Assignment 1: Cryptanalysis Set-10

Name: Rimjhim Singh
Roll number: 1910110317

Problem 1:

The following ciphertext is obtained by eavesdropping a communication which used Vigenere Cipher. Break the cipher by obtaining the used key to obtain the following ciphertext

Ciphertext:

“QFLVLQQNZMAFBWNWLCEXLKZERPDTSLBMOIESEACWHAVETEDZNIQEYEUXAPQOMN
DTZDBGNIOLTLLVVLICQOCXKGITFEAETWFOXSTSNLVTAIIERPXMDUEOGAYEIYASFDFZ
CKWNXLTNELACOASPRPEPWUGZFLOGIKPPJZUEAMJBSCXEGPVGQVMPSEDCHAVNZM
AFBWNWNWADDQUWPNZMAFBWNWHSINSQFYPFOEDXIJPTSZNPDIFZPLATZAAWJGZO
ETYNGNQLEIZYQFXMYLRJMQLOXSLTNLVWEXSPRMPAGNWTYABFIFPYXNOXACLAVES
EMLAAYYYTTZQUWISCJIDLYMWREFMMTBGNUFMIEBCTEXDLRPXIVAYDTNRAPQOMNLL
DJALAQDDUNSIKPLPDPTYWXWRPWENEZGJSCEHPZZAARELTTZVGBEASOEZVLDIDPSJ
DBWIWNLNMPQFIEYJDTQNWNIYEACCIFCIXPNEDIDHEEZNNPIHNSAPREJSFKAYLSBFIF
PYXDUAPZHKWTEIZYYMXMEDCLYIDOSMPIYPFLNMNLBWJTAJOPOTZRMLDICFSTYOSL
LPYOXPVGJGLWLPOYMWREFMPYBSJKWPMPYBLDICPSFWBAOXSLTLDMJEIDZFBFJAP
WNLNCPXJAWPYTOTNXAVPYTESQFCWDTMFW BSJIZFSWJ”

1. **Kasiski test**:
   Kasiski's test is a method of attacking polyalphabetic substitution ciphers, such as the Vigenère cipher. The first step to decode this ciphertext is to use Kasiski’s test to guess the length of the keyword used.

   We find **4 grams** which are repeating in the ciphertext. Then we look at the distances between their occurrences. The kasiski’s table is attached below. By analyzing the table we can notice that factors 3, 6 and 9 are the most common factors throughout. It is more likely that the keyword is of length 9, instead of 3 or 6.
   Hence our best guess for the length of the keyword is 9, (m=9)

		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
NZMA	180	Y	Y	Y	Y	Y			Y	Y		Y			Y			Y		Y				
ZMAF	180	Y	Y	Y	Y	Y			Y	Y		Y			Y			Y		Y				
MAFB	180	Y	Y	Y	Y	Y			Y	Y		Y			Y			Y		Y				
AFBW	180	Y	Y	Y	Y	Y			Y	Y		Y			Y			Y		Y				
FBWN	180	Y	Y	Y	Y	Y			Y	Y		Y			Y			Y		Y				
BWNW	180	Y	Y	Y	Y	Y			Y	Y		Y			Y			Y		Y				
NZMA	198	Y	Y			Y			Y		Y							Y				Y		
ZMAF	198	Y	Y			Y			Y		Y							Y				Y		
MAFB	198	Y	Y			Y			Y		Y							Y				Y		
AFBW	198	Y	Y			Y			Y		Y							Y				Y		
FBWN	198	Y	Y			Y			Y		Y							Y				Y		
BWNW	198	Y	Y			Y			Y		Y							Y				Y		
APQO	315		Y		Y		Y		Y						Y						Y			
PQOM	315		Y		Y		Y		Y						Y						Y			
QOMN	315		Y		Y		Y		Y						Y						Y			
BFIF	207		Y						Y														Y	
FIFP	207		Y						Y														Y	
IFPY	207		Y						Y														Y	
FPYX	207		Y						Y														Y	
EIZY	261		Y						Y															
YMWR	252	Y	Y	Y		Y	Y		Y			Y		Y				Y			Y			
MWRE	252	Y	Y	Y		Y	Y		Y			Y		Y				Y			Y			
WREF	252	Y	Y	Y		Y	Y		Y			Y		Y				Y			Y			
REFM	252	Y	Y	Y		Y	Y		Y			Y		Y				Y			Y			
MPYB	9		Y						Y															
LDIC	45		Y		Y				Y						Y									
OXSL	342	Y	Y			Y			Y									Y	Y					
XSLT	342	Y	Y			Y			Y									Y	Y					
WNLN	189		Y				Y		Y												Y			
ERP	84	Y	Y	Y		Y	Y					Y		Y							Y			
HAV	144	Y	Y	Y		Y		Y	Y			Y				Y		Y						Y
NZM	180	Y	Y	Y	Y	Y			Y	Y		Y			Y			Y		Y				
ZMA	180	Y	Y	Y	Y	Y			Y	Y		Y			Y			Y		Y				
MAF	180	Y	Y	Y	Y	Y			Y	Y		Y			Y			Y		Y				
AFB	180	Y	Y	Y	Y	Y			Y	Y		Y			Y			Y		Y				
FBW	180	Y	Y	Y	Y	Y			Y	Y		Y			Y			Y		Y				
BWN	180	Y	Y	Y	Y	Y			Y	Y		Y			Y			Y		Y				
WNW	180	Y	Y	Y	Y	Y			Y	Y		Y			Y			Y		Y				
WNW	182	Y					Y						Y	Y										
NZM	198	Y	Y			Y			Y		Y							Y				Y		
ZMA	198	Y	Y			Y			Y		Y							Y				Y		
MAF	198	Y	Y			Y			Y		Y							Y				Y		
AFB	198	Y	Y			Y			Y		Y							Y				Y		

2. **Index of coincidence:**

The next step is to confirm the length of the keyword. This can be done by computing the Index of coincidence.

The following table depicts the Index of Coincidence of a Random String of English Letters:

TABLE 1.1
Probabilities of occurrence of the 26 letters

letter	probability	letter	probability
A	.082	N	.067
B	.015	O	.075
C	.028	P	.019
D	.043	Q	.001
E	.127	R	.060
F	.022	S	.063
G	.020	T	.091
H	.061	U	.028
I	.070	V	.010
J	.002	W	.023
K	.008	X	.001
L	.040	Y	.020
M	.024	Z	.001

$$I_c(x) \approx \sum_{i=0}^{25} p_i^2 \approx 0.065$$

Ref: D.R. Stinson, Cryptography: Theory and Practice, Third Edition, CRC Press, 2006.

Index of coincidence uses the following formula:

$$IC = \sum_{i=A}^{i=Z} \frac{n_i(n_i - 1)}{N(N-1)}$$

Where ni is the number of occurrences of the letter i in the text and
N is the total number of letters.

When the index of coincidence is calculated for this ciphertext for keyword **length = 9** (m=9) we get the following value:

L=9 IC ≈ 0.06619 ± 0.009

IC = 0.06784
This value is fairly close to 0.065, which is what we were looking for.
Index of coincidence has confirmed our initial guess using the kasiski test. Hence it can be concluded that 9 is the correct length for the keyword.

3. Frequency Analysis:

The next step in our decrypting process is to perform frequency analysis to obtain the actual secret keyword.
Frequency analysis is the study of the distribution (and count) of the letters in a text. Analysis of frequencies helps cryptanalysis and decrypting substitution-based ciphers (Vigenere Cipher in this case) using the fact that some letters apparitions are varying in a given language: in English, letters E, T, or A are common while Z or Q are rare.

Letters by frequency of appearance in English:

E	12.7 %	T	9.1 %	A	8.2 %
O	7.5 %	I	7.0 %	N	6.7 %
S	6.3 %	H	6.1 %	R	6.0 %
L	4.0 %	D	4.3 %	C	2.8 %
U	2.8 %	M	2.4 %	W	2.4 %
F	2.2 %	G	2.0 %	Y	2.0 %
P	1.9 %	B	1.5 %	V	1.0 %
K	0.8 %	J	0.2 %	X	0.2 %
Q	0.1 %	Z	0.1 %		

Frequency Analysis of substrings in the Ciphertext:

1. The text is split into segments containing 9 letters. Shown in the Images below.
2. Then 9 substrings are made which contain the letters in each of the 9 columns.
3. Then we perform frequency analysis on each of the nine substrings.
4. The frequency analysis of each of these strings lead us to the keyword which is **"ALLISWELL"**

```
QFLVLQQNZ
OEZVLDIDP
NNPIHNSAP
NEDIDHEEZ
ACCIFCIXP
DTQNWNIYE
NMPQFIEYJ
SJDBWIWNL
TTZVGBEAS
SBFIFPYXD
HPZZAAREL
ENEZGJSCE
PTYWXWRPW
UNSIKPLPD
LDJALAQDD
NRAPQOMNL
RPXIVAYDT
REJSFKAYL
UAPZHKWTE
MMTBGNUFM
MPYBLDICP
TESQFCWDT
TOTNXAVPY
NCPXJAWPY
```

```
FBFJAPWNL
TLDMJEIDZ
SFWBAOXSL
MPYBSJKWP
IZYYMXMED
LPOYMWREF
OXPVGJGLW
STYOSLLPY
TZRMLDICF
BWJTAJOPO
IYPFLNMNL
CLYIDOSMP
IEBCTEXDL
IDLYMWREF
MAFBWNWLC
AETWFOXST
RPEPWUGZF
NELACOASP
FZCKWNXLT
AYEIYASFD
RPXMDUEOG
SNLVTAIIE
OCXKGITFE
UEAMJBSCX
TLLVVLICQ
```

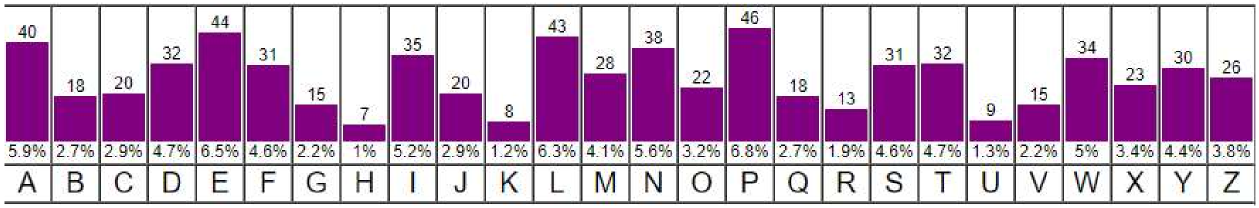
```
TZDBGNIOL
UXAPQOMND
EDZNIQEYE
EACWHAVET
TSLBMOIES
EXLKZERPD
LOGIKPPJZ
EGPVGQVMP
TZQUWISCJ
IZYQFXMYL
EMLAAYYYT
OXACLAVES
ABFIFPYXN
RMPAGNWTY
TNLVWEXSP
RJMQLOXSL
ETYNGNQLE
SEDCHAVNZ
TZAAWJGZO
NPDIFZPLA
EDXIJPTSZ
INSQFYPFO
MAFBWNWHS
ADDQUWPNZ
MAFBWNWNW
```

Additional method to perform frequency analysis on ciphertext:

The table shown below shows the frequency of occurrence of letters in the ciphertext compared to the frequency of occurrence of letters in the English language. The percentages show the frequency with which each letter occurs in the ciphertext.

So we begin by substituting the most frequent letter in the ciphertext by the most frequent letter in the English language and continued to do so. Just blind substitution does not help to actually decrypt the message, we also have to carefully and continuously analyze the plaintext after each substitution and check if it makes sense.

If a substitution does not help in shaping the text into something meaningful, we do not perform that substitution and look for something else.



After a lot of substitutions we get to the final meaningful plaintext (in this method keyword was not used)

4. **Recovered Plaintext**:

The decoded plain text after using the key: “ALLISWELL” is:

quantumcomputersaremachinesthatusethepropertiesofquantumphysicstostoredataandperform computationsthiscanbeextremelyadvantageousforcertaintaskswheretheycouldvastlyoutperform evenourbestsupercomputersclassicalcomputerswhichincludesmartphonesandlaptopsencodeinf ormationinbinarybitsthatcaneitherbesorsinaquantumcomputerthebasicunitofmemoryisaquantu mbitorqubitqubitsaremadeusingphysicalsystemssuchasthespinofanelectronortheorientationofa photonthesesystemscanbeinmanydifferentarrangementsallatonceapropertyknownasquantumsu perpositionqubitscanalsobeinextricablylinkedtogetherusingaphenomenoncalledquantumentangl ementtheresultisthataseriesofqubitscanrepresentdifferentthingssimul taneously

As it can be seen that the text has been decrypted into meaningful English words. The only thing that remains is to add spaces after each word to get the fully recovered text. The fully retrieved text is:

Quantum computers are machines that use the properties of quantum physics to store data and perform computations this can be extremely advantageous for certain tasks where they could vastly outperform even our best supercomputers classical computers which include smartphones and laptops encode information in binary bits that can either be sors in a quantum computer the basic unit of memory is a quantum bit or qubit qubits are made using physical systems such as the spin of an electron or the orientation of a photon the sea systems can be in many different arrangements all at once a property known as quantum superposition qubits can also be inextricably linked together using a phenomenon called quantum entanglement the result is that a series of qubits can represent different things simultaneously

This is the final plaintext!