Classical Cryptography

Dept. of Computer Science & Engg.
IIT Kharagpur, India

Teacher: Dipanwita Roy Chowdhury

Classical Techniques

- Broadly falls under two categories:

1. Substitution ciphers
   - Each letter of group of letters of the plaintext are replaced by some other letter or group of letters, to obtain the ciphertext.
2. Transposition ciphers
   - Letters of the plaintext are permuted in some form.

Substitution Ciphers

1. Caesar Cipher

- Earliest known substitution cipher.
- Replace each letter of the alphabet with the letter *three places* after that alphabet.
- Alphabets are assumed to be wrapped around (Z is followed by A, etc.).

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

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 D

P: HAPPY NEW YEAR

C: KDSSB QHZ BHDU

- We can generalize the idea by replacing each letter by the kth following letter.
- If we assign a number to each letter (A=1, B=2, etc), then

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

$$C = E(P) = E_k(P) = (P + k) \bmod 26$$

$$P = D(C) = D_k(C) = (C - k) \bmod 26$$

- <u>Drawback</u>:
  - Brute force attack is easy
  - Try out all the 25 possible keys

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

2. Mono-alphabetic Cipher

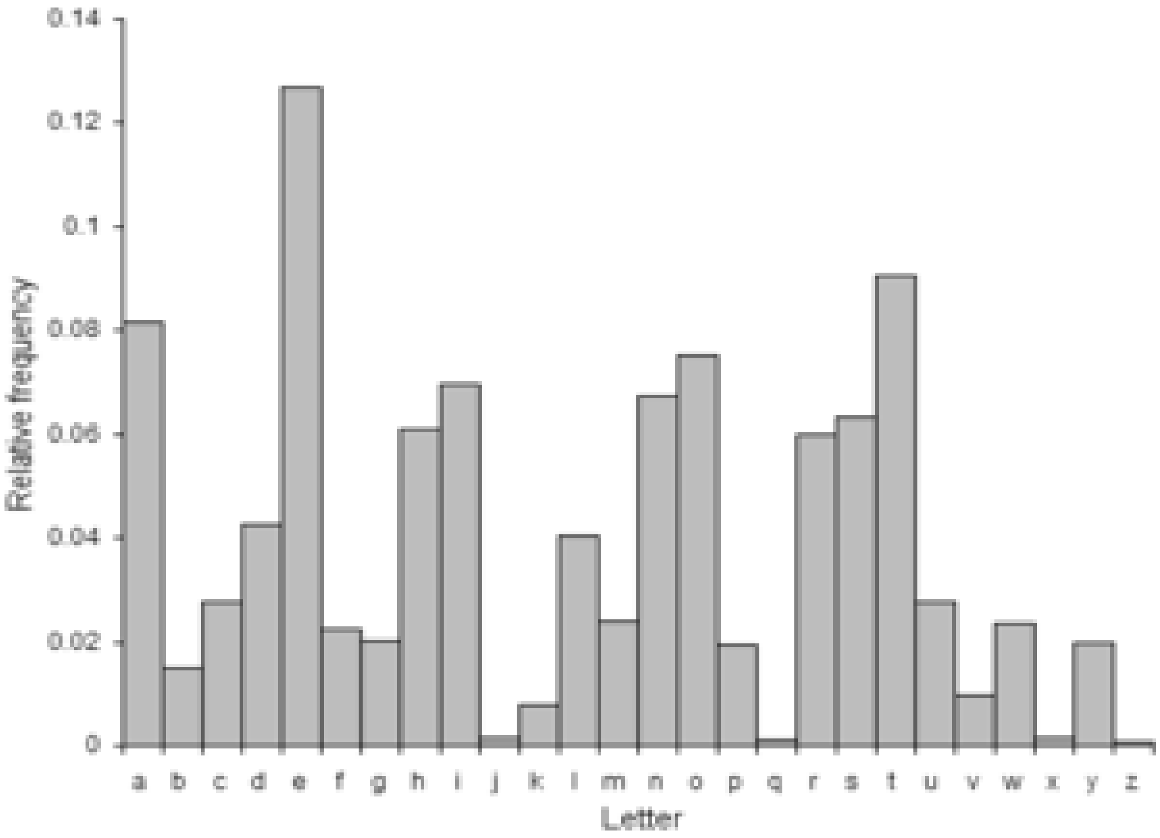
- Allow any arbitrary substitution.
- There can be 26! or $4x10^{26}$ possible keys.
- A typical key may be:

  (ZAQWSXCDERFVBGTYHNMJUIKLOP)
- Drawback:
  - We can make guesses by observing the relative frequency of letters in the text.
  - Compare it with standard frequency distribution charts in English (say).
  - Also look at the frequency of digrams and trigrams, for which tables are also available.
  - Easy to break in general.

Language Redundancy and Cryptanalysis

- human languages are **redundant**
- eg "th lrd s m shphrd shll nt wnt qu"
- 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

Relative Frequency Analysis



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:

```
UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZ
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
```

3. Poly-alphabetic Cipher

- Use different mono-alphabetic substitutions as we proceed through the plaintext message.
- Vigenere cipher is the best known cipher of this class.
  - Consists of 26 Caesar ciphers, with shifts of 0 to 25.
  - Each cipher is denoted by a key letter, which is the ciphertext letter that substitutes for the plaintext letter ‘a’.
  - To encrypt a message, a key is needed that is as long as the message (usually, a repeating keyword).
  - Decryption is just the reverse.

Drawback:

- Key and the plaintext share the same frequency distribution of letters.
- The best thing would have been to use a keyword which is as large as the plaintext, and has no statistical relationship to it..

Transposition Cipher

- Many techniques were proposed under this category.
- A simple scheme:
  - Write out the plaintext in a rectangle, row by row, and read the message column by column, by permuting the order of the columns.
  - Order of the column becomes the *key*.

An example

P: we are attending a lecture at IIT Kharagpur

Key:	4	3	1	2	5	6	7
	w	e	a	r	e	a	t
	t	e	n	d	i	n	g
	a	l	e	c	t	u	r
	e	a	t	I	I	T	K
	h	a	r	a	g	p	u
	r	-	-	-	-	-	-

C: anetr- rdcIa- eelaa- wtaehr eitIg- anuTp- tgrKu-

Drawback:

- The ciphertext has the same letter frequency as the original plaintext.
- Guessing the number of columns and some probable words in the plaintext holds the key.

Q2

Hill Ciphers

- Lester Hill, 1929. Not used much, but is historically significant: first time linear algebra used in crypto

- Use an *n* x *n* matrix M. Encrypt by breaking plaintext into blocks of length *n* (padding with x's if needed) and multiplying each by M (mod 26).
- Decryption is done by reversing the process, multiplying each block by M inverse (mod 26)

Hill Ciphers

An example with n = 3

$$\begin{pmatrix} c1 \\ c2 \\ c3 \end{pmatrix} = \begin{pmatrix} k11 & k12 & k13 \\ k21 & k22 & k23 \\ k31 & k32 & k33 \end{pmatrix} \begin{pmatrix} p1 \\ p2 \\ p3 \end{pmatrix}$$

C = KP mod 26

Plaintext “paymoremoney” and key

$$(K) = \begin{pmatrix} 17 & 17 & 5 \\ 21 & 18 & 21 \\ 2 & 2 & 19 \end{pmatrix}$$

The 1st 3 letters “pay” – (15 0 24)

K (15 0 24) = (375 819 486) mod 26 = (11 13 18) = LNS

Plaintext “paymoremoney” --- ciphertext “LNSHDLEWMTRW”

Q2

Hill Ciphers

- Decryption

$K\,K^{-1} = I$

- $(K^{-1}) = \begin{pmatrix} 4 & 9 & 15 \\ 15 & 17 & 6 \\ 24 & 0 & 17 \end{pmatrix}$

- $P = K^{-1}\,C \bmod 26 = K^{-1}\,K\,P = P$