalysis:
  – Conditional probability – e.g., the letter with the highest frequency right after the letter `H`?

Shift Cipher

› Take the following example cipher text

```
BPMZM WVKM EIA IV COTG LCKSTQVO
EQBP NMIBPMZA ITT ABCJJG IVL JZWEV
IVL BPM WBPMZ JQZLA AIQL QV AW  UIVG EWZLA
OMB WCB WIN BWEV
OMB WCB, OMB WCB , OMB WCB  WIN BWEV
IVL PM EMVB EQBP I YCIKS IVL I EILLTM IVL I YCIKS
QV I NTCZZG WN MQL MZLWEV
BPIB XWWZ TQBBTM COTG LCKS TQVO
EMVB EIVLMZQVO NIZ IVL VMIZ
JCB IB MDMZG XTIKM BPMG AIQL BW PQA NIKM
VWE OMB WCB, OMB WCB, OMB WCB WN PMZM
IVL PM EMVB EQBP I YCIKS IVL I EILLTM IVL I YCIKS
IVL I DMZG CVPIXXG BMIZ
```

– We need to compare the frequency distribution of this text with standard English

Shift Cipher

› Underlying Plain Text

› Cipher Text

– The shift of `E` seems to be either 4, 8,17,18 or 23
– The shift of `A` seems to be either 1, 8,12,21 or 22

Shift Cipher

- Hence the key is probably equal to 8
- We can now decrypt the cipher text to reveal

```
There once was an ugly duckling
With feathers all stubby and brown
And the other birds said in so many words
Get out of town
Get out, get out, get out of town
And he went with a quack and a waddle and a quack
In a flurry of eiderdown
That poor little ugly duckling
Went wandering far and near
But at every place they said to his face
Now get out, get out, get out of here
And he went with a quack and a waddle and a quack
And a very unhappy tear
```

Monoalphabetic Cipher

- rather than just shifting the alphabet
- could shuffle (jumble) the letters arbitrarily
- each plaintext letter maps to a different random ciphertext letter
- hence key is 26 letters long

```
Plain:  abcdefghijklmnopqrstuvwxyz
Cipher: DKVQFIBJWPESCXHTMYAUOLRGZN

Plaintext:  ifwewishtoreplaceletters
Ciphertext: WIRFRWAJUHYFTSDVFSFUUFYA
```

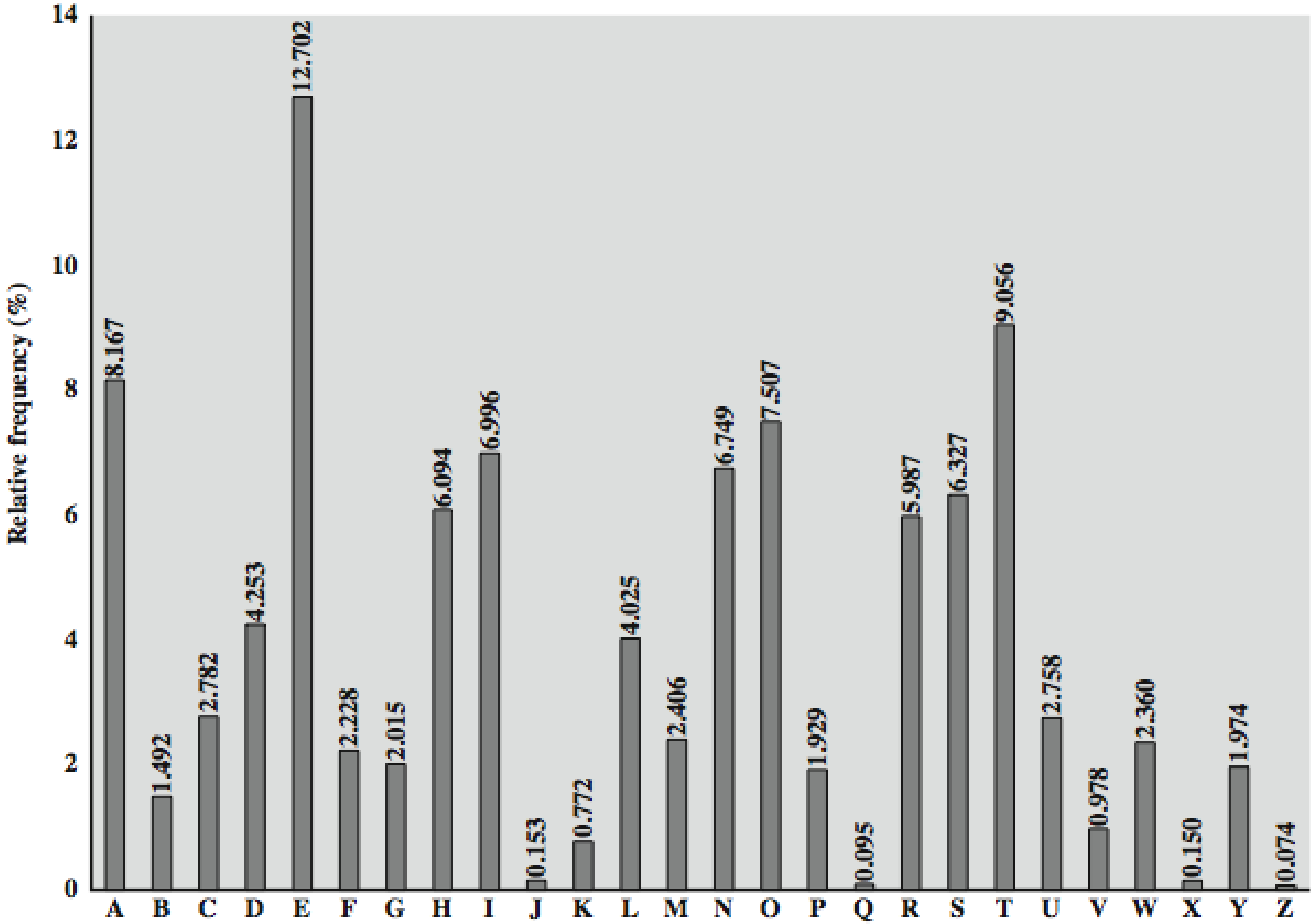
Monoalphabetic Cipher Security

- now have a total of $26! \approx 4.03 \cdot 10^{26} \approx 2^{88}$ keys
- This is far too large a number to brute force search using modern computers
- with so many keys, might think is secure?
- problem is language characteristics

Language Redundancy and Cryptanalysis

- Human languages are **redundant**
  - "th lrd s m shphrd shll nt wnt" letters are not equally commonly used
- E is by far the most common letter
  - followed by T,R,N,I,O,A,S
- Z,J,K,Q,X are fairly rare
- have tables of single, double & triple letter frequencies for various languages

English Letter Frequencies



Use in Cryptanalysis

- monoalphabetic substitution ciphers do not change relative letter frequencies
  - discovered by Arabian scientists in 9^{th} century
- calculate letter frequencies for ciphertext
- compare counts/plots against known values
- identify each letter tables of common double/triple letters help

Example Cryptanalysis

› ciphertext:

UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZ
VUEPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSX
EPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMOHMQ

› count relative letter frequencies (see text)

- guess P and Z are e and t
- guess ZW is th and hence ZWP is the

› proceeding with trial and error finally get:

it was disclosed yesterday that several informal
but direct contacts have been made with political
representatives of the viet cong in moscow

Playfair Cipher

- Not even the large number of keys in a monoalphabetic cipher provides security
- Playfair Cipher
  - improves security by encrypting multiple letters
  - invented by Charles Wheatstone in 1854, but named after his friend Baron Playfair

Playfair Key Matrix

- A 5X5 matrix of letters based on a keyword
  - fill in letters of keyword (sans duplicates)
  - fill rest of matrix with other letters
- 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

Encrypting and Decrypting

- Plaintext is encrypted two letters at a time
  - if a pair is a repeated letter, insert filler like 'X'
  - if both letters fall in the same row, replace each with letter to right (wrapping back to start from end)
  - if both letters fall in the same column, replace each with the letter below it (wrapping to top from bottom)
  - otherwise each letter is replaced by the letter in the same row and in the column of the other letter of the pair

Security of Playfair Cipher

- Security much improved over monoalphabetic
  - have 26 x 26 = 676 digrams
  - Needs a 676 entry frequency table to analyse (verses 26 for a monoalphabetic)
  - Needs correspondingly more ciphertext
- Widely used for many years
  - eg. by US & British military in WW1
- it **can** be broken
  - given a few hundred letters
  - still has much of plaintext structure

Vigenère Cipher

› The problem with the Caesar and Substitution cipher: Each plaintext letter is always encrypted to the same ciphertext letter
  – Hence underlying statistics of the language could be used to break the cipher

› From the early 1800's onwards cipher designers tried to break this link between the plain and cipher texts

› The most famous cipher from the 1800's is the Vigenère cipher
  – Believed to be unbreakable for a number of years

Polyalphabetic Ciphers

- improves security using multiple cipher alphabets
  - make cryptanalysis harder with more alphabets to guess and flatter frequency distribution
- use a key to select which alphabet is used for each letter of the message
- use each alphabet in turn
- repeat from start after end of key is reached

Vigenère Cipher

- Vigenère cipher again identifies letter with 0, ..., 25
- The secret key is a short sequence of letters
  - e.g. a word
- Encryption adds the plaintext letter to a key letter
  - with the key letters used in rotation
- If the key is SESAME, encryption works as follows,

T H I S I S A T E S T M E S S A G E	Message
S E S A M E S E S A M E S E S A M E	Keystream
L L A S U W S X W S F Q W W K A S I	Ciphertext

Vigenère Table

	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 Table

	H	E	L	L	O	message
	7(H)	4 (E)	11(L)	11(L)	14(O)	message
+	23(X)	12(M)	2 (C)	10(K)	11(L)	key
=	30	16	13	21	25	message + key
=	4(E)	16(Q)	13(N)	21(V)	25(Z)	message + key (mod 26)
	E	Q	N	V	Z	→ ciphertext
	4(E)	16(Q)	13(N)	21(V)	25(Z)	ciphertext
-	23(X)	12(M)	2 (C)	10(K)	11(L)	key
=	-19	4	11	11	14	ciphertext - key
=	7 (H)	4 (E)	11(L)	11(L)	14(O)	ciphertext - key (mod 26)
	H	E	L	L	O	→ message

Security of Vigenère Ciphers

- have multiple ciphertext letters for each plaintext letter
  - letter frequencies are obscured
  - not all knowledge of the plaintext structure is lost
  - The breaking process depends on determining the length of the keywords

Autokey Cipher

- enfource a key as long as the message
  - Vigenère proposed the **autokey** cipher
  - keyword is prefixed to message as key
- Decryption
  - knowing keyword can recover the first few letters
  - use these in turn on the rest of the message
- have frequency characteristics to attack
- given key deceptive

```
key:        deceptivewearediscoveredsav
plaintext:  wearediscoveredsaveyourself
ciphertext: ZICVTWQNGKZEIIGASXSTSLVVWLA
```

Vernam Cipher

- ultimate defense against cryptanalysis
  - use a key as long as the plaintext
  - with no statistical relationship to it
- invented by AT&T engineer Gilbert Vernam in 1918
  - originally proposed using a very long but eventually repeating key

Vernam Cipher

	H	E	L	O	message
	7(00111)(H)	4(00100)(E)	11(01011)(L)	14(01110)(O)	message
(Xor)	23(10111)(X)	12(01100)(M)	2(00010)(C)	11(01011)(L)	key
=	16(10000)	8(01000)	9(01001)	5(00101)	
=	16(Q)	8(I)	9(J)	5(F)	
	Q	I	J	F	→ ciphertext
	16(10000)(Q)	8(01000)(I)	9(01001)(J)	5(00101)(F)	ciphertext
(Xor)	23(10111)(X)	12(01100)(M)	2(00010)(C)	11(01011)(L)	key
=	7(00111)	4(00100)	11(01011)	14(01110)	
=	7(H)	4(E)	11(L)	14(O)	
	H	E	L	O	→ message

One-Time Pad

- using a random key that was truly as long as the message, with no repetitions
  - unbreakable
    - since ciphertext bears no statistical relationship to the plaintext
- only use the key **once**
- fundamental difficulties
  - making large quantities of random keys
  - The problem of key distribution and protection
    - where for every message to be sent, a key of equal length is needed by both sender and receiver.

Transposition Techniques

- classical **transposition** or **permutation** ciphers
  - hide the message by rearranging the letter order
  - without altering the actual letters used
- can recognise these
  - have the same frequency distribution as the original text

Rail Fence Cipher

- Write message letters out diagonally over a number of rows
- read off cipher row by row
- write message out as

```
m   e   m   a   t   r   h   t   g   p   r   y
  e   t   e   f   e   t   e   o   a   a   t
```

- giving ciphertext

  MEMATRHTGPRYETEFETEOAAT

Row Transposition Ciphers

- complex transposition
- write letters of message out in rows over a specified number of columns
- reorder the columns according to some key before reading off the rows

Key: 4312567

Column Out:	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

Product Ciphers

- substitutions or transpositions are not secure
  - language characteristics
- using <u>several ciphers in succession</u> to make harder
  - two substitutions make a more complex substitution
  - two transpositions make more complex transposition
  - but a substitution followed by a transposition makes a new much harder cipher
- this is bridge from classical to modern ciphers

Product Ciphers

› Key Transport:

01	02	03	04	05	06	07	08	09	10
11	12	13	14	15	16	17	18	19	20

15	11	19	18	16	03	07	14	02	20
04	12	09	06	01	05	17	13	10	08

› Plantext : `thestrengthofthispig`

t	h	e	s	t	r	e	n	g	t
h	o	f	t	h	i	s	p	i	g

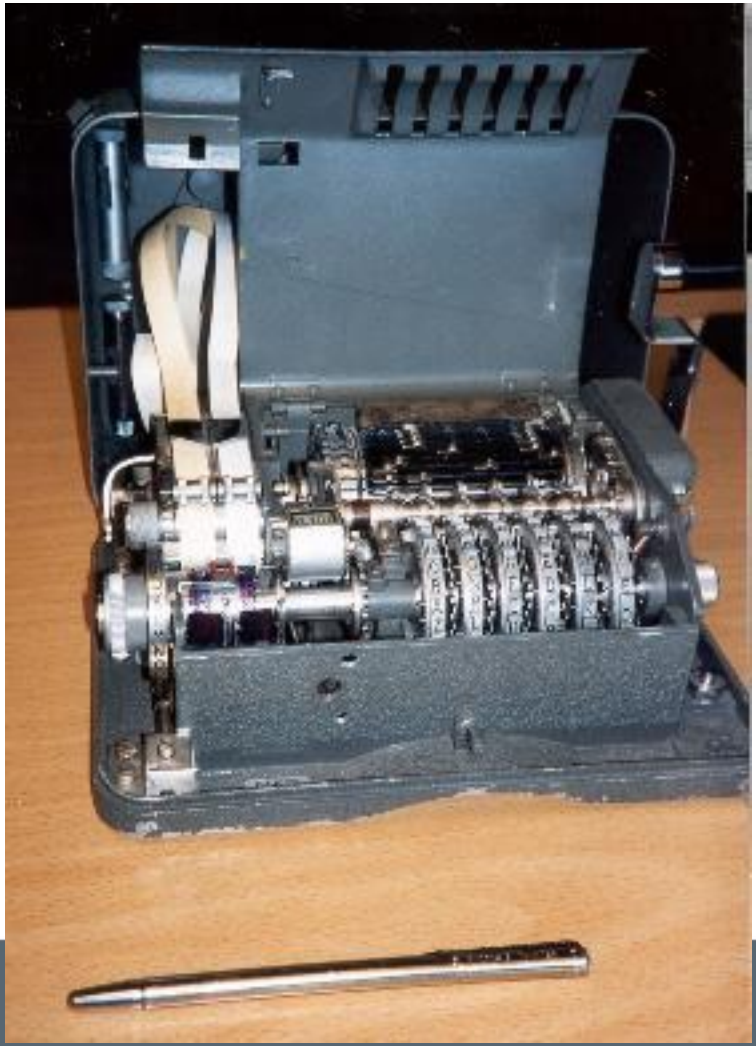
H	H	I	P	I	E	E	T	H	G
S	O	G	R	T	T	S	F	T	N

› Ciphertext : `HHIPIEETHGSOGRTTSFTN`

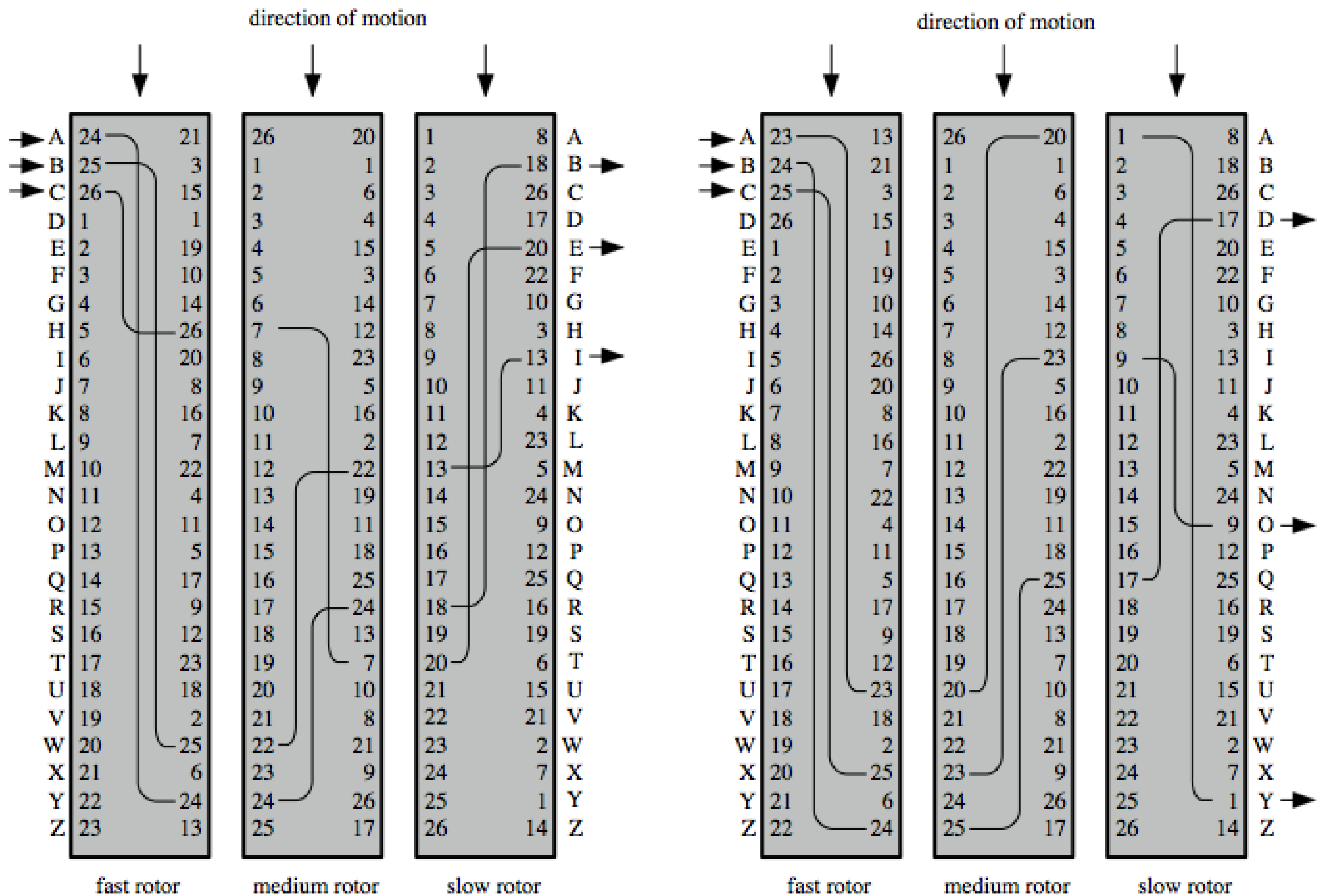
Rotor Machines

- Before modern ciphers, rotor machines were most common complex ciphers in use
- Widely used in WW2
  - German Enigma, Allied Hagelin, Japanese Purple
- Implemented a very complex, varying substitution cipher
- series of cylinders
  - each giving one substitution,
  - rotates and changes after each letter was encrypted
- with 3 cylinders have $26^3=17576$ alphabets

Hagelin Rotor Machine



Rotor Machine Principles



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

Steganography

- an alternative to encryption
- hides existence of message
  - using only a subset of letters/words in a longer message marked in some way
  - using invisible ink
  - hiding in LSB in graphic image or sound file
- drawbacks
  - high overhead to hide relatively few info bits
- advantage
  - obscure encryption use

Summary

- classical cipher techniques and terminology
- monoalphabetic substitution ciphers
- cryptanalysis using letter frequencies
- Playfair cipher
- polyalphabetic ciphers
- transposition ciphers
- product ciphers and rotor machines
- steganography