

Experiment 01

Breaking the Mono-alphabetic Substitution Cipher using Frequency analysis method.

Roll No.	70
Name	Ishika Sanjay Devare
Class	D15-A
Subject	Internet Security Lab
LO Mapped	LO1: To apply the knowledge of symmetric cryptography to implement classical ciphers.

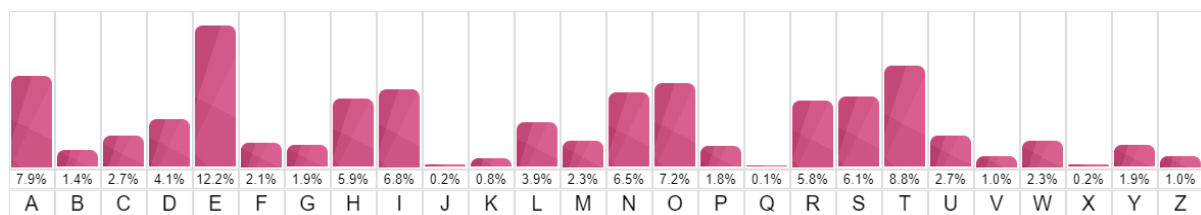
Aim: To understand the process of Breaking the Mono-alphabetic Substitution Cipher using Frequency analysis method.

Introduction:

In cryptography, frequency analysis is the study of the frequency of letters or groups of letters in a ciphertext. The method is used as an aid to breaking substitution ciphers.

Frequency analysis consists of counting the occurrence of each letter in a text. Frequency analysis is based on the fact that, in any given piece of text, certain letters and combinations of letters occur with varying frequencies. For instance, given a section of English language, letters E, T, A and O are the most common, while letters Z, Q and X are not as frequently used.

The following chart shows the frequency of each letter of the alphabet for the English language:



We can assume that most samples of text written in English would have a similar distribution of letters. However, this is only true if the sample of text is long enough. A very short text may lead to a significantly different distribution.

When trying to decrypt a cipher text based on a substitution cipher, we can use a frequency analysis to help identify the most recurring letters in a cipher text and hence make hypothesis of what these letters have been encoded as (e.g., E, T, A, O, etc). This will help us decrypt some of the letters in the text. We can then recognise patterns/words in the partly decoded text to identify more substitutions.

What is Ciphertext?

Ciphertext is encrypted text transformed from plaintext using an encryption algorithm. Ciphertext can't be read until it has been converted into plaintext (decrypted) with a key.

Types of Ciphers-

There are various types of ciphers, including:

1) **Substitution ciphers-** Replace bits, characters, or character blocks in plaintext with alternate bits, characters or character blocks to produce ciphertext.

A substitution cipher may be monoalphabetic or polyalphabetic:

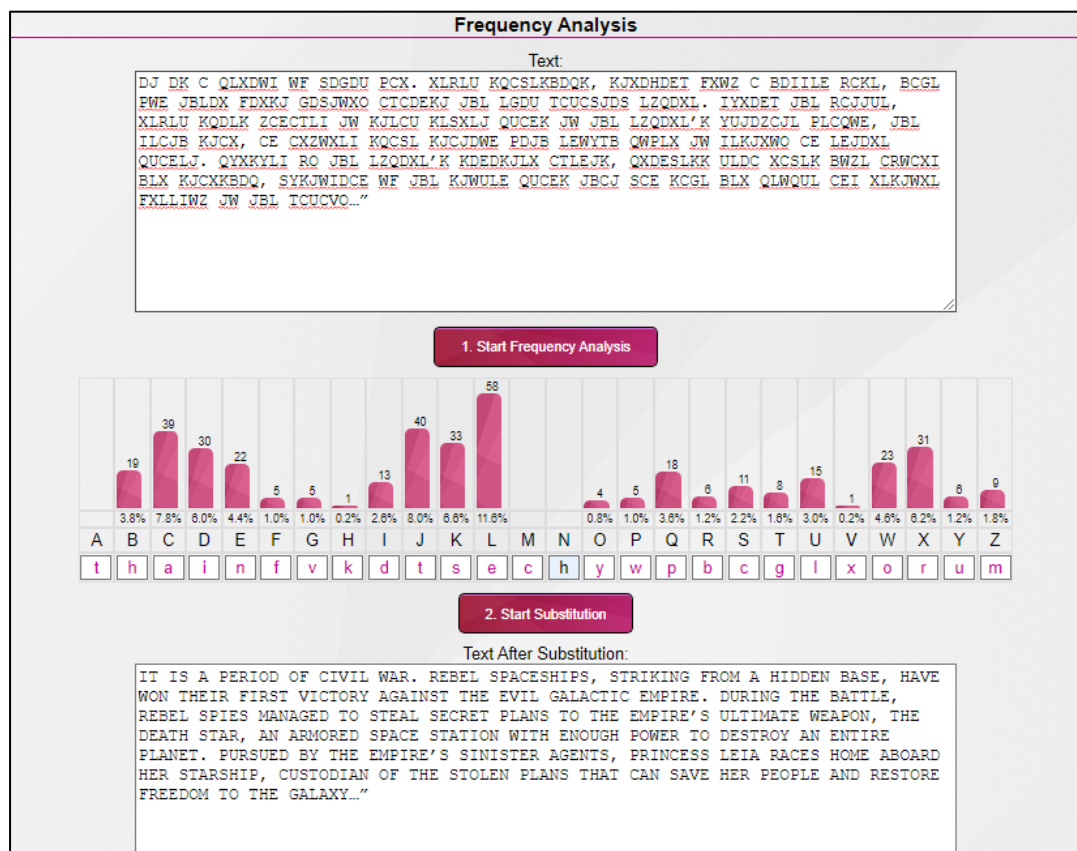
Single alphabet is used to encrypt the entire plaintext message. For example, if the letter A is enciphered as the letter K, this will be the same for the entire message.

A more complex substitution using a mixed alphabet to encrypt each bit, character or character block of a plaintext message. For instance, the letter A may be encoded as the letter K for part of the message, but later it might be encoded as the letter W.

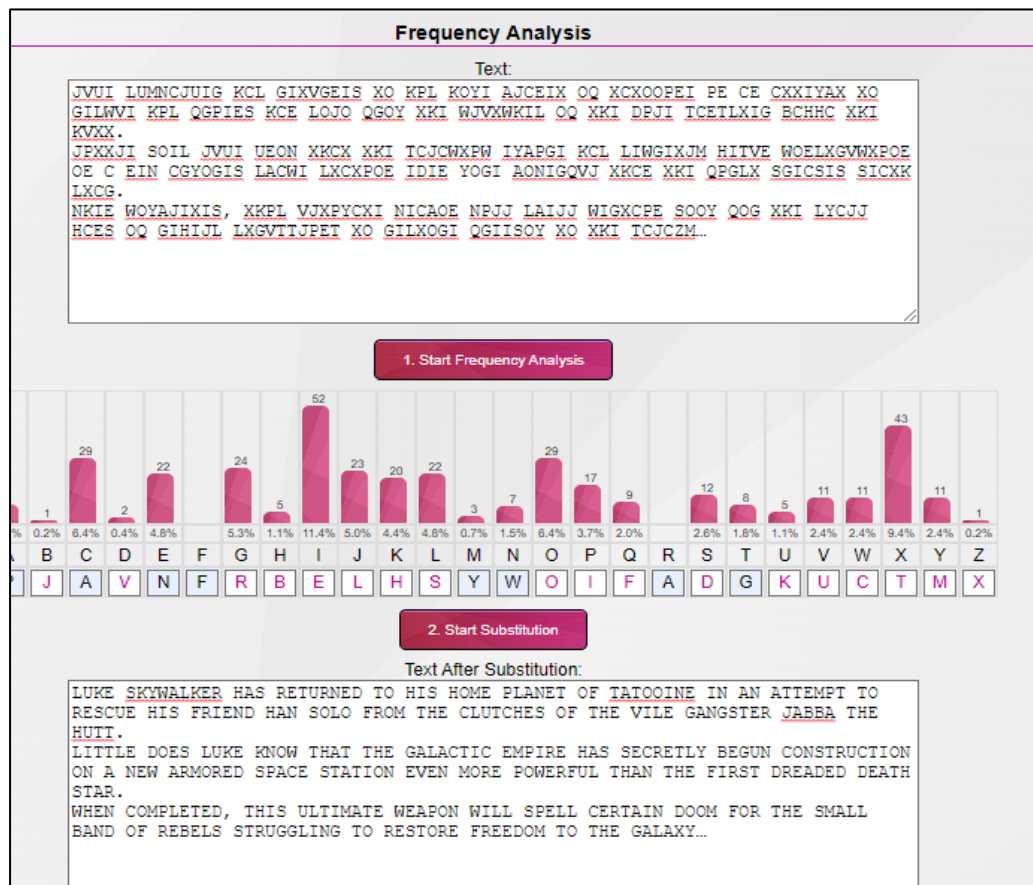
- 2) Transposition ciphers. Unlike substitution ciphers that replace letters with other letters, transposition ciphers keep the letters the same, but rearrange their order according to a specific algorithm. For instance, in a simple columnar transposition cipher, a message might be read horizontally but would be written vertically to produce the ciphertext.
- 3) Polygraphic ciphers- Substituting one letter for another letter, a polygraphic cipher performs substitutions with two or more groups of letters. This masks the frequency distribution of letters, making frequency analysis attacks much more difficult.
- 4) Permutation ciphers- In this cipher, the positions held by plaintext are shifted to a regular system so that the ciphertext constitutes a permutation of the plaintext.
- 5) Private-key cryptography- In this cipher, the sender and receiver must have a pre-shared key. The shared key is kept secret from all other parties and is used for encryption, as well as decryption.
- 6) Public-key cryptography- In this cipher, two different keys -- public key and private key -- are used for encryption and decryption. The sender uses the public key to perform the encryption, but the private key is kept secret from the receiver.

Results:

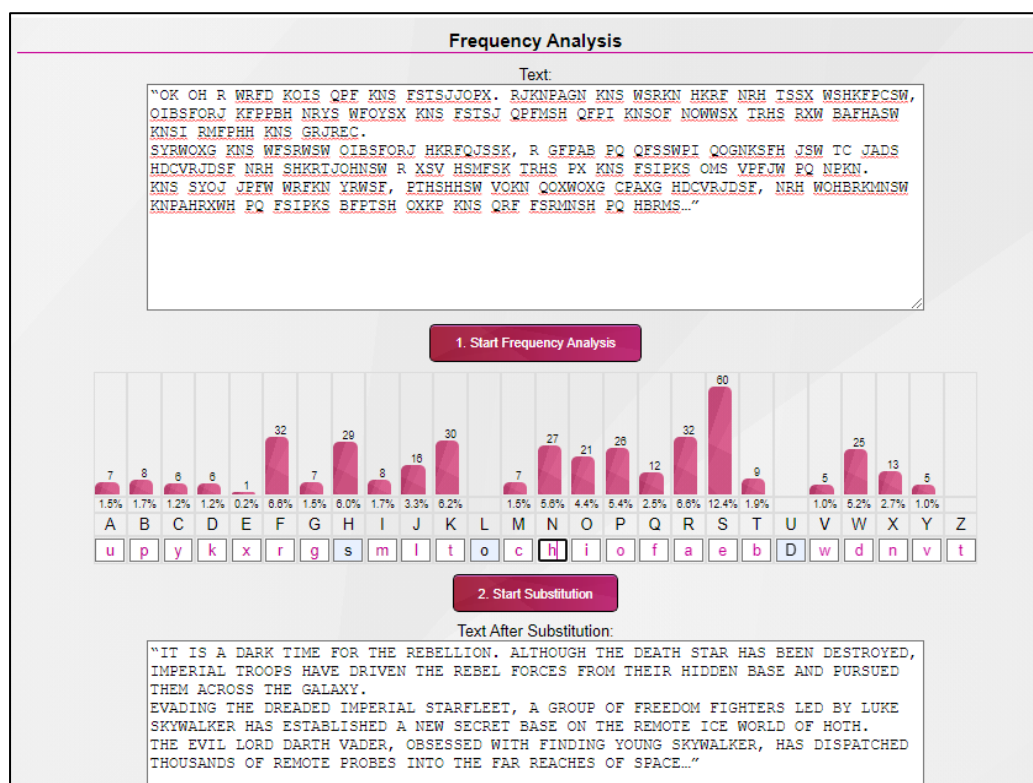
Cipher 1:



Cipher 2:



Cipher 3:



Cipher 4:

Frequency Analysis

Text:

"ZRIET IH PQFTHZ IQ ZRI XBGBOZIO HTQBZT. HTWIFBG ZRLPHBQV HLGFB HYHITSH RBWT VTOGBFTV ZRIIF IQLTQZILQH ZL GTBWT ZRI FTEPKGIO. ZRIH HTEBFBZIHZ SLWTSTQZ, PQVIF ZRI GTBVTFHRIE LD ZRI SYHZTFILPH OLPQZ VLLAP, RBH SBVI IZ VIDDIOBGZ DLF ZRI GISIZTV QFSKIF LD CTVI AQIXRZH ZL SBIQZBIQ EIBOT BQV LFVIF IQ ZRI XGBBJY. HTQBZLF BSIVBGB, ZRI DLESTF NETTQ LD QBKLL, IH FTZPEQIQX ZL ZRI XBGBOZIO HTQBZT ZL WLZT LQ ZRI OFIZIOBG IHHT LD OFTBZIQX BQ BFSY LD ZRI FTEPKGIO ZL BHHIHZ ZRI LWTFMRTGSTV CTVI..."

1. Start Frequency Analysis

Letter	Frequency (%)	Count
A	1.7%	8
B	0.2%	1
C	6.3%	29
D	0.4%	2
E	4.8%	22
F	0.0%	0
G	5.2%	24
H	1.1%	5
I	11.4%	52
J	5.0%	23
K	4.4%	20
L	4.8%	22
M	0.7%	3
N	1.5%	7
O	6.3%	29
P	3.7%	17
Q	2.0%	9
R	0.0%	0
S	2.6%	12
T	1.7%	8
U	1.1%	5
V	2.4%	11
W	2.4%	11
X	9.4%	43
Y	2.4%	11
Z	0.2%	1

2. Start Substitution

Text After Substitution:

"THERE IS UNREST IN THE GALACTIC SENATE. SEVERAL THOUSAND SOLAR SYSTEMS HAVE DECLARED THEIR INTENTIONS TO LEAVE THE REPUBLIC. THIS SEPARATIST MOVEMENT, UNDER THE LEADERSHIP OF THE MYSTERIOUS COUNT DOOKU, HAS MADE IT DIFFICULT FOR THE LIMITED NUMBER OF JEDI KNIGHTS TO MAINTAIN PEACE AND ORDER IN THE GALAXY. SENATOR AMIDALA, THE FORMER QUEEN OF NABOO, IS RETURNING TO THE GALACTIC SENATE TO VOTE ON THE CRITICAL ISSUE OF CREATING AN ARMY OF THE REPUBLIC TO ASSIST THE OVERWHELMED JEDI..."

Cipher 5:

Frequency Analysis

Text:

"FX IWBBJX PB NB PB PWX GBBD. VSP FWO, JBGX JRO, FWX GBBD? FWO IWBBJX PWUJ RJ BSA NBRK? RDL FWXO GRO FXKK RJM FWO IKUGV FWX NUNWJXP GBSDPRUD? FWO, 38 OXRAJ RNB, EKO FWX RPKRDPUI? FWO LBKJ AUIX CKRO FXQRJ? FX IWBBJX PB NB PB PWX GBBD UD FWUJ LXIRLX RDL LB PWX BFWXA FWUDNJ, DBP VXIRSJX FWXO RAX XRJO, VSP VXIRSJX FWXO RAX WRAL, VXIRSJX FWRP NBRK FUKK JXATX PB BANRDUZX RDL GXRJSAX FWX VXJP BE BSA XDKANUKJ RDL JMUUKJ, VXIRSJX FWRP IWRKKXDNX UJ BDX FWRP FX RAX FUKKUDN PB RIIXCP, BDX FX RAX SDFUKKUDN PB CBJPCBDX, RDL BDX FWUW FX UDPXDL PB FUD, RDL FWX BFWXAJ, FBB..."

1. Start Frequency Analysis

Letter	Frequency (%)	Count
A	2.7%	15
B	7.6%	43
C	0.7%	4
D	5.0%	28
E	0.4%	2
F	2.8%	16
G	1.4%	8
H	0.0%	0
I	2.7%	15
J	5.0%	28
K	3.2%	18
L	2.1%	12
M	0.4%	2
N	2.1%	12
O	2.5%	14
P	7.6%	43
Q	0.2%	1
R	6.5%	37
S	1.9%	11
T	0.2%	1
U	3.7%	21
V	1.4%	8
W	5.8%	33
X	10.8%	60
Y	0.2%	1
Z	0.0%	0

2. Start Substitution

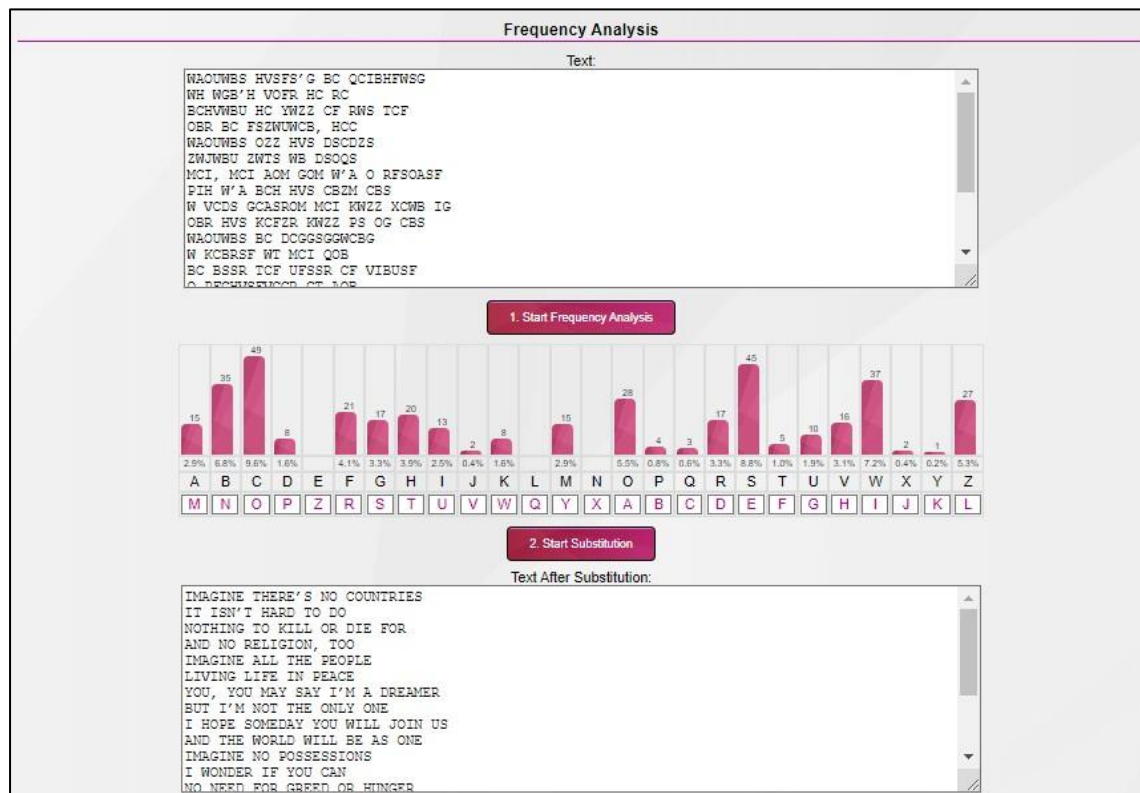
Text After Substitution:

"WE CHOOSE TO GO TO THE MOON. BUT WHY, SOME SAY, THE MOON? WHY CHOOSE THIS AS OUR GOAL? AND THEY MAY WELL ASK WHY CLIMB THE HIGHEST MOUNTAIN? WHY, 38 YEARS AGO, FLY THE ATLANTIC? WHY DOES RICE PLAY TEXAS? WE CHOOSE TO GO TO THE MOON IN THIS DECADE AND DO THE OTHER THINGS, NOT BECAUSE THEY ARE EASY, BUT BECAUSE THEY ARE HARD, BECAUSE THAT GOAL WILL SERVE TO ORGANISE AND MEASURE THE BEST OF OUR ENERGIES AND SKILLS, BECAUSE THAT CHALLENGE IS ONE THAT WE ARE WILLING TO ACCEPT, ONE WE ARE UNWILLING TO POSTPONE, AND ONE WHICH WE INTEND TO WIN, AND THE OTHERS, TOO..."

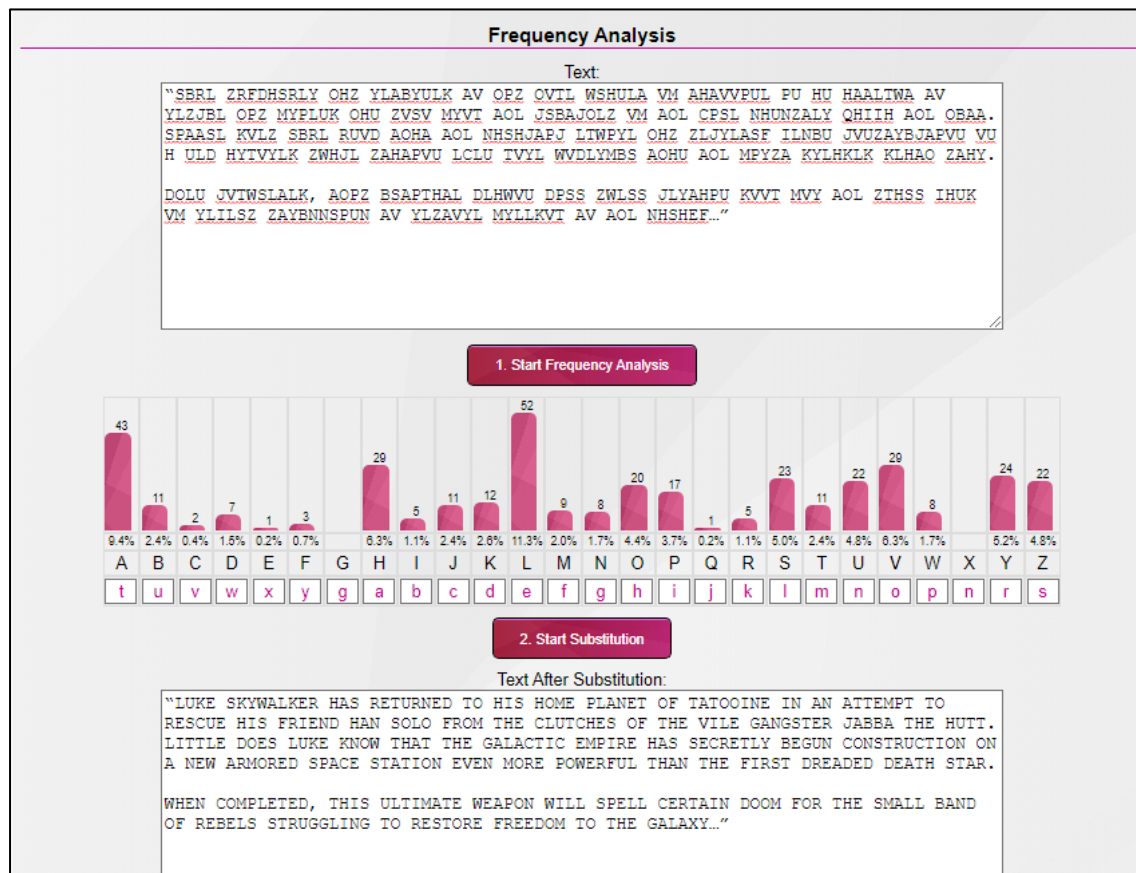
Cipher 6:



Cipher 7:



Cipher 8:



Conclusion: We successfully understood the process of Breaking the Mono-alphabetic Substitution Cipher using Frequency analysis method.