Introduction to Classical Cryptography

February 10, 2009

Internetwork Security - EE548

- Cryptography is the science or art of secret writing.
- The fundamental objective of cryptography is to enable two people (Alice and Bob) to communicate over an insecure channel in such a way that an opponent (Oscar) cannot understand what is being said.

- Plaintext : the information that Alice wants to send to Bob.
- Alice encrypts the plaintext, using a predetermined key, and send the resulting ciphertext to Bob over the public channel.
- Upon receiving the ciphertext
  - Oscar cannot determine what the plaintext was
  - But Bob knows the encryption key, can decrypt the ciphertext and get the plaintext.

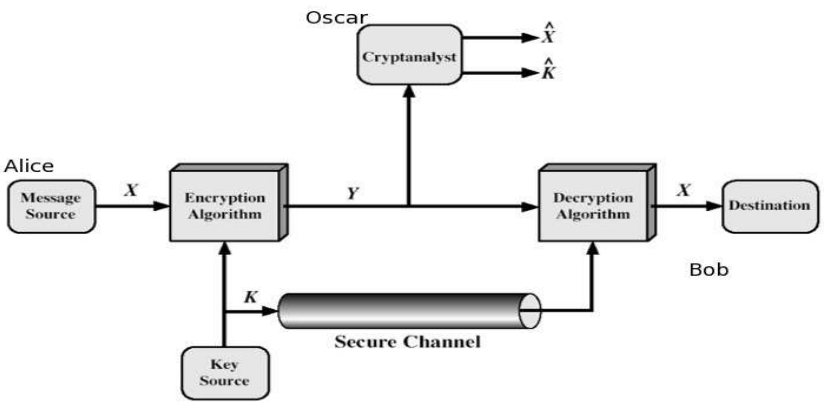


Figure 1: Communication Channel.

- Cryptology - two competing areas:
  - Cryptography - Art of converting information to a form that will be unintelligible to an unintended recipient, carried out by **cryptographer**.
  - Cryptanalysis - Art of breaking cryptographic systems, carried out by **cryptanalyst**.
- Two main types of cryptography in use today:
  - **Symmetric** or **secret key** cryptography
  - **Asymmetric** or **public key** cryptography

Internetwork Security

- Internetwork security: consists of measures to deter, prevent, detect, and correct security violations that involve the transmission of information across interconnected networks.

- Three aspects of information security:
  1. **Security attack**: Any action that compromises the security of information owned by an organisation.
  2. **Security mechanism**: A mechanism that is designed to detect, prevent or recover from a security attack.
  3. **Security service**: A service that enhances the security of the data processing systems and information transfers - makes use of one or more security mechanisms.

- One useful classification of security services is as follows:
  - **Confidentiality**: Requires that information be accessible only by authorised parties.
  - **Authentication**: Requires that the origin of a message be correctly identified with the assurance that the identity is not false.
  - **Integrity**: Requires that system assets and transmitted information be capable of modification only by authorised parties.

- **Non-repudiation**: Requires that neither the sender nor receiver of a message be able to deny the transmission.
- **Access Control**: Requires that access to information resources be controlled by or for the target system.

- Security attack consist of four general categories:
  1. **Interruption**: A system asset is destroyed or becomes unavailable - an attack on the **availability** of a system.
  2. **Interception**: An unauthorised party gains access to an asset. This is an attack on **confidentiality**.
  3. **Modification**: An unauthorised party not only gains access to an asset but tampers with it. This is an attack on **Integrity**.

4. **Fabrication**: An unauthorised party inserts counterfeit objects into the system. This is an attack on **authenticity**.

- **Passive attack**: Passive attacks involve monitoring data transmissions without modification of the data stream. The main aim is to **prevent** this form of attack as detection is difficult.
- **Active attack**: Active attacks involve some modification of the data stream or the creation of a false stream. With active attacks the main aim is to **detect and recover** as prevention is difficult.

Conventional Encryption

- Also termed single-key or symmetric encryption

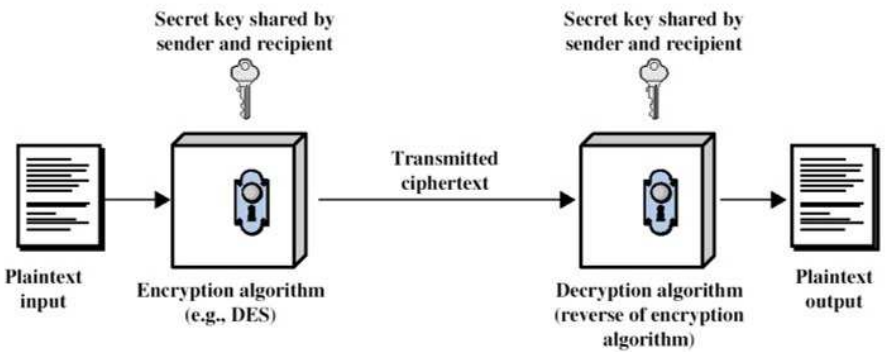


Figure 2: Simplified model of conventional encryption.

Cryptosystem

- Cryptosystem is a five tuple $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$
  - –Plaintext Space ($\mathcal{P}$): set of all possible plaintexts
  - –Cipherext Space ($\mathcal{C}$): set of all possible ciphertexts
  - –Key Space ($\mathcal{K}$): set of all possible keys
  - – $\mathcal{E}$: set of all possible encryption rules and $\mathcal{D}$: set of all possible decryption rules
- For each $k \in \mathcal{K}$, there is an encryption rule $e_k \in \mathcal{E}$ and a corresponding decryption rule $d_k \in \mathcal{D}$ such that $d_k(e_k(x)) = x$ for every plaintext $x \in \mathcal{P}$

- A practical cryptosystem should satisfy
  - Each encryption function e_k and each decryption function d_k should be efficiently computable.
  - An opponent, upon seeing the ciphertext string y, should be unable to determine the key k that was used, or the plaintext string x

- The process of attempting to compute the key k, given a string of ciphertext y, is called **cryptanalysis**.
  - If opponent can determine k, then he can decrypt y just as Bob would, using d_k.
  - Determining k should be as difficult as determining the plaintext string x, given the ciphertext string y.

Shift Cipher

- $Z_{26} = \{0, 1, 2 \ldots, 24, 25\}$
- $\mathcal{P} = \mathcal{C} = \mathcal{K} = Z_{26}$
- For $k \in \mathcal{K}$,

  $e_k(x) = (x + k) \bmod 26$ for $x \in \mathcal{P}$

  $d_k(y) = (y - k) \bmod 26$ for $y \in \mathcal{C}$
- Caesar Cipher is a particular case (for $k = 3$)

Example

- Plaintext is ordinary English text
- Correspondence between alphabetic characters and integer: $A = 0, B = 1, \ldots, Y = 24, Z = 25$.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Q	R	S	T	U	V	W	X	Y	Z
16	17	18	19	20	21	22	23	24	25

Encryption

- key $k = 11$
- Plaintext is "wewillmeetatmidnight"
- corresponding sequence of integers:

  22 4 22 8 11 11 12 4 4 19 0 19 12 8 3 13 8 6 7 19.
- we add 11 (key) to each value (reducing modulo 26):

  7 15 7 19 22 22 23 15 15 4 11 4 23 19 14 24 19 17 18 4.
- convert the sequence of integers to alphabetic characters:

  Ciphertext is "HPHTWWXPPELEXTOYTRSE"

Decryption

- ciphertext : "HPHTWWXPPELEXTOYTRSE".
- convert the ciphertext to sequence of integers:

  7 15 7 19 22 22 23 15 15 4 11 4 23 19 14 24 19 17 18 4.
- subtract 11 from each value (reducing modulo 26):

  22 4 22 8 11 11 12 4 4 19 0 19 12 8 3 13 8 6 7 19.
- convert the sequence of integers to alphabetic characters:

  Plaintext is "wewillmeetatmidnight"

Caesar Cipher

- Caesar Cipher is the earliest known (and the simplest). It involves replacing each letter of the alphabet with the letter standing three places further down. This is then wrapped around on itself when the end is reached. For example:

Key:	k=3
Plaintext:	meetmeaftertheparty
Ciphertext:	PHHWPHDIWHUWKHSDUWB

Shift Cipher is not Secure

- **Brute-force** cryptanalysis easily performed on the shift cipher by trying all 25 possible keys.
- Given a ciphertext string, Oscar successively try the decryption process with $k = 0, 1, 2$, etc. until get a meaningful text.

- Ciphertext : JBCRCLQRWCRVNBJENBWRWN

$$
\begin{aligned}
k = 0 &\rightarrow \text{jbcrclqrwcrvnbjenbwrwn} \\
k = 1 &\rightarrow \text{iabqbkpqvbqumaidmavqvm} \\
k = 2 &\rightarrow \text{hzapajopuaptlzhclzupul} \\
k = 3 &\rightarrow \text{gyzozinotzoskygbkytotk} \\
k = 4 &\rightarrow \text{fxynyhmnsynrjxfajxsnsj} \\
k = 5 &\rightarrow \text{ewxmxglmrxmqiweziwrmri} \\
k = 6 &\rightarrow \text{dvwlwfklqwlphvdyhvqlqh} \\
k = 7 &\rightarrow \text{cuvkvejkpvkojucxjupkpg} \\
k = 8 &\rightarrow \text{btujudijoujnftbwftojof} \\
k = 9 &\rightarrow \text{astitchintimesavesnine}
\end{aligned}
$$

- The key is $k = 9$

Substitution Cipher

- $\mathcal{P} = \mathcal{C} =$ set of 26-letter English alphabet

  $\mathcal{P} = \{a, b, c, \ldots, y, z\}$
  $\mathcal{C} = \{A, B, C, \ldots, Y, Z\}$

- $\mathcal{K}=$ set of all possible permutations of 26 alphabetic characters.

- For each permutation $\phi \in \mathcal{K}$,

  $e_\phi(x) = \phi(x)$ for $x \in \mathcal{P}$
  $d_\phi(y) = \phi^{-1}(y)$ for $y \in \mathcal{C}$, where ϕ^{-1} is the inverse permutation of ϕ.

Example

- Encryption function is the permutation ϕ :

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
X	N	Y	A	H	P	O	G	Z	Q	W	B	T	S	F	L

q	r	s	t	u	v	w	x	y	z
R	C	V	M	U	E	K	J	D	I

- Decryption function is the inverse permutation ϕ^{-1}:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
d	l	r	y	v	o	h	e	z	x	w	p	t	b	g	f

Q	R	S	T	U	V	W	X	Y	Z
j	q	n	m	u	s	k	a	c	i

- **Key**: $k = \phi$
- **Ciphertext**:
  MGZVYZLGHCMHJMYXSNHAHYCDLMHA
- Find the plaintext???

- **Monoalphabetic Cipher:** Each alphabetic character is mapped to a unique alphabetic character
- We use arbitrary monoalphabetic substitution, so there are 26! or $4 \times 10^{26} \approx 2^{88}$ possible permutations, which is a very large number. Thus brute force is infeasible.
- However we will see later that a Substitution Cipher is insecure against frequency analysis.

Vigenère Cipher

- **Polyalphabetic cipher**: use different monoalphabetic substitutions while moving through the plaintext.
- Let m be a positive integer
- $\mathcal{P} = \mathcal{C} = \mathcal{K} = (Z_{26})^m$
- For $k = (k_1, k_2, \ldots, k_m) \in \mathcal{K}$,

$$e_k(x_1, x_2, \ldots, x_m) = (x_1 + k_1, x_2 + k_2, \ldots, x_m + k_m)$$
$$d_k(y_1, y_2, \ldots, y_m) = (y_1 - k_1, y_2 - k_2, \ldots, y_m - k_m)$$

- All above operations are performed in Z_{26}

Example

- Correspondence between alphabetic characters and integer: $A = 0, B = 1, \ldots, Y = 24, Z = 25$.
- $m = 6$.
- Keyword is "CIPHER", this corresponds to the numerical equivalent $k = (2, 8, 15, 7, 4, 17)$

- Plaintext : "thi**s**crypto**s**y**s**temi**s**not**s**ecure".
- Encryption: add modulo 26

19	7	8	18	2	17	24	15	19	14	18	24	18	19	4	12
2	8	15	7	4	17	2	8	15	7	4	17	2	8	15	7
21	15	23	25	6	8	0	23	8	21	22	15	20	1	19	19

8	18	13	14	19	18	4	2	20	17	4
4	17	2	8	15	7	4	17	2	8	15
12	9	15	22	8	25	8	19	22	25	19

- Ciphertext:
  "VPX**Z**GIAXIV**W**P**U**BTTM**J**PWI**Z**ITWZT".

- **Transposition techniques**: So far all the ciphers we have looked at involved only substitution. A very different kind of mapping is achieved using transposition.
- In its simplest form, the **rail fence** technique involves writing down the plaintext as a sequence of columns and the ciphertext is read off as a sequence of rows. For example, if we use a rail fence of depth 2 with the plaintext *meet me after the party is over* we get:

m	e	m	a	t	r	h	p	r	y	s	v	r
e	t	e	f	e	t	e	a	t	i	o	e	

- Ciphertext is *mematrhprysvretefeteatioe* which is simply the first row concatenated with the second.

Transposition/Permutation Cipher

- Let m be a positive integer
- $\mathcal{P} = \mathcal{C} = (Z_{26})^m$
- $\mathcal{K}$= set of all possible permutations of $\{1, 2, \ldots, m\}$
- For each permutation $\pi \in \mathcal{K}$,

$$e_\pi(x_1, x_2, \ldots, x_m) = (x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(m)})$$
$$d_\pi(y_1, y_2, \ldots, y_m) = (y_{\pi^{-1}(1)}, y_{\pi^{-1}(2)}, \ldots, y_{\pi^{-1}(m)})$$

- π^{-1} being the inverse permutation of π

Example

- $m = 6$.
- key is the following permutation π:

x	1	2	3	4	5	6
$\pi(x)$	3	5	1	6	4	2

- inverse permutation π^{-1}:

x	1	2	3	4	5	6
$\pi^{-1}(x)$	3	6	1	5	2	4

- Plaintext : "defendthehilltopatsunset"

- partition the plaintext into group of six letters:

  defend | thehil | ltopat | sunset

- rearrange according to π:

  fnddee | eitlhh | oaltpt | nestsu

- Ciphertext: "FNDDEEEITLHHOALTPTNESTSU"
- Decryption can be done using π^{-1}

Cryptanalysis

- **Brute-force** cryptanalysis easily performed on the Shift cipher by trying all 25 possible keys.
- Three characteristics of the problem facilitate the successful use of the brute force approach:
  1. The encryption scheme is known.
  2. There are only a limited no. of keys.
  3. The plaintext is easily recognisable.
- Most cases, key size tends to be the main problem for brute-force attacks.

- **Monoalphabetic Ciphers**: If instead of using only the 25 possible keys, arbitrary substitution is used as in Substitution cipher, then there are 26! or $4 \times 10^{26} \approx 2^{88}$ (10 orders of magnitude greater than the keyspace for DES!) possible keys and hence brute force is infeasible.
- We now show the frequency analysis on Substitution cipher

Frequency Analysis

- Suppose we have a long ciphertext, the challenge is to decipher it
- Let we know the text is in English and has been encrypted using a monoalphabetic substitution cipher
- searching all possible keys is impractical as the keyspace is of size 26!

- In English, e is the most common letter, followed by t, then a, and so on, as shown in the Figure 3

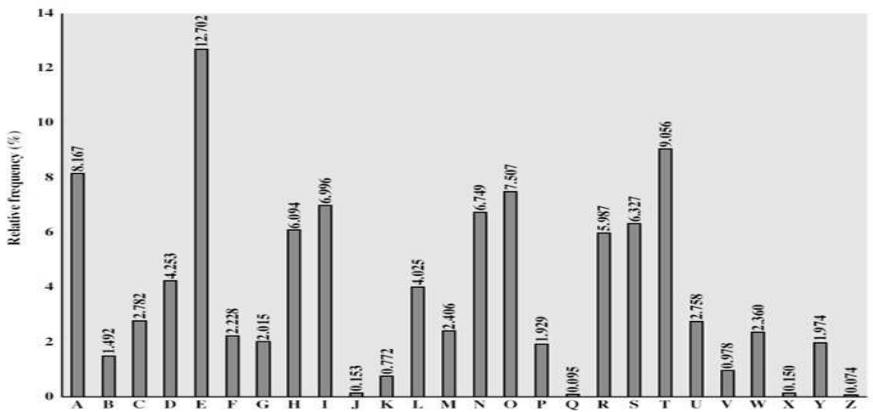


Figure 3: Relative Frequency of letters in the English Language.

- examine the ciphetext in question, and work out the frequency of each letter
- if most common letter in the ciphertext is, for example, J then it would seems likely that this is a substitution for e
- if the second most common letter in the ciphertext is P, then this is probably a substitution for t, and so on
- however, regularities of the language may be exploited, e.g. relative frequency
- frequency analysis requires logical thinking, intuition, flexibility and guesswork

- **Homophones** : Assigning more than one ciphertext symbol to each plaintext
- **Example:** Let $\mathcal{P} = \{a, b\}$, $H(a) = \{00, 10\}$, and $H(b) = \{01, 11\}$.

$$\begin{aligned} aa &\longrightarrow \{0000, 0010, 1000, 1010\} \\ ab &\longrightarrow \{0001, 0011, 1001, 1011\} \\ ba &\longrightarrow \{0100, 0110, 1100, 1110\} \\ bb &\longrightarrow \{0101, 0111, 1101, 1111\} \end{aligned}$$

- Frequency of occurrence of ciphertext symbols are uniform, at the expense of data expansion.

- Two methods are used in substitution ciphers to lessen the extent to which the structure of the plaintext survives in the ciphertext:
  1. **Polyalphabetic ciphers**: Using different monoalphabetic substitutions while moving through the plaintext. The "Vigenère" cipher is the best example
  2. **Multiple letter encryption**: e.g. "Playfair".

Playfair Cipher

- Use the keyword CHARLES (Charles Wheatstone invented the cipher).
- Draw up a 5×5 matrix with the keyword first, removing any repeating letters as follows:

c	h	a	r	l
e	s	b	d	f
g	i/j	k	m	n
o	p	q	t	u
v	w	x	y	z

- Plaintext: "meet me at the bridge".
  - Split the sentence into digrams removing spaces, 'x' used to make even number of letters:

    me et me at th eb ri dg ex
  - Repeating plaintext letters that are in the same pair are separated with a filler letter, such as x:

    "balloon" would be treated as ba lx lo on

– Two plaintext letters in the same row are each replaced by the letter to the right, with the first element of the row circularly following the last.

eb is replaced by sd
ng is replaced by gi (or gj as preferred)

– 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.

dt would be replaced by my
ty would be replaced by yr

– 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.

me becomes gd

- Ciphertext therefore is :

"gd do gd rq pr sd hm em bv"

Rotor Machine

- Multiple stage encryption can result in an algorithm that is significantly more difficult to cryptanalyse.
- Before the arrival of DES the most important application of this principle was a class of systems known as Rotor machines. Rotor machines were used by both Germany (Enigma) and Japan (purple) in WW2 and the breaking of both codes was a significant factor in the war's outcome.

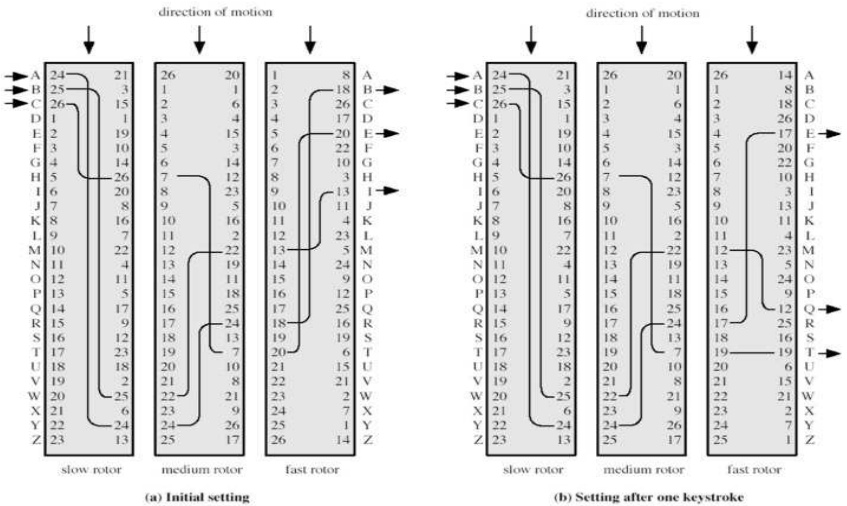


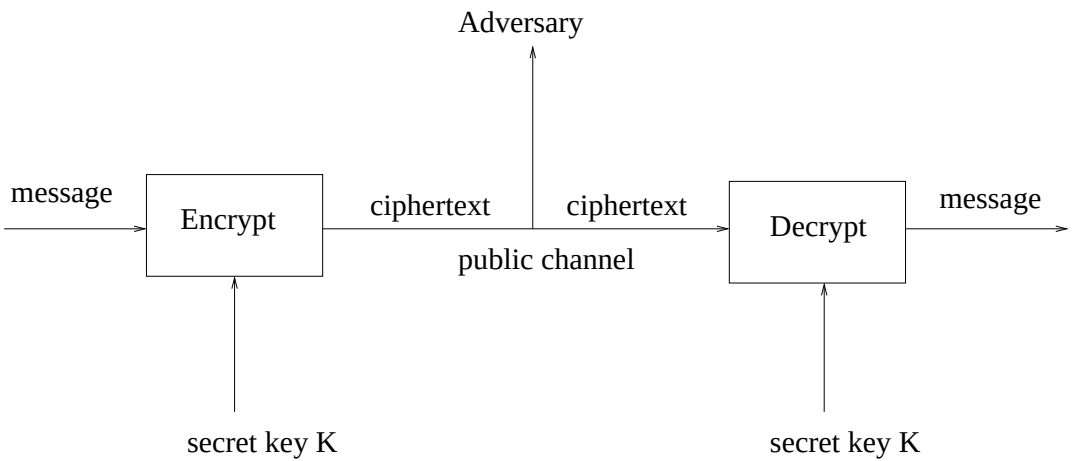
Figure 4: Three Rotor machine with wiring represented by numbered contacts.

- The system consists of a number of independently rotating cylinders through which electrical pulses can flow.
- Each cylinder has 26 input pins and 26 output pins with internal wiring that connects each input to a unique output, e.g. a monoalphabetic substitution (for simplicity only three connections are shown in the figure).

- Now, consider a machine consisting of one cylinder. After each input is depressed, the cylinder moves one position thus creating a different monoalphabetic cipher. After 26 letters of plaintext (a full rotation) the cylinder is back to its original position so we have a polyalphabetic substitution cipher of period 26.
- A single cylinder is trivial but the use of multiple cylinder is formidable.

- Figure 4 is a three cylinder system and the output pins of one cylinder are connected to the input pins of the next one. The LHS of figure 4 shows a position where the input from the operator to the first pin (plaintext letter A) is routed through three cylinders to appear at the output of the second pin (ciphertext B).
- With multiple cylinders the one furthest from the operator rotates one pin position with each operator keystroke. For every rotation of the middle cylinder the inner/slow cylinder rotates one pin position.
  $26 \times 26 \times 26 = 17,576$ substitution alphabets before repeating.

Symmetric Key Encryption



- Block ciphers and stream ciphers are two types of symmetric key cryptosystems

Block Cipher

- plaintext string $x = x_1 x_2 x_3, \cdots$
- Successive plaintext elements are encrypted using the same key k:

Key:	k			
Plaintext:	x_1	x_2	x_3	$\cdots$
Ciphertext:	$e_k(x_1)$	$e_k(x_2)$	$e_k(x_3)$	$\cdots$

- ciphertext string $y = y_1 y_2 y_3, \cdots = e_k(x_1) e_k(x_2) e_k(x_3) \cdots$
- **Examples:** DES, Rijndael (The AES), IDEA, RC6, and many more....

Stream Cipher

- plaintext string $x = x_1 x_2 x_3, \cdots$
- generate a keystream $k_1, k_2, k_3 \cdots$ from the key k:

Key:	k			
keystream:	k_1	k_2	k_3	$\cdots$
Plaintext:	x_1	x_2	x_3	$\cdots$
Ciphertext:	$e_{k_1}(x_1)$	$e_{k_2}(x_2)$	$e_{k_3}(x_3)$	$\cdots$

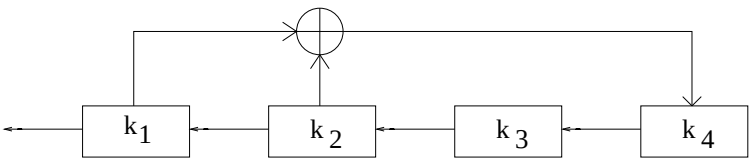
- ciphertext string
  $y = y_1 y_2 y_3, \cdots = e_{k_1}(x_1) e_{k_2}(x_2) e_{k_3}(x_3) \cdots$

Example

- Consider message to be a bit stream $m_1, m_2, \ldots$
- Let $k_1, k_2, \ldots$ be a sequence of pseudorandom bits.
- Encrypt: $c_i = m_i \oplus k_i$.
- The cipher is $c_1, c_2, \ldots$
- Decrypt: $m_i = c_i \oplus k_i$

- Security depends upon the sequence $k_1, k_2, \ldots$
- If k_i is a true random sequence, then the cipher is called an one-time pad.
- One-time pad possesses *perfect secrecy*.
- One-time pads are impractical.
- Use a pseudorandom generator. Secret key of the system is the "seed" of the pseudorandom generator.

Linear Feedback Shift Register (LFSR)



- An LFSR of length m consists of m stages numbered $1, 2, \ldots, m$, each storing one bit and having one input and one output; together with a clock which controls the movement of data.
- The vector $(k_1, k_2, \cdots, k_m)$ would be used to initialize the shift register

- During each unit of time the following operations would be performed concurrently

  (i) k_1 would be tapped as the next keystream bit

  (ii) $k_2, \cdots, k_m$ would each be shifted one stage to the left

  (iii) the "new" value of k_m would be computed to be

$$\sum_{j=1}^{m-1} c_j k_{j+1}$$

  the linear feedback is carried out by tapping certain stages of the register (as specified by the constants c_j having the value "1") and computing a sum modulo 2 (which is an exclusive-or).

Modern Cryptography

Modern cryptography started in 1976 with the publication of the following paper.

- W. Diffie and M.E.Hellman. "New directions in cryptography". IEEE Transactions on Information Theory, 22 (1976) 644-654.

Importance of Cryptography

- The fast paced development of digital computers.
- Increasing dependence of industry on computers.
- Internet.
- Significant development of communication networks.
- New network problems: sensor network.

Cryptographic Primitives

- One-way function.
- Symmetric encryption.
- Asymmetric encryption.
- Digital Signatures.
- Hash Functions.
- Protocols.
- Other primitives.

Cryptographic Security

- **Kerckhoff's Principle:** Assume that the adversary knows the algorithm that is used. The secret is *only* the secret key.
- **Attack Models:**
  - Ciphertext only attack: The opponent possesses a string of ciphertext, y
  - Known plaintext attack: The opponent possesses a string of plaintext, x, and the corresponding ciphertext, y

– Chosen plaintext attack: The opponent has obtained temporary access to the encryption machinery. Hence he can choose a plaintext string, x, and construct the corresponding ciphertext string, y.

– Chosen ciphertext attack: The opponent has obtained temporary access to the decryption machinery. Hence he can choose a ciphertext string, y, and construct the corresponding plaintext string, x.

- **Adversarial Goal:**
  - Key recovery.
  - Distinguishing attack.
  - Malleability.
  - Other application specific security goals.