

Experiment No. 01

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

Theory:

Mono-alphabetic Substitution Cipher:

A mono-alphabetic substitution cipher is a type of substitution cipher where each letter of the plaintext is replaced with a fixed corresponding letter from the cipher alphabet. In other words, it involves mapping each letter of the alphabet to a different letter. The key to the cipher is the mapping between the plaintext alphabet and the cipher alphabet.

For example, using a simple mono-alphabetic substitution cipher with a fixed key, the mapping might look like this:

Plaintext: ABCDEFGHIJKLMNOPQRSTUVWXYZ

Cipher: XYZABCDEFGHIJKLMNPQRSTUVWXYZ

Advantages of Mono-alphabetic Substitution Cipher:

Ease of Implementation: Mono-alphabetic ciphers are relatively easy to implement and understand, making them accessible for educational purposes or simple encryption needs.

Initial Security: Mono-alphabetic ciphers can provide some basic level of security against casual attempts at decryption, especially if the cipher alphabet is randomly generated.

Disadvantages of Mono-alphabetic Substitution Cipher:

Vulnerable to Frequency Analysis: The biggest weakness of mono-alphabetic substitution ciphers is that each letter in the plaintext is always mapped to the same letter in the ciphertext. This leads to patterns in the ciphertext, making it susceptible to frequency analysis.

Limited Key Space: The key space of mono-alphabetic substitution ciphers is relatively small since there are only $26!$ (factorial) possible key combinations. This makes brute-force attacks feasible, especially with the aid of frequency analysis.

Lack of Perfect Secrecy: Unlike more complex ciphers like the one-time pad, mono-alphabetic substitution ciphers do not provide perfect secrecy. Once the key is discovered, the entire message can be decrypted.

Frequency Analysis Method:

Frequency analysis is a technique used to break mono-alphabetic substitution ciphers or ciphers with relatively weak encryption methods. It takes advantage of the fact that certain letters or combinations of letters occur with predictable frequency in natural languages like English.

The steps in a frequency analysis attack are as follows:

Collect Ciphertext: Obtain the encrypted message that you want to decrypt.

Analyze Frequency: Count the occurrences of each letter in the ciphertext. Certain letters will have higher frequencies due to their prevalence in the language.

Map Frequencies: Map the most frequently occurring letters in the ciphertext to the most frequently occurring letters in the language (e.g., 'E' in English).

Compare Context: Use the context of the message to identify other words and patterns to gradually piece together the key and the original plaintext.

Trial and Error: In more complex cases, frequency analysis may not fully decrypt the entire message, but it can significantly reduce the key space, allowing for manual trial and error to find the correct decryption.

Frequency analysis is particularly effective against longer ciphertexts because it provides more data for analyzing letter frequencies. To counter frequency analysis, more secure ciphers, such as poly-alphabetic ciphers or modern cryptographic algorithms like AES, were developed, which are not vulnerable to this type of attack.

Cipher 1:

Frequency Analysis

Text:

```
DJ DK C QLXDWI WF SDGDU PCX. XLRUQ KQCSLKBDQK, KJXDHDET FXWZ C BDIILE RCKL, BCGL PWE
JBLDX FDXXKJ GDSJWXO CTCDEKJ JBL LGDU TCUCSJDS LZQDXL. IYXDET JBL RCJJUL, XLRUQ KQDLK
ZCECTLI JW KJLCU KLSKLJ QUCEK JW JBL LZQDXL'K YUJDZCOL PLCQWE, JBL ILCJB KJCK, CE
CXZWXLI KQCSL KJCDW E PDJB LEWVIB QWFLX JW ILKQWXO CE LEJDXL QUCELJ. QYXKJL RO JBL
LZQDXL'K KDEDKJLX CTLEJK, QXDESLKK ULDC XCSLK BWZL CRWCXI BLX KJCKKBQ, SYKJWIDCE WF
JBL KJWULE QUCEK JBCJ SCE KCGL BLX QLWQUL CEI XLRQWXL FXLLIWZ JW JBL TCUCVO...
```

1. Start Frequency Analysis

Cipher Letter	Text Letter	Percentage
A	H	3.8%
B	A	7.8%
C	I	6.0%
D	N	4.4%
E	F	2.6%
F	V	1.0%
G	D	0.2%
H	T	1.0%
I	S	1.3%
J	E	8.0%
K	R	40
L	M	33
M	O	11.6%
N	P	6.6%
O	B	4.6%
P	C	3.6%
Q	G	3.0%
R	L	2.2%
S	X	1.6%
T	U	1.2%
U	V	0.8%
V	W	0.2%
W	Y	4.6%
X	Z	6.2%
Y	Q	1.2%
Z	U	1.8%

2. Start Substitution

Text After Substitution:

```
IT IS A PERIOD OF CIVIL WAR. REBEL SPACESHIPS, STRIDING FROM A HIDDEN BASE,
HAVE WON THEIR FIRST VICTORY AGAINST THE EVIL GALACTIC EMPIRE. DURING THE
BATTLE, REBEL SPIES MANAGED TO STEAL SECRET PLANS TO THE EMPIRE'S ULTIMATE
WEAPON, THE DEATH STAR, AN ARMORED SPACE STATION WITH ENOUGH POWER TO DESTROY
AN ENTIRE PLANET. PURSUED BY THE EMPIRE'S SINISTER AGENTS, PRINCESS LEIA RACES
HOME ABOARD HER STARSHIP, CUSTODIAN OF THE STOLEN PLANS THAT CAN SAVE HER
PEOPLE AND RESTORE FREEDOM TO THE GALAXY...
```

Cipher 2:

Frequency Analysis

Text:

```
"JVUI LUMNCUIJG KCL GIXVGEIS XO KPL KOYI AJCEIX OQ XCXOOPEI PE CE CXXIYAX XO GILWVI
KPL QGPIES KCE LOJO QGOY XXI WJVXWKL OQ XXI DPJ1 TCETLXIG BCHHC XXI KVXX.
JPXXJI SOIL JVUI UEOX XKCX XXI TCJCWXPW IYAPGI KCL LIWGIXJM HITVE WOELXGVWXPOE OE C
EIN CGYOGIS LACWI LXCPXOE IDIE YOGI AONIGQVJ XKCE XXI QPGLX SGICSISS SICXK LXCG.
NKIE WOYAJIXIS, XKPL VJXPYCXI NICAOE NPJJ LAIJJ WIGXCPX SOOY QOG XXI LYCJJ HCES OQ
GIHIJL LXGVTTJPET XO GILXOGI QGIISOY XO XXI TCJCZM..."
```

1. Start Frequency Analysis

Cipher Letter	Text Letter	Percentage
A	P	1.7%
B	A	0.2%
C	A	6.3%
D	V	0.4%
E	N	4.8%
F	R	5.2%
G	B	1.1%
H	E	11.3%
I	L	5.0%
J	H	4.3%
K	S	4.8%
L	Y	0.7%
M	W	1.5%
N	O	6.3%
O	I	3.7%
P	F	2.0%
Q		
R	D	2.8%
S	G	1.7%
T	K	1.1%
U	U	2.4%
V	C	2.4%
W	T	9.3%
X	M	2.4%
Y	X	0.2%
Z		

2. Start Substitution

Text After Substitution:

```
"LUKE SKYWALKER HAS RETURNED TO HIS HOME PLANET OF TATOOINE IN AN ATTEMPT TO
RESCUE HIS FRIEND HAN SOLO FROM THE CLUTCHES OF THE VILE GANGSTER ABBA THE HUTT.
LITTLE DOES LUKE KNOW THAT THE GALACTIC EMPIRE HAS SECRETLY BEGUN CONSTRUCTION ON
A NEW ARMORED SPACE STATION EVEN MORE POWERFUL THAN THE FIRST DREADED DEATH STAR.
WHEN COMPLETED, THIS ULTIMATE WEAPON WILL SPELL CERTAIN DOOM FOR THE SMALL BAND
OF REBELS STRUGGLING TO RESTORE FREEDOM TO THE GALAXY..."
```

Cipher 3:

Frequency Analysis

Text:

```
"OK OH R WRFD KOIS QPF KNS FSTSJJOPX. RJKNPAGN KNS WSRKN HKRF NRH TSSX
WSHKFPCSW, OIBSFORJ KFPPBH NRYS WFOYSX KNS FSTSJ QPFMSH QFPI KNSOF NOWWSX TRHS
EXW BAFHASH KNSI RMFPHH KNS GRJREC.
SYRWXG KNS WFSRMW OIBSFORJ HKRFQJSSK, R GFPAB PQ QFSSWPI QOGNKSFH JSW TC
JADS HDCVRJDSE NRH SHKRJJOHNSW R XSV HSMFSK TRHS PX KNS FSIPKS OMS VPFW PQ
NPKN.
KNS SYOJ JPFW WRFKQ YRWSF, PTHSHRSW VOKN QOXWGX CPAWXG HDCVRJDSE, NRH
WOHBRMMNSW KNPAAHRXWH PQ FSIPKS BFFPTSH OXKP KNS QRF FSRMNSh PQ HBRMS..."
```

1. Start Frequency Analysis

Character	Count	Percentage
A	7	1.5%
B	8	1.7%
C	6	1.2%
D	6	1.2%
E	1	0.2%
F	32	6.8%
G	29	6.0%
H	7	1.5%
I	8	1.7%
J	16	3.3%
K	30	6.2%
L	7	1.5%
M	27	5.6%
N	21	4.4%
O	26	5.4%
P	12	2.5%
Q	32	6.6%
R	60	12.4%
S	9	1.9%
T	5	1.0%
U	5	1.0%
V	25	5.2%
W	13	2.7%
X	5	1.0%
Y	5	1.0%
Z	0	0.0%

2. Start Substitution

Text After Substitution:

```
"IT IS A DARK TIME FIR THE REBELLIIM. ALTHIUHH THE DEATH STAR HAS BEEN
DESTRIYED, IMPERIAL TRIPS HAVE DRIVEM THE REBEL FIRCES FRIM THEIR HIDDEM BASE
AMD PURSUED THEM ACRISS THE HALAXY.
EVADIMH THE DREADED IMPERIAL STARFLEET, A HRIUP IF FREDIM FIRHTERS LED BY
LUKE SKYWALKER HAS ESTABLISHED A NEW SECRET BASE IM THE REMITE ICE WIRLD IF
HITH.
THE EVIL LIRD DARTH VADER, IBSESSSED WITH FIMDIMH YIUMH SKYWALKER, HAS
DISPATCHED THIUSAMDS IF REMITE PRIBES IMTI THE FAR REACHES IF SPACE..."
```

Cipher 4:

Text:

```
"ZRTFT IH PQFTHZ IQ ZRT XBGBOZIO HTQBZT. HTWTFBG ZRLPHBQV HLGBF HYHZTSH RBWT
VTOGBFTV ZRTIF IQZTQZILQH ZL GTBWT ZRT FTEPKGIO.
ZRIH HTEFBZIHZ SLWISTQZ, PQVTF ZRT GTBVIFHRIE LD ZRT SYHZTFILPH OLPQZ VLLAP,
RBH SBVT IZ VIDDIOPGZ DLF ZRT GSIZITV QPSKTF LD CTVI AQIXRZH ZL SBIQZBIQ ETBOT
BQV LFVTF IQ ZRT XBGBOJY.
HTQBZLF BSIVBGB, ZRT DLFTF NPTTQ LD QBKLL, IH FTZPFQIQX ZL ZRT XBGBOZIO
HTQBZT ZL WLZT LQ ZRT OFIZIOBG IHHPT LD OFTBZIQX BQ BFSY LD ZRT FTEPKGIO ZL
BHHIHZ ZRT LWTFMRTGSTV CIVI..."
```

1. Start Frequency Analysis

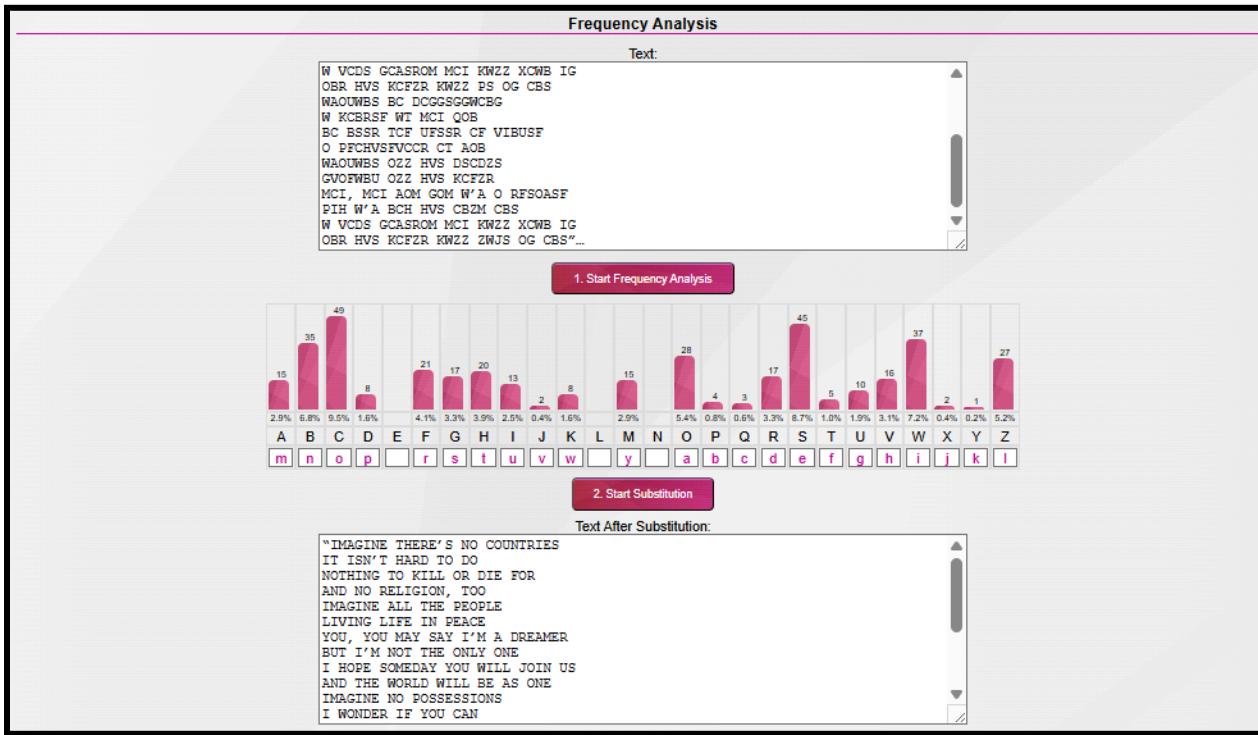
Character	Count	Percentage
A	2	0.4%
B	33	6.8%
C	2	0.4%
D	9	1.9%
E	5	1.0%
F	25	5.1%
G	15	3.1%
H	28	5.3%
I	32	6.6%
J	1	0.2%
K	4	0.8%
L	27	5.6%
M	1	0.2%
N	1	0.2%
O	13	2.7%
P	13	2.7%
Q	25	5.1%
R	20	4.1%
S	12	2.5%
T	56	11.5%
U	15	3.1%
V	6	1.2%
W	6	1.2%
X	4	0.8%
Y	44	9.1%
Z	0	0.0%

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



Answer in brief for the below questions:

1. What is the primary weakness of monoalphabetic cipher?
 - A. The **primary weakness of a monoalphabetic cipher** lies in its vulnerability to **frequency analysis**. Since each plaintext letter is always replaced by the same ciphertext letter, the statistical patterns of the language remain preserved. For example, in English, the letter **E** is the most common, followed by **T, A, O, I, N**. An attacker can study the frequency of letters in the ciphertext and match them with these common frequencies. Once a few substitutions are guessed, the rest of the message can usually be reconstructed easily.

2. How can you decode a message encrypted with a monoalphabetic cipher without knowing the key?
 - A. A **message encrypted with a monoalphabetic cipher can be decoded without knowing the key** by using cryptanalysis. The most common method is **frequency analysis**, where the attacker counts how often each symbol occurs in the ciphertext and

compares it with known letter frequencies in the language. Repeated patterns also help; for example, one-letter words are likely to be **A** or **I**, and common digraphs such as **TH**, **HE**, **IN** appear frequently. Once partial substitutions are made, the remaining letters can be revealed through pattern recognition and context-based guessing. Modern tools automate this process, making it even faster.

3. Can a monoalphabetic cipher be used to encode numbers and symbols as well as letters?
 - A. Monoalphabetic ciphers are not limited to letters; they **can be extended to encode numbers and symbols** as well. This is done by enlarging the substitution alphabet to include digits and punctuation marks. For example, the digit **0** might be substituted by a symbol like **#**, or a period might be replaced by **+**. However, this extension does not remove the cipher's weakness because numbers and symbols also follow predictable patterns of use (for instance, years like 2024 contain repeating digits, and punctuation like commas and full stops occur regularly). Thus, frequency analysis can still be applied.

4. What is a substitution table, and how is it used in monoalphabetic ciphers?
 - A. A **substitution table** is a chart that defines the mapping between plaintext characters and their ciphertext equivalents in a monoalphabetic cipher. It essentially acts as the “key” to the cipher. For encryption, each plaintext letter is replaced by the corresponding ciphertext symbol according to the table. For decryption, the process is reversed. For instance, in a Caesar cipher shifted by three places, **A** is mapped to **D**, **B** to **E**, and so on. Cryptanalysts also build partial substitution tables while breaking a cipher by gradually filling in the guessed mappings. This table form makes both encryption and decryption systematic and easy to follow.