Unit 2: Encryption Techniques

Symmetric cipher model

Symmetric encryption, also referred to as conventional encryption or single-key Encryption.

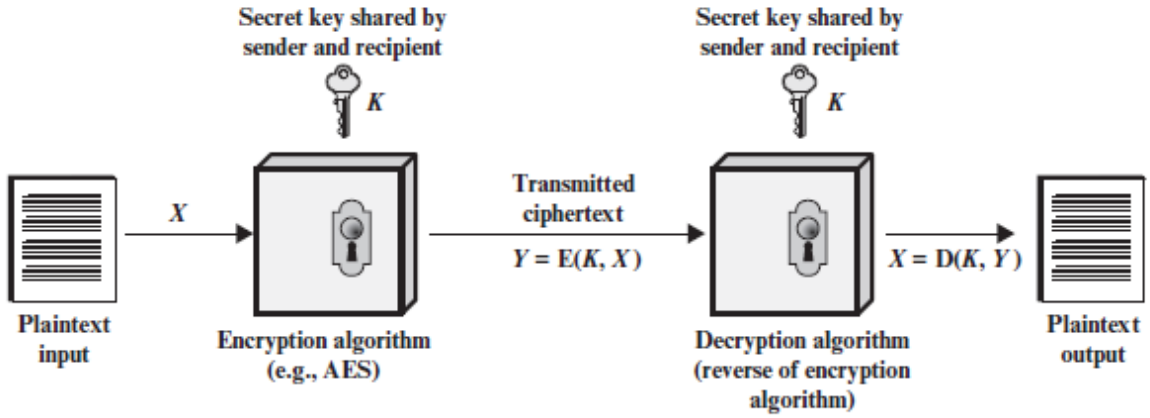

Figure 3.1 Simplified Model of Symmetric Encryption

A symmetric encryption scheme has five ingredients (Figure):

■ **Plaintext:** This is the original intelligible message or data that is fed into the algorithm as input.

■ **Encryption algorithm:** The encryption algorithm performs various substitutions and transformations on the plaintext.

■ **Secret key:** The secret key is also input to the encryption algorithm. The key is a value independent of the plaintext and of the algorithm. The algorithm will produce a different output depending on the specific key being used at the time. The exact substitutions and transformations performed by the algorithm depend on the key.

■ **Ciphertext:** This is the scrambled message produced as output. It depends on the plaintext and the secret key. For a given message, two different keys will produce two different ciphertexts. The ciphertext is an apparently random stream of data and, as it stands, is unintelligible.

■ **Decryption algorithm:** This is essentially the encryption algorithm run in reverse. It takes the ciphertext and the secret key and produces the original plaintext.

There are two requirements for secure use of conventional encryption:

1. We need a strong encryption algorithm. At a minimum, we would like the algorithm to be such that an opponent who knows the algorithm and has access to one or more ciphertexts would be unable to decipher the ciphertext or figure out the key. This requirement is usually stated in a stronger form: The opponent should be unable to decrypt ciphertext or discover the key even if he or she is in possession of a number of ciphertexts together with the plaintext that produced each ciphertext.

2. Sender and receiver must have obtained copies of the secret key in a secure fashion and must keep the key secure. If someone can discover the key and knows the algorithm, all communication using this key is readable.

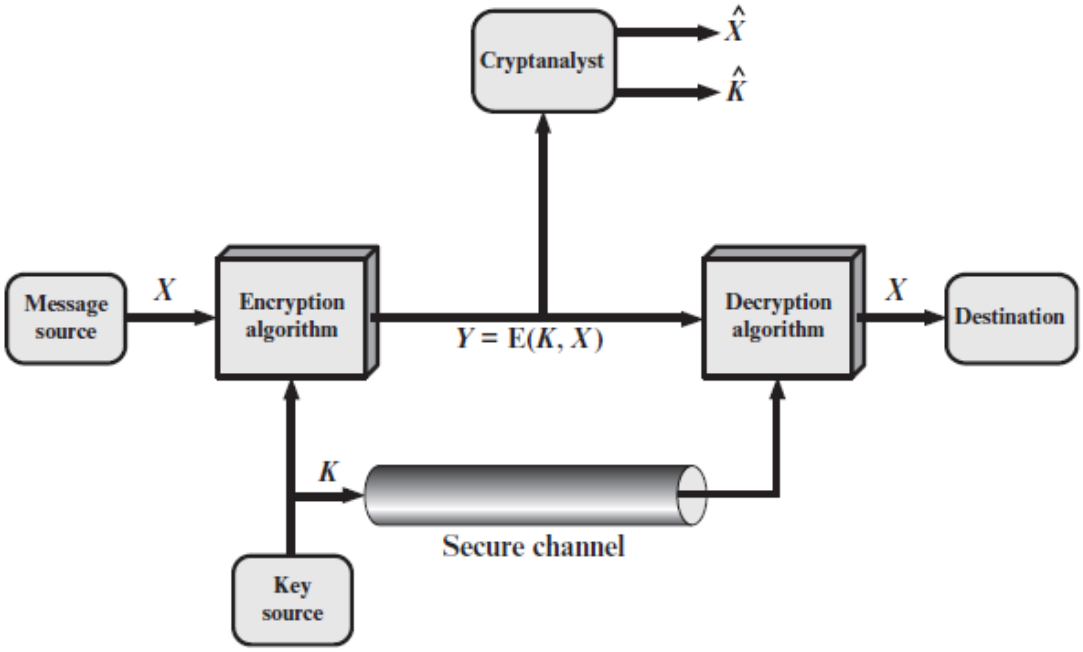

Figure 3.2 Model of Symmetric Cryptosystem

Let us take a closer look at the essential elements of a symmetric encryption scheme, using Figure 3.2. A source produces a message in plaintext, $X = [X_1, X_2, c, X_M]$. The M elements of X are letters in some finite alphabet.

Traditionally, the alphabet usually consisted of the 26 capital letters. Nowadays, the binary alphabet {0, 1} is typically used.

For encryption, a key of the form $K = [K_1, K_2, c, K_J]$ is generated. If the key is generated at the message source, then it must also be provided to the destination by means of some secure channel.

Alternatively, a third party could generate the key and securely deliver it to both source and destination.

With the message X and the encryption key K as input, the encryption algorithm forms the ciphertext $Y = [Y_1, Y_2, c, Y_N]$. We can write this as

$Y = E(K, X)$

This notation indicates that Y is produced by using encryption algorithm E as a function of the plaintext X, with the specific function determined by the value of the key K.

The intended receiver, in possession of the key, is able to invert the transformation:

$X = D(K, Y)$

An opponent, observing Y but not having access to K or X, may attempt to recover X or K or both X and K. It is assumed that the opponent knows the encryption (E) and decryption (D) algorithms. If the opponent is interested in only this particular message, then the focus of the effort is to recover X by generating a plaintext estimate $\hat{X}$.

Often, however, the opponent is interested in being able to read future messages as well, in which case an attempt is made to recover K by generating an estimate $\hat{K}$.

Cryptography

Cryptographic systems are characterized along three independent dimensions:

1. 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 (i.e., that all operations are reversible). Most systems, referred to as *product systems*, involve multiple stages of substitutions and transpositions.

2. 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.

3. 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.

Cryptanalysis and Brute-Force Attack

Typically, the objective of attacking an encryption system is to recover the key in use rather than simply to recover the plaintext of a single ciphertext. There are two general approaches to attacking a conventional encryption scheme:

■ **Cryptanalysis:** Cryptanalytic attacks rely on the nature of the algorithm plus perhaps some knowledge of the general characteristics of the plaintext or even some sample plaintext–ciphertext pairs. This type of attack exploits the characteristics of the algorithm to attempt to deduce a specific plaintext or to deduce the key being used.

■ **Brute-force attack:** The attacker tries every possible key on a piece of ciphertext until an intelligible translation into plaintext is obtained. On average, half of all possible keys must be tried to achieve success.

Types of Attacks on Encrypted Messages

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 ■ 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 ■ Ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key

Substitution Techniques

- Various conventional encryption schemes or substitution techniques are as under:

Caesar cipher

- The encryption rule is simple; replace each letter of the alphabet with the letter standing 3 places further down the alphabet.
- The alphabet is wrapped around so that Z follows A.
- Example:

  Plaintext: MEET ME AFTER THE PARTY

  Ciphertext: PHHW PH DIWHU WKH SDUWB
- Here, the key is 3. If different key is used, different substitution will be obtained.
- Mathematically, starting from a=0, b=1 and so on, Caesar cipher can be written as:

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

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

- This cipher can be broken
  - If we know one plaintext-cipher text pair since the difference will be same.
  - By applying Brute Force attack as there are only 26 possible keys.

Monoalphabetic Substitution Cipher

- Instead of shifting alphabets by fixed amount as in Caesar cipher, any random permutation is assigned to the alphabets. This type of encryption is called monoalphabetic substitution cipher.
- For example, A is replaced by Q, B by D, C by T etc. then it will be comparatively stronger than Caesar cipher.
- The number of alternative keys possible now becomes 26!.
- Thus, Brute Force attack is impractical in this case.
- However, another attack is possible. Human languages are redundant i.e. certain characters are used more frequently than others. This fact can be exploited.
- In English ‘e’ is the most common letter followed by ‘t’, ‘r’, ‘n’, ’o’, ‘a’ etc. Letters like ‘q’, ‘x’, ‘j’ are less frequently used.
- Moreover, digrams like ‘th’ and trigrams like ‘the’ are also more frequent.
- Tables of frequency of these letters exist. These can be used to guess the plaintext if the plaintext is in uncompressed English language.

Playfair Cipher

- In this technique multiple (2) letters are encrypted at a time.
- This technique uses a 5 X 5 matrix which is also called key matrix.

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 plaintext is encrypted **two letters at a time**:
  - Break the plaintext into pairs of two consecutive letters.
  - If a pair is a repeated letter, insert a filler like 'X'in the plaintext, eg. "Balloon" is treated as "ba lx lo on".
  - If both letters fall in the same row of the key matrix, replace each with the letter to its right (wrapping back to start from end), eg. "AR" encrypts as "RM".
  - 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".
  - Otherwise 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)
- Security is much improved over monoalphabetic as here two letters are encrypted at a time and hence there are 26 X 26 =676 diagrams and hence it needs a 676 entry frequency table.
- However, it can be broken even if a few hundred letters are known as much of plaintext structure is retained in cipher text.

Polyalphabetic Ciphers

- Another way to improve on the simple monoalphabetic technique is to use different
- monoalphabetic substitutions as one proceeds through the plaintext message.
- The general name for this approach is **polyalphabetic substitution cipher**. All these
- techniques have the following features in common:

  1. A set of related monoalphabetic substitution rules is used.

  2. A key determines which particular rule is chosen for a given transformation.

The Vigenère cipher

- This is a type of polyalphabetic substitution cipher (includes multiple substitutions depending on the key). In this type of cipher, the key determines which particular substitution is to be used.
- To encrypt a message, a key is needed that is as long as the message. Usually, the key is a repeating keyword.
- For example, if the keyword is *deceptive*, the message "we are discovered save yourself" is encrypted as follows:

  Key: *deceptivedecept*

  Plaintext: wearediscovered

  Ciphertext: ZICVTWQNGRZGVTW
- Encryption can be done by looking in the Vigenere Table where ciphertext is the letter key's row and plaintext's column or by the following formula:

$C_i = (P_i + K_{i \bmod N}) \bmod 26$

- 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.
- The strength of this cipher is that there are multiple ciphertext letters for each plaintext letter, one for each unique letter of the keyword.
- Thus, the letter frequency information is obscured however, not all knowledge of the plaintext structure is lost.

Autokey system

The periodic nature of the keyword can be eliminated by using a nonrepeating keyword that is as long as the message itself. Vigenère proposed what is referred to as an **autokey system**, in which a keyword is concatenated with the plaintext itself to provide a running key. For our example,

key: *deceptivewearediscoveredsav*

plaintext: wearediscoveredsaveyourself

ciphertext: ZICVTWQNGKZEIIGASXSTSLVVWLA

Vernam Cipher

- This system works on binary data (bits) rather than letters.
- The technique can be expressed as follows:

$C_i = P_i \oplus K_i$

Where

P_i = i^{th} binary digit of plaintext.

K_i = i^{th} binary digit of key.

C_i = i^{th} binary digit of ciphertext.

$\oplus$ = exclusive-or (XOR) operation

- Thus, the ciphertext is generated by performing the bitwise XOR of the plaintext and the key.
- Decryption simply involves the same bitwise operation:

$$P_i = C_i \oplus K_i$$

- The essence of this technique is the means of construction of the key.
- It was produced by the use of a running loop of tape that eventually repeated the key, so that in fact the system worked with a very long but repeating keyword.

Hill Cipher

The Hill Algorithm This 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, c, z = 25). For *m* = 3, the system can be described as

$$c_1 = (k_{11}p_1 + k_{21}p_2 + k_{31}p_3) \bmod 26$$
$$c_2 = (k_{12}p_1 + k_{22}p_2 + k_{32}p_3) \bmod 26$$
$$c_3 = (k_{13}p_1 + k_{23}p_2 + k_{33}p_3) \bmod 26$$

This can be expressed in terms of row vectors and matrices:[6]

$$(c_1\ c_2\ c_3) = (p_1\ p_2\ p_3)\begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \bmod 26$$

or

$$\mathbf{C} = \mathbf{PK} \bmod 26$$

where **C** and **P** are row vectors of length 3 representing the plaintext and ciphertext, and **K** is a 3 * 3 matrix representing the encryption key. Operations are performed mod 26.

Transposition Techniques

- A very different kind of mapping is achieved by performing some sort of permutation on the plaintext letters. This technique is referred to as a transposition cipher.
- The simplest such cipher is the **rail fence** technique.

Rail Fence Technique

- Encryption involves writing plaintext letters diagonally over a number of rows, then read off cipher row by row.
- For example, the text “meet me after the party” can be written (in 2 rows) as:

m e m a t r h p r y
e t e f e t e o a t

- Ciphertext is read from the above row-by-row: MEMATRHPRYETEFETEAT
- This scheme is very easy to cryptanalyze as no key is involved.
- Transposition cipher can be made significantly more secure by performing more than one stage of transposition. The result is a more complex permutation that is not easily reconstructed.

Columnar technique

A more complex scheme is to write the message in a rectangle, row by row, and read the message off, column by column, but permute the order of the columns. The order of the columns then becomes the key to the algorithm. For example,

```
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
```

Ciphertext: TTNAAPTMTSUOAODWCOIXKNLYPETZ

Thus, in this example, the key is 4312567. To encrypt, start with the column that is labeled 1, in this case column 3. Write down all the letters in that column.

Proceed to column 4, which is labeled 2, then column 2, then column 1, then columns 5, 6, and 7.

Difference between Symmetric and Asymmetric key cryptography

Symmetric key Cryptography	Asymmetric Key Cryptography
Symmetric key cryptography uses the same secret (private) key to encrypt and decrypt its data	Asymmetric key cryptography uses a public and a private key to encrypt and decrypt its data
The secret key must be known by both parties.	The public key is known to anyone with which they can encrypt the data but it can only be decoded by the person having the private key
In key distribution process, key information may have to be shared which decreases the security.	Here, the need for sharing key with key distribution center is eliminated.
Symmetric key encryption is faster than asymmetric key.	It is Slower than symmetric key encryption.
Basic operations used in encryption/ decryption are transposition and substitution.	It uses mathematical functions.

Steganography

- Plaintext message may be hidden in one of two ways.
  - Conceal the existence of the message-Steganography.
  - Render the message unintelligible to outsiders by various transformations of the text-Cryptography
- A simple but time consuming form of steganography is the one in which an arrangement of words or letters within an apparently normal text spells out the real message.
- For example, the sequence of first letters of each word of the overall message spells out the hidden message.
- Some other techniques that have been used historically are listed below:
  - **Character marking:** Selected letters of printed or typewritten text are overwritten in pencil. The marks are ordinarily not visible unless the paper is held at an angle to bright light.
  - **Invisible ink:** A number of substances can be used for writing but leave no visible trace until heat or some chemical is applied to the paper.
  - **Pin punctures:** Small pin punctures on selected letters are ordinarily not visible unless the paper is held up in front of a light.
  - **Typewriter correction ribbon:** Used between lines typed with a black ribbon, the results of typing with the correction tape are visible only under a strong light.
- Although these techniques may seem ancient, they have modern equivalents.

- For example, suppose an image has a resolution of 2048 X 3072 pixels where each pixel is denoted by 24 bits (Kodak CD photo format).

- The least significant bit of each 24-bit pixel can be changed without greatly affecting the quality of the image.
- The result is that you can hide a 2.3-megabyte message in a single digital snapshot.
- There are now a number of software packages available that take this type of approach to steganography.
- Steganography has a number of drawbacks when compared to encryption.
  - It requires a lot of overhead to hide a relatively few bits of information.
  - Once the system is discovered, it becomes virtually worthless.
- The advantage of steganography is that it can be employed by parties who have something to lose if the fact of their secret communication is discovered.