CS641: Level 1
Team: TrojanHorse

Members:
Ayush Kumar, 180174
Rishav Kumar, 180612
Nilay Majorwar, 180483

January 23, 2020

Reaching the ciphertext:

Reaching the ciphertext was straightforward:

1. Speak `go` at the first screen.
2. `read` the instructions on the boulder in the second screen.
3. As in the instructions, speak `enter` at the second screen to enter the level's main chamber.
4. The chamber is empty, except for a locked door and a glass panel. `read` the glass panel to get the ciphertext.

Ciphertext on glass panel:

```
Age qlmd dbvdhdt vqd nrhxv iqljsdh gn vqd ilmdx. Lx age ilb xdd vqdhd rx
bgvqrbw gn rbvdhdxv rb vqd iqljsdh. Xgjd gn vqd olvdh iqljsdhx kroo sd jghd
rbvdhdxvrbw vqlb vqrx gbd, r lj xdhrgex. Vqd igtd exdt ngh vqrx jdxxlwd rx l
xrjpod xesxvrvevrgb irpqdh rb kqriq trwrvx qlmd sddb xqrnvdt sa 6 polidx.
Ngh vqrx hgebt plxxkght rx wrmdb sdogk, krvqgev vqd uegvdx.

emTc88Qqjt
```

Identifying the encryption method:

There were some clear clues that the method used was a substitution cipher - one of the most important clues was that **some words were repeated throughout the text**. For instance, the word `vqd` has been used 6 times, and `gn` and `rx` have been used 3 times each. Such a repetition in small-length ciphertext is almost impossible in linear cipher (with a proper, dense key), permutation cipher or modern encryption methods. So, it was easy to guess that the given ciphertext is just (mainly) a substitution over the plaintext.

The main clues:

1. The method is most certainly a classic alphabet-based encryption method, since the ciphertext has evenly distributed spaces like in English.
2. The length of the words also indicates alphabet-based encryption, as most words are 3-4 characters in length, like in English.
3. As mentioned above, there is repetition of some words like `vqd`, `gn` and `rx` in the ciphertext, which heavily points towards substitution cipher.

Breaking the cipher:

1. **Undoing the substitution:**

   The first step was obviously to run a frequency analysis of the ciphertext.

   The last word of the ciphertext was not included in the frequency analysis. The word was completely isolated from rest of the text and had numbers in between, and thus did not seem to be part of the main English part of plaintext. Even if it were somehow a part of the English plaintext, including the word would not have had any significant impact on frequency analysis results.

   Results:

   i. **D**: 14.69% (42 times)

   ii. **X**: 9.79% (28 times)

   iii. **V**: 9.44% (27 times)

   ...

   As the percentage gaps taper off, rest of the results are unreliable. Since `D` occurs well more often than others, taking `E` $\rightarrow$ `D` is a fair assumption.

   Now, as `vqd` occurs 6 times, we can guess that `vqd` $\rightarrow$ `the`, which gives us `T` $\rightarrow$ `V` and `H` $\rightarrow$ `Q`.

   It would have been reasonable to try `A` $\rightarrow$ `X`, except that `x` occurs consecutively in the word `plxxkght` (in the last line of ciphertext), which is not possible with `a`.

   Consider the word `vqrx` that occurs in the third line of the ciphertext. With the current substitutions, we have `vqrx` $\rightarrow$ `THrx`, where small letters denote unsubstituted letters and capitals denote substituted letters. Since the subword `rx` also occurs 3 times in the ciphertext as a full word, we can guess that `rx` $\rightarrow$ `is`, which gives us `I` $\rightarrow$ `R` and `S` $\rightarrow$ `X`.

   With these substitutions, the ciphertext (partially) decodes to the following:

   age HlmE EbTEhEt THE nIhST iHljsEh gn THE ilmES. lS age ilb SEE
   THEhE IS bgTHIbw gn IbTEhEST Ib THE iHljsEh. SgjE gn THE olTEh
   iHljsEhS kIoo sE jghE IbTEhESTIbw THlb THIS gbE, I lj SEhIgeS.
   THE igtE eSEt ngh THIS jESSlwE IS l SIjpoE SesSTITeTIgb iIpHEh
   Ib kHIiH tIwITS HlmE sEEb SHInTEt sa 6 poliES. ngh THIS
   hgebt plSSkght IS wImEb sEogk, kITHgeT THE uegTES.

   Now some words are starting to pop out, like `IbTEhESTIbw` $\equiv$ `INTERESTING` and `SesSTITeTIgb` $\equiv$ `SUBSTITUTION`. This gives us 6 more substitutions.

   Proceeding like this, we get our almost complete decryption key as

   `YN_EU_ORCMWAVFLPHIBDQTGS__` (i.e., `A` decrypts to `Y` and so on.)

   Since some letters, namely {c, f, y, z} do not appear in the ciphertext, there is no way to logically find the corresponding substitutions. The blanks in the key represent these gaps.

2. **Getting the password:**

   The decryption key, when applied on the first paragraph of ciphertext, gives:

   `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 6 places. For this round password is given below, without the quotes.`

   The last word (the password), when decrypted, gives:

   `uvD_88Hhmd`

   The encrypted password has the character `c`, which we do not know the substitution of. Thus, there is a blank in the decrypted password, which denotes the missing character. Since there are only 4 characters we do not know the substitution of, there are only **4 possibilities for the blank, i.e.** $\{j, k, x, z\}$.

   The decrypted message says that the cipher was a simple substitution cipher, but with digits shifted by 6. But since the message itself was encrypted the same way, the 6 is not the actual number in the message (our decryption key ignores numbers).

   If the original number in the message is d, then after encryption, it will get shifted by d, and the number in the ciphertext and our decrypted version will be $2d \ mod \ 10$. Thus, $2d \ mod \ 10 = 6$, which gives us $d = 3$ *or* 8. Then, the number 8 in the decrypted password should actually be either 5 or 0, i.e. we have **2 possibilities for the number, i.e.** $\{5, 0\}$.

   Thus, we have a total of just $4 \times 2 = 8$ possibilities of the key to try. Trying them one by one, the possibility **uvDz55Hhmd** succeeds, and we enter the chamber of Level 2.

– – – – – – – – – – – – – – – – – – – –