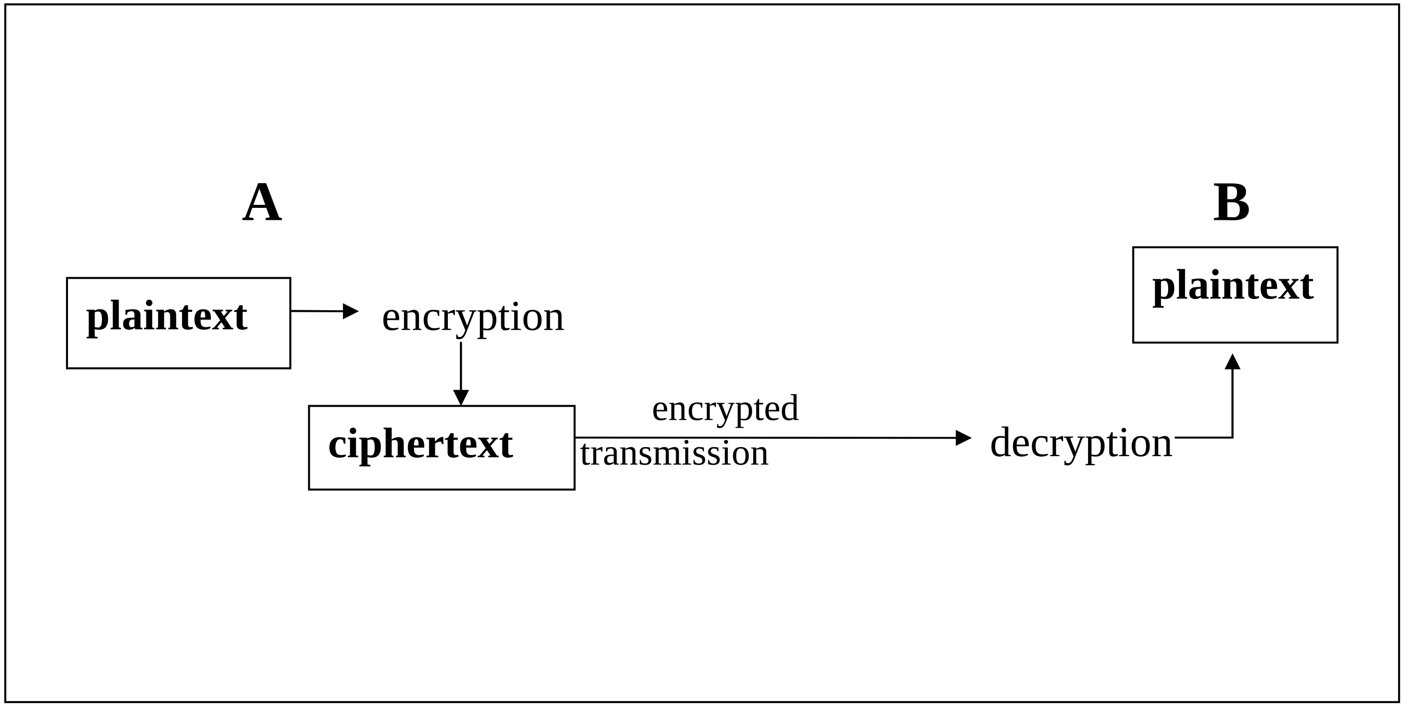
Chapter 2

Basic Encryption and Decryption

Encryption / Decryption

- A: sender; B: receiver
- Transmission medium
- An interceptor (or intruder) may *block*, *intercept*, *modify*, or *fabricate* the transmission.



Encryption / Decryption

- **Encryption**: A process of encoding a message, so that its meaning is not obvious. (= encoding, enciphering)
- **Decryption**: A process of decoding an encrypted message back into its original form. (= decoding, deciphering)
- A **cryptosystem** is a system for encryption and decryption.
- **Plain text**: The original form of a message
- **Cipher text**: The encrypted form of a message

Plaintext / ciphertext

- P: plaintext
- C: ciphertext
- E: encryption
- D: decryption
- C = E (P)
- P = D (C)
- P = D (E(P))

Cryptosystems

- **Symmetric** encryption:

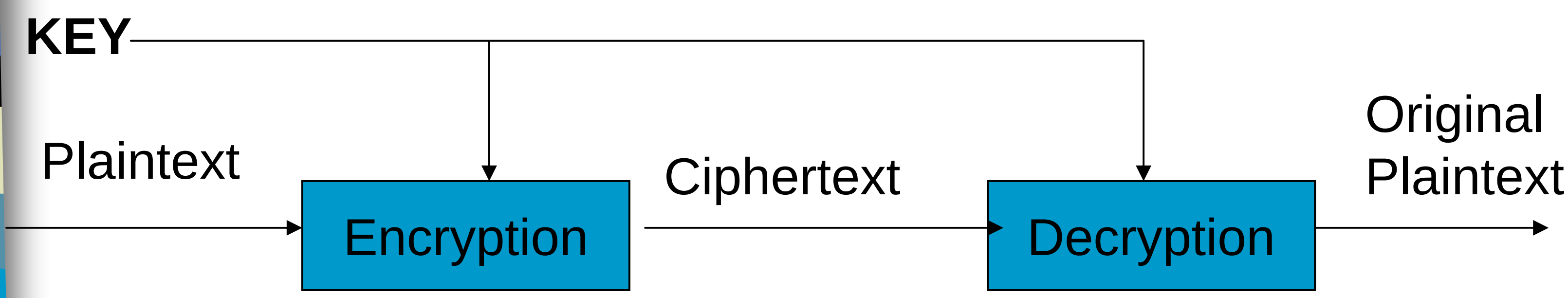
  $P = D(Key, E(Key, P))$

- **Asymmetric** encryption:

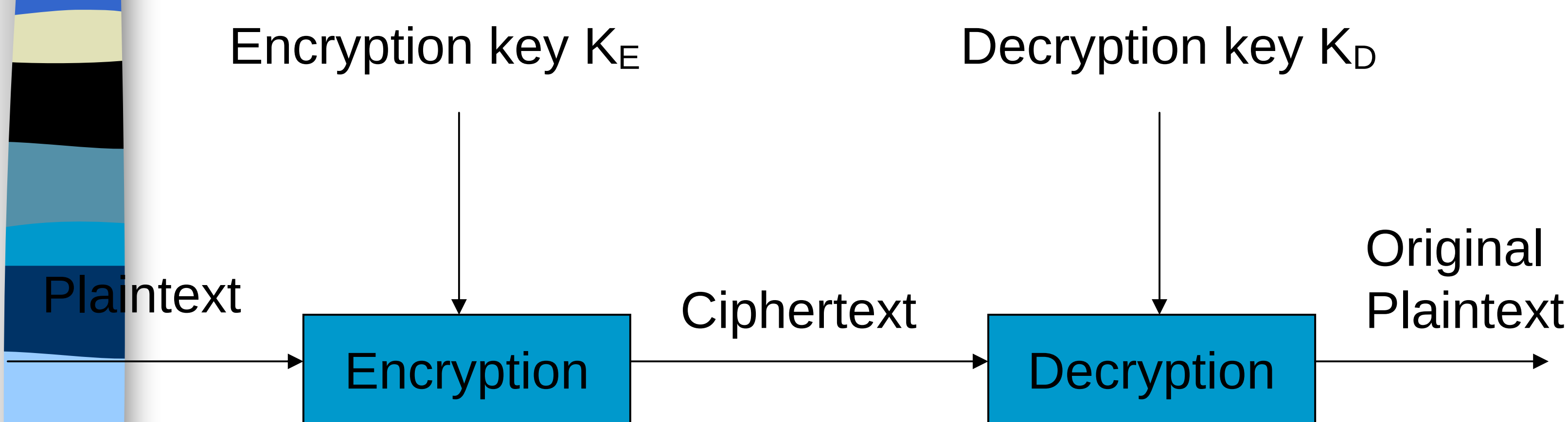
  $P = D(Key_D, E(Key_E, P))$

- **Symmetric cryptosystem**: A cryptosystem that uses symmetric encryption.
- **Asymmetric cryptosystem**

Encryption Algorithms (contd.)



Symmetric Cryptosystem



Asymmetric Cryptosystem

Keys

- Use of a key provides additional security.
- Exposure of the encryption **algorithm** does not expose the future messages (as long as the **key**(s) are kept secret).

Terminology

- **Cryptography:** The practice of using encryption to conceal text. (**cryptographer**)
- **Cryptanalysis:** The study of encryption and encrypted messages, with the goal of finding the hidden meanings of the messages. (**cryptanalyst**)
- **Cryptology** = cryptography + cryptanalysis

Cryptanalysis

- A cryptanalyst may work with various data (intercepted messages, data items known or suspected to be in a ciphertext message), known encryption algorithms, mathematical or statistical tools and techniques, properties of languages, computers, and plenty of ingenuity and luck.

1. Attempt to break a single message
2. Attempt to recognize patterns in encrypted messages
3. Attempt to find general weakness in an encryption algorithm

Modular arithmetic

- Known as mod n for example n=26 (since we have 26 English letters)
- A .. Z: 0 .. 25
- *Perform operations on letters for ex.:*
  - *A+3=D*
  - *D-3=a*
  - *A-1= Z*

Two forms of encryption

- **Substitutions**

  One letter is exchanged for another

  For example: ABC$\mapsto$DEF
- **Transpositions (= permutations)**

  The order of the letters is rearranged

  For example: ABC$\mapsto$CBA

Simple Encryption Algorithms

- Substitution
  – Monoalphabetic (Ceaser, Using a key, Random)
  – Polyalphabetic (two or more, Using a key, Perfect)
- Permutation
  – Simple
  – Double

Ceaser Cipher

The Caesar Cipher

- Each letter is translated to the letter a fixed number of letters after it in the alphabet.
- Uses a shift of 3
- Plain text $\mathbf{p_i}$ is used to obtain cipher text $\mathbf{c_i}$ as follows:

$$\mathbf{c_i = E(p_i) = p_i + 3}$$

Ceaser Cipher

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
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

- Example
  - Plain: This is the message
  - Cipher: wklv lv wkh phvvdjh

Other monoalphabetic substitutions

- Using Shift (**permutation)**
  - **(lambda) = 25 – lambda**
- Using a key

  A key is a word that controls the ciphering.

  Example: The key is (key)

	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
K	E	Y	A	B	C	D	F	G	H	I	J	L	M	N	O	P	Q	R	S	T	U	V	W	X	Z

Another Example Key is Juliette

JULIETABCDFGHKMNOPQRSVWXYZ

Note:

- Use Large keys
- Use Keys that contain letters from the end of the alphabet

Other monoalphabetic substitutions

- Random monoalphabetic substitution
- Example

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	W	Z	Q	J	D	V	K	X	P	B	O	L	F	R	M	N	C	S	U	H	E	Y	T	G	I

Note: Each letter should appear only once

Cryptanalysis example

- Consider the following ciphertext:

WKLV PHVVDJH LV QRW WRR KDUG WR EUHDN

Try to make cryptanalysis to conclude the full or part of the original message

Comparaison of the security of all monoalphabetic ciphering

- All easy to perform
- All unsecure enough
- In order of highest to lowest security
  - Random (need to know all letters substitutions)
  - Using a key (need to know all the key)
  - Ceaser or shift (need to know the size of the shift)

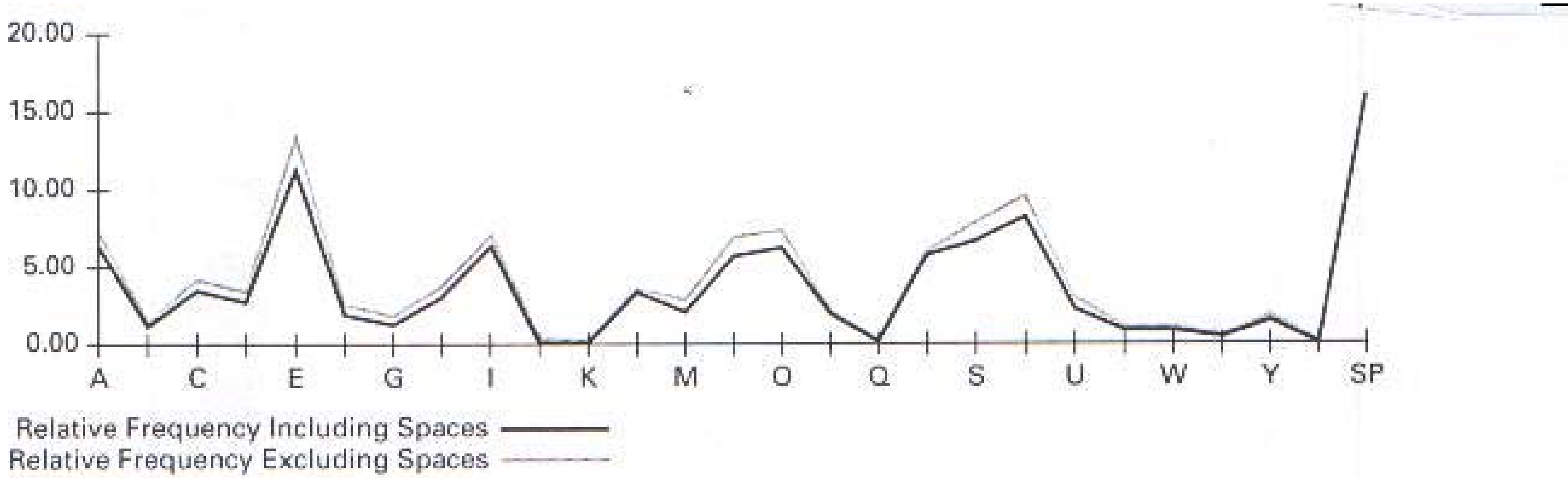
Weakness of the security of all monoalphabetic ciphering

- Blanks are not encrypted (blanks remain blanks)
- As English has few small words, the two and three letters words are easily guessed after few trials
- The same two letters are encrypted as same two other letters and not all two letters are repeated.

Frequency distribution

- It shows the counts and **relative frequencies** of letters in English.
- For example the frequencies of Ceaser's cipher are shifted three letters from the normal distribution
- Frequencies in a sample cipher: Compare the sample cipher against the normal frequency distribution.

Frequency distribution



Relative Frequencies of Characters in English Text

Another disadvantage of monoalphabetic cipher

- It always has lot o peaks and valleys and
- With comparison of the normal distribution the exact substitution can be guessed
- To solve this we have Polyalphabetic cipher

Polyalphabetic substitutions

- By combining distributions that are high with the ones that are low
- Using multiple substitution tables
- Example :One table for odd positions, One table for even positions

Polyalphabetic substitutions

- EXAMPLE:
  - ABCDEFGHIJKLMNOPQRSTUVWXYZ
- S1:DEFGHIJKLMNOPQRSTUVWXYZABC
- S2: ZYXWVUTSRQPONMLKJIHGFEDCBA
  - P= IMPOSSIBLE
  - C= LNSLVHMYOV

Polyalphabetic substitutions (Vigenère Tableaux)

- Table 2-1: p.51
- Example: p.50
- P= I am I exist that is certain
- Key: Secure
- C= aeocv pmunk zevej vitnr ar
- Exercise: Complete the encryption of the following message
  – Plain: BUT SOFT WAS LIGHT
  – KEY: JULIETTE

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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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

Cryptanalysis of Polyalphabetic cipher.

- Find the number of alphabets used
- Break cipher into chains that were ciphered with the same alphabet
- Solve each piece as a monoalphabetic substitution
  - Two tools can help:
    - Kasiski method
    - Index of coincidence

Cryptanalysis of Polyalphabetic substitutions (kasiski method)

- The **Kasiski method**: a method to find the number of alphabets used for encryption
- Works on duplicate fragments in the ciphertext
- The distance between the repeated patterns must be a multiple of the keyword length

Kasiski Method

- Relies on regularity of English
- Not only letters but also groupings
  - Ending: th, ing, ed, ion, tion, ation, etc.
  - Beginning: im, in, un, re, etc.
  - Patterns: eek, oot, our, etc.

Steps of The Kasiski Method

- Identify all repeated patterns in the ciphertext
- For each pattern:

1. Write down the starting position
2. Compute the distance between successive starting positions
3. The distance must be a multiple of key length
4. Determine all factors of difference (eg. 16 means the key is either 1, or 2 or 4 or 8 characters long)
5. Divide the cipher into chains (based on key length)
6. Columns in each chain must have been encrypted using the SAME key character
7. Use frequency distribution to determine the letter

Index of coincidence

- A measure of the variation between frequencies in a distribution
- To measure the *nonuniformity* of a distribution
- To rate how well a particular distribution matches the distribution of letters in English

Index of Coincidence

- **Measure of roughness** (or the **variance**): a measure of the size of the peaks and valleys.
- The var of a perfectly flat distribution = 0.
- *The **variance** can be estimated by counting the number of pairs of identical letters and dividing by the total number of pairs possible*

Index of Coincidence

- **IC** (index of coincidence): a way to approximate variance from observed data.
- 0.0384 (perfect) ≤ IC < 0.068 (English)

$$IC = \sum_{\lambda=a}^{\lambda=z} \frac{\text{Freq}_\lambda * (\text{Freq}_\lambda - 1)}{n * (n-1)}$$

Combined use of IC with Kasiski method

- All the chains from the Kasiski method, if the key length was correct, should have distributions close to 0.068.

Table 2-6 Number of Enciphering Alphabets Versus Index of Coincidence

Alphabets	1	2	3	4	5	10	Large
IC	.068	.052	.047	.044	.044	.041	.038

Analyzing a polyalphabetic cipher

1. Use the Kasiski method to predict the likely numbers of enciphering alphabets.
2. Compute the IC to validate the predictions
3. (When 1 and 2 show promises) Generate chains and calculate IC's for each chains

The “Perfect” Substitution Cipher

- Use an infinite *nonrepeating* sequence of alphabets
- A key with an infinite number of nonrepeating digits
- Examples: one-time pads, the Vernam cipher, …

Vernam Cipher

- Sample function:

  c = (p + random( )) mod 26

- Example: p.42
- Binary Vernam Cipher

  p.43: How would you decipher a ciphertext encrypted by Binary Vernam Cipher?

  Ans.: p = (c – random( )) mod 26

  Example (p.42): c = 19 (‘t’), random( ) = 76

  p = (19-76) mod 26 = 21

Random Number Generators

- A pseudo-random number generator is a computer program that generates numbers from a predictable, repeating sequence.
- Example: The **linear congruential random number generator**

  $r_{i+1} = (a * r_i + b) \bmod n$

  Note: 12 is *congruent* to 2 (modulo 5), since (12-2) mod 5 = 0
- Problem? Its dependability; probable word attack

Breakability of an encryption

- An encryption algorithm may be **breakable**, meaning that given enough time and data, an analyst could determine the algorithm.
- Suppose there exists 10^{30} possible decipherments for a given cipher scheme. A computer performs 10^{10} operations per second. Finding the decipherment would require 10^{20} seconds (or roughly 10^{12} years).

Summary

- Substitutions and permutations together form a basis for the most widely used encryption algorithms