

**Name:** Shrey Pendurkar

**Class:** D15C

**Batch:** C

**Roll No:** 64

## **CNS - Experiment 1**

**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.

**Answer in brief for the below questions:**

**1. What is the primary weakness of monoalphabetic cipher?**

The main weakness of a monoalphabetic cipher lies in its vulnerability to frequency analysis. In this kind of substitution cipher, each letter in the plaintext is always replaced by the same letter in the ciphertext throughout the message. Because of this fixed relationship, the statistical frequency of letters and common letter patterns of the plaintext language are preserved in the encoded message. For example, if 'E' is the most common letter in English, whatever letter replaces 'E' in the cipher will also be the most common letter in the ciphertext. Cryptanalysts can study these patterns and compare them to known letter frequencies in the target language to break the code without needing to know the key. This is why monoalphabetic ciphers, while simple, are considered insecure for protecting information.

**2. How can you decode a message encrypted with a monoalphabetic cipher without knowing the key?**

Decoding a message encrypted with a monoalphabetic cipher without knowing the key typically involves a technique called frequency analysis. This method takes advantage of the fact that every language has characteristic letter frequencies. For instance, in English, some letters like 'E', 'T', and 'A' appear more often than others. By analyzing the frequency of each character in the ciphertext and matching these frequencies to those found in the target language, a cryptanalyst can make educated guesses about which ciphertext letters correspond to which plaintext letters. Over time, by also spotting common patterns and digraphs (i.e., two-letter combinations like 'TH' or 'HE'), the attacker can usually reconstruct the entire substitution table and fully decode the message.

- Ciphertext: "XLI UYMGQ XLI QEPR"
- Suppose in English, "E" is the most common letter, and "X" appears frequently in the ciphertext.
- By matching the frequency of letters in the ciphertext with frequencies of letters in English, a cryptanalyst might guess that "X" corresponds to "T", "L" to "H", and so on, gradually revealing the original phrase: "THE QUICK THE BROWN".

### **3. Can a monoalphabetic cipher be used to encode numbers and symbols as well as letters?**

Yes, a monoalphabetic cipher scheme can be extended beyond just letters to include numbers and special symbols. Traditional monoalphabetic substitution ciphers only operate on letters (usually the 26 letters of the English alphabet), but there's nothing in principle preventing the extension of the substitution rule to cover other characters. To do this, you would simply expand the substitution table so that every symbol in your plaintext alphabet—including digits, punctuation marks, and other symbols—has a unique substitute in the ciphertext alphabet. Each number or symbol would be consistently replaced according to this expanded mapping, granting flexibility, though it does not increase security against frequency analysis by much.

- Let's say the plaintext is "MEET AT 10:00!"
- A substitution table assigns "M" → "Q", "E" → "R", "T" → "Y", "1" → "5", "0" → "3", ":" → "#", "!" → "@".
- The ciphertext becomes: "QRR YQ 53#33@"

### **4. What is substitution table, and how is it used in monoalphabetic ciphers?**

A substitution table in the context of monoalphabetic ciphers is essentially a reference chart or mapping that defines how each character from the plaintext alphabet is converted into a character in the ciphertext alphabet. The table contains two rows: one for all possible plaintext symbols and one for their corresponding ciphertext substitutes. During encryption, the sender looks up each character of the plaintext in the table and replaces it with the mapped character to form the ciphertext. Similarly, for decryption, the receiver uses the table in reverse to recover the original message. The substitution table is the 'key' to the cipher, and both the sender and receiver must agree upon it—and keep it secret—to successfully communicate using this technique.

- Suppose the English alphabet is mapped to another sequence:
  - Plaintext: A B C D E F G ... Z
  - Ciphertext: Q W E R T Y U ... M
- To encrypt "CAB", use the table:
  - "C" → "E", "A" → "Q", "B" → "W"
  - Result: "EQW"

## Screenshots:

**Frequency Analysis**

Text:

```
"DJ DK C QLXDWI WF SDGDU PCX. XRLU KQCSLKBDQK, KJXDHDET FWZ C BDIILE RCKL, BCGL PWE JBLDX FDXKJ GDSJWXO CTCDEKJ JBL LGDU TCUCSJDS LZQDXL. IYXDET JBL RCJJUL, XRLU KQDLK ZCECTLI JW KJLCU KLSXLJ QUCEK JW JBL LZQDXL'K YUJDZCJL PLCOWE, JBL ILCJB KUCX, CE CXZWXLJ KQCSL KUCJDWE PDJB LEWYTB QWPLX JW ILKJXWO CE LEJDXL QUCELJ. QYXKYLI RO JBL LZQDXL'K KDEDKJLX CTLEJK, QXDESLKK ULDC XCSLK BWZL CRWXCXI BLX KJCKKBDO, SYKJWIDCE WF JBL KJWULE QUCEK JBCJ SCE KCGL BLX QIWQUL CEI XLKJWXL FXLLIWZ JW JBL TCUCVO..."
```

1. Start Frequency Analysis

Letter	Count	Percentage
A	19	3.8%
B	39	7.8%
C	30	6.0%
D	22	4.4%
E	5	1.0%
F	5	1.0%
G	1	0.2%
H	13	2.6%
I	40	8.0%
J	33	6.6%
K	58	11.6%
L	4	0.8%
M	5	1.0%
N	18	3.6%
O	6	1.2%
P	11	2.2%
Q	8	1.6%
R	15	3.0%
S	1	0.2%
T	23	4.6%
U	6	1.2%
V	31	6.2%
W	9	1.8%
X	6	1.2%
Y	1	0.2%
Z	1	0.2%

2. Start Substitution

Text After Substitution:

```
"IT IS A PERIOD OF CIVIL WAR. REBEL SPACESHIPS, STRIKING 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 LETA RACES HOME ABOARD HER STARSHIP, CUSTODIAN OF THE STOLEN PLANS THAT CAN SAVE HER PEOPLE AND RESTORE FREEDOM TO THE GALAXY..."
```

**Frequency Analysis**

Text:

```
"JVUI LUMNCJUIIG KCL GIXVGEIS XO KPL KOYI AJCEIX OQ XCXOPEI PE CE CXXIYAX XO GILWVI KPL QGPIES KCE LOJO QGOY XKI WJWVKIL OQ XKI DPJI TCETLXIG BCHHC XKI KVXX. JPXXJI SOIL JVUI UEON XKCX XKI TCJCWXPW IYAPGI KCL LIWGIXJM HITVE WOELXGVWXFOE OE C EIN CGYOGIS LACWI LXCPXOE IDIE YOGI AONIQVJ XKCE XKI QPGLX SGICCSIS SICXK LXCG. NKIE WOYAJIXIS, XKPL VUXPYCXI NICAOE NPJJ LAIJU WIGXCPE SOOY QOG XKI LYCJJ HCES OQ GIHIJL LXGVTTJPET XO GILXOGI QGIISOY XO XKI TCJCZM..."
```

1. Start Frequency Analysis

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

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 JABBA 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..."
```

### 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 QPF KNSOF NOWNSX TRHS RXW BAFHASW
KNSI RMFPHH KNS GRJREC.
SYRWOXO KNS WFSRWWS OIBSFORJ HKRFQJSSK, R GFPAB PQ QFSSWPI QOGNKSFH JSW TC JADS
HDCVRJDSE NRH SHKRJOHNSW R XSV HSMFSK TRHS PX KNS FSIPKS QMS VPFJW PQ NPKN.
KNS SYOJ JFWW WRKRN YRWSF, PTHSHHSW VOKN QOXWOXG CPAVG HDCVRJDSE, NRH WOBRKMNSW
KNPAHRXWH PQ FSIPKS BFFTSK OXKP KNS QRF FSRMNSH PQ HBRMS..."
```

1. Start Frequency Analysis

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
7	8	6	6	1	32	7	29	8	16	30	1	27	21	28	12	32	60	9	5	5	25	13	5		
1.4%	1.7%	1.2%	1.2%	0.2%	6.6%	1.4%	6.0%	1.7%	3.3%	6.2%	1.4%	5.8%	4.3%	5.4%	2.5%	6.6%	12.4%	1.9%	1.0%	5.2%	2.7%	1.0%			
<input type="checkbox"/> u	<input type="checkbox"/> p	<input type="checkbox"/> y	<input type="checkbox"/> k	<input type="checkbox"/> x	<input type="checkbox"/> r	<input type="checkbox"/> g	<input type="checkbox"/> s	<input type="checkbox"/> m	<input type="checkbox"/> l	<input type="checkbox"/> t	<input type="checkbox"/> c	<input type="checkbox"/> h	<input type="checkbox"/> i	<input type="checkbox"/> o	<input type="checkbox"/> f	<input type="checkbox"/> a	<input type="checkbox"/> e	<input type="checkbox"/> b	<input type="checkbox"/> w	<input type="checkbox"/> d	<input type="checkbox"/> n	<input type="checkbox"/> v	<input type="checkbox"/>		

2. Start Substitution

Text After Substitution:

```
"IT IS A DARK TIME FOR THE REBELLION. ALTHOUGH THE DEATH STAR HAS BEEN DESTROYED,
IMPERIAL TROOPS HAVE DRIVEN THE REBEL FORCES FROM THEIR HIDDEN BASE AND PURSUED
THEM ACROSS THE GALAXY.
EVADING THE DREADED IMPERIAL STARFLEET, A GROUP OF FREEDOM FIGHTERS LED BY LUKE
SKYWALKER HAS ESTABLISHED A NEW SECRET BASE ON THE REMOTE ICE WORLD OF HOTH.
THE EVIL LORD DARTH VADER, OBSESSED WITH FINDING YOUNG SKYWALKER, HAS DISPATCHED
THOUSANDS OF REMOTE PROBES INTO THE FAR REACHES OF SPACE..."
```

### Frequency Analysis

Text:

```
"ZRTFT IH PQFTHZ IQ ZRT XBGBOZIO HTQBZT. HTWTFBG ZRLPHBQV HLGBF HYHZTSH RBWT
VTOGBFTV ZRTIF IQ2TQZILQH ZL GTBWT ZRT FTEPKGIO.
ZRIH HTEBFBZIH SLWTSTQZ, PQVTF ZRT GTBVFHRIE LD ZRT SYHZTFLPH OLPQZ VLLAP, RBH
SBVT IZ VIDDIOPGZ DLF ZRT GISIZTV QPSKTF LD CTIV AQIXRZH ZL SBIQZBIQ ETBOT BQV
LWTFV IQ ZRT XBGBOY.
HTQBZLF BSIVEGB, ZRT DLFSTF NPTTQ LD QBKLL, IH FTZPFQIQX ZL ZRT XBGBOZIO HTQBZT
ZL WLZT LQ ZRT OFIZIOBG IHHPT LD OFTBZIQX BQ BFSY LD ZRT FTEPKGIO ZL BHIIHZ ZRT
LWTFMRIGSTV CTVI..."
```

1. Start Frequency Analysis

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
2	33	2	9	5	25	15	26	32	1	4	27	1	1	13	13	20	12	58	15	6	6	4	44		
0.4%	6.8%	0.4%	1.8%	1.0%	5.1%	3.1%	5.3%	6.6%	0.2%	0.8%	5.5%	0.2%	0.2%	2.7%	2.7%	5.1%	4.1%	2.5%	11.5%	3.1%	1.2%	1.2%	0.8%	9.0%	
<input type="checkbox"/> k	<input type="checkbox"/> a	<input type="checkbox"/> j	<input type="checkbox"/> f	<input type="checkbox"/> p	<input type="checkbox"/> r	<input type="checkbox"/> l	<input type="checkbox"/> s	<input type="checkbox"/> i	<input type="checkbox"/> x	<input type="checkbox"/> b	<input type="checkbox"/> o	<input type="checkbox"/> w	<input type="checkbox"/> q	<input type="checkbox"/> c	<input type="checkbox"/> u	<input type="checkbox"/> n	<input type="checkbox"/> h	<input type="checkbox"/> m	<input type="checkbox"/> e	<input type="checkbox"/> d	<input type="checkbox"/> v	<input type="checkbox"/> g	<input type="checkbox"/> y	<input type="checkbox"/> t	

2. Start Substitution

Text After Substitution:

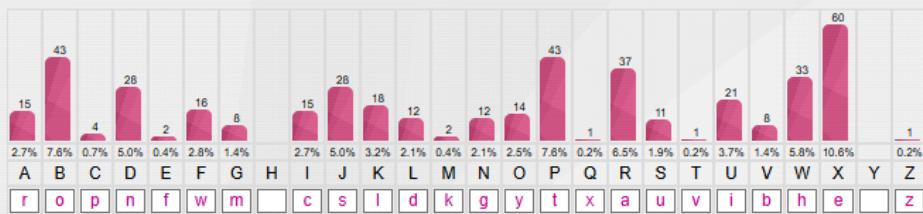
```
"THERE IS UNREST IN THE GALACTIC SENATE. SEVERAL THOUSAND SOLAR SYSTEMS HAVE
DECLARED THEIR INTENTIONS TO LEAVE THE REPUBLIC.
THIS SEPARATIST MOVEMENT, UNDER THE LEADERSHIP OF THE MYSTERIOUS COUNT DOOKU, HAS
MADE IT DIFFICULT FOR THE LIMITED NUMBER OF JEDI KNIGHTS TO MAINTAIN PEACE AND
ORDER IN THE 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..."
```

### Frequency Analysis

Text:

"FX IWBBJX PB NB PB PWX GBBD. VSP FWO, JBGX JRO, PWX GBBD? FWO IWBBJX PWUJ RJ BSA NBRK? RDL PWXO GRO FXKK RJM FWO IKUGV PWX WUNWXJP GBSDPRUD? FWO, 35 OXRAJ RNB, EKO PWX RPKRDPUI? FWO LBXJ AUIX CKRO PXQRJ? FX IWBBJX PB NB PB PWX GBBD UD PWUJ LXIRLX RDL LB PWX BPWXA PWUDNJ, DBP VXIRSJK PWXO RAX XRJO, VSP VXIRSJK PWXO RAX WRAL, VXIRSJK PWXP NBRK FUKK JXATX PB BANRDUZX RDL GXRJSAX PWX VXJP BE BSA KDXANUXJ RDL JMUKKJ, VXIRSJK PWXP IWRKXXDNX UU BDX PWXP FX RAX FUKKUDN PB RIIKCP, BDX FX RAX SDFUKKUDN PB CBJPCBDX, RDL BDX FWUW FX UDPXDL PB FUD, RDL PWX BPWXAJ, PBB."

**1. Start Frequency Analysis**



**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, 35 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 ORGANIZE 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."