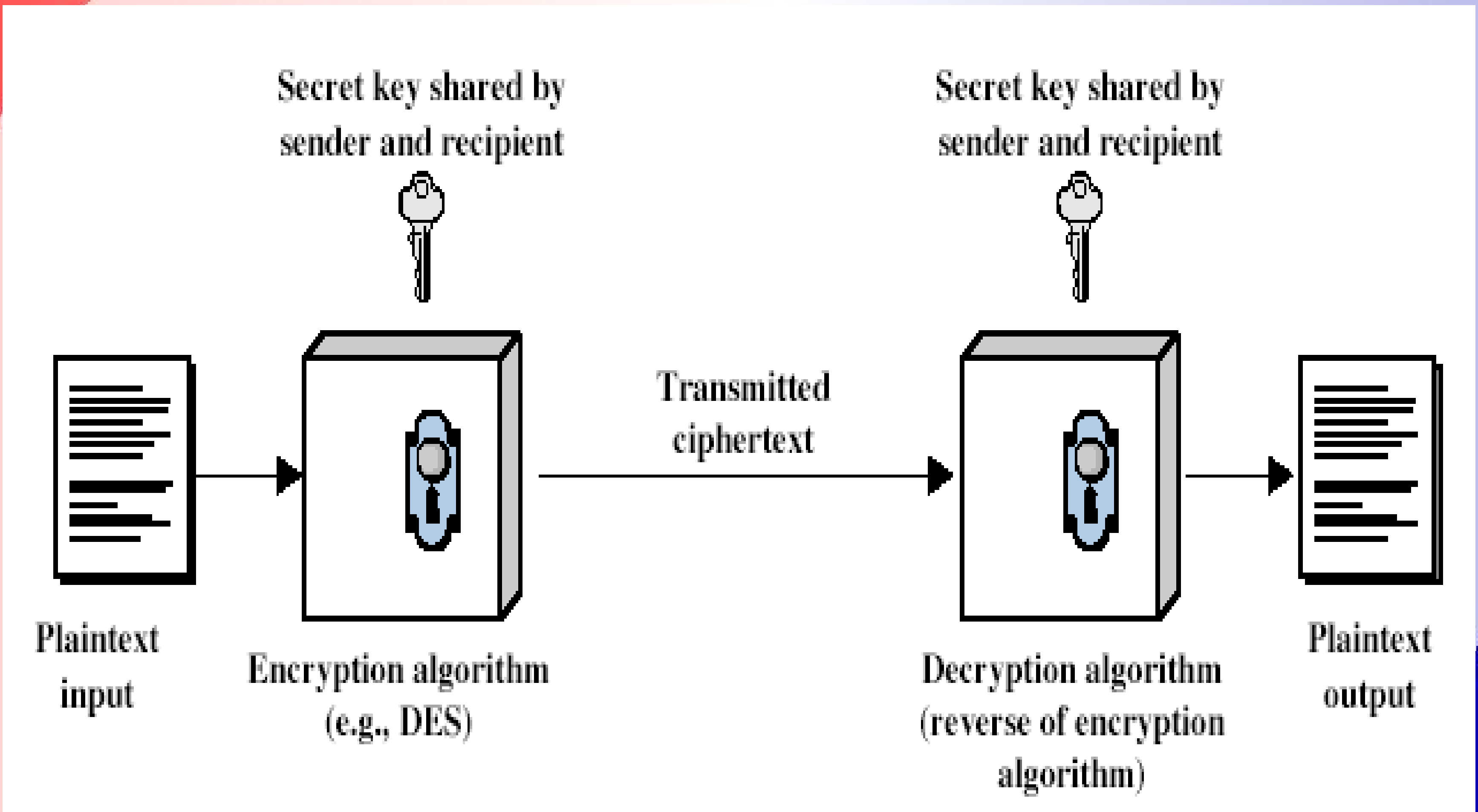
Information Security: an Introduction

Basic Terms in Security

- **plaintext** - the original message
- **ciphertext** - the coded message
- **key** - info used in cipher known only to sender/receiver
- **encipher (encrypt)** - converting plaintext to ciphertext
- **decipher (decrypt)** - recovering ciphertext from plaintext
- **cryptography** - study of encryption principles/methods
- **cryptanalysis (codebreaking)** - the study of principles/ methods of deciphering ciphertext *without* knowing key
- **cryptology** - the field of both cryptography and cryptanalysis

Symmetric Cipher Model



Symmetric-Key Algorithms

- **DES** – The Data Encryption Standard
- **AES** – The Advanced Encryption Standard
- **Cipher Modes**
  - Stream Ciphers
  - Block Ciphers

Requirements

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

    $Y = E_K(X)$

    $X = D_K(Y)$
- Assume encryption algorithm is known
- Implies a secure channel to distribute key

Classical Encryption Techniques

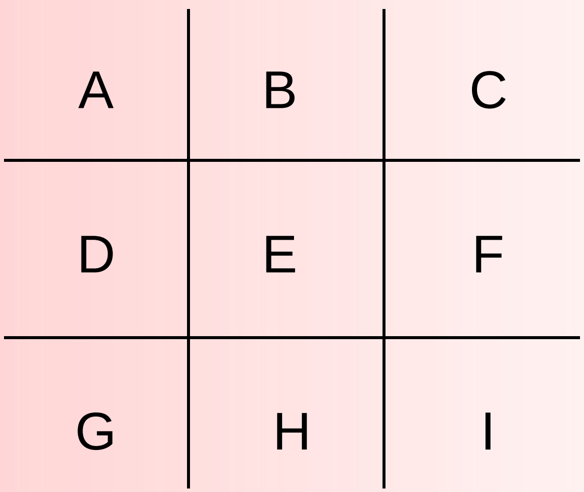
- **Substitution Ciphers**
- **Transposition Ciphers**

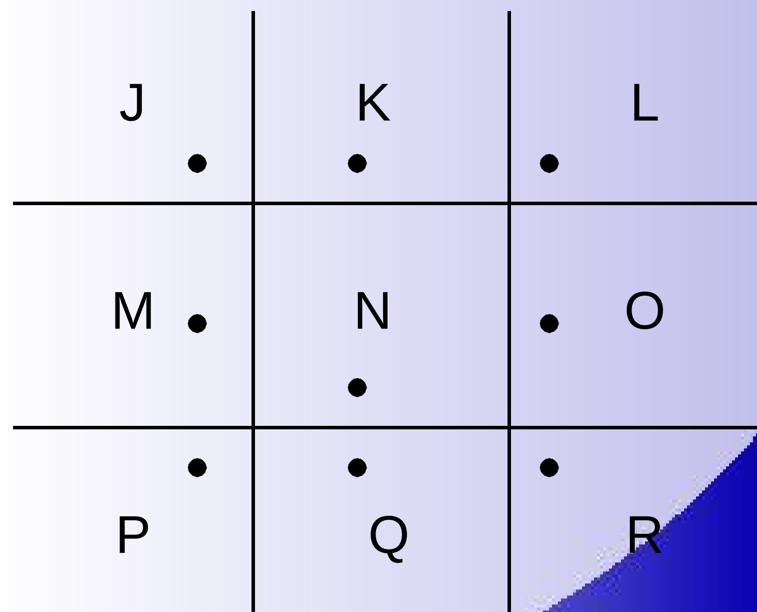
Substitution Ciphers

- 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

Pigpen Cipher

- Pigpen cipher is a variation on letter substitution
- Alphabets are arranged as follows:

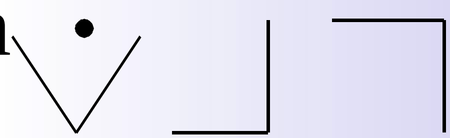




Pigpen Cipher diagram (cont’d)

Pigpen Cipher

- Alphabets will be represented by the corresponding diagram
- E.g., WAG would be



- This is a weak cipher

ADFGVX Cipher

- This is a variation on substitution cipher and is a strong cipher

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

ADFGVX Cipher

- Rules:
  - – Remove spaces and punctuation marks from message
  - – For each letter or number substitute the letter pair from the column and row heading
  - – Next, use a transposition operation on the pair of letters using a key word (which the receiver knows)
  - – Rearrange the columns of the new arrangement in alphabetical order
  - – Finally, arrange the letters from consecutive columns

ADFGVX Cipher

- E.g., Message = SEE ME IN MALL
  - SEEMEINMALL
  - VDXDXDGXXDVGAXGXDVDADA
  - Use keyword of INFOSEC
  - Arrange the stage 1 ciphertext characters in a fresh grid with keyword as the column heading
  - Ciphertext is written in column order from left to right

ADFGVX Cipher

I	N	F	O	S	E	C
V	D	X	D	X	D	G
X	X	D	V	G	A	X
G	X	D	V	D	A	V

ADFGVX Cipher

C	E	F	I	N	O	S
G	D	X	V	D	D	X
X	A	D	X	X	V	G
V	A	D	G	X	V	D

ADFGVX Cipher

- Ciphertext is:
  GXVDAAXDDVXGDXXDVVXGD
- Recipient reverses the process using the same keyword and gets the plaintext
- Reason for this cipher using the name ADFGVX is that in Morse code these characters all have dissimilar patterns of dots and dashes

Caesar Cipher

- Earliest known substitution cipher by Julius Caesar
- first attested use in military affairs
- replaces each letter by 3rd letter after it

example:

```
meet me after the college
PHHW PH DIWHU WKH froohjh
```

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
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
```

- Then have Caesar cipher as:

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

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

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 "sdd qgmj tskw sjw twdgfy lg mk"

Affine Cipher

- Each letter is assigned a number. “a” = 0, “b” = 1, “c” = 2, …
- The key to an affine cipher is a pair of numbers (a, b).
- The greatest common divisor (GCD) of a and 26 must be 1.
- Let p be the number of the plaintext letter and c the number of the ciphertext letter.
- $c = (a\,p + b) \bmod 26$
- $p = (a^{-1}(c - b)) \bmod 26$

Note: $a\,a^{-1} = 1 \pmod{n}$

Affine: example

- $a = 3$, $b = 7$
- Find the equations for encryption and decryption.
- Encrypt the message “the dog”
- Decrypt the message “TIVUJWL”

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! =~ 4 x 10^26 keys
- so many keys! must be secure against ciphertext cryptanalysis!

Language Redundancy and Cryptanalysis

- human languages are **redundant**
- letters are not equally commonly used
- in English **e** is by far the most common letter
- then T,R,N,I,O,A,S
- other letters are fairly rare
- cf. Z,J,K,Q,X
- have tables of single, double & triple letter frequencies

Cryptanalysis of Monoalphabetic Ciphers

- Use the technique to break the Caesar cipher to break monoalphabetic cipher
- Guess-> substantiate -> correct or contradiction
- Frequency Distributions: in English, some letters are used more frequently than others.
- E, T, and A occur far more frequency than J, Q, and Z for

An Example

ENCRYPTION IS A MEANS OF ATTAINING SECURE COMPUTATION
OVER INSECURE CHANNELS
BY USING ENCRYPTION WE DISGUISE THE MESSAGE SO THAT
EVEN IF THE TRANSMISSION IS DIVERTED
THE MESSAGE WILL NOT BE REVEALED

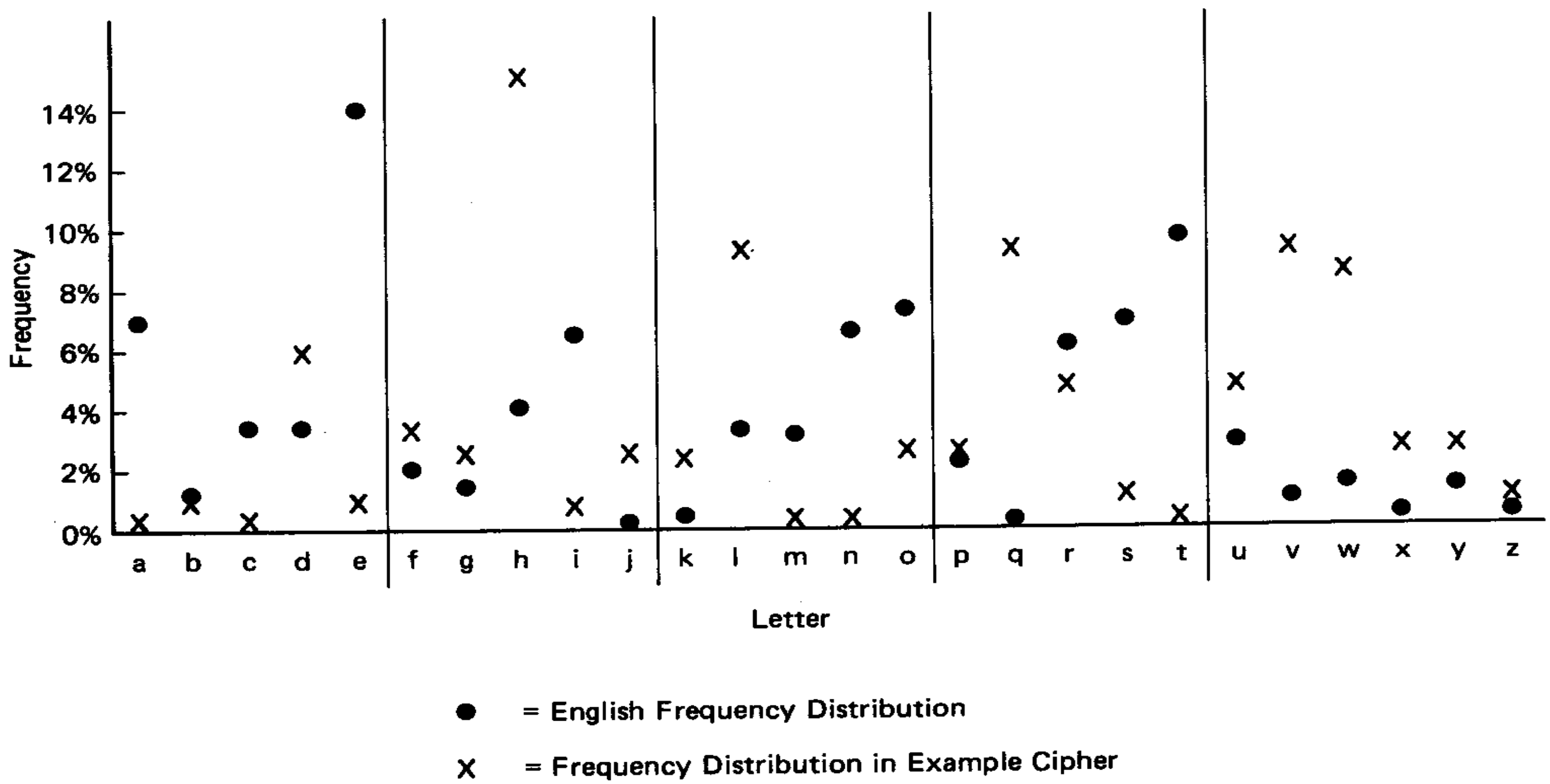
hqfubswlrq lv d phdqv ri dwwdlqlqj vhfxuh frpsxwdwlrq
ryhu lqvhfxuh fkdqqhov
eb xvlqj hqfubswlrq zh glvjxlvh wkh phvvdjh vr wkdw
hyhq li wkh wudqvplvvlrq lv glyhuwhg
wkh phvvdjh zloo qrw eh uhyhdohg

Frequency of Example

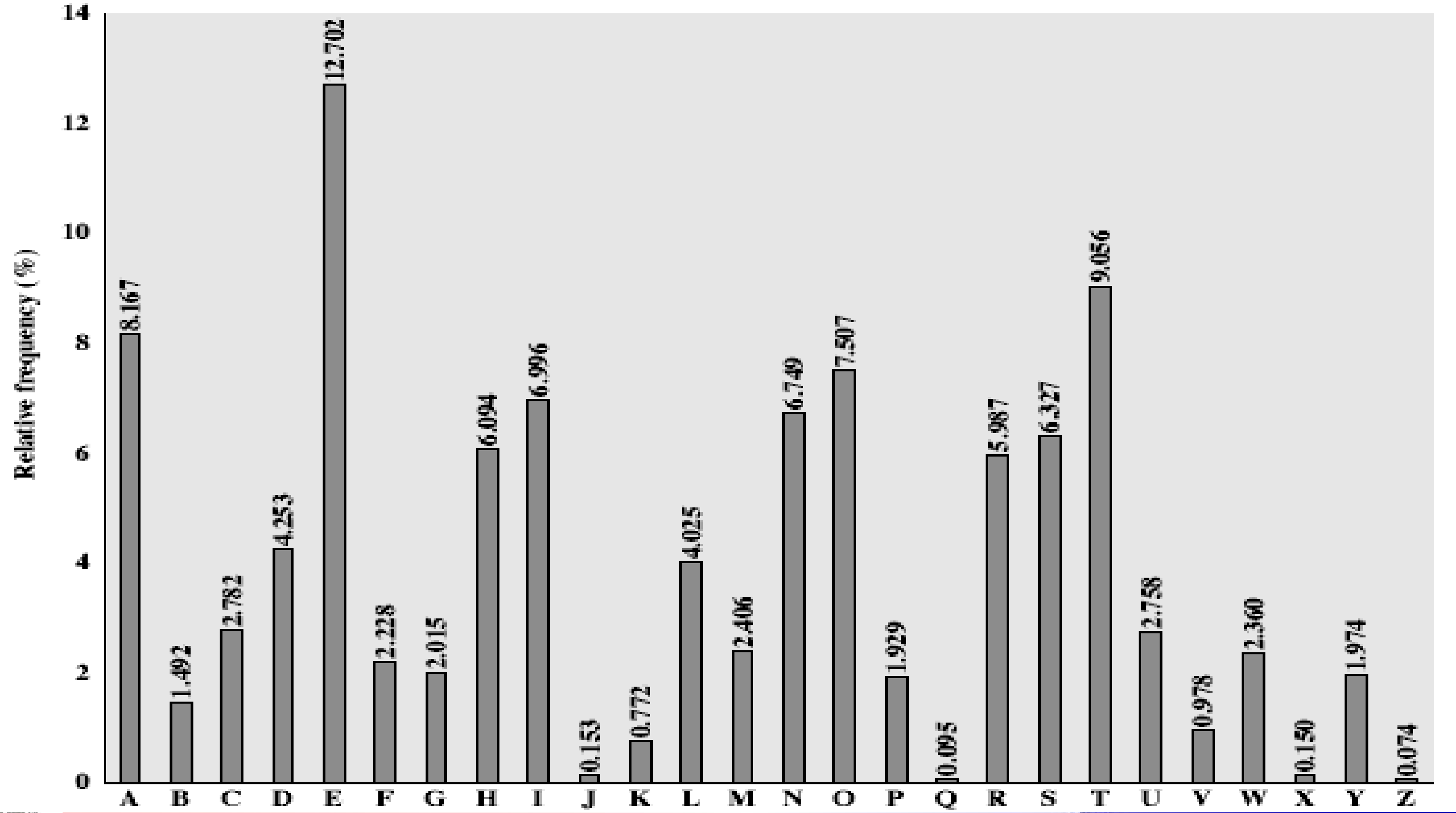
TABLE 2.2 FREQUENCIES IN EXAMPLE CIPHER.

Letter	Count	Percent	Letter	Count	Percent
a	0	0.00	n	0	0.00
b	3	1.80	o	4	2.41
c	0	0.00	p	5	2.99
d	11	6.59	q	16	9.58
e	2	1.20	r	9	5.39
f	6	3.61	s	3	1.80
g	4	2.40	t	0	0.00
h	26	15.56	u	8	4.79
i	2	1.20	v	17	10.18
j	5	2.99	w	14	8.38
k	5	2.99	x	5	2.99
l	16	9.58	y	4	2.40
m	0	0.00	z	2	1.20
ALL	167				

Frequencies of Sample Cipher against Normal Text



English Letter Frequencies



Use in Cryptanalysis

- key concept - 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
- if Caesar cipher look for common peaks/troughs
  - peaks at: A-E-I triple, NO pair, RST triple
  - troughs at: JK, X-Z
- for monoalphabetic must identify each letter
  - tables of common double/triple letters help

Example Cryptanalysis

- Given ciphertext:

```
UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZ
VUEPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSX
EPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMOHMQ
```

- count relative letter frequencies
- guess P & Z are e and t
- guess ZW is th and hence ZWP is the
- proceeding with trial and error finally we get:

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

The Hill Cipher

- the Hill cipher is a polygraphic substitution cipher based on linear algebra
- Each letter is treated as a digit in base 26: A = 0, B =1, and so on.
- Consider the message ‘GOD', and the key “PAYMOREMONEY” in letters:
- Ciphertext = Key x Plaintext mod 26
- C = KP mod 26

$$K = \begin{matrix} 17 & 17 & 5 \\ 21 & 18 & 21 \\ 2 & 2 & 19 \end{matrix}$$

- Since ‘G' is 6, 'O' is 14 and ‘D' is 3, the message is the vector:

$$P = \begin{matrix} 6 \\ 14 \\ 3 \end{matrix}$$

- Thus the enciphered vector is given by:

$$\begin{bmatrix} 17 & 17 & 5 \\ 21 & 18 & 21 \\ 2 & 2 & 19 \end{bmatrix} \times \begin{bmatrix} 6 \\ 14 \\ 3 \end{bmatrix} \bmod 26 = \begin{bmatrix} 355 \\ 441 \\ 97 \end{bmatrix} \bmod 26$$

K P C

$$\begin{bmatrix} 17 \\ 25 \\ 19 \end{bmatrix}$$

corresponds to a ciphertext of 'RZT'

Decryption

P = $K^{-1} \times C \text{ MOD } 26$

Playfair Cipher

- not even the large number of keys in a monoalphabetic cipher provides security
- one approach to improving security was to encrypt multiple letters
- the **Playfair Cipher** is an example
- 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
  – (I and J aren't distinguished)
- fill in letters of keyword (sans duplicates)
- fill rest of matrix with other letters
- eg. using the keyword MONARCHY

```
MONAR
CHYBD
EFGIK
LPQST
UVWXZ
```

Encrypting and Decrypting

- plaintext encrypted two letters at a time:
  1. each letter is replaced by the one in its row in the column of the other letter of the pair, eg. “hs" encrypts to "BP", and “ea" to "IM" or "JM" (as desired). Except when that doesn’t work!
  2. if a pair is a repeated letter, insert a filler like 'X', eg. "balloon" transformed to "ba lx lo on"
  3. if both letters fall in the same row, replace each with letter to right (wrapping back to start from end), eg. “ar" encrypts as "RM"
  4. if both letters fall in the same column, replace each with the letter below it (again wrapping to top from bottom), eg. “mu" encrypts to "CM"

Security of the Playfair Cipher

- security much improved over monoalphabetic
- since have 25 x 25 = 625 diagrams
- would need a 625 entry frequency table to analyse (verses 26 for a monoalphabetic)
- and correspondingly more ciphertext
- was widely used for many years (eg. US & British military in WW1)
- it **can** be broken, given a few hundred letters since still has much of plaintext structure

Solve following Example

Playfair encryption technique with

- KEY: SHERRY
- Plaintext 1 : wireless
- Plaintext 2 : monday

ANS

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

Wireless wi re le sx sz VK YR KR RU YU

Monday mo nd ay IT QA FS

Polyalphabetic Ciphers

- another approach to improving security is to use multiple cipher alphabets called **polyalphabetic substitution ciphers**
- makes 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

- simplest polyalphabetic substitution cipher is the **Vigenère Cipher**
- effectively multiple caesar ciphers
- key is multiple letters long K = k1 k2 ... kd
- i^{th} letter specifies i^{th} alphabet to use
- use each alphabet in turn
- repeat from start after d letters in message
- decryption simply works in reverse

Example

- write the plaintext out
- write the keyword repeated above it
- use each key letter as a caesar cipher key
- encrypt the corresponding plaintext letter
- eg using keyword *deceptive*

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

Security of Vigenère Ciphers

- have multiple ciphertext letters for each plaintext letter
- hence letter frequencies are obscured
- but not totally lost
- start with letter frequencies
  - see if look monoalphabetic or not
- if not, then need to determine number of alphabets, since then can attack each

Autokey Cipher

- ideally want a key as long as the message
- Vigenère proposed the **autokey** cipher
- with keyword is prefixed to message as key
- knowing keyword can recover the first few letters
- use these in turn on the rest of the message
- but still have frequency characteristics to attack
- eg. given key *deceptive*

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

Example paragraph…

This is an unusual paragraph. I'm curious how quickly you can find out what is so unusual about it. It looks so plain you would think nothing was wrong with it. In fact, nothing is wrong with it! It is unusual though. Study it, and think about it, but you still may not find anything odd. But if you work at it a bit, you might find out! Try to do so without any coaching! You probably won't, at first, find anything particularly odd or unusual or in any way dissimilar to any ordinary composition. That is not at all surprising, for it is no strain to accomplish in so short a paragraph a stunt similar to that which an author did throughout all of his book, without spoiling a good writing job, and it was no small book at that. By studying this paragraph assiduously, you will shortly, I trust, know what is its distinguishing oddity. Upon locating that "mark of distinction," you will probably doubt my story of this author and his book of similar unusuality throughout. It is commonly known among book-conscious folk and proof of it is still around. If you must know, this sort of writing is known as a lipogram, but don't look up that word in any dictionary until you find out what this is all about.—Unknown

Classical Transposition Ciphers

M	E	G	A	B	U	C	K
7	4	5	1	2	8	3	6
p	l	e	a	s	e	t	r
a	n	s	f	e	r	o	n
e	m	i	l	l	i	o	n
d	o	l	l	a	r	s	t
o	m	y	s	w	i	s	s
b	a	n	k	a	c	c	o
u	n	t	s	i	x	t	w
o	t	w	o	a	b	c	d

Plaintext

pleasetransferonemilliondollarsto
myswissbankaccountsixtwotwo

Ciphertext

AFLLSKSOSELAWAIATOOSSCTCLNMOMANT
ESILYNTWRNNTSOWDPAEDOBUOERIRICXB

Transposition Ciphers

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

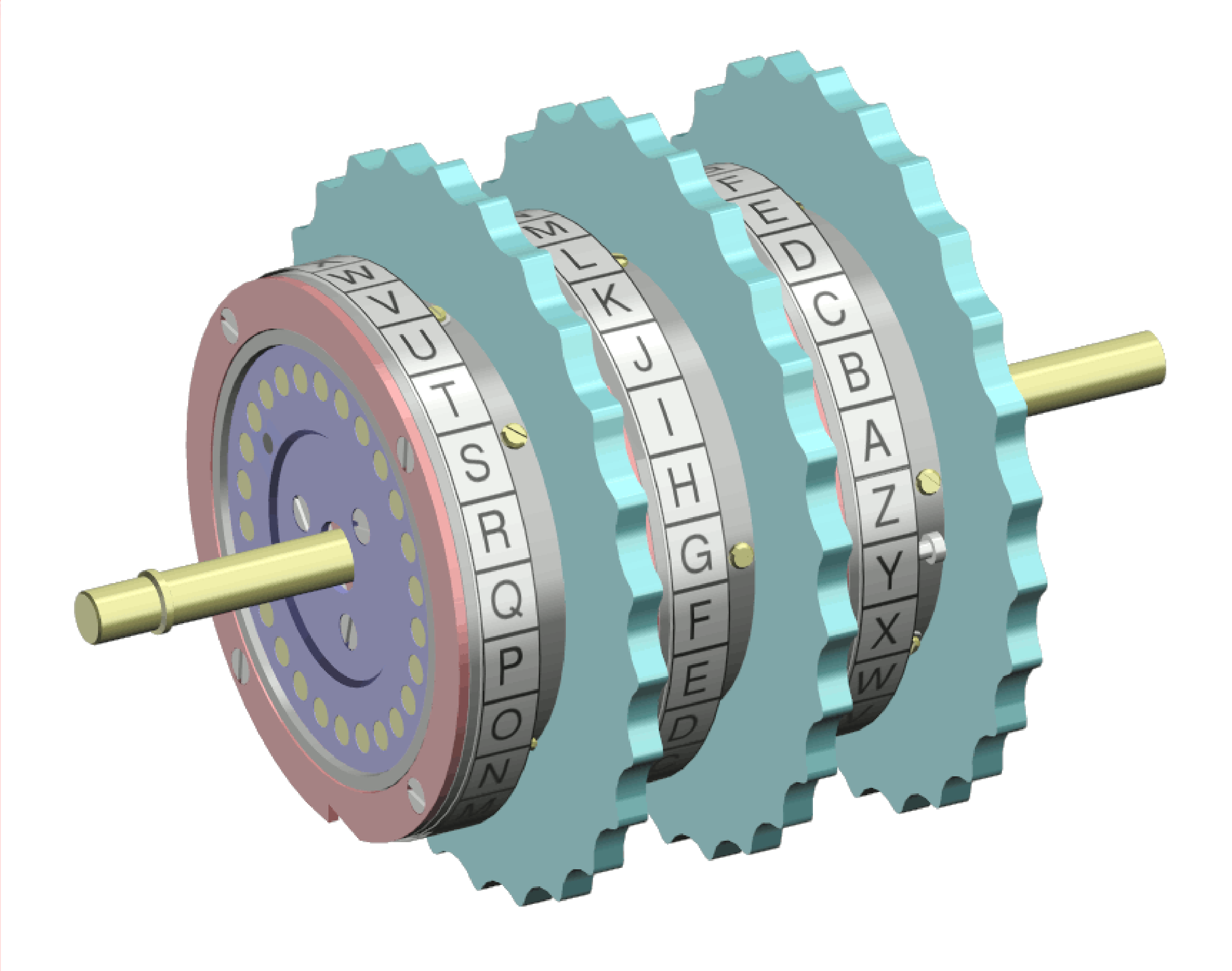
Row Transposition Ciphers

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

```
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
```

Rotor Machines

- before modern ciphers, rotor machines were most common product cipher
- were widely used in WW2
  - German Enigma, Allied Hagelin, Japanese Purple
- implemented a very complex, varying substitution cipher
- used a series of cylinders, each giving one substitution, which rotated and changed after each letter was encrypted
- with 3 cylinders have 26^3=17576 alphabets
  - 3! rearrangements of cylinders in Enigma


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

One-Time Pads

In 1917

Message 1:	1001001	0100000	1101100	1101111	1110110	1100101	0100000	1111001	1101111	1110101	0101110
Pad 1:	1010010	1001011	1110010	1010101	1010010	1100011	0001011	0101010	1010111	1100110	0101011
Ciphertext:	0011011	1101011	0011110	0111010	0100100	0000110	0101011	1010011	0111000	0010011	0000101
Pad 2:	1011110	0000111	1101000	1010011	1010111	0100110	1000111	0111010	1001110	1110110	1110110
Plaintext 2:	1000101	1101100	1110110	1101001	1110011	0100000	1101100	1101001	1110110	1100101	1110011

Theoretically unbreakable (Claude Shannon)

- Encryption: $C = P \oplus K$
- Decryption: $P = C \oplus K$

One-Time Pad cont…

- if a truly random key as long as the message is used, the cipher will be secure
- called a One-Time pad
- is unbreakable since ciphertext bears no statistical relationship to the plaintext
- since for **any plaintext** & **any ciphertext** there exists a key mapping one to other
- unconditional security! why look any further??

Example…

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

- 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

4 (E) 16 (Q) 13 (N) 21 (V) 25 (Z) ciphertext

− 19 (T) 16 (Q) 20 (U) 17 (R) 8 (I) possible key

−15 0 −7 4 17 ciphertext – key

11 (L) 0 (A) 19 (T) 4 (E) 17 (R) ciphertext-key (mod 26)

L A T E R → message

Difficulties with One-Time Pad

- Making Large quantities of random keys
- Key distribution and protection

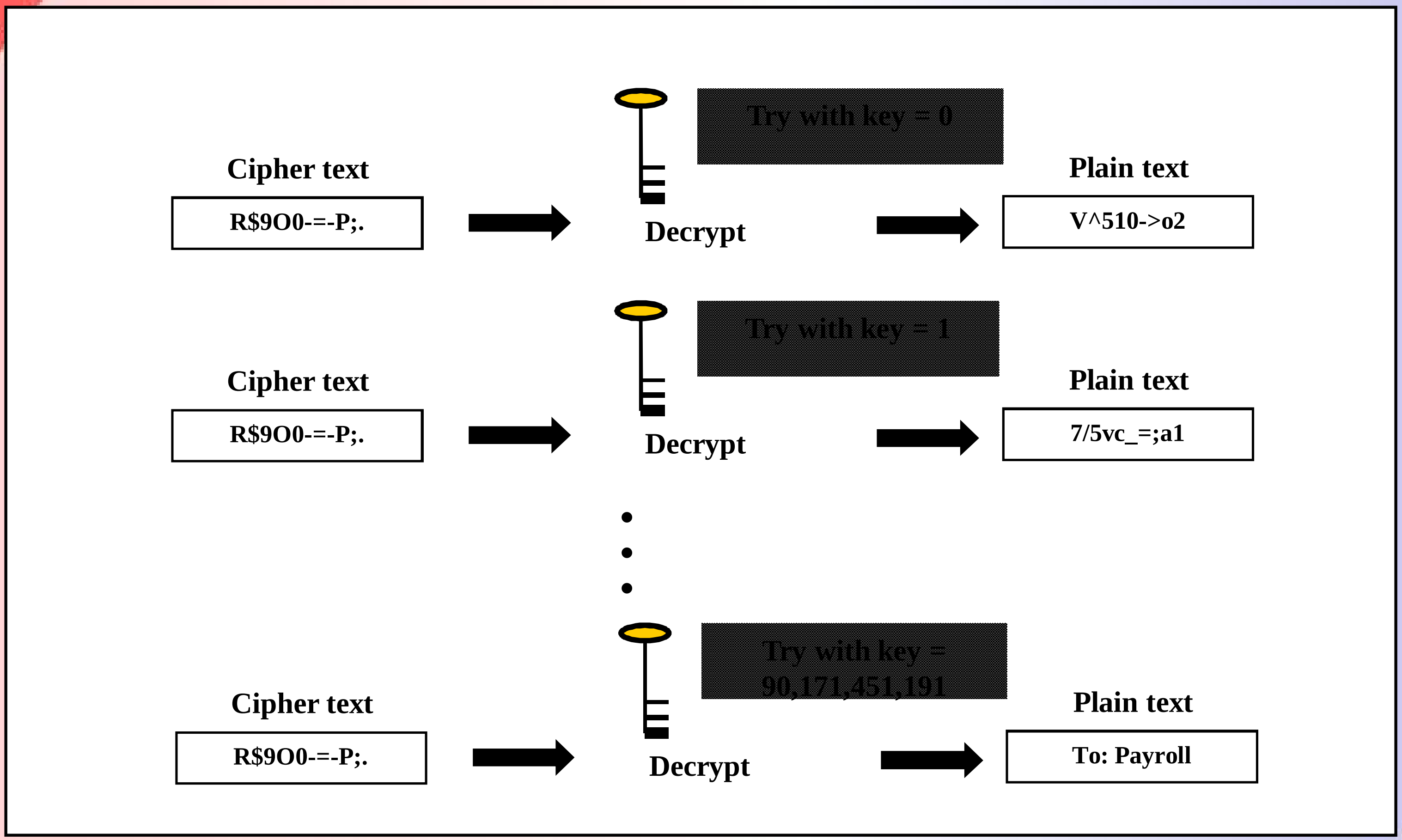
Symmetric Encryption

- or conventional / private-key / single-key
  - "ciphers"
- 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

Brute Force Attack

- Attacker tries all possible keys one by one
- Can be successful if key length is small
- Start with Key = 0, then Key = 1, etc.

Brute Force Attack



Key Range

- Specifies the number of possible keys
- Bigger the key range, more difficult is the attack
- In practice, at least 64, 128, 256 bit keys are used

Key Range

A 2-bit binary number has four possible states:
00
01
10
11

If we have one more bit to make it a 3-bit binary number, the number of possible states also doubles to eight, as follows:
000
001
010
011
100
101
110
111

In general, if an *n bit* binary number has *k* possible states, an *n+1 bit* binary number will have *2k* possible states.

Key Sizes and Range

Key size = 40 bits

00 00 00 00 00
00 00 00 00 01
…
FF FF FF FF FF

Key size = 64 bits

00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 01
…
FF FF FF FF FF FF FF FF

Key size = 128 bits

00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 01
…
FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF

Brute Force Search

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

Key Size (bits)	Number of Alternative Keys	Time required at 1 encryption/μs	Time required at 10^6 encryptions/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

Types of Cryptanalytic Attacks

adversary needs strongest attack

adversary's attacks can be weaker

- **Ciphertext only**
  - only know algorithm or ciphertext, statistical analysis can identify plaintext, or worse: the key
- **Known plaintext**
  - Know algorithm or cipher text or one or plaintext-ciphertext pairs formed with secret key
- **Chosen plaintext**
  - Know algorithm, ciphertext, plaintext message chosen by cryptanalyst , together with its corresponding ciphertext generated with the secret key
- **Chosen ciphertext**
  - select ciphertext and obtain plaintext to attack cipher
- **Chosen text**
  - select either plaintext or ciphertext to en/decrypt to attack cipher

Summary

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

Thank you….