Classical Ciphers

AK Bhateja

Classical Cryptosystems

- Substitution Cipher
- Transposition Cipher
- Vigenere Cipher
- Hill Cipher
- Playfair Cipher

Cryptography

- Classical cryptography (say, before the 1950s).
- Kerckhoff's Principle for cipher design: The encryption scheme is not kept secret, and only the key is kept secret. Why?
  - algorithms can be leaked or reverse engineered

    It is much easier to maintain secrecy of a short key than to maintain secrecy of an algorithm (program: large),
  - in case the key is exposed, it is much easier for the honest parties to change the key than to replace the algorithm being used
- Cipher: set of steps (an algorithm) for performing both an encryption, and the corresponding decryption

Classical Systems

- Caesar's cipher - written in approx 110 AD
- Each letter in the plaintext is 'shifted' a certain number of places down the alphabet.
- Encryption function

$$y = (x + k) \bmod 26$$

- Decryption function

$$x = (y - k) \bmod 26$$

Example:

plaintext: defend the east wall of the castle

ciphertext: efgfoe uif fbtu xbmm pg uif dbtumf

Affine Cipher

- The 'key' for the Affine cipher consists of 2 numbers, a and b.
- No. of alphabet (m = 26).
- a should be chosen to be relatively prime to m.
- Encryption function:

  $y = (ax + b) \bmod m$
- Decryption function:

  $x = a^{-1} (y - b) \bmod m$

Example: Affine Cipher

- Encryption:

  $a = 5$ and $b = 7$, $y = (5*x + 7)(\text{mod } 26)$.

  Plain text: 'defend the east wall of the castle'

  Use ('*a*'= 0, '*b*'=1, ..., '*z*'=25), first letter '*d*' = 3

  $y = (5 \times 3 + 7) \text{ mod } 26 \equiv 22$

  since 'w' = 22, 'd' is transformed into 'w'

  Cipher text: 'wbgbuwyqbbhtynhkkzgyqbrhtykb'

- Decryption:

  inverse of 5 modulo 26 is 21, i.e. $5 \times 21 = 1 \ (\text{mod } 26)$.

  $x = 21 \times (22 - 7) \text{ mod } 26 \equiv 3$ i.e. '*d*'

Simple Substitution Cipher

- Substituting every plaintext character for a different ciphertext character
- To make the key easy, use a key word, e.g. 'zebra‘. The key:

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

Plain text: DEFEND THE EAST WALL OF THE CASTLE

- ciphertext: RACALR SFA AZQS VZJJ MC SFA BZQSJA
- Key space: $26! \approx 2^{88.4}$

Cryptanalysis of Substitution Cipher (frequency analysis)

- Standard languages do not have uniform probabilities
- In English
  - Order Of Frequency of single letters:

    E T A O I N S H R D L U
  - E has probability 0.12 (12%)
  - Order Of Frequency of Digraphs

    TH ER ON AN RE HE IN ED ND HA AT EN ES
  - Order Of Frequency Of Trigraphs

    THE AND THA ENT ION TIO FOR NDE HAS NCE EDT

Transposition Cipher

- Rearranging the order of the letters
- The key for the columnar transposition cipher is a keyword e.g. GERMAN.
- The row length that is used is the same as the length of the keyword.
- Plain text: DEFEND THE EAST WALL OF THE CASTLE

G	E	R	M	A	N
D	E	F	E	N	D
T	H	E	E	A	S
T	W	A	L	L	O
F	T	H	E	C	A
S	T	L	E	X	X

- Reorder the columns s.t. the letters in the key word are ordered alphabetically.

A	E	G	M	N	R
N	E	D	E	D	F
A	H	T	E	S	E
L	W	T	L	O	A
C	T	F	E	A	H
X	T	S	E	X	L

- The ciphertext is read off along the columns:
  NALCXEHWTTDTTFSEELEEDSOAXFEAHL

Vigenère Cipher

- It is a polyalphabetic substitution cipher
- In Vigenere cipher each plaintext letter has multiple corresponding ciphertext letters
- The Vigenère Cipher was developed by mathematician Blaise de Vigenère in the 16th century.

Vigenère Cipher

- Def: Given m, a positive integer and $K = (k_1, k_2, \ldots, k_m)$ a key where each $k_i \in Z_{26}$, the Vigenere cipher is defined as:
- Encryption: $c_i = p_i + k_{i \text{ (mod } m)} \pmod{26}$
- Decryption: $p_i = c_i - k_{i \text{ (mod } m)} \pmod{26}$
- Example: Consider 'CODE' as the key and CRYPTANALYSIS as the plaintext

Plaintext:	C	R	Y	P	T	A	N	A	L	Y	S	I	S
Key	C	O	D	E	C	O	D	E	C	O	D	E	C
Ciphertext	E	F	B	T	V	O	Q	E	N	M	V	M	U

Cryptanalysis of Vigenère Cipher

- The key space of the Vigenere cipher is 26^m, m is key size
- Brute force techniques infeasible for sufficiently large values of m.
- Cryptanalysis of the Vigenere cipher has 2 main steps:
  - identify the period of the cipher (the length of the key)
    - Kasiski method
    - Index of Coincidence
  - finding the specific key

Kasiski Method

- Published by Friedrich Kasiski in 1863
- The Kasiski examination involves looking for strings of three or more characters that are repeated in the ciphertext.
- Find the distances between consecutive occurrences of the strings (are likely to be multiples of the length of the keyword)
- Find the greatest common divisor of all the distances.
- If a repeated substring in a plaintext is encrypted by the same substring in the keyword, then the ciphertext contains a repeated substring and the distance of the two occurrences is a multiple of the keyword length.
- Not every repeated string in the ciphertext arises in this way; but, the probability of a repetition by chance is small.

Example: Kasiski Method

- Intercepted message:

  VHVSSP**QUCE**MRVBVBBB**VHVS**URQGIBDUGRNICJ **QUCE**RVUAXSSR
- The gap between the "VHVS" pair is 18, implies key length may be 18, 9, 6, 3 or 2. The gap between the "QUCE" pair is 30, implies key length 30, 15, 10, 6, 5, 3 or 2.
- So looking at both together the most likely key length is 6 or possibly 3 (though in practice this is unlikely).

Index of Coincidence (Friedman Test)

- Invented by William F. Friedman in 1922
- Putting two texts side-by-side and counting the number of times that identical letters appear in the same position in both texts.
- The index of coincidence provides a measure of how likely it is to draw two matching letters by randomly selecting two letters from a given text.
- It is a ratio of the total and the expected count for a random source model.

Index of Coincidence

- The index of coincidence (IC): the probability of having two identical letters from the text is.

$$IC = \frac{\sum_{i=1}^{n} f_i \left(f_i - 1 \right)}{N(N-1)}$$

Where f_i is the frequency count of ith letter in the ciphertext of length N.

- $IC_{English} = 0.0686$, $IC_{Random} \approx 1/26 = 0.038466$
- For a ciphertext encrypted by a monoalphabetic cipher IC will be the same as for the original plaintext
- For polyalphabetic ciphers (like Vigenère) it is between $IC_{English}$ and IC_{Random}.

Finding length of the key

- This procedure of breaking up the ciphertext and calculating the I.C. for each subsequence is repeated for all the key lengths we wish to test.
- If IC for a particular length say k is very close to $IC_{English}$ stop and declare the length of the key is k.

Example: Vigenère Cipher

- Vignere cipher of size 313 characters

CHREEVOAHMAERATBIAXXWTNXBEEOPH
BSBQMQEQERBWRVXUOAKXAOSXXWEAHB
WGJMMQMNKGRFVGXWTRZXWIAKLXFPSK
AUTEMNDCMGTSXMXBTUIADNGMGPSREL
XNJELXVRVPRTULHDNQWTWDTYGBPHXT
FALJHASVBFXNGLL**CHR**ZBWELEKMSJIK
NBHWRJGNMGJSGLXFEYPHAGNRBIEQJT
AMRVLCRREMNDGLXRRIMGNSNRW**CHR**QH
AEYEVTAQEBBIPEEWEVKAKOEWADREMX
MTBHH**CHR**TKDNVRZ**CHR**CLQOHPWQAIIW
XNRMGWOIIFKEE

Finding length by Kasiski Method

- The text CHR, starts at 1, 166, 236, 276 and 286.
- The distances between the occurrences are 10, 70, 110, 120, 165, 235, 275 and 285.
- Thus k = gcd(10,70, 110, 120, 165, 235, 275, 285) = 5.

Verifying the length of key by IC

CHREEVOAHMAERATBIAXXWTNXBEEOPH
BSBQMQEQERBWRVXUOAKXAOSXXWEAHB
WG

A	B	C	E	G	H	I	K	M	N
7	6	1	8	1	4	1	1	2	1
O	P	Q	R	S	T	U	V	W	X
4	1	3	4	2	2	1	2	4	7

Finding length by IC

original: CHREEVOAHMAERATBIAXXWTNXBEEOPH…

if key were length 2:

sequence 1: C R E O H A R T I X W N B E P ...

sequence 2: H E V A M E A B A X T X E O H …

if key were length 3:

sequence 1: C E O M R B X T B O ...

sequence 2: H E A A A I X N E P …

sequence 3: R V H E T A W X E H …

- For k = 1, 2, 3, 4 IC ≈ 0.04
- For k = 5, IC = 0.065 ($\approx IC_{English}$)

Hill Cipher

- Invented by Lester S. Hill in 1929
- Hill cipher is a polygraphic substitution cipher based on linear algebra
- Let K ($n \times n$ matrix) be key, P: plaintext vector, C: ciphertext vector
- Encryption: $C = K \times P \bmod N$
- Decryption: $P = K^{-1} \times C \bmod N$

N is cardinality of the character set

Hill Cipher: Example

Consider N = 26 and $K = \begin{bmatrix} 3 & 2 \\ 3 & 5 \end{bmatrix}$

Plaintext: ATTACK IS TONIGHT

P = [A T]^T = [0 19]^T

$$C = \begin{bmatrix} 3 & 2 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} 0 \\ 19 \end{bmatrix} \bmod 26 = \begin{bmatrix} 12 \\ 17 \end{bmatrix} = \begin{bmatrix} M \\ R \end{bmatrix}$$

Hill Cipher: Example

Let $$\begin{bmatrix} 3 & 2 \\ 3 & 5 \end{bmatrix}\begin{bmatrix} 15 & 20 \\ 17 & 9 \end{bmatrix} \bmod 26 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\therefore \quad K^{-1} = \begin{bmatrix} 15 & 20 \\ 17 & 9 \end{bmatrix}$$

Example: Hill Cipher

Consider N = 26 and $K = \begin{bmatrix} 2 & 4 & 5 \\ 9 & 2 & 1 \\ 3 & 17 & 7 \end{bmatrix}$

Plaintext: ATTACK IS TONIGHT

Break the message into chunks of 3

First chunk: ATT = $[0 \;\; 19 \;\; 19]^T$ = P

$$C = \begin{bmatrix} 2 & 4 & 5 \\ 9 & 2 & 1 \\ 3 & 17 & 7 \end{bmatrix} \begin{bmatrix} 0 \\ 19 \\ 19 \end{bmatrix} \bmod 26 \equiv \begin{bmatrix} 15 \\ 5 \\ 14 \end{bmatrix} = \begin{bmatrix} P \\ F \\ O \end{bmatrix}$$

Example: Hill Cipher

Consider n = 26 and $K = \begin{bmatrix} 2 & 4 & 5 \\ 9 & 2 & 1 \\ 3 & 17 & 7 \end{bmatrix}$

Plaintext: ATTACK IS TONIGHT

Break the message into chunks of 3

First chunk: ATT = [0 19 19]^T = P

$$C = \begin{bmatrix} 2 & 4 & 5 \\ 9 & 2 & 1 \\ 3 & 17 & 7 \end{bmatrix} \begin{bmatrix} 0 \\ 19 \\ 19 \end{bmatrix} \bmod 26 \equiv \begin{bmatrix} 15 \\ 5 \\ 14 \end{bmatrix} = \begin{bmatrix} P \\ F \\ O \end{bmatrix}$$

Cryptanalysis of Hill Cipher

- Key space: 26^{n^2}
- known-plaintext attack
- Suppose it is 2 by 2 hill cipher
- In standard english, the most frequent digraph is ‘TH', followed by ‘HE'.
- Suppose in the cipher text the most frequent digraph is ‘KX', followed by ‘VZ'
- Guess: TH $\rightarrow$ KX and HE $\rightarrow$ VZ

  or [19, 7] $\rightarrow$ [10, 23] and [7, 4] $\rightarrow$ [21, 25]

Cryptanalysis of Hill Cipher

- Let K be the key, then $K \times P = C$

$$\text{i.e. } K \times \begin{bmatrix} 19 & 7 \\ 7 & 4 \end{bmatrix} = \begin{bmatrix} 10 & 21 \\ 23 & 25 \end{bmatrix} \bmod 26$$

- Since $K = C \times P^{-1}$

$$\text{Find } P^{-1} = \begin{bmatrix} 4 & 19 \\ 19 & 19 \end{bmatrix} \bmod 26$$

$$\therefore K = \begin{bmatrix} 10 & 21 \\ 23 & 25 \end{bmatrix} \times \begin{bmatrix} 4 & 19 \\ 19 & 19 \end{bmatrix} \bmod 26 = \begin{bmatrix} 23 & 17 \\ 21 & 2 \end{bmatrix}$$

- Decrypt & if it is not correct, try other combinations of common pair of digraphs, until the correct key

Playfair cryptosystem

- The first practical digraph substitution cipher
- Invented in 1854 by Charles Wheatstone, but was named after Lord Playfair who promoted the use of the cipher
- It uses a 5×5 table containing a key word or phrase.
- No. of digraphs: 25×25
- Example: key word "HELLO WORLD"

  Remove duplicate letters

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

- Plaintext: "hide the gold"
- Split into digraphs: HI DE TH EG OL D

Encryption process: Playfair

- If both letters are the same (or only one letter is left), add an "X" after the first letter.
- If both letters are in the same column, take the letter below each one (going back to the top if at the bottom)
- If both letters are in the same row, take the letter to the right of each one (going back to the left if at the farthest right)
- If neither of the preceding two rules are true, form a rectangle with the two letters and take the letters on the horizontal opposite corner of the rectangle
- "HI DE TH EG OL DX" with the key of "hello world" would be "LF GD MW DN WO AV".