INFORMATION SECURITY
LECTURE 3
CRYPTANALYSIS OF SYMMETRIC CIPHERS

Khalid Abdullah

MSc computer and Network Security

Computer Science Dept.

University Of Zakho

Last Lecture

- **Some basic definitions**
  - Cipher
  - Breaking the code
  - Cryptology
- **Symmetric cryptography**
- **Substitution Techniques**
  - Caesar Cipher
  - Polyalphabetic Ciphers (Vigenere Cipher)

OUTLINE

- Vernam or one time pad
- Cryptanalysis of symmetric ciphers
- Cryptography Terminology
- Frequency distribution table
- Kasiski method

One-Time Pad or Vernam Cipher

- The one-time pad, which is a provably secure cryptosystem,was developed by Gilbert Vernam in 1918.

- The message is represented as a binary string (a sequence of 0's and 1's using a coding mechanism such as ASCII coding.

- The key is a truly random sequence of 0's and 1's of the same length as the message.

- The encryption is done by adding the key to the message , bit by bit.

This process is often called *exclusive or,* and is denoted by *XOR. The symbol ⊕ is used.*

A	B	C = A ⊕ B
0	0	0
0	1	1
1	0	1
1	1	0

ONE-TIME PAD OR VERNAM CIPHER

- **Example:** Let the message be IF then its ASCII code **be**(1001001 1000110) and the key be (1110110 0110001).The ciphertext can be found exoring message and key bits.
- *Encryption:*

1001001 1000110	plaintext
1110110 0110001	key
0111111 1110111	ciphertext

Decryption:

0111111 1110111	ciphertext
1110110 0110001	Key
1001001 1000110	Plainteaxt

One-Time Pad or Vernam

Joseph Mauborgne, proposed an improvement to the Vernam cipher that yields the ultimate in security.

Properties:

- using a random key that is as long as the message, so that the key need not be repeated.
- the key is to be used to encrypt and decrypt a single message, and then is discarded.
- It 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 (Such a scheme, known as a **one-time pad, is unbreakable).**

THE ONE-TIME PAD OFFERS COMPLETE SECURITY BUT, IN PRACTICE, HAS TWO FUNDAMENTAL DIFFICULTIES:

- There is the practical problem of making large quantities of random keys. Supplying truly random characters in this volume is a significant task.
- *In other words: There are problems to generate a large amount of random numbers.*
- Even more daunting is the problem of ***key distribution*** and ***protection***. For every message to be sent, a key of equal length is needed by both sender and receiver. Thus, a mammoth key distribution problem exists.
- Because of these difficulties, the one-time pad is of limited utility, and is useful primarily for low-bandwidth channels requiring very high security.

Two forms of encryption

- **Substitutions**
  One letter is exchanged for another
  Examples: monoalphabetic substitution ciphers, polyalphabetic substitution ciphers
- **Transpositions (= permutations)**
  The order of the letters is rearranged
  Examples: columnar transpositions

CRYPTANALYSIS OF SYMMETRIC CIPHERS

- ***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.

Cryptanalysis Steps to break ciphertext

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

Cryptanalysis of the Caesar Cipher

- Deduction based on guesses versus frequency distribution letter.
- It easy to Decrypt the encrypted message by brute force attack (try all possible key which is 25).

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).

Frequency of occurrence of letters in English Text

Letter	Frequency	Letter	Frequency	Letter	Frequency	Letter	Frequency
E	12.7	H	6.1	W	2.3	K	0.08
T	9.1	R	6.0	F	2.2	J	0.02
A	8.2	D	4.3	G	2.0	Q	0.01
O	7.5	L	4.0	Y	2.0	X	0.01
I	7.0	C	2.8	P	1.9	Z	0.01
N	6.7	U	2.8	B	1.5		
S	6.3	M	2.4	V	1.0		

GROUPING OF DIGRAMS AND TRIGRAMS BASED ON THEIR FREQUENCY IN ENGLISH.

Digram	TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU, EA, NG, AS, OR, TL, IS, ET, IT, AR, TE, SE, HL, OF
Trigram	THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR, DTH

Cryptanalisis of Vigenère cipher

- The idea behind Vigenère’s cipher, similarly to other polyalphabetic ciphers, is to disguise plaintext letter frequencies, which interferes with a straightforward application of frequency analysis.

- For instance, if W is the most frequent letter in a ciphertext whose plaintext is in English, one might suspect that W corresponds to E, because E is the most frequently used letter in English.

- However, with the Vigenère cipher, E can be encrypted with different ciphertext letters in different points in the message, thus making simple frequency analysis difficult.

Vigenère cipher

- Strength of Vigenère's cipher hinges (depend) on the secrecy of its key length.
- Indeed, if one guesses correctly the length of the key, then ciphertext can be viewed as produced by interwoven Caesar ciphers (each of which can be easily broken!).
- Thus, one possible weakness of Vigenère's cipher is the repeating nature of its key.
- The Kasiski test or Ciphertext autocorrelation can help to determine the key length.

KASISKI TEST

- Take advantage of the fact that certain common words like "the"will, by chance, be encrypt using the same key letters, leading to repetitions in the ciphertext.

- Example. A message encoded with key DECEPTIVE might not encrypt the Letters "*E*" the same way each time "*E*" appears in the plaintext:
- Plaintext: **WE AREDISCOVERED SAVE YOUR SELF**
- Key: **D E CEPTIVEDECEPT IVED ECEP TIVE**
- Cipphertext: **Z I CVTWQNGRZGVTW AVZH CQYG L MGJ**

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

KASISKI TEST

Kasiski test

Example. The same message encrypted with the keyword ABCD results in:

Key: **ABCDAB CD ABCDA BCD ABCDABCDABCD**

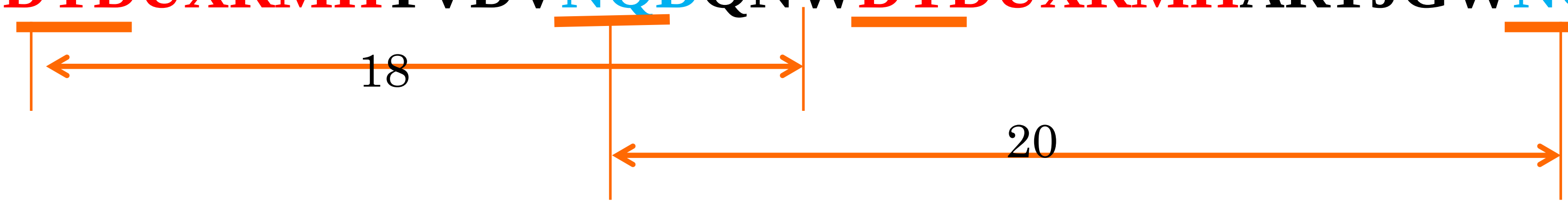
Plaintext: **CRYPTO IS SHORT FOR CRYPTOGRAPHY**

Ciphertext: **CSASTP KV SIQUT GQU CSASTPIUAQJB**

- Notes.
- 1. Shorter keys help the test.
- 2. Longer messages help the test (corresponding ciphertext usually contains more repetitions)

KASISKI TEST

- Kasiski test
- Ciphertext: **DYDUXRMHTVDVNQDQNWDYDUXRMHARTJGWNQD**



- Repetition at distance 18: key length could be {1,2,3,6,9,18}
- Repetition at distance 20: key length could be {1,2,4,5,10,20}

- Taking the intersection of these sets, one could argue that the Key length is probably 2 (why not 1?)

Kasiski test

- If two identical sequences of plaintext letters occur at distance that is an integer multiple of the keyword length, they will generate identical ciphertext sequences.
- In the foregoing example, two instances of the sequence "red" are separated by nine character positions.
- Consequently, in both cases, r is encrypted using key letter e, e is encrypted using key letter p, and d is encrypted using key letter t. Thus, in both cases the ciphertext sequence is VTW.
- the keyword DECEPTIVE, the letters in positions 1, 10, 19, and so on are all encrypted with the same monoalphabetic cipher.

KASISKI TEST

- Thus, we can use the known frequency characteristics of the plaintext language to attack each of the monoalphabetic ciphers separately.
- Example: Let us assume we have intercepted the following ciphertext

- LIOMWGFEGGDVWGHHCQUCRHRWAGWIOWQLKGZETKKMEVLWPCZVGTHVTSGXQOVGCSVETQLTJSUMVWVEUVLXEWSLGFZMVVWLGYHCUSWXQHKVGSHEEVFLCFDGVSUMPHKIRZDMPHHBVWVWJWLXGFWLTSHGJOUEEHHVUCFVGOWICQLTJSUXGLW

String	First Index	Second index	Difference
JSU	68	168	100
SUM	69	117	48
VWV	72	132	60
MPH	119	127	8

Kasiski Test

- The greatest common divisor of differences is 4. which mean that the key length is multiple of 4.
- First try M= 4 .
- Divide the ciphertext into four pieces.
- Pieces C1 is made of characters 1,5,9,13,17….
- Pieces C2 is made of characters 2,6,10,14,18…..
- Pieces C3 is made of characters 3,7,11,15,19…
- Pieces C4 is made of characters 4,8,12,16,20….

C1:LWGWCRAOKTEPGTQCTJVUEGVGUQGECVPRPVJGTJEUGCJG

P1:

C2:IGGGQHGWGKVCTSOSQSWVVWFVYSHSVFSHZHWWFSOHCOQSL

P2:

C3:OFDHURWQZKLZHGVVLUVLSZWHWKHFDDUKDHVIWHUHFWLUW

P3:

C4:MEVHCWILEMWVVXGETMEXLMLCXVELGMIMBWXLGEVVITX

P4:

According the most frequency letter:

We assume that **V** or **G** = e

When G =e then the key letter will be C

We assume that the W (16 Frequency letter)= T

Then the key letter will be D.

We assume H = T as well, then the third letter key is o.

Finally the L could be a,o,i,n,s,h,r which all close to each other. Then the key letter could be s,x,d,y,t,e,u

Letter	Frequency	Letter	Frequency
L	12	W	16
I	4	G	17
O	5	F	6
M	6	E	10
C	8	Q	6
V	18	H	13

We have these letter C D O or s,x,d,y,t, e,u

Which word can we create from these letters and make sense? CODE

OUTLINE

- Vernam or one time pad
- Cryptanalysis of symmetric ciphers
- Cryptography Terminology
- Frequency distribution table
- Kasiski method

What Next

- transposition cipher
- Column transposition cipher
- Double Transposition Cipher