

Cryptography and Network Security

Lecture 2

Classical Encryption Techniques

Prof. Mahmood Kh. Ibrahim

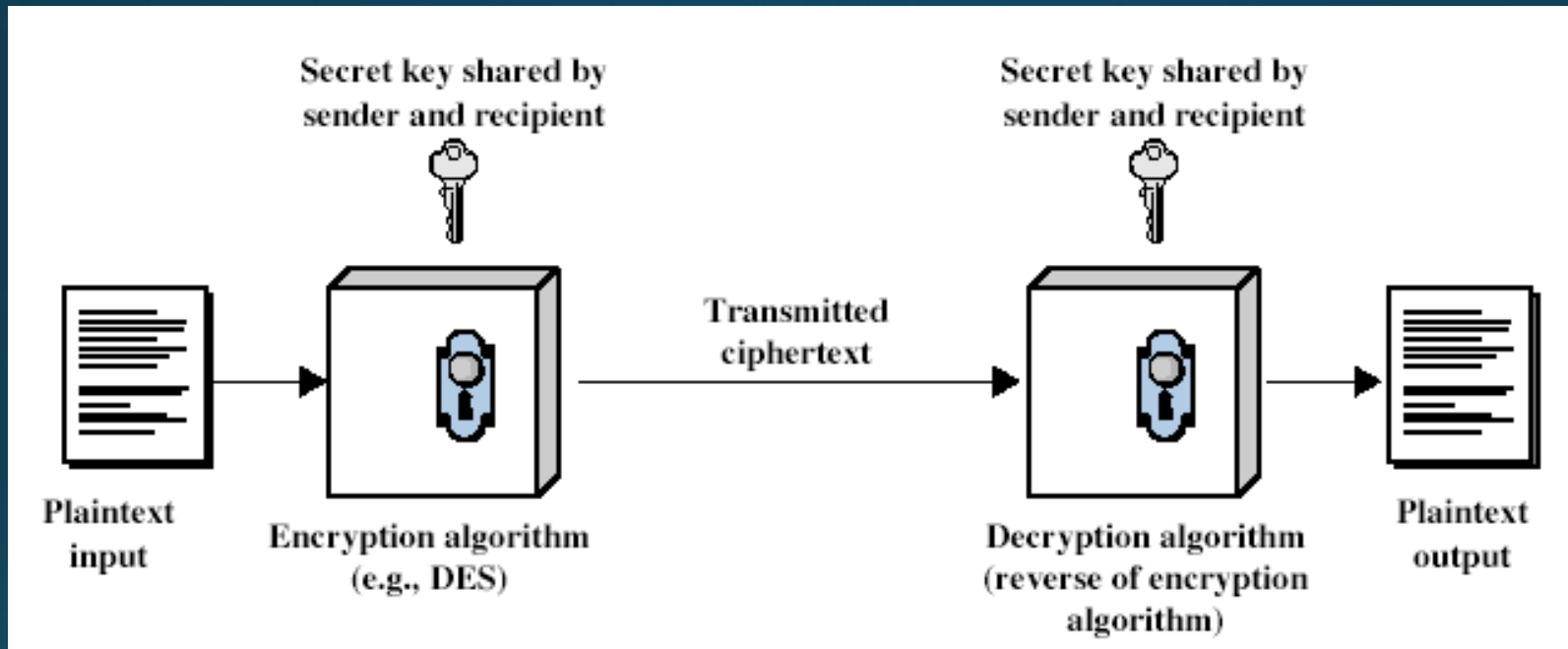
Symmetric Encryption

- or 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

Some Basic Terminology

- **plaintext** - original message
- **ciphertext** - coded message
- **cipher** - algorithm for transforming plaintext to ciphertext
- **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)** - study of principles/ methods of deciphering ciphertext *without* knowing key
- **cryptology** - field of both cryptography and cryptanalysis

Symmetric Cipher Model



Requirements

- two requirements for secure use of symmetric encryption:
 - a strong encryption algorithm
 - a secret key known only to sender / receiver
- mathematically have:
$$Y = E_K(X)$$
$$X = D_K(Y)$$
- assume encryption algorithm is known
- implies a secure channel to distribute key

Cryptography

- 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

- objective to recover key not just message
- general approaches:
 - cryptanalytic attack
 - brute-force attack

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 since the ciphertext provides insufficient information to uniquely determine the corresponding plaintext

- **computational security**

- given limited computing resources (eg 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 know / recognise plaintext

Key Size (bits)	Number of Alternative Keys	Time required at 1 decryption/ μ s	Time required at 10^6 decryptions/ μ s
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 = 5.4×10^{24} years	5.4×10^{18} years
168	$2^{168} = 3.7 \times 10^{50}$	$2^{167} \mu$ s = 5.9×10^{36} years	5.9×10^{30} years
26 characters (permutation)	$26! = 4 \times 10^{26}$	$2 \times 10^{26} \mu$ s = 6.4×10^{12} years	6.4×10^6 years

Classical Substitution Ciphers

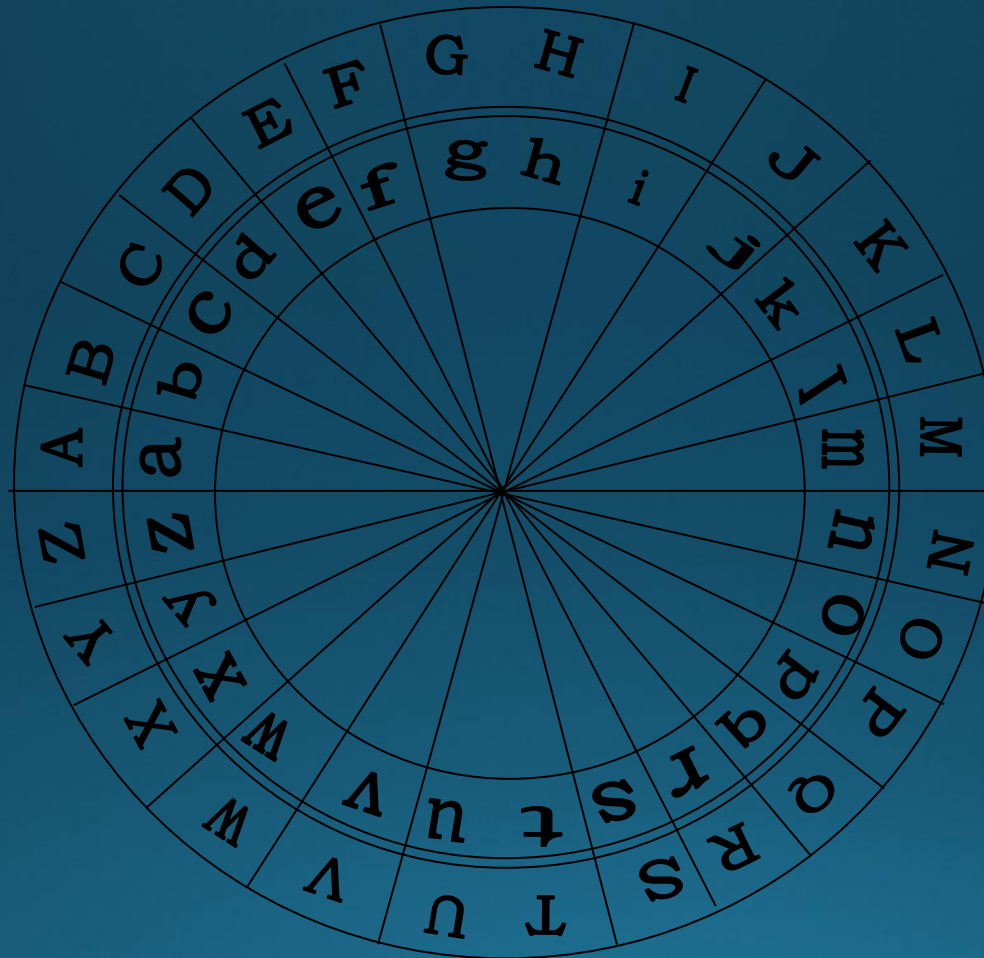
- where letters of plaintext are replaced by other letters or by numbers or symbols
- or if plaintext is viewed as a sequence of bits, then substitution involves replacing plaintext bit patterns with ciphertext bit patterns

Caesar Cipher

- earliest known substitution cipher, by Julius Caesar
- 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

Simple Substitution Ring



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:

$$c = E(p) = (p + k) \bmod (26)$$

$$p = D(c) = (c - k) \bmod (26)$$

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"
- with a shift of 2 (key C).
- "easy to break",

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

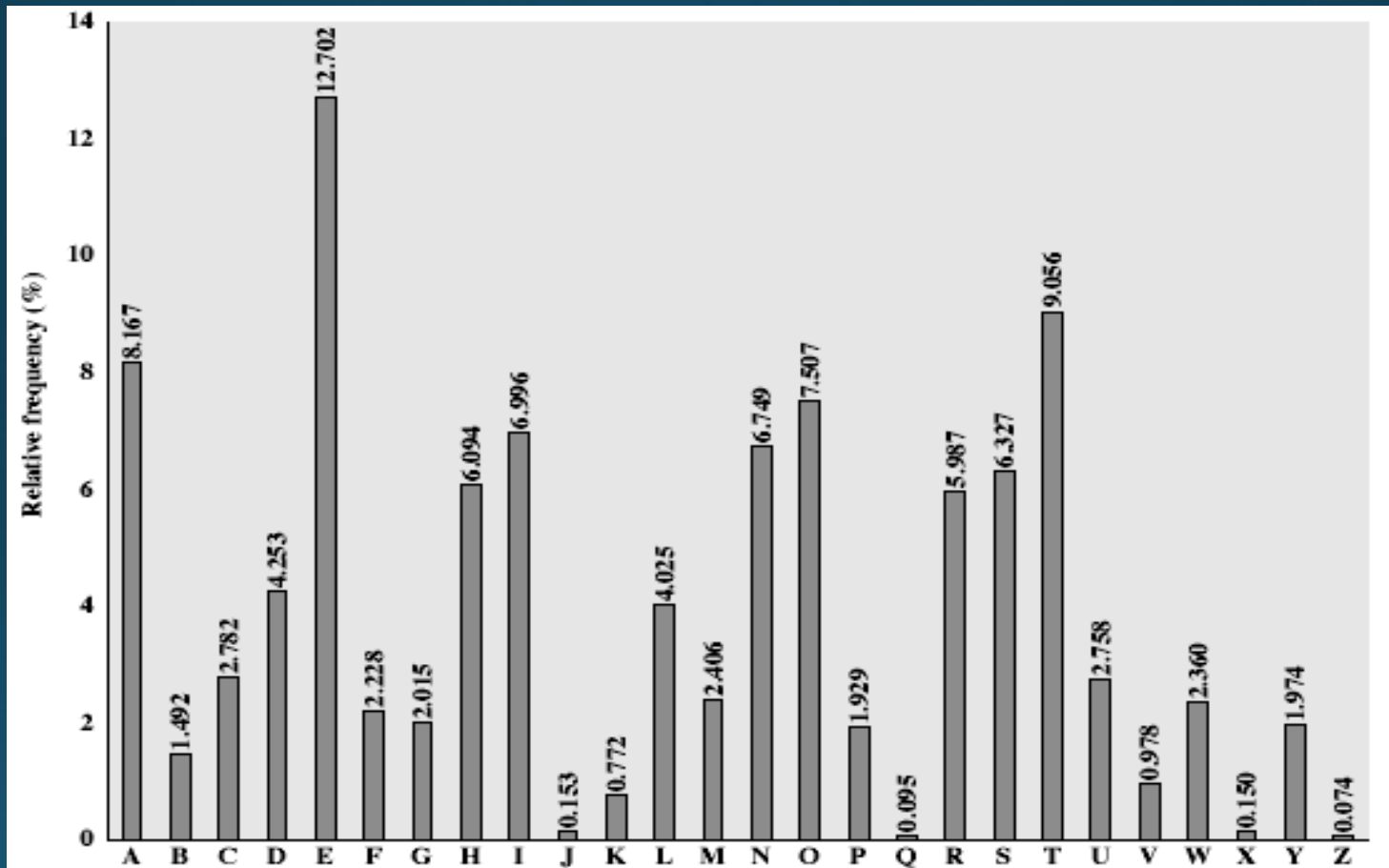
Monoalphabetic Cipher Security

- now have a total of $26! = 4 \times 10^{26}$ keys
- with so many keys, might think is secure
- but would be **!!!WRONG!!!**
- problem is language characteristics

Language Redundancy and Cryptanalysis

- human languages are **redundant**
- eg "th lrd s m shphrd shll nt wnt"
- letters are not equally commonly used
- in English E is by far the most common letter
 - followed by T,R,N,I,O,A,S
- other letters like 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

- key concept - monoalphabetic substitution ciphers do not change relative letter frequencies
- discovered by Arabian scientists in 9th 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:

```
UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAI  
Z VUEPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSX  
EPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMOHMQ
```

- count relative letter frequencies (see text)
- guess P & 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
- 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
- fill in letters of keyword (sans duplicates)
- fill rest of matrix with other letters
- eg. 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
 1. if a pair is a repeated letter, insert filler like 'X'
 2. if both letters fall in the same row, replace each with letter to right (wrapping back to start from end)
 3. if both letters fall in the same column, replace each with the letter below it (again wrapping to top from bottom)
 4. 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
- since have $26 \times 26 = 676$ digrams
- would need a 676 entry frequency table to analyse (verses 26 for a monoalphabetic)
- and correspondingly more ciphertext
- was widely used for many years
 - eg. by US & British military in WW1
- it can be broken, given a few hundred letters
- since still has much of plaintext structure

Polyalphabetic Ciphers

- **polyalphabetic substitution ciphers**
- improve 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

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

Rail Fence cipher

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

- eg. 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
```

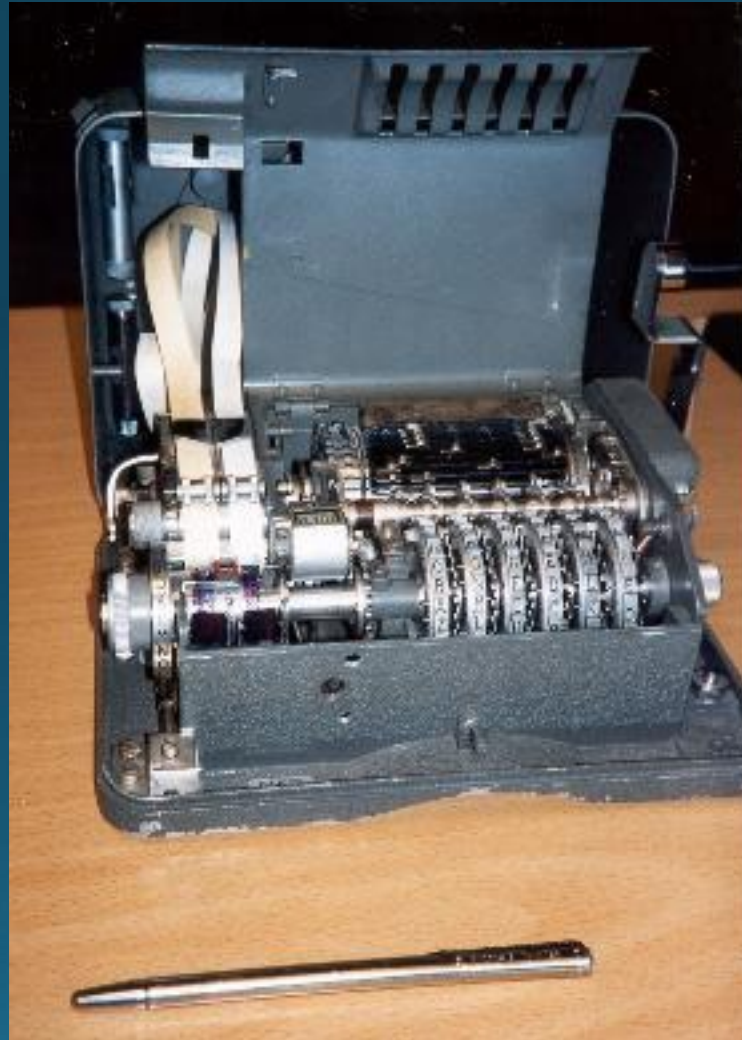
Product Ciphers

- ciphers using substitutions or transpositions are not secure because of language characteristics
- hence consider using several ciphers in succession to make harder, but:
 - 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

Rotor Machines

- before modern ciphers, rotor machines were most common complex ciphers in use
- 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

Hagelin Rotor Machine



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 graphic image or sound file
- has drawbacks
 - high overhead to hide relatively few info bits

Summary

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