Chapter 1
`Introduction to Cryptography

The word ***cryptography*** comes from two Greek words meaning "*secret writing*" and is the art and science of information hiding. This field is very much associated with mathematics and computer science with application in many fields like computer security, electronic commerce, telecommunication, etc.

So cryptography is a subject that should be of interest to many people, especially because we now live in the Information Age, and our secrets can be transmitted in so many ways – email, cell phone, etc. – and all these channels need to be protected [simon singh].

Secrecy and Encryption

In the ancient days, cryptography was mostly referred to as ***encryption*** – the mechanism to convert the readable ***plaintext*** into unreadable (incomprehensible) text i.e. ***ciphertext***, and ***decryption*** – the opposite process of encryption i.e. conversion of ciphertext back to the plaintext. Though the consideration of cryptography was on message confidentiality (encryption) in the past, nowadays cryptography considers the study and practices of authentication, digital signatures, integrity checking, and key management, etc.

Encryption mostly provides the secrecy of massage being transmitted over the communication network. This is called confidentiality of massage. The only sender knows the keys and can decipher the massage.

Cryptology

Cryptanalysis is the breaking of codes. Cryptanalysis encompasses all of the techniques to recover the plaintext and/or key from the ciphertext.

The combined study of cryptography and cryptanalysis is known as ***cryptology***. Though most of the time we use cryptography and cryptology in the same way.

Objective of cryptography

Encryption and Decryption

Encryption is the process of encoding a message so that its meaning is not obvious i.e. converting information from one form to some other unreadable form using some algorithm called ***cipher*** with the help of secret message called ***key***. The converting text is called is ***plaintext*** and the converted text is called ***ciphertext***.

Decryption is the reverse process, transforming an encrypted message back into its normal, original form. In decryption process also the use of key is important.

Alternatively, the terms ***encode*** and ***decode*** or ***encipher*** and ***decipher*** are used instead of ***encrypt*** and ***decrypt***. That is, we say that we encode, encrypt, or encipher the original message to hide its meaning. Then, we decode, decrypt, or decipher it to reveal the original message.

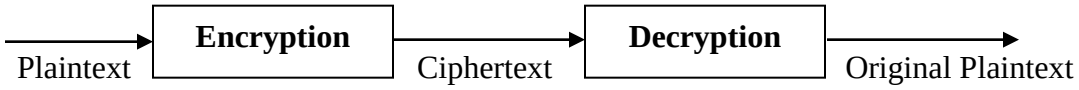

Fig: *Encryption-Decryption*

The use of encryption techniques is being used since very long period as it can be noted from the technique called ***Caesar's cipher*** used by Julius Caesar for information passing to his soldiers. Encryption techniques have also been extensively used in military purposes to conceal the information from the enemy. Nowadays to gain the confidentiality encryption is being used in many areas like communication, internet banking, digital right management, etc.

Key

A ***key*** is a parameter or a piece of information used to determine the output of cryptographic algorithm. While doing the encryption, key determines the transformation of plaintext to the cipher text and vice versa. Keys are also used in other cryptographic processes like message authentication codes and digital signatures. Most of the cryptographic systems depend upon the key and thus the secrecy of the key is very important and is one of the difficult problems in practice. Another important issue for the key is its length. Since key is the sole entity that defines the strength of the security (normally algorithm used is public) we need to select the key in a way such that attacker should take long enough to try all possibilities. To prevent the key from being guessed the choice of the key must be random.

Cipher

A ***cipher*** is an algorithm for performing encryption and decryption. The operation of cipher depends upon the special information called key. Without knowledge of the key, it should be difficult, if not nearly impossible, to decrypt the resulting cipher into readable plaintext. There are many types of encryption techniques that have advanced from history, however the distinction of encryption technique can be broadly categorized in terms of number of key used and way of converting plaintext to the ciphertext.

Cryptosystem

Cryptosystem is a 5-tuple/quintuple (E, D, M, K, C), where M set of plaintexts, K set of keys, C set of ciphertexts, E set of encryption functions e: $M \times K \rightarrow C$ and D set of decryption functions d: $C \times K \rightarrow M$.

Example: *Caesar Cipher*

M = {sequences of letters}

K = {i | i is an integer and $0 \leq i \leq 25$}

E = {E_k | $k \in K$ and for all letters m, $E_k(m) = (m + k) \bmod 26$ }

D = {D_k | $k \in K$ and for all letters c, $D_k(c) = (26 + c – k) \bmod 26$}

C = M

Cryptographic system characteristics

Cryptographic systems are characterized along three independent dimensions:

The type of operations used for transforming plaintext to ciphertext. All encryption algorithms are based on two general principles: substitution, in which each element in the plaintext (bit, letter, group of bits or letters) is mapped into another element, and transposition, in which elements in the plaintext are rearranged. The fundamental requirement is that no information be lost (that is, that all operations are reversible). Most systems, referred to as product systems, involve multiple stages of substitutions and transpositions.

The number of keys used. If both sender and receiver use the same key, the system is referred to as symmetric, single-key, secret-key, or conventional encryption. If the sender and receiver use different keys, the system is referred to as asymmetric, two-key, or public-key encryption.

The way in which the plaintext is processed. A block cipher processes the input one block of elements at a time, producing an output block for each input block. A stream cipher processes the input elements continuously, producing output one element at a time, as it goes along.

Classical Cryptosystem

Historical pen and paper ciphers used in the past are sometimes known as classical ciphers.
These are the very old or quite old cryptosystem that were used in pre computer age.these crypto system are too weak now days and can be broken easily with computer.
But we even studied these cryptosystem because they illustrate basic of the concepts of cryptography.

Substitution Cipher

In substitution ciphers the letters are systematically replaced by other letters or symbols.

1. Caesar Cipher

It is the simple shift monoalphabetic classical cipher where each letter is replaced by a letter 3 position (actual Caesar cipher) ahead using the circular alphabetic ordering i.e. letter after Z is A.

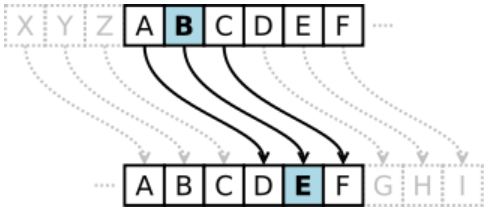

Fig: Caesar Cipher

So when we encode HELLO WORLD, the cipher text becomes KHOORZRUOG. Here we number each English alphabet starting from 0 (A) to 25 (Z). Each letter of the clear message is replaced by the letter whose number is obtained by adding the key (a number from 0 to 25) to the letter's number modulo 26. See the picture to visualize the Caesar cipher. The encryption can also be represented using modular arithmetic by first transforming the letters into numbers, according to the scheme, A = 0, B = 1, ..., Z = 25. Encryption of a letter c by a shift k can be described mathematically as,

$$c = E_k(m) = (m + k) \mod 26$$

Decryption is performed similarly,

$$m = D_k(c) = (c + 26 - k) \mod 26$$

Similarly, consider some examples of Caesar cipher;

```
Plaintext:  meet me after the toga party
```

Ciphertext: PHHW PH DIWHU WKH WRJD SDUWB

```
Plaintext:  the quick brown fox jumps over the lazy dog
```
Ciphertext: WKH TXLFN EURZQ IRA MXPSV RYHU WKH ODCB GRJ

Attacking the Cipher

Caesar Cipher is quite easily broken even with ciphertext only. One can attack the cipher text using exhaustive search by trying all possible keys until you find the right one. Exhaustive search is best suited if the key space is small and we have only 26 possible keys in Caesar cipher. Another approach of attacking the cipher is statistical analysis where we compare the ciphertext to 1-gram model of English.

Caesar's Problem

The main problem with Caesar's Cipher is that the key is too short and can be found by exhaustive search. Again statistical frequencies not concealed well i.e. they look too much like regular English letters. So the solution can be to increase the key length (can be done using multiple letters in key) so that cryptanalysis gets harder.

Transposition Cipher

In transposition ciphers the letters are systematically arranged so that the actual position of letters is gets changed making the text garble.

2. Rail-Fence Cipher

The Rail Fence Cipher is a form of transposition cipher that derives its name from the way in which it is encoded. In the rail fence cipher, the plaintext is written downwards and diagonally on successive "rails" of an imaginary fence, then moving up when we reach the bottom rail. When we reach the top rail, the message is written downwards again until the whole plaintext is written out. The message is then read off in rows.

For example, using 3 "rails" and a message of 'WE ARE DISCOVERED FLEE AT ONCE', the cipherer writes out:

```
W . . . E . . . C . . . R . . . L . . . T . . . E
. E . R . D . S . O . E . E . F . E . A . O . C .
. . A . . . I . . . V . . . D . . . E . . . N . .
```

Then reads off:

```
WECRL TEERD SOEEF EAOCA IVDEN
```

Similarly, if we have 3 "rails" and a message of THIS IS THE PLAINTEXT, the cipherer writes out (we are not showing diagonal move here just write in down rail a step ahead):

T S T P I E

H I H L N X

I S E A T T

The ciphertext is T S T P I E H I H L N X I S E A T T

The problem with Rail Fence Cipher is that the rail fence cipher is not very strong; the number of practical keys is small enough that a cryptanalyst can try them all by hand. To decrypt we get the number of letters to be skipped. For this if the number of rail is n key is $\lceil total\ letters\ in\ ciphertext\ /\ n \rceil$ so in our e.g. n = 3 and key is 18/3 = 6 i.e. skip 6 letters from the letter you are reading every time to get plaintext (remember to go circular that is if count ends continue from the starting letter leaving the read letter). See below:

We have

T	S	T	P	I	E	H	I	H	L	N	X	I	S	E	A	T	T
1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6

selected letter with index 1 THI. Now choose the letter with index 2, see below

T	S	T	P	I	E	H	I	H	L	N	X	I	S	E	A	T	T
1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6

Continue like this until you read off all the characters.

3. Vigenere Cipher: Substitution Cipher (Polyalphateic)

It is like Caesar cipher, but uses a phrase for e.g. for the message THE BOY HAS THE BALL and the key VIG, encipher using Caesar cipher for each letter:

key VIGVIGVIGVIGVIGV

plain THEBOYHASTHEBALL

cipher OPKWWECIYOPKWIRG

Here, generally, we repeatedly write key above the plaintext and use the Caesar cipher for each letter in the plaintext where key for each letter being processed is taken from the repeated key letter just above it. This process is simplified by using the table as below called Tableau

Key

Plaintext

	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

Fig: Vigenere Tableau

Period: length of key. In example above it is 3.

Tableau: Table used to encipher and decipher. In tableau Vigènere cipher has key letters on top, plaintext letters on the left or vice versa. It is also possible to have key on top (left) plaintexts in middle and ciphertexts in left (top).

Assuming key on top and the plaintext on left, Decryption is performed by finding the position of the ciphertext letter in a column, corresponding to the key letter, of the table, and then taking the label of the row in which it appears as the plaintext letter. For example, in column V (key letter), the ciphertext letter O appears in row T, which taken as the first plaintext letter. The second letter is decrypted by looking up P in column I of the table; it appears in row H, which is taken as the plaintext letter. This process continues until we find the plaintext letters for all the ciphertext letters

4. One-Time Pad (simple XOR)

It is a variant of a Vigenère cipher with a random key at least as long as the message. Since it has very high key length it is provably unbreakable. *Joseph Mauborgne* proposed this concept. He suggested using a random key that is as long as the message, so the key need not be repeated. In addition, the key is to be used to encrypt and decrypt a single message, and then is discarded. Each new message requires a new key of the same length as the new message. In One-time pad keys must be random, or we can attack the cipher by trying to regenerate the key approximations, such as using pseudorandom number generators to generate keys, are not random. This approach produces random output that bears no statistical relationship to the plaintext. Because the ciphertext contains no information whatsoever about the plaintext, there is simply no way to break the code.

5. 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 ciphertext digrams

The Playfair algorithm is based on the use of a 5 x 5 matrix of letters constructed using a keyword. Here keyword is MONARCHY then the matrix is:

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

The matrix is constructed by filling in the letters of the keyword (minus duplicates) from left to right and from top to bottom, and then filling in the remainder of the matrix with the remaining letters in alphabetic order. Plaintext is encrypted two letters at a time, according to the following rules:

1. Repeating plaintext letters that are in the same pair are separated with a filler letter, such as x, so that balloon would be treated as ba lx lo on.

2. 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.

3. 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.

4. 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).

6. Hill Cipher

Another interesting multi letter cipher is the Hill cipher, developed by the mathematician Lester Hill in 1929. The encryption algorithm takes m successive plaintext letters and substitutes for them m ciphertext letters. The substitution is determined by m linear equations in which each character is assigned a numerical value (a = 0, b = 1 ... z = 25).

For example, consider the plaintext "paymoremoney" and use the encryption key

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

The first three letters of the plaintext are represented by the vector

$$\begin{pmatrix}15\\ 0\\ 24\end{pmatrix} \text{ then } K.\begin{pmatrix}15\\ 0\\ 24\end{pmatrix}=\begin{pmatrix}375\\ 819\\ 487\end{pmatrix}\bmod 26=\begin{pmatrix}11\\ 13\\ 18\end{pmatrix}=LNS$$

the ciphertext for the entire plaintext is LNSHDLEWMTRW.

Hence in general the hill cipher can be expressed as

$C= E(K, P) = KP \bmod 26$

$P = D(K, P) = K\text{-}^{1}C \bmod 26 = K^{1}KP = P$

As with Playfair, the strength of the Hill cipher is that it completely hides single-letter frequencies. Indeed, with Hill, the use of a larger matrix hides more frequency information. Thus a 3 x 3 Hill cipher hides not only single-letter but also two-letter frequency information.

Cryptanalytic Attacks **(asked many times in exam)**

Cryptanalysis (from the Greek *kryptós*, "hidden", and *analýein*, "to loosen" or "to untie") is the study of methods for obtaining the meaning of encrypted information, without access to the secret information which is normally required to do so. Typically, this involves finding a secret key.

Cryptanalysis usually excludes methods of attack that do not primarily target weaknesses in the actual cryptography, such as bribery, physical coercion, burglary, keystroke logging, and social engineering, although these types of attack are an important concern and are often more effective than traditional cryptanalysis.

Cryptanalysis can be performed under a number of assumptions about how much can be observed or found out about the system under attack. It it is normally assumed that the general algorithm is known; this is Kerckhoffs' principle of "the enemy knows the system". There can be many types of attacks and broadly we categorize them as attack models:

Type of Attack	**Known to Cryptanalyst**
Ciphertext only	• Encryption algorithm • Ciphertext
Known plaintext	• Encryption algorithm • Ciphertext • One or more plaintext-ciphertext pairs formed with the secret key
Chosen plaintext	• Encryption algorithm • Ciphertext • Plaintext message chosen by cryptanalyst, together with its corresponding ciphertext generated with the secret key
Chosen ciphertext	• Encryption algorithm • Ciphertext • Purported ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key
Chosen text	• Encryption algorithm • Ciphertext • Plaintext message chosen by cryptanalyst, together with its corresponding ciphertext generated with the secret key • Purported ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key

Bruit Force attacks

A **brute-force attack** involves trying every possible key until an intelligible translation of the ciphertext into plaintext is obtained. On average, half of all possible keys must be tried to achieve success.