CS 641 : ASSIGNMENT : PART 1

A PREPRINT

Srinjay Kumar
170722

Abhyuday Pandey
170039

Kumar Shivam
170354

January 23, 2020

1 Initial Sequence

The sequence of keywords used to get to the cipher text are :
go, enter, read

2 Problem

The first problem of our assignment is to decipher the ciphertext *"Nwy dejp pmcplpz cdp sxlrc adegipl ws cdp aejpr. Er nwy aem rpp cdplp xr mwcdxmv ws xmcplprc xm cdp adegipl. Rwgp ws cdp qecpl adegiplr fxqq ip gwlp xmcplprcxmv cdem cdxr wmp, x eg rplxwyr. Cdp awzp yrpz swl cdxr gprrevp xr e rxgbqp ryircxcycxwm axbdpl xm fdxad zxvxcr dejp ippm rdxscpz in 2 bqeapr. Swl cdxr lwymz berrfwlz xr vxjpm ipqwf, fxcdwyc cdp hywcpr."*.

3 Frequency Analysis

The frequency analysis for our cipher-text is shown in Table 1.

The expected percentage of characters in any general English text is shown in Table 2.

4 Analytical Decryption

- **Step 1** According to the expected frequency analysis the highest frequency is of character **e**. But in our cipher-text, the highest frequency is of character **p**. So, we can safely assume that the letter **p** from our ciphertext is actually letter **e**.

- **Step 2** Next, we look for single letter words in the cipher text. The two most probable options for the single letter words in the english language are **a** and **i**. Therefore, by doing some hit and trial, we get the ciphertext letters **x** and **e** to be **i** and **a** respectively.

- **Step 3** We now see the words that occur most permanently. The word **cdp** occurs quite frequently in the ciphertext. Since **p** has been guessed as **e**, therefore **cdp** must be **the** and hence **c** and **d** must be **t** and **h** respectively.

- **Step 4** We now have a word in the ciphertext **dejp**. Since, 3 letters have already been decrypted in the above word as "ha*e". We can safely mark **j** as **v**.

- **Step 5** A two letter word starting by **a** in the start of the sentence must be as or at. Therefore the word **er** must be **as**. So **r** is **s**.

- **Step 6** The word "*aves" translates to "caves". Therefore, **a** is **c**.
- **Step 7** The word must be **can** because ca has already been decrypted. Therefore, **m** is **n**.
- **Step 8** There is a phrase in the ciphertext "i a* se*i**s" which certainly translates to "i am serious". So we have **g,l,w,y** as **m,r,o,u** respectively.
- **Step 9** The first word of the cipher if You. Therefore, **N** is **Y**.
- **Step 10** The word **ws** is **of**. Therefore, **s** is **f**.
- **Step 11** The phrase "entere* the first cham*er" translates to "entered the first chamber". So we have **z** and **i** are **d** and **b** respectively.

After some further decryption by hand, we have the following information in Table 3.

The decrypted text is "You have entered the first chamber of the caves. As you can see there is nothing of interest in the chamber. Some of the later chambers will be more interesting than this one, i am serious. The code used for this message is a simple substitution cipher in which digits have been shifted by 2 places. For this round password is given below, without the quotes.".
Decrypting the password gives "cyLe81Lecy". But there is still a catch. A line in the deciphered text reads "digits have been shifted by 2 places". Since the digit in the cipher text is already shifted by itself therefore we have two possibilities. Assuming the required number to be x, we have the equation $2x \equiv 2 \mod 10$, we have two possible values of x, i.e. 1 and 6.

Therefore, we now have 4 possibilities.

- cyLe70Lecy
- cyLe92Lecy
- cyLe47Lecy
- cyLe25Lecy

Testing these passwords, we get "**cyLe70Lecy**" is our password.

5 Decryption by Script

We also wrote a script in Haskell for any general substitution cipher. The underlying algorithm is very simple.

- Initialize a mapping between letters with expected frequencies.
- Perform random swaps in mappings and check the number of words matching in the dictionary.
- In case, the number decreases, backtrack.
- Repeat

The details for the execution of the script are mentioned in readme.txt.

After executing the script on the ciphertext, we get the decrypted plaintext as "you have entered the first chamber of the caves. as you can see there is nothing of interest in the chamber. some of the later chambers will be more interesting than this one, i am serious. the code used for this message is a simple substitution cipher in which digits have been shifted by x places. for this round password is given below, without the kuotes." which is quite close to the actual plaintext.

References

- Letter Frequency

Table 1: Expected Frequency Analysis

Letter	Percentage
A	8.17%
B	1.49%
C	2.78%
D	4.25%
E	12.70%
F	2.29%
G	2.01%
H	6.09%
I	6.97%
J	0.15%
K	0.77%
L	4.02%
M	2.40%
N	6.74%
O	7.50%
P	1.92%
Q	0.09%
R	5.987%
S	6.37%
T	9.05%
U	2.75%
V	0.98%
W	2.36%
X	0.15%
Y	1.97%
Z	0.07%

Table 2: Frequency Analysis

Letter	Frequency	Percentage
P	42	14.69%
R	28	9.79%
C	27	9.44%
X	26	9.09%
D	22	7.69%
W	19	6.64%
L	16	5.59%
E	15	5.24%
M	15	5.24%
A	9	3.15%
Y	9	3.15%
I	8	2.8%
G	8	2.8%
S	7	2.45%
Z	7	2.45%
Q	6	2.1%
V	5	1.75%
F	5	1.75%
B	4	1.4%
J	4	1.4%
N	3	1.05%
H	1	0.35%

Table 3: Actual Key

Ciphertext	Actual
P	E
E	A
X	I
C	T
D	H
J	V
R	S
M	N
G	M
L	R
W	O
Y	U
N	Y
S	F
A	C
Z	D
I	B
B	P
F	W
H	Q
Q	L
V	G
K	-
O	-
T	-
U	-