

## SUBSTITUTION TECHNIQUES

- A substitution technique is one in which the letters of plaintext are replaced by other letters or by numbers or symbols.
- Substitution ciphers can be categorized as either—

### **i) Monoalphabetic ciphers or ii) polyalphabetic ciphers.**

- In monoalphabetic substitution, the relationship between a symbol in the plaintext to a symbol in the ciphertext is always one-to-one.
- In polyalphabetic substitution, each occurrence of a character may have a different substitute. The relationship between a character in the plaintext to a character in the ciphertext is one-to-many.

Various substitution ciphers are

- (i) Caesar Cipher
- (ii) Mono alphabetic cipher
- (iii) Playfair cipher
- (iv) Hill cipher
- (v) Poly alphabetic cipher
- (vi) Vignere cipher

### **(i) CAESAR CIPHER (OR) SHIFT CIPHER**

Caesar cipher was proposed by Julius Caesar. The Caesar cipher involves replacing each letter of the alphabet with the letter standing 3 places further down the alphabet.

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

plain: 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  
 cipher: 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

Let us assign a numerical equivalent to each letter:

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

Note that the alphabet is wrapped around, so that letter following 'z' is 'a'.

For each plaintext letter  $p$ , substitute the cipher text letter  $c$  such that

$$c = E(3, p) = (p+3) \bmod 26$$

Decryption is

$$p = D(3, c) = (c-3) \bmod 26$$

The general Caesar algorithm is

$$C = E(k, p) = (p + k) \bmod 26$$

where  $k$  takes on a value in the range 1 to 25.

The decryption algorithm is simply

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

If it is known that a given cipher text is a Caesar cipher, then a brute-force cryptanalysis is easily performed: simply try all the 25 possible keys.

### Cryptanalysis of Caesar Cipher

1. The encryption and decryption algorithms are known
2. There are only 25 possible keys. Hence brute force attack takes place
3. The language of the plaintext is known and easily recognizable

#### (ii) MONOALPHABETIC CIPHER

- Each plaintext letter maps to a different random cipher text letter
- Here, 26! Possible keys are used to eliminate brute force attack

There is, however, another line of attack. If the cryptanalyst knows the nature of the plaintext (e.g., non-compressed English text), then the analyst can exploit the regularities of the language.

UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZ  
 VUEPHZHMDZSHZOWSFPAPPDTSVPPQUZWYMXUZUHSX  
 EPYEPOPDZSZUFFPOMBZWPFPUPZHMDJUDTMOHMQ

As a first step, the relative frequency of the letters can be determined and compared to a standard frequency distribution for English

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				

Continued analysis of frequencies plus trial and error should easily yield a solution.

### (iii) PLAYFAIR CIPHER

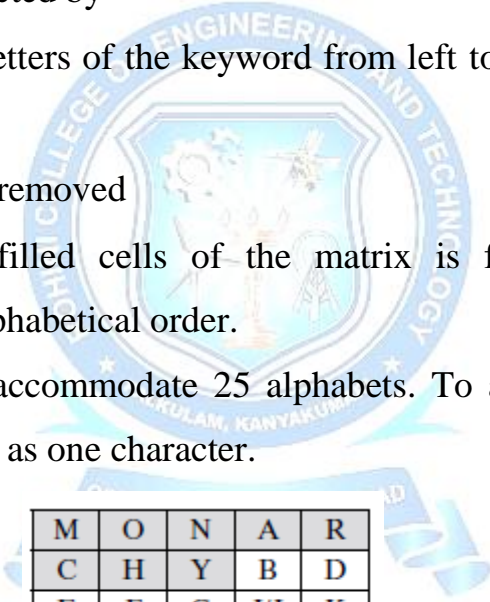
The best known multiple letter encryption cipher is the playfair, which treats digrams in the plaintext as single units and translates these units into cipher text digrams. The playfair algorithm is based on the use of 5x5 matrix of letters constructed using a keyword.

Let the keyword be “monarchy”.

The matrix is constructed by

- Filling in the letters of the keyword from left to right and from top to bottom
- Duplicates are removed
- Remaining unfilled cells of the matrix is filled with remaining alphabets in alphabetical order.

The matrix is 5x5. It can accommodate 25 alphabets. To accommodate the 26<sup>th</sup> alphabet I and J are counted as one character.



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

Rules for encryption

- Repeating plaintext letters that would fall in the same pair are separated with a filler letter such as ‘x’.
- Two plaintext letters that fall in the same row of the matrix are each replaced by the letter to the right, with the first element of the row circularly following the last. For example, ar is encrypted as RM.

- Two plaintext letters that fall in the same column are each replaced by the letter beneath, with the top element of the column circularly following the last. For example, mu is encrypted as CM.
- Otherwise, each plaintext letter in a pair is replaced by the letter that lies in its own row and the column occupied by the other plaintext letter. Thus, hs becomes BP and ea becomes IM (or JM, as the encipherer wishes).

### Example

Plain text: Balloon

Ba ll oo n

Ba lx lo on

Ba → I/JB

lx → SU

lo → PM

on → NA



### Strength of playfair cipher

- Playfair cipher is a great advance over simple mono alphabetic ciphers.
- Since there are 26 letters,  $26 \times 26 = 676$  diagrams are possible, so identification of individual digram is more difficult.
- Frequency analysis is much more difficult.

### Disadvantage

Easy to break because it has the structure and the resemblance of the plain text language

#### (iv) HILL CIPHER

It is a multi-letter cipher. It is developed by Lester Hill. The encryption algorithm takes  $m$  successive plaintext letters and substitutes for them  $m$  cipher text letters. The substitution is determined by  $m$  linear equations in which each character is assigned numerical value ( $a=0, b=1 \dots z=25$ ). For  $m=3$  the system can be described as follows:

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

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix} \pmod{26}$$

$$C = KP \pmod{26}$$

$C$  and  $P$  are column vectors of length 3 representing the cipher and plain text respectively.

Consider the message 'ACT', and

$$\begin{pmatrix} 0 \\ 2 \\ 19 \end{pmatrix}$$

The key below (or GYBNQKURP in letters)

$$\begin{pmatrix} 6 & 24 & 1 \\ 13 & 16 & 10 \\ 20 & 17 & 15 \end{pmatrix}$$

Thus the enciphered vector is given by:

$$\begin{pmatrix} 6 & 24 & 1 \\ 13 & 16 & 10 \\ 20 & 17 & 15 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \\ 19 \end{pmatrix} = \begin{pmatrix} 67 \\ 222 \\ 319 \end{pmatrix} \equiv \begin{pmatrix} 15 \\ 14 \\ 7 \end{pmatrix} \pmod{26}$$

which corresponds to a [ciphertext](#) of 'POH'

## Decryption

Decryption algorithm is done as  $\mathbf{P}=\mathbf{K}^{-1}\mathbf{C} \pmod{26}$

In order to decrypt, we turn the ciphertext back into a vector, then simply multiply by the inverse matrix of the key matrix (IFKVIVVMI in letters).

$$\begin{pmatrix} 6 & 24 & 1 \\ 13 & 16 & 10 \\ 20 & 17 & 15 \end{pmatrix}^{-1} \equiv \begin{pmatrix} 8 & 5 & 10 \\ 21 & 8 & 21 \\ 21 & 12 & 8 \end{pmatrix}$$

Cipher text of 'POH'

$$\begin{pmatrix} 8 & 5 & 10 \\ 21 & 8 & 21 \\ 21 & 12 & 8 \end{pmatrix} \begin{pmatrix} 15 \\ 14 \\ 7 \end{pmatrix} \equiv \begin{pmatrix} 260 \\ 574 \\ 539 \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 2 \\ 19 \end{pmatrix} \pmod{26}$$

Now gets us back the plain text 'ACT'

## Merits and Demerits

- Completely hides single letter and 2 letter frequency information.
- Easily attacked with known plain text attack

## (v)POLYALPHABETIC CIPHERS

Poly alphabetic cipher is a simple technique to improve mono-alphabetic technique.

The features are

- A set of related mono-alphabetic substitution rules are used
- A key determines which particular rule is chosen for a given transformation.

### Example: **Vigenere Cipher**

Each of the 26 ciphers is laid out horizontally, with the key letter for each cipher to its left. A normal alphabet for the plaintext runs across the top. The process of encryption is simple: Given a key letter x and a plaintext letter y, the cipher text is at the intersection of the row labelled x and the column labelled y; in this case, the cipher text is V. To encrypt a message, a key is needed that is as long as the message. Usually, the key is a repeating keyword.

Key=deceptive

Plain text= we are discovered save yourself

e.g., key = d e c e p t i v e d e c e p t i v e d e c e p t i v e

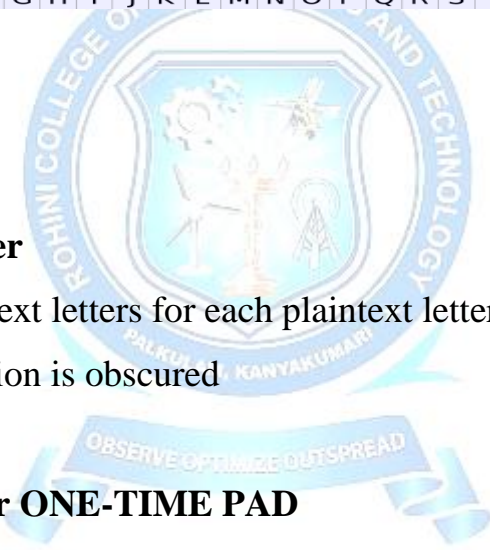
PT = w e a r e d i s c o v e r e d s a v e y o u r s e l f

CT = Z I C V T W Q N G R Z G V T W A V Z H C Q Y G L M G J

Decryption is equally simple. The key letter again identifies the row. The position of the cipher text letter in that row determines the column, and the plaintext letter is at the top of that column.



	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	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
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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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



### Strength of Vigenere cipher

- o There are multiple ciphertext letters for each plaintext letter.
- o Letter frequency information is obscured

### (vi) VERNAM CIPHER or ONE-TIME PAD

It is an unbreakable cryptosystem. It represents the message as a sequence of 0s and 1s. This can be accomplished by writing all numbers in binary, for example, or by using ASCII. The key is a random sequence of 0's and 1's of same length as the message. Once a key is used, it is discarded and never used again.

The system can be expressed as follows:

$$C_i = P_i \oplus K_i$$

$C_i$  -  $i$ th binary digit of cipher text  $P_i$  -  $i$ <sup>th</sup> binary digit of plaintext  $K_i$  -  $i$ th binary digit of key

$\oplus$  – exclusive OR operation

Thus the cipher text is generated by performing the bitwise XOR of the plaintext and the key. Decryption uses the same key. Because of the properties of XOR, decryption simply involves the same bitwise operation:

$$P_i = C_i \oplus K_i$$

### **Advantages**

- It is unbreakable since cipher text bears no statistical relationship to the plaintext
  - Not easy to break
- 

### **Drawbacks**

- Practically impossible to generate a random key as to the length of the message
- The second problem is that of key distribution and key protection.

Due to the above two drawbacks, one time pad is of limited use and is used for low band width channel which needs high security.

