CS641A Assignment-1

Sherlocked

Bholeshwar Khurana (170214) **Yaghyesh Chouhan (170813)** **Sarthak Dubey (180674)**

1 How we reached the cipher text:

1. We were standing at the base of a hill. We found a trail and followed it using the command **"go"**.
2. As we reached the end of the trail, we found a door. The door was closed but we noticed some message written on some rocks near the door. We read the message using the command **"read"**.
3. The rocks had something written about "The Great Caves" and we decided to enter the caves. To open the door in front of us, we used the command **"enter"** and entered the caves.
4. We realised that we have entered a small chamber. We found another closed door with something written on a glass chamber near it. We read the glass chamber by using the command **"read"**.
5. We found the ciper text:

Tlm usdw wrvwywb vuw hcygv pusfkwy lh vuw psdwg. Sg tlm psr gww vuwyw cg rlvucra lh crvwywgv cr vuw pusfkwy. Glfw lh vuw nsvwy pusfkwyg qcnn kw flyw crvwywgvcra vusr vucg lrw, c sf gwyclmg. Vuw plbw mgwb hly vucg fwggsaw cg s gcfznw gmkgvcvmvclr pczuwy cr qucpu bcacvg usdw kwwr guchvwb kt 2 znspwg. Hly vucg ylmrb zsggqlyb cg acdwr kwnlq, qcvulmv vuw jmlvwg.

tmHu42Dggm

2 How we cracked the cipher text:

To know which type of encryption was used to encrypt the text, we found out the frequency of each letter and the index of coincidence of the text (using the code *"freq_analysis.cpp"* which is provided along). The index of coincidence (IC) is a measure of how similar a frequency distribution is to the uniform distribution of English alphabets and if IC is close to 0.065, we conclude that the cipher is probably **Substitution Cipher**. The IC came out to be 0.07 which is quite close to 0.065, so we started with solving substitution cipher.

1. The letter 'w' appeared most frequently, and the difference in its frequency (42) with the second most frequent letter 'g' (30) was significant. So we substituted **w** with **e**.
2. The frequencies of the second and third most frequent letters were close, so we decided to guess the letters at this point.
3. There was a word 'gee', and we were confident of the letter e. The most common match would be 'see', so we substituted **g** with **s** in the ciphertext. Note that the other possibility could be 'bee' but 'see' has higher frequency than 'bee', so we tried with 'see' first.
4. The trigram 'vue'(here mapping of v and u is unknown) appeared quite frequently in the text (7 times), so we safely assumed it to be 'the' and substituted **v** with **t** and **u** with **h**.
5. Continuing with a similar analysis, we found the quadgram 'thcs' (mapping of c is unknown) to be quite frequent, so assuming it to be 'this' we substituted **c** with **i**.

6. There are two single-lettered words in the ciphertext 'c' and 's'. Since 'c' is already mapped to 'i', hence 's' must be mapped to 'a'. So, we substituted **s** with **a**.
7. The words 'i af' (mapping of f is unknown), the only meaningful words could be 'i am', so we substituted **f** with **m**.
8. Using 'messaae' (mapping of second last a is unknown), and assuming it to be 'message', we substituted **a** with **g**.
9. Using 'simzne' (mapping of z and n is unknown) and assuming it to be 'simple', we substituted **z** with **p** and **n** with **l**.
10. Using 'latey' (mapping of y is unknown) and assuming it to be 'later', we substituted **y** with **r**.
11. Using 'pipher' (mapping of first p is unknown) and assuming it to be 'cipher', we substituted **p** with **c**.
12. Using 'smkstitmtilr cipher' and assuming it to be 'substitution cipher', we substituted **m** with **u**, **k** with **b**, **l** with **o** and **r** with **n**.
13. Using 'Tou' and assuming it to be 'You', we substituted **t** with **y**.
14. Using 'hade' and assuming it to be 'have', we substituted **d** with **v**.
15. Using 'entereb' and assuming it to be 'entered', we substituted **b** with **d**.
16. Using 'hirst' and assuming it to be 'first', we substituted **h** with **f**.
17. Using 'qhich' and assuming it to be 'which', we substituted **q** with **w**.
18. Using 'juotes' and assuming it to be 'quotes', we substituted **j** with **q**.
19. After carrying out the above operations, we got the following mapping:

    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
    g d i v * u s f * q b o u l * c w n a y h t e * n p

    where mapping of *s is still unknown. We got the following decrypted text:

    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.

    yuFh42Vssu

20. The decrypted text reads that the digits have been shifted by 2 places, but we have not deciphered the digits yet. Assuming the original shift in digits to be x, we have

$$x + x = 2$$

$$\Rightarrow x = 1$$

21. Hence, the digits were shifted by 1 place. Using this, we obtain the password: **yuFh31Vssu** and thus advance to the next chapter.