What is Cryptography?

by

Amit Konar,

Dept. of Math and CS, UMSL

Definition: Cryptosystem

Cryptography means secret writing and it is the art of concealing meaning.

A **Cryptosystem** is a 5-tuple(E, D,M,K,C), where

M is the set of plain texts,

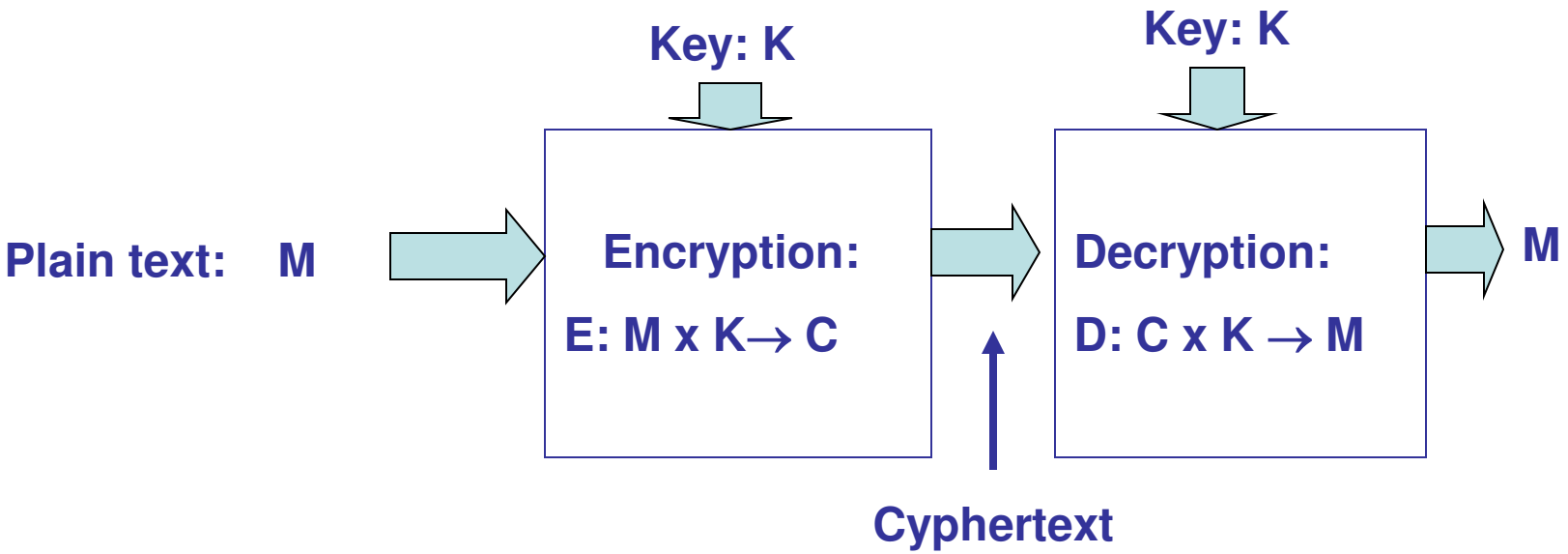
K is the set of keys

C is the set of ciphertexts,

E: M x K$\rightarrow$C is the set of enciphering functions, and

D: C x K$\rightarrow$M is the set of deciphering functions.

A Simple Cryptosystem



Example: M= A B;

K=3;

C= D E

Example: Caesar Cipher

This cipher is a cryptosystem with

M={all sequence of roman letters}

= HELLO= [7 4 11 11 14

K={i | an integer, such that 0<=i<=25}=3,say.

E={E_k |k$\in$K and for all m$\in$M,E_k(m) =(m +k) mod 26}= [10 mod 26, 7 mod 26….]

D={D_k|k $\in$K and for all c$\in$C, D_k(c)= (26+c-k)mod26} = {(26+10-3) mod 26,…}

Classical Cryptosystems

Transposition Ciphers: HELLO WORLD changed to HLOOL

ELWRD, resulting the ciphertex tHLOOLELWRD.

Substitution Ciphers:

key: V IGVIG V IGVIGV I G

plaintext: THEBOY HASTHEBAG

Ciphertext: OPKWWE C IY OPK W I M

Goal of Cryptography

1. **Ciphertext Only Attack:** The adversary has only the ciphertext. Her goal is to find the corresponding plaintext. If possible, she may find the key too.
2. **Plaintext Attack:** The adversary has the ciphertext and the plaintext. Her goal was to find the key used.
3. **Chosen Plaintext Attack:** The adversary may ask that specific plaintexts be enciphered. She is given the corresponding ciphertexts. Her goal is to find the key that was used.

Models Used for Attack

Attacks use both mathematics and statistics.

The statistical methods make assumptions about the statistics of the plaintext language and examine the ciphertext to corelate the properties with those assumptions. These assumptions are collectively called **models** of the language.

n-Gram Models

N-gram models represent **frequency of occurrence of n-consecutive letters** in the English language.

Example: 1-gram model describes frequency of occurrence of a character A, B, C etc. in the English language. Their selection is independent.

Table of character frequencies in the English language

a 0.080

b 0.015

c 0.030

d 0.040

e 0.130

f 0.020

.....

z 0.002

2-gram(digrams) models with one character as h

he 0.0305

ho 0.0043

wh 0.0026

eh 0.0002

How to get these digram models?

Paragraph: Hello! Does **he** know you? ….

Frequency of **he**: 2, freq. of other digrams: **oe: 1**, **ou: 1**, **ow: 1**, **lo: 1**. **Normalized freq**. of **he**= 2/(2+1+1+1)= 2/5.

How digrams can be used to retrieve plaintexts in transposition ciphers?

Recall that a transposition cipher rearranges characters in the plaintext to form ciphertext. Characters are not changed.

Example: Plaintext: HELLO WORLD
Ciphertext: **HLOOL**ELWRD

Consult Konheim's digram table and note that **freq(he) = 0.0305 > freq(ho) (=0.0043) > freq(hl) >freq(lh) etc. So, the first two leters of the plaintext may be HE.** This is possible if we write

HLOOL	by	HE
ELWRD		LL
		OW
		OR
		LD

Reading the letters across and down produces HELLO WORLD.

How 1-grams can be used to retrieve plaintext in Substitution Ciphers

Example: Caesar cipher is a substitution cipher. Given plaintexts: HELLO WORLD, and ciphertexts: KHOOR ZRUOG, key=3.

Let

$$\Phi(i)=\sum_{0<=c<=25} f(c)\ p(c-i)$$

be the co-relation of the freq. of each letter in the ciphertext with the character freq. in English. f(c)= freq. of a char. in the ciphertext; p(c-i)= prob. of occurrence of the i-position shifted character from c.

Main Steps to retrieve plaintexts in Substitution Ciphers

Example (Contd.): Plaintexts: HELLO WORLD

Key=3; Ciphertexts:KHOOR ZRUOG

Freq. of letters in ciphertexts

G 0.1 H 0.1 K 0.1 O 0.3 R 0.2 U 0.1 Z 0.1

$\Phi(i)=\sum_{0<=c<=25} f(c)\, p(c-i)$

$= 0.1\, p(6-i) + 0.1\, p(7-i) + 0.1\, p(10-i) +$

$0.3\, p(14-i) + 0.2\, p(17-i) + 0.1\, p(25-i)$

Find 0<=i<=25 such that **Φ(i) is maximized**.

Computing Φ(i) for varying i

Part of Table Showing Computed Results for Φ(i)

……..	**Deciphered Plaintext**
Φ(6) = 0.0660	EBIL TLOIA
Φ(10)= 0.0635	AXEEH PHKEW
Φ(3) = 0.0575	***HELLO WORLD***
Φ(14)= 0.0535	WTAAD LDGAS

Ciphertexts: KHOOR ZRUOG

Conclusion: i=3 as Φ(i) for other i does not produce meaningful texts.

Vigenere Cipher

Given a **vignere table, how to encipher?**

Plaintext:	A	B	E	G	H	S	T	Y
Key								
G	G	H	K	M	N	Y	Z	E
I	I	J	M	O	**P**			
V							**O**	
Key:	VIG	VIG	VIG	VIG	VIG			
Plaintest:	THE	BOY	HAS	THE	BAG			
Ciphertext:	**OPK**	WWE	CIY	OPK	WIM			

Attacking Vigenere Cipher by Index of Coincidence

What is index of Coincidence?

The index of coincidence measures the differences in the frequencies of the letters in the ciphertexts.

It is defined as the **probability that two randomly chosen letters from the ciphertext will be the same**.

Formal Definition of Index of Coincidence

Let

F_c be the freq. of the cipher character c,

N be the length of the ciphertext.

Then index of Coincidence(IC) is given by

$$IC = \frac{1}{N(N-1)} \sum_{c=0}^{25} F_c (F_c - 1)$$

The lower the IC, the less the variation in the characters, and longer the period

Period: gap between repeatations

Table of IC

Period:	1	2	3	Large
Expected IC:	0.06	0.05	0.04	0.038

Hence, lower the IC, longer the period.

Main Steps: Guessing Plaintext from Ciphertext

1. List repetitive words (digrams, 3-grams etc.), determine their positions in the ciphertext and hence evaluate gap length.
2. Determine factors of gap length. (eg. If gap length is 10, factors are: 2, 5).
3. Determine the length of longest repetitive word. The key length should be equal or more than the longest repetitive word.

Main Steps (Contd.)

4. Identify the next longest repetitive n-gram, and note its gap length.
5. Determine the greatest common divisor (GCD) of the gap length for the first two longest n-grams. Guess this to be the length of the Key.
6. Divide the ciphertext into strings of 6 character, and write them row-wise.

Mian Steps: Retrieving Plaintexts (contd.)

7. Suppose you obtain:

A D Q Y S M

I U S B O X

K K T M I B

.....

I U I X

Evaluate Column-wise Index of Coincidence for 6 alphabet sequence.

Retrieval of Plaintext (Contd.)

8. Check from Index of Coincidence values: whether period of the alphabet sequences are different. This can be performed by matching with the table of IC versus periods. If no, fine, else new period may be presumed.

Retrieval (Contd.)

9. Determine frequency of character in each alphabet.

Column	A	B	C	D	E	F	G	H	….
Alpha#1	3	1	0	0	4	0	1	1	
Alpha#2	1	0	0	2	2	2	1	0	
Alpha#3	1	2	0	0	0	0	0	0	
Alpha#4									
Alpha#5									
Alpha#6									
Standard:	H	M	M	M	H	M	M	H	……..

1. The first alphabet matches with standard.
2. Given the gap between B and I, the third alphabet seems to shifted with I mapping to A.
3. A similar gap occurs in the sixth alphabet between O and V, suggesting V maps to A.

This makes three changes in every 6-lettered ciphertexts.

Add more heuristics to obtain final solution: plaintexts

1. In the last row one group of 3-characters is AJE, which should be replaced by ARE.
2. In English Q is always followed by U.
3. Now isolate words from the running string of characters using a standard English dictionary.