Classical Cryptography

Prof. Ashok K Bhateja, IIT Delhi

Classical Cryptosystems

- Substitution Cipher
- Transposition Cipher
- Vigenère Cipher
- Hill Cipher
- Playfair Cipher

Classical Systems: Caesar Cipher

- 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)(mod 26).

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

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

  y = (5 × 3 + 7) mod 26 ≡ 22

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

  Cipher text: 'wbgbuwyqbbhtynhkkzgyqbrhtykb’

- Decryption:

  inverse of 5 modulo 26 is 21, i.e., 5×21 = 1 (mod 26).

  x = 21 × (22 – 7) mod 26 ≡ 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

Letter Probability in Standard English

Letter	Probability	Letter	Probability
A	0.082	N	0.067
B	0.015	O	0.075
C	0.028	P	0.019
D	0.043	Q	0.001
E	0.127	R	0.060
F	0.022	S	0.063
G	0.020	T	0.091
H	0.061	U	0.028
I	0.070	V	0.010
J	0.002	W	0.023
K	0.008	X	0.001
L	0.040	Y	0.020
M	0.024	Z	0.001

Transposition Cipher

1. Rearranging the order of the letters
2. The key for the columnar transposition cipher is a keyword e.g., GERMAN.
3. The row length that is used is the same as the length of the keyword.
4. Plain text: DEFEND THE EAST WALL OF THE CASTLE
5. Reorder the columns s.t., the letters in the key word are ordered alphabetically.
6. The ciphertext is read off along the columns:

   NALCXEHWTTDTTFSEELEEDSOAXFEAHL

After step 4

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

After step 5

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

Vigenère Cipher

- It is a polyalphabetic substitution cipher
- In Vigenère 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: $c_i = p_i + k_{i \pmod m} \pmod{26}$
- Decryption: $p_i = c_i - k_{i \pmod m} \pmod{26}$
- 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 Vigenère 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:

  VHVSSPQUCEMRVBVBBBVHVSURQGIBDUGRNICJQUCERVUAXSSR

- 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
- IC indicate how much like random, or how different from random, is this text?
- 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

$$IC = \frac{\sum_{i=1}^{n} \binom{f_i}{2}}{\binom{N}{2}} = \frac{\sum_{i=1}^{n} f_i(f_i - 1)}{N(N-1)}$$

Where n is number of different alphabets, f_i is the frequency count of ith letter in the ciphertext of length N.

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

where $\boldsymbol{p_i}$ is the probability of ith alphabet.

For large N, $IC_{English} \approx \sum_{i=1}^{n} p_i^2 \approx 0.0686$ & $IC_{Random} = p_i = \frac{1}{26} = 0.038466$

Index of Coincidence

- $IC_{\text{English}} = 0.0686, \quad IC_{\text{Random}} \approx 1/26 = 0.038466$
- For a ciphertext encrypted by a monoalphabetic cipher *IC* will be the same as for the original plaintext
- In monoalphabetic cipher the individual probabilities will be permuted, but the $\sum p_i^2$ will be unchanged. So, this is an Invariant.
- For polyalphabetic ciphers (like Vigenère) it is between IC_{English} and IC_{Random}

Index of Coincidence

Example: Encrypted text, size $N = 205$ and number of alphabets $n = 26$

VVQGY TVVVK ALURW FHQAC MMVLE HUCAT WFHHI PLXHV UWSCI GINCM UHNHQ RMSUI MHWZO DXTNA EKVVQ GYTVV QPHXI NWCAB ASYYM TKSZR CXWRP RFWYH XYGFI PSBWK QAMZY BXJQQ ABJEM TCHQS NAEKV VQGYT VVPCA QPBSL URQUC VMVPQ UTMML VHWDH NFIKJ CPXMY EIOCD TXBJW KQGAN

A	B	C	D	E	F	G	H	I	J	K	L	M
11	6	11	3	5	5	6	13	8	4	7	5	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
7	2	8	14	6	7	9	8	18	10	8	9	3

$$IC = \frac{1}{N(N-1)} \sum_{i=1}^{26} f_i(f_i - 1) = \quad 0.041989$$

- This is neither a plain English text nor random

Finding length of the key

- This procedure of breaking up the ciphertext and calculating the *IC* for each subsequence is repeated for all the key lengths estimated by Kasiski Method.
- 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

Vigenere cipher of size 313 characters

CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQER

BWRVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTR

WIAKLXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSREL

XNJELXVRVPRTULHDNQWTWDTYGBPHXTFALJ HASVBF

XNGLL**CHR**ZBWELEKMSJIK NBHWRJGNMGJSGLXFEYPH

AGNRBIEQJTAMRVLCRREMNDGLXRRIMGNSNRW**CHR**QH

AEYEVTAQEBBIPEEWEVKAKOEWADREMXMTBH H**CHR**T

DNVRZ**CHR**CLQOHPWQAIIWXNRMGWOIIFKEE

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

CHREEVOAHMAERATBIAXXWTNXBEEOPHBS
BQMQEQERBWRVXUOAKXAOSXXWEAHBWG

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…

For key 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 …

For key 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 $= 0.065$ ($\approx IC_{English}$)

Mutual Index of Coincidence

- Suppose x = $x_1, x_2 \ldots, x_n$, and y = $y_1, y_2 \ldots, y_{n'}$ are strings of n and n' alphabetic characters, respectively. The mutual index of coincidence of x and y, denoted MIC(x, y), is the probability that a random element of x is identical to a random element of y.

$$MIC(\text{x}, \text{y}) = \frac{\sum_{i=0}^{25} f_i f_i'}{nn'}$$

Where f_i and $f_i\,'$ are the frequency count of i^{th} letter in x and y respectively.

Index of alphabet	0	1	…	25
Probability	p_0	p_1		p_{25}

If k_i is a key

Index of alphabet	$0 + k_i$	$1 + k_i$	…	$25 + k_i$
Probability	p_0	p_1		p_{25}

Probability that character is *A* in cryptogram:

It is the probability of j^{th} character in the cryptogram where $j + k_i = 0$ or $j = -k_i \pmod{26}$

i.e. P (character is *A* in cryptogram) = p_{-k_i}.

Suppose K = (k_1 , k_2, . . . , k_m) is the keyword.

- To estimate $MIC(z_i, z_j)$

  Consider a random character in z_i and a random character in z_j.

  Probability that both characters are A is $p_{0-k_i} \cdot p_{0-k_j}$

  Probability that both are B is $p_{1-k_i} \cdot p_{1-k_j}$ etc.

$$MIC(z_i, z_j) = \sum_{h=0}^{25} p_{h-k_i} . p_{h-k_j} = \sum_{h=0}^{25} p_h p_{h+k_i-k_j}$$

  The value of this estimate depends only on the difference $k_i - k_j$ (mod 26), which is called the relative shift of z_i and z_j.

$$\sum_{h=0}^{25} p_h p_{h+l}$$

Relative shift $l = k_i - k_j$ (mod 26)

Mutual Index of Coincidence

- If the relative shift of string x & y is not zero, then *MIC* (x, y) vary between 0.031 and 0.045;
- A relative shift of zero yields an estimate of 0.065

Mutual Index of Coincidence

- Suppose we fix x, and consider the effect of encrypting y by e_0, e_1, e_2, …, e_{25} Denote the resulting strings by y^0, y^1, y^2 …, y^{25}.
- It is easy to compute the indices MIC (x, y^g) where $0 \leq g \leq 25$. This can be done using the formula

$$MIC(\mathrm{x}, y^g) = \frac{\sum_{i=0}^{25} f_i f'_{i-g}}{nn'}$$

- When $g = l$, the MIC should be close to 0.065, since the relative shift of x and y^l is zero. However, for values of $g \neq l$, the MIC should vary between 0.031 and 0.045.
- In this way, relative shifts of any two of the substrings y_i can be obtained. This leaves only 26 possible keywords, which can easily be obtained by exhaustive key search.

Example: Let the keyword length be 5.

K = $(k_1, k_2, k_3, k_4, k_5)$

Find MIC (y_i, y_j^g), for

$1 \le i < j \le 5$,

$0 \le g \le 25$.

$k_1 - k_2 = 9$

$k_1 - k_5 = 16$

$k_2 - k_3 = 13$

$k_2 - k_5 = 7$

$k_3 - k_5 = 20$

$k_4 - k_5 = 11$

So, the key is likely to be

$(k_1, k_1 + 17, k_1 + 4, k_1 + 21, k_1 + 10)$ for some $k_1 \in Z_{26}$.

Taken from Cryptography Theory & practice by Stinson

i	j	value of $MI_c(\mathbf{y}_i, \mathbf{y}_j^g)$								
1	2	.028	.027	.028	.034	.039	.037	.026	.025	.052
		.068	.044	.026	.037	.043	.037	.043	.037	.028
		.041	.041	.034	.037	.051	.045	.042	.036	
1	3	.039	.033	.040	.034	.028	.053	.048	.033	.029
		.056	.050	.045	.039	.040	.036	.037	.032	.027
		.037	.036	.031	.037	.055	.029	.024	.037	
1	4	.034	.043	.025	.027	.038	.049	.040	.032	.029
		.034	.039	.044	.044	.034	.039	.045	.044	.037
		.055	.047	.032	.027	.039	.037	.039	.035	
1	5	.043	.033	.028	.046	.043	.044	.039	.031	.026
		.030	.036	.040	.041	.024	.019	.048	**.070**	.044
		.028	.038	.044	.043	.047	.033	.026	.046	
2	3	.046	.048	.041	.032	.036	.035	.036	.030	.024
		.039	.034	.029	.040	**.067**	.041	.033	.037	.045
		.033	.033	.027	.033	.045	.052	.042	.030	
2	4	.046	.034	.043	.044	.034	.031	.040	.045	.040
		.048	.044	.033	.024	.028	.042	.039	.026	.034
		.050	.035	.032	.040	.056	.043	.028	.028	
2	5	.033	.033	.036	.046	.026	.018	.043	**.080**	.050
		.029	.031	.045	.039	.037	.027	.026	.031	.039
		.040	.037	.041	.046	.045	.043	.035	.030	
3	4	.038	.036	.040	.033	.036	.060	.035	.041	.029
		.058	.035	.035	.034	.053	.030	.032	.035	.036
		.036	.028	.046	.032	.051	.032	.034	.030	
3	5	.035	.034	.034	.036	.030	.043	.043	.050	.025
		.041	.051	.050	.035	.032	.033	.033	.052	.031
		.027	.030	**.072**	.035	.034	.032	.043	.027	
4	5	.052	.038	.033	.038	.041	.043	.037	.048	.028
		.028	.036	**.061**	.033	.033	.032	.052	.034	.027
		.039	.043	.033	.027	.030	.039	.048	.035	

Finding key letters

- Let keyword length be m.
- Divide the ciphertext into m substrings,
- Shift each substring by 0, 1, 2, … , 25 and compute the MIC values.

  i.e., Compute values of $MIC(z_i, z_j^g)$, where $1 \le i < j \le m$, $0 \le g \le 25$.
- For each i and j, look for values of $MIC(z_i, z_j^g)$ that are close to 0.065.
- If there is a unique such value (for a given (i, j) pair), then the value of g is the value of the relative shift. i.e., $k_i - k_j = g$.
- Solve all such equations, which leaves only 26 possible key words, which can be obtained with some heuristics/guess/exhaustive key search.

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

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

Plaintext: ATTACK IS TONIGHT

$$P = [A \;\; T]^T = [0 \;\; 19]^T$$

$$C = \begin{bmatrix} 3 & 2 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} 0 \\ 19 \end{bmatrix} \bmod 26 = \begin{bmatrix} 12 \\ 17 \end{bmatrix} = \begin{bmatrix} M \\ R \end{bmatrix}, \quad \text{AT} \rightarrow \text{MR}$$

Decryption: $K^{-1} = \begin{bmatrix} 15 & 20 \\ 17 & 9 \end{bmatrix}$, $\because \begin{bmatrix} 3 & 2 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} 15 & 20 \\ 17 & 9 \end{bmatrix} \bmod 26 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$P = \begin{bmatrix} 15 & 20 \\ 17 & 9 \end{bmatrix} \begin{bmatrix} 12 \\ 17 \end{bmatrix} \bmod 26 = \begin{bmatrix} 0 \\ 19 \end{bmatrix} = \begin{bmatrix} A \\ T \end{bmatrix}, \quad \text{MR} \rightarrow \text{AT}$$

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 $K \times P = C$

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

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

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

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

- 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”

  Plaintext: “HIDE THE GOLD”
- Split into digraphs: HI DE TH EG OL D

Remove duplicate letters

Key Matrix:

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

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”.