

MAS 433: Cryptography

Lecture 3
Classical Ciphers (Part 2)

Wu Hongjun

Lecture Outline

- Shift cipher (Caesar cipher)
- Substitution cipher, frequency cryptanalysis
- Vigenere Cipher
- Transposition (permutation) cipher

How to improve substitution cipher?

- Problem with substitution cipher
 - The frequency of occurrence of letters in plaintext is not randomized by the substitution cipher
- How to hide the statistics of a language after encryption?
 - Two main approaches:
 - Homophonic substitution
 - Polyalphabetic substitution

How to improve substitution cipher?

- Homophonic substitution
 - the substitution is no longer bijective
 - may consider it as one-to-many mapping
 - Example:
 - suppose that there are 1000 symbols in ciphertext
 - ‘E’ is encrypted to one of 127 ciphertext symbols (randomly)
 - ‘Z’ is encrypted to one ciphertext symbol
 -
 - so the symbols in ciphertext appear uniformly distributed
 - How to break it (for a long message)?
 - Try to find out what ciphertext symbols related to one letter

How to improve substitution cipher?

- Polyalphabetic substitution
 - Multiple substitution alphabets
 - to hide the statistical feature of a language
 - Vigenere is the best-known polyalphabetic cipher
 - Inventor: Giovan Battista Bellaso, 1553
 - Believed unbreakable until 19th century

Vigenere Cipher

- Use a number of shift ciphers
- Definition

Plaintext: $P = (\mathbb{Z}_{26})^n$

Ciphertext: $C = (\mathbb{Z}_{26})^n$

Key: $K = (\mathbb{Z}_{26})^m$ (key consists of m letters)

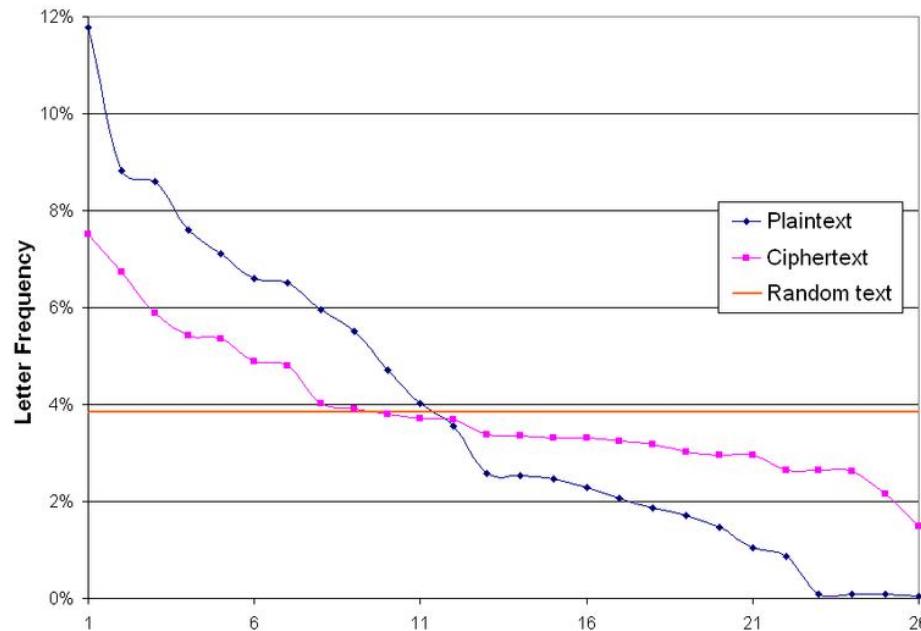
Encryption: $C_i = (P_i + K_i \text{ mod } m) \text{ mod } 26$

Decryption: $P_i = (C_i - K_i \text{ mod } m) \text{ mod } 26$

- Example
 - Plaintext: C R Y P T O G R A P H Y
 - Key: L U C K L U C K L U C K
 - Ciphertext: N L A Z E I I B L J J I

Vigenère Cipher

- The Vigenère cipher masks the characteristic letter frequencies of English plaintexts, but some patterns remain (wiki diagram)



Cryptanalysis of Vigenere Cipher

- Cryptanalysis approach
 - Find the length of the key
 - Then for each letter in the key, the problem becomes a simple shift cipher
 - use frequency cryptanalysis to break it (determining one letter is enough)
- How to find the key length m ?
 - Two methods
 - Kasiski test
 - Index of coincidence

Cryptanalysis of Vigenere Cipher

- Kasiski test
 - Based on the observation:
 - two identical segments of plaintext will be encrypted to the same ciphertext if their distance is the multiple of m
 - Algorithm:
 1. Search for pairs of identical segments of length at least 3
 2. Record distances between the two segments: $\Delta_1, \Delta_2, \dots$
 3. m is a divisor of $gcd(\Delta_1, \Delta_2, \dots)$

Cryptanalysis of Vigenere Cipher

- Kasiski test (contd.)
 - Example:

P: t h e s u n a n d t h e m a n i n t h e m o o n
Key: K I N G K I N G K I N G K I N G K I N G K I N G
C: D P R Y E V N T N B U K W I A O X B U K W W B T

distance = 8

Cryptanalysis of Vigenere Cipher

- Index of coincidence
 - simple idea
 - If the value of m is guessed correctly, then the distribution of the ciphertext letters $\{C_{\beta+m \times i}\}$ would be close to the distribution of the English letters for any constant β
 - since $\{C_{\beta+m \times i}\}$ are generated from the same substitution table
 - If the value of m is guessed wrongly, then the distribution of the ciphertext letters $\{C_{\beta+m \times i}\}$ would be random

Cryptanalysis of Vigenere Cipher

- Index of coincidence (contd.)

Definition: Suppose $X = x_1x_2 \cdots x_n$ is a string of n alphabetic characters. The index of coincidence of X , denoted $I_c(x)$, is defined to be the probability that two random elements of X are identical.

Denote the numbers of A,B,C, ..., Z in X by $f_0, f_1, f_2, \dots, f_{25}$ (respectively). We can choose two elements of X in $\binom{n}{2}$ ways. For each i , $0 \leq i \leq 25$, there are $\binom{f_i}{2}$ ways of choosing both elements to be i . Thus

$$I_c = \frac{\sum_{i=0}^{25} \binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)} \approx \frac{\sum_{i=0}^{25} f_i^2}{n^2} = \sum_{i=0}^{25} p_i^2$$

Cryptanalysis of Vigenere Cipher

- Index of coincidence (contd.)
 - If X is a string of English language text, $I_c = 0.065$

- If X is a random string, then

$$I_c = 1/26 = 0.038$$

| letter | probability | letter | probability |
|--------|-------------|--------|-------------|
| A | .082 | N | .067 |
| B | .015 | O | .075 |
| C | .028 | P | .019 |
| D | .043 | Q | .001 |
| E | .127 | R | .060 |
| F | .022 | S | .063 |
| G | .020 | T | .091 |
| H | .061 | U | .028 |
| I | .070 | V | .010 |
| J | .002 | W | .023 |
| K | .008 | X | .001 |
| L | .040 | Y | .020 |
| M | .024 | Z | .001 |

Cryptanalysis of Vigenere Cipher

- Index of coincidence (contd.)

Algorithm: Guess the value of m . For $\beta = 0, 1, 2, 3, \dots m - 1$, compute the value of I_c for each set $\{C_{\beta+m \cdot i}\}$. If the values of I_c are all close to 0.065, then the value of m is guessed correctly.

Cryptanalysis of Vigenere Cipher

- Example: ciphertext from a Vigenere cipher

CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQRBW
RVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWIAK
LXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSRELXNJELX
VRVPRTULHDNQWTWDTYGBPHXTFALJHASVBFXNGLLCHR
ZBWELEKMSJIKNBHWRJGNMGJSGLXFEYPHAGNRBIEQJT
AMRVLCRREMNDGLXRRIMGNSNRWCHRQHAEYEVTAQEBBI
PEEWEVKAKOEWADREMXTBHHCRTKDNRVZCHRCLQOHP
WQAIIWGXNRMGWOIIFKEE

Cryptanalysis of Vigenere Cipher

- Example (contd.)
 - Kasiski test
 - Ciphertext string CHR occurs in five places, beginning at positions 1, 166, 236, 276 and 286. Thus the distances are 165, 235, 275, 285. $gcd(165,235,275,285) = 5$
 - Indices of coincidences
 - $m = 1, I_c = 0.045$
 - $m = 2, I_c = 0.046, 0.041$
 - $m = 3, I_c = 0.043, 0.050, 0.047$
 - $m = 4, I_c = 0.042, 0.039, 0.045, 0.040$
 - **$m = 5, I_c = 0.063, 0.068, 0.069, 0.061, 0.072$**

Transposition (permutation) cipher

- Definition:

Let m be a positive integer. Let $\mathcal{P} = \mathcal{C} = (\mathbb{Z}_{26})^m$ and let \mathcal{K} consist of all permutations of $\{1, \dots, m\}$. For a key (i.e., a permutation) π , we define

$$e_\pi(x_1, \dots, x_m) = (x_{\pi(1)}, \dots, x_{\pi(m)})$$

and

$$d_\pi(y_1, \dots, y_m) = (y_{\pi^{-1}(1)}, \dots, y_{\pi^{-1}(m)}),$$

where π^{-1} is the inverse permutation to π .

Transposition (permutation) cipher

- Example:

Suppose $m = 6$ and the key is the following permutation π :

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

plaintext: shesellsseashellsbytheseashore

- 1) Partition the plaintext into groups of 6 letters

shesel | lsseas | hellsb | ythese | ashore

- 2) Rearrange the 6 letters in each group

EESLSH | SALSES | LSHBLE | HSYEET | HRAEOS

ciphertext: EESLSHSALSESLSHBLEHSYEETHRAEOS

Transposition (permutation) cipher

- Security
 - Vulnerable to divide-and-conquer attack
 - Guess & determine part of the key (permutation table)
 - For example: Let $\pi(x) = \alpha$; $\pi(y) = \alpha+1$;
part of the values of x and y can be determined by trying all the values of x and y with frequency cryptanalysis
 - Then gradually recover the whole permutation table

Summary of classical ciphers

- Shift ciphers
- Substitution ciphers
 - frequency cryptanalysis
- Vigenere cipher
 - Kasiski test
 - Index of coincidence
- Transposition (permutation) cipher

Recommended reading for classical ciphers

- Cryptography Theory and Practice
 - Section 1.1.1, 1.1.2, 1.1.4 and 1.1.6
 - Section 1.2.2, 1.2.3