

SOLUTIONS REPORT

MODERN CRYPTOLOGY (CS641)

COMPUTER SCIENCE AND ENGINEERING

Levels 1-5

Team: **team58**

Aman Deep Singh (15807084)

Ayush Bansal (160177)

Gunjan Jalori (170283)

Date: March 8, 2020

1 Chapter 1 (The Entry)

There are 5 sub-levels in the chapter, first 4 of these don't have any cipher which needs to be decrypted.

The last sub-level is a **Substitution Cipher**, the answer to - "how it was recognised and solved" is explained in the subsection after the following list of commands.

Below is the solution to each of the sub-levels:

1. go
2. read
3. enter
4. read
5. cyLe70Lecy

1.1 Substitution Cipher

The ciphertext given was:

Nwy dejp pmcplpz cdp sxlrc adegip1 ws cdp aejpr. Er nwy aem rpp cdplp xr mwcdxmv ws xmcplprc xm cdp adegip1. Rwgpr ws cdp qecpl adegip1r fxqq ip gwlp xmcplprcxmv cdem cdxr wmp, x eg rplxwyr. Cdp awzp yrpz swl cdxr gprrevp xr e rxgbqp ryircxcycxwm axbdpl xm fdxad zxxvcr dejp ippm rdxscpz in 2 bqearp. Swl cdxr lwymz berrfwlz xr vxjpm ipqw, fxcdwyc cdp hywcpr.

For identifying what kind of cipher is applied in the above text, we will use the **Index of Coincidence**.

The detailed explanation on *Index of Coincidence* can be found in ??.

The *Index of Coincidence* of the above ciphertext is about 0.07, which is approximately same as a valid English text, this suggests that the cipher used is *Mono-alphabetic* such as *Substitution Cipher*.

For Solving the *Substitution Cipher*, the following steps were employed:

1. Calculate the frequency of each of the characters in the ciphertext, ignoring anything which is not an english alphabet.
2. The Character with the highest frequency is most probably 'e' or 'a', which can be placed in its place and identified further.
3. As the places get revealed, played hangman to find out what the other characters might be looking at one-letter, 2-letter, 3-letter words with highest number of characters revealed.
4. Built the decryption key by keeping a map of characters as they are being replaced.
5. Finally used the decryption key to decrypt the code given for the solution.

The code used in this part is in the file - `break_substitution.py`.

The Steps employed in the hangman game and building the key are mentioned below:

```
key = {}
key['p'] = 'e'      # Because 'p' has very high frequency
key['r'] = 's'      # 'r' has very high frequency,
                    # _ee word exists, matches with "see" not "bee"
key['i'] = 'b'      # _e word exists, matches with "be"
key['n'] = 'y'      # b_ word exists, matches with "by"
key['m'] = 'n'      # bee_ word exists, matches with "been"
key['w'] = 'o'      # _ne word exists, matches with "one"
```

```

key['s'] = 'f'      # o_ word exists, 'n' is already taken, matches with "of"
key['l'] = 'r'      # fo_ word exists, matches with "for"
key['y'] = 'u'      # yo_ word exists, matches with "you"
key['g'] = 'm'      # so_e word exists, matches with "some"
key['z'] = 'd'      # use_ word exists, 'r' is already taken, matches with "used"
key['c'] = 't'      # en_ered word exists, matches with "entered"
key['d'] = 'h'      # t_e word exists, matches with "the"
key['x'] = 'i'      # f_rst word and _ (single letter word) exist,
                    # matches with "first" and "i"
key['e'] = 'a'      # single letter word exists, 'i' is already taken, matches with "a"
key['j'] = 'v'      # ha_e word exists, matches with "have"
key['a'] = 'c'      # _hamber word exists, matches with "chamber"
key['v'] = 'g'      # nothin_ word exists, matches with "nothing"
key['f'] = 'w'      # _hich word exists, matches with "which"
key['q'] = 'l'      # be_ow and wi__ word exists, matches with "below" and "will"
key['b'] = 'p'      # sim_le and ci_her word exists, matches with "simple" and "cipher"
key['h'] = 'q'      # _uotes word exists, matches with "quotes"

```

The plaintext revealed after using the above decryption key 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.

```

# For the case of integer digits, "1" must be subtracted from each digit, as mentioned
# text after decryption, it was "2" but it itself was shifted so
# x+x = 2, this gives x = 1

```

So, final plaintext 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 1 places. For this round password is given below, without the quotes.

Using the above decryption key and the logic for digit, we can decipher the code for the answer as well:

Code: anQp81Qpan
Solution: cyLe70Lecy

2 Chapter 2 (The Caveman)

There are 2 sub-levels in the chapter, first one doesn't have any cipher which needs to be decrypted.

The second sub-level is a **Vigenere Cipher**, the answer to - "how it was recognised and solved" is explained in the subsection after the following list of commands.

The detailed explanation on *Vigenere Cipher* can be found in ??.

Below is the solution to each of the sub-levels:

1. read
2. the_cave_man_be_pleased

2.1 Vigenere Cipher

The ciphertext given was:

```
Lg ccud qh urg tgay ejbwdk, wmgf su bgud nkudnk lrd vjfbg. Yrhfm qvd vng sfuuxytj
"vkj_ecwo_ogp_ej_rnfkukf" wt iq urtuwjm. Ocz iqa jdag vio uzthsivi pqx vkj pgdy encpggt.
Uy hopg yjg fhkz arz hkscv ckoq pgfn vu wwygt nkioe zttft djcth.
```

For identifying what kind of cipher is applied in the above text, we will use the **Index of Coincidence**.

The detailed explanation on *Index of Coincidence* can be found in ??.

The *Index of Coincidence* of the above ciphertext is about 0.042, which is closer to the uniform distribution of English text, this suggests that the cipher is *Poly-alphabetic* such as *Vigenere Cipher*, it may be some other Poly-alphabetic cipher as well but we still have to give it a shot.

For solving the *Vigenere Cipher*, the following steps were employed:

1. Remove all characters from the text which are not part of the English alphabets and capitalize all characters.
2. Partition the text according to different key lengths and sort them according to the *Index of Coincidences* achieved, since higher the IC, closer it is to valid English Text.
3. For each keylen, perform frequency analysis to get the best key possible with the given length.
4. Try out all the keys retrieved and see which one gives some valid English text.

The code used in this part is in the file - `break_vigenere.py`.

The plaintext revealed after using the above decryption key is:

```
Be wary of the next chamber, there is very little joy there. Speak out the password
"the_cave_man_be_pleased" to go through. May you have the strength for the next chamber.
To find the exit you first will need to utter magic words there.
```

From the above, the solution is revealed: `the_cave_man_be_pleased`.

3 Chapter 3 (The Holes)

There are 4 sub-levels in the chapter, first 3 of these don't have any cipher but there are different tricks which need to be employed to get to the final sub-level.

The last sub-level is a **Permutation-Substitution Cipher**, the answer to - "how it was recognised and solved" is explained in the subsection after the following list of steps/commands.

Below are the solution steps to get out of the final chamber:

1. Type `enter` to go to sub-level 2.
2. At sub-level 2, you try to `put` your hand in the small hole, it is bitten, denoting there is someone there.
3. Type `enter` to go to sub-level 3, here there are a lot of mushrooms growing on the ground.
4. Type `pick` to pluck some mushrooms and come back to sub-level 2.
5. Type `give` to give the mushrooms to whatever is there in the small hole.
6. There is a spirit here, who gives you the code `thrnxtzy` which can reveal a hidden door at the entrance chamber (sub-level 1).
7. Go back to sub-level 1 and type `thrnxtzy`, this reveals a hidden door with a glass panel beside it.
8. Type `read` to get the ciphertext and code.
9. Type `jyg_izuqo_rr`, which is the decoded plaintext from the cipher provided.

3.1 Permutation-Substitution Cipher

The ciphertext given was:

```
cpiftgt ef oldo ukuq vtyp vv ptttqkk dp txe tkcnmbi uxkfft ueukwuqe ad uwv ttdo. da tocmc,
qqc qgcu woyg cx cpifteud wat tvkbd vu owk zclc dp txe vthr uccfgg. keb dteuof ut gle
dzcc rtc wv ukkyyc xxuo edw. mqgu zec dtyac uldw cgev evyu xvo tee moo mt gle dkcur.
tm evyoi qtzc cxz o mlcuauc, vw wetd kkcc gwhego! cf da foedokm, aibet ccd ktbkqyo:
```

For identifying what kind of cipher is applied in the above text, we used the following techniques:

- The **Index of Coincidence** of the above ciphertext is about 0.057, which is very close to that of valid English text, this suggests that the cipher used is *Mono-alphabetic* such as *Substitution Cipher*, the detailed explanation on *Index of Coincidence* can be found in ??.
- The **Chi-squared Statistic** of the above ciphertext is about 157 against *uniform distribution*, this suggests that the cipher used is *not Poly-alphabetic* since it is not closer to uniform distribution, the detailed explanation on *Chi-squared Statistic* can be found in ??.
- The **Chi-squared Statistic** of the above ciphertext is about 958 against *valid English text*, this suggests that the cipher used is **not** a *Simple Permutation* of letters.
- Based on the above, we try out different forms of *Mono-alphabetic* ciphers first instead of *Poly-alphabetic*.

Firstly, we will try to solve the cipher assuming it is *Simple Substitution Cipher*. This doesn't seem to work, since we are not able to get any valid English text from it.

The code for solving the Substitution Cipher uses the **n-gram** approach and it is in the file: `ngram_score.py`.

Since a *Simple Substitution Cipher* doesn't work here, it could be some other form of *Mono-alphabetic* cipher.

Lets make some observations about the ciphertext:

- The occurrences of double letter phrases in words is very frequent and at very odd places, see the below text:
cpiftgt ef oldo ukuq vtyp vv ptttqkk dp txe tkcnmbi uxkfft ueukwuqe ad uwv tt do.
da towc, qqc qgcu woyg cx cpifteud wat tvkbd vu owk zelc dp txe vthr uccfgg. keb
dteuof ut gle dzcc rtc ww ukkyyc xxuo edw. mgu zec dtyac uldw cgev evyu xvo tee
moo mt gle dkcur. tm evyoi qtzc cxz o mlcuauc, vw wetd kkcc gwhego! cf da foedokm,
aibet ccd ktbfkqyo:
- The character ‘o’ appears as a single letter, if we assume it to be ‘a’ or ‘i’ (according to english text), then the word ‘moo’ will coincide to ‘_aa’ or ‘_ii’ which will not come out to be a valid English word.
- A simple substitution solver doesn’t give us a valid result for the ciphertext.

The above observations suggest the following things:

- The letters in the ciphertext needs to be permuted before applying Substitution, this permutation can be a block permutation or matrix permutation or maybe something entirely different.
- The cipher is *Poly-alphabetic* (less-likely).

Firstly, we will try out **block-permutation** along with **Substitution**, i.e. **Simple Permutation-Substitution Cipher**.

For solving the *Simple Permutation-Substitution Cipher*, the following steps were employed:

1. Remove all characters from the text which are not part of the English alphabets, noting their position in the text since they will have to be added back at the end.
2. Calculate the *block length* (for permutation) using the idea that block length will be a *factor of the total number of characters*.
3. For each permutation get the permuted text from the ciphertext, and insert all the special characters at their designated places in the text.
4. Apply a Simple Substitution Cipher Solver to the text and see whether it gives a valid English Text.
5. If no permutation gives a successful result, try other block length till we reach some valid English Text or run out of factors.

The code used in this part is in the file - `break_perm-subs.py`.

The *total number of characters* in the ciphertext is 270.

Another observation to make here is that the code to be deciphered: `uhs_xafmf_no` has total of 10 characters.

This suggests that the block length will be a factor of both 270 and 10, so it can be 2, 5 or 10, and we will try each of these one by one.

Using the method described above, we got each permutation of ciphertext corresponding to block-length 2 first, but none of them returned a valid English text after solving the Substitution, so we changed block-length to 5.

Using block-length 5 revealed the following plaintext after a certain permutation of ciphertext was solved using Substitution Cipher Solver:

breaker of this code will be blessed by the squeaky spirit residing in the hole. go ahead, and find away of breaking the spell on him cast by the evil jaffar. the spirit of the cave man is always with you. find the magic wand that will let you out of the caves. it would make you a magician, no less than jaffar! to go through, speak the password:

At this point, we got a valid English text from the ciphertext, so we won't be moving forward and trying out different block-length or other ciphers.

Using the permutation and decryption key retrieved from solving above cipher, we can decipher the code for the answer as well:

The decryption key retrieved from the Solver gave us 2 possible solutions, since we did not have mapping for all 26 characters:

Code: `uhs_xafmf_no`

Solution 1: `jyg_izuqo_rr`

Solution 2: `jyg_ixuqo_rr`

Finally, Solution 1 was the answer.

4 Chapter 4 (The Spirit)

This level is tricky as we had to go back to the previous level and perform some task before we could proceed further.

There are 3 sub-tasks here, firstly we have to retrieve a *Magic Wand*, secondly we have to *free the Spirit*, finally solve the **DES Cipher** to advance to next level.

Below are the solution steps for each of the tasks listed above:

- Retrieving the Magic Wand:
 1. Type **enter** to go ahead in the chamber.
 2. Type **dive** to take a dive into the lake.
 3. At this point, we see an object looking like a wand, but trying to pull it directly causes us to drown, so first go back to surface and take a deep breath.
 4. Type **dive** and **pull** the wand.
- Freeing the Spirit:
 1. At this point, we can't figure out any way out, the screen in the chamber door is also blank and wand does not help us here.
 2. We can recall that there was an old man/spirit before who helped us in chamber 3 and mentioned that he was trapped by someone, he could be freed by the magic wand.
 3. We go back to chamber 3, **wave** our wand in front of the hole where the old man's spirit was.
 4. The spirit is freed and says that he will help us along the way.
- Solving the **DES Cipher**:
 1. After entering the 4th chamber, type **read**.
 2. The screen is still blank, but the spirit tells us what is supposed to be there:

This is a magical screen. You can whisper something close to the screen and the corresponding coded text would appear on it after a while. So go ahead and try to break the code! The code used for this is a 4-round DES, so it should be easy for you!! Er wait ... maybe it is a 6-round DES ... sorry, my memory has blurred after so many years. But I am sure you can break even 6-round DES easily. A 10-round DES is a different matter, but this one surely is not 10-round ... (long pause) ... at least that is what I remember. One thing that I surely remember is that you can see the coded password by whispering 'password'. There was something funny about how the text appears, two letters for one byte or something like that. I do not recall more than that. I am sure you can figure it out though ...

4.1 3-Round DES Cipher

Earlier, we were trying to figure out how many rounds of DES Cipher is applied here, so that we can devise the algorithm for the same.

The task became easier when a hint was revealed stating that the DES Cipher is *3 Round*.

We will break *3 Round DES* using **Differential Cryptanalysis** as we had discussed in class.

The idea behind *Differential Cryptanalysis* was to get rid of the unknown value (i.e. the Key) so that an equation can be formed over the non-linear step (i.e. sBoxes).

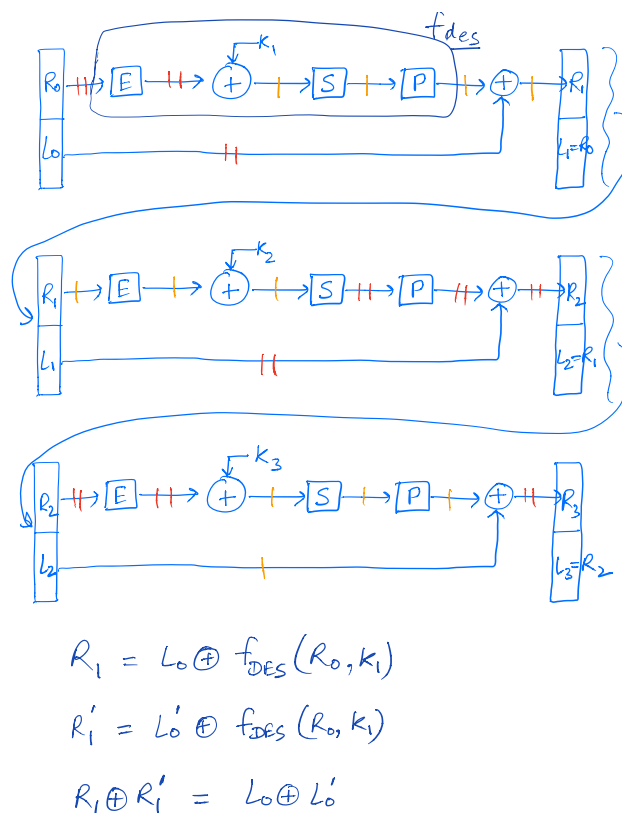


Figure 1: 3 Round DES

Important points about the figure above:

- *Double Red Line* implies we know both individual values of *Differential Cryptanalysis*.
- *Single Golden Line* implies we know the differential value of 2 inputs, not the individual values themselves.
- The pairs of inputs taken during Differential Cryptanalysis must have equal 32 bits on the right side.

Solving 3 Round DES:

1. We will start by figuring out the 48-bit Key for the 3rd Round.
2. We know the differential value, just after and before the S-Boxes, thus, there will be 4 possible pairs of input values to the S-Box as studied in class.
3. This narrows down the search for key over each 6 bits to 4 possibilities.
4. We will pick another input here, take the intersection of possibilities and narrow down the key.
5. After sufficient tries, we get the 48-bit 3rd round key.
6. Now only 8 bits of the key remain, these can be easily brute-forced and figured out using the other round values.

The code used in this part is as follows:

- `constants.py`: Contains the constants for the DES.
- `des.py`: Defines the DES Class encoding all the functions related to it.
- `utils.py`: Defines the common utility functions related to DES and Key generation.

- `break_des.py`: Defines the main function which uses all the utilities and DES class to break the 3 Round DES according the steps described above.
- `generate_input.py`: Generates a pair of input which have equal 32 bits on the right side.

The key retrieved for the 3rd round of the DES is as follows:

[61, 28, 9, 54, 55, 9, 28, 51]

here, each value represents the 6-bits of the 48-bit key.

After doing brute force on the remaining part of the key, we get the following value:

147

The Final Key (64-bit including the parity bits) for the DES is as follows:

0111101001011100001010000011011011110010011010100110101011101000

The encrypted password: `gnushmilfrplulktkrtrtgtjojfqqpt`

The decrypted password: `rirfiirqnujpopirgkholonsqntpkqqi`

5 Chapter 5 (The Fall)

This level is a bit similar to last one in the way of decoding the cipher, our *Magic Wand* will come handy in this level.

There are 6 sub-levels in the chapter, first 5 of these don't have any cipher but there are trick which needs to be employed to reach the final sub-level.

The last sub-level is an **AES-type Cipher**, the answer to - "how it was recognised and solved" is explained in the subsection after the following list of steps/commands.

Below is the solution to each of the sub-levels:

1. go further in the passage as you enter the chamber 5.
2. wave your wand as you begin to fall, as anything else will lead to your death as you reach the end.
3. dive into the water, as there is nothing around you but water.
4. go further into the passage you reach after diving into the water.
5. read the glass panel beside the closed door.

Once we reach the sub-level 6, we get the following information from the spirit about what's written on the screen:

"This is another magical screen. And this one I remember perfectly... Consider a block of size 8 bytes as 8×1 vector over F_{128} - constructed using the degree 7 irreducible polynomial $x^7 + x + 1$ over F_2 . Define two transformations: first a linear transformation given by invertible 8×8 key matrix A with elements from F_{128} and second an exponentiation given by 8×1 vector E whose elements are numbers between 1 and 126. E is applied on a block by taking the i th element of the block and raising it to the power given by i th element in E . Apply these transformations in the sequence $EAEAE$ on the input block to obtain the output block. Both E and A are part of the key. You can see the coded password by simply whispering 'password' near the screen..."

5.1 AES-type Cipher

Firstly, lets talk discuss some observations made about the problem:

- Each element of the input 8×1 vector is part of the finite polynomial field F_{128} , so each element (byte) will lie between 0 and 127 (both inclusive).
- The input and output is encoded as in last level, i.e. one byte is represented by 2 characters, each lying between 'f' and 'u'. If byte value is 123, then its encoded value will be - 'mq'.
- Exponentiation process has a 8×1 vector, each element of the input is raised to corresponding power in the vector, the elements of this exponent vector will lie between 1 and 126 (both inclusive).
- Linear Transformation Matrix A will be an 8×8 matrix with each element from the finite polynomial field F_{128} , so each element will lie between 0 and 127 (both inclusive).
- Exponentiation is a byte to byte operation and this operation is independent for each byte.
- Linear Transformation is a mixed byte type operation and the mixing will depend on the actual values of the Matrix.
- Both A and E are unknown to us and direct brute-force for the solution would require approx. 128^{72} or 2^{504} computations.

For breaking the cipher, we will perform the **Chosen Plaintext Attack** on the cipher.

Before moving towards the actual solution, we will prove the following:

- Matrix A is **Lower Triangular Matrix**.
- In the 8×1 output vector, i^{th} byte will depend on all bytes upto $(i-1)^{th}$ byte of the input, i.e. changing i^{th} byte of input will change all bytes from i^{th} to the 8^{th} byte of output.

Lemma 1

Matrix A is **Lower Triangular Matrix**

Proof 1

Suppose the matrix A is as follows:

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 \\ b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & b_8 \\ c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 & c_8 \\ d_1 & d_2 & d_3 & d_4 & d_5 & d_6 & d_7 & d_8 \\ e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 \\ f_1 & f_2 & f_3 & f_4 & f_5 & f_6 & f_7 & f_8 \\ g_1 & g_2 & g_3 & g_4 & g_5 & g_6 & g_7 & g_8 \\ h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 \end{bmatrix}$$

Part 1: Using the **eighth** byte.

Since, we are performing a chosen plaintext attack, lets take inputs of the following format:

$$ffffffffffffffffXX$$

here, XX will be a non-zero byte value i.e. from 1 to 127 (both inclusive).

The ciphertext for the above plaintext format comes out to be:

$$ffffffffffffffffXX$$

All first seven bytes remain zero (ff), even though the last byte of the input was non-zero, this is true for any value of the eighth byte.

The 8×1 vector retrieved after performing the first Linear Transformation will be as follows:

$$\begin{bmatrix} a_8 \cdot v & b_8 \cdot v & c_8 \cdot v & d_8 \cdot v & e_8 \cdot v & f_8 \cdot v & g_8 \cdot v & h_8 \cdot v \end{bmatrix}^T$$

here, v is the value of the eighth byte of Exponentiation.

The value of first seven bytes must result in zero for all possible input values of eighth byte, this will be true only if the **first seven values of the last column of A are zero**.

So, the updated matrix A will be as follows:

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & 0 \\ b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & 0 \\ c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 & 0 \\ d_1 & d_2 & d_3 & d_4 & d_5 & d_6 & d_7 & 0 \\ e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & 0 \\ f_1 & f_2 & f_3 & f_4 & f_5 & f_6 & f_7 & 0 \\ g_1 & g_2 & g_3 & g_4 & g_5 & g_6 & g_7 & 0 \\ h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 \end{bmatrix}$$

Part 2: Using the **seventh** byte.

Now, lets take input of the following format:

ffffffffffffXXff

The ciphertext for the above plaintext format comes out to be:

ffffffffffffXXXX

First six bytes remain zero (ff), even though the seventh byte of the input was non-zero, this is true for any value of the seventh byte.

The 8x1 vector retrieved after performing the first Linear Transformation will be as follows:

$$\begin{bmatrix} a_7 \cdot v & b_7 \cdot v & c_7 \cdot v & d_7 \cdot v & e_7 \cdot v & f_7 \cdot v & g_7 \cdot v & h_7 \cdot v \end{bmatrix}^T$$

here, v is the value of the seventh byte of Exponentiation.

The value of first six bytes must result in zero for all possible input values of seventh byte, this will be true only if the **first six values of the seventh column of A are zero**.

So, the updated matrix A will be as follows:

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & 0 & 0 \\ b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & 0 & 0 \\ c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & 0 & 0 \\ d_1 & d_2 & d_3 & d_4 & d_5 & d_6 & 0 & 0 \\ e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & 0 & 0 \\ f_1 & f_2 & f_3 & f_4 & f_5 & f_6 & 0 & 0 \\ g_1 & g_2 & g_3 & g_4 & g_5 & g_6 & g_7 & 0 \\ h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 \end{bmatrix}$$

Part 3: Generalizing for the rest of the bytes.

Taking the following byte formats one at a time and performing the same steps we did above, we get the following ciphertext formats:

$$\begin{aligned} \text{EAEAE}(ffffffffffffXXffff) &= fffffffffffffXXXXXX \\ \text{EAEAE}(ffffffffffffXXffff) &= fffffffffffffXXXXXXXX \\ \text{EAEAE}(ffffffffXXffffffffff) &= fffffffXXXXXXXXXXXXX \\ \text{EAEAE}(ffffXXffffffffff) &= fffffXXXXXXXXXXXXX \\ \text{EAEAE}(ffXXffffffffff) &= ffXXXXXXXXXXXXX \end{aligned}$$

Performing the same calculations, the finally updated matrix A that we receive will be as follows:

$$\begin{bmatrix} a_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ b_1 & b_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ c_1 & c_2 & c_3 & 0 & 0 & 0 & 0 & 0 \\ d_1 & d_2 & d_3 & d_4 & 0 & 0 & 0 & 0 \\ e_1 & e_2 & e_3 & e_4 & e_5 & 0 & 0 & 0 \\ f_1 & f_2 & f_3 & f_4 & f_5 & f_6 & 0 & 0 \\ g_1 & g_2 & g_3 & g_4 & g_5 & g_6 & g_7 & 0 \\ h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 \end{bmatrix} \quad (1)$$

The above matrix is a **Lower Triangular Matrix**.

Lemma 2

In the 8x1 output vector, i^{th} byte will depend on all bytes upto $(i-1)^{th}$ byte of the input, i.e. changing i^{th} byte of input will change all bytes from i^{th} to the 8th byte of output.

Proof 2

The Matrix A will be as defined in eq. (1).

Let the input 8x1 vector be as follows:

$$\begin{bmatrix} i_1 & i_2 & i_3 & i_4 & i_5 & i_6 & i_7 & i_8 \end{bmatrix}^T$$

After applying the Linear Transformation using the matrix A , we will get the following 8x1 vector:

$$\begin{bmatrix} (a_1 \cdot i_1) \\ (b_1 \cdot i_1) + (b_2 \cdot i_2) \\ (c_1 \cdot i_1) + (c_2 \cdot i_2) + (c_3 \cdot i_3) \\ (d_1 \cdot i_1) + (d_2 \cdot i_2) + (d_3 \cdot i_3) + (d_4 \cdot i_4) \\ (e_1 \cdot i_1) + (e_2 \cdot i_2) + (e_3 \cdot i_3) + (e_4 \cdot i_4) + (e_5 \cdot i_5) \\ (f_1 \cdot i_1) + (f_2 \cdot i_2) + (f_3 \cdot i_3) + (f_4 \cdot i_4) + (f_5 \cdot i_5) + (f_6 \cdot i_6) \\ (g_1 \cdot i_1) + (g_2 \cdot i_2) + (g_3 \cdot i_3) + (g_4 \cdot i_4) + (g_5 \cdot i_5) + (g_6 \cdot i_6) + (g_7 \cdot i_7) \\ (h_1 \cdot i_1) + (h_2 \cdot i_2) + (h_3 \cdot i_3) + (h_4 \cdot i_4) + (h_5 \cdot i_5) + (h_6 \cdot i_6) + (h_7 \cdot i_7) + (h_8 \cdot i_8) \end{bmatrix}$$

The above matrix clearly implies that first byte is independent of other bytes, second byte depends on first byte and itself, third byte depends on first, second bytes and itself and so on.

Now, we have established certain properties about Linear Transformation and Exponentiation, we can move on to cracking password and breaking the cipher.

Firstly, we will just decode the ciphertext of the password and get the corresponding plaintext without finding the actual values of the matrix A and exponent E .

Nextly, we will use *Chosen Plaintext Attack* to retrieve the values of A and E .

5.1.1 Cracking Password

We will make use of the result of lemma 2.

Given a 8-byte ciphertext, we will determine the plaintext bytes one-by-one beginning from the first byte.

The Algorithm used is as follows:

1. Divide the password into chunks of 8 bytes, if last chunk is not complete 8 bytes then pad it with zeroes (ff), and solve each chunk separately.
2. Initialize **known_plaintext** as an empty string, since nothing is known right now.
3. For each (i^{th}) byte of the ciphertext, do the following:
 - (a) Enumerate over all possibilities of byte values - 0 to 127, let this value be **current_byte**.
 - (b) Generate a plaintext as - **known_plaintext** + **current_byte** + **padding**, here padding will be applied to make the input plaintext 8 bytes.
 - (c) Get the corresponding ciphertext, let its i^{th} byte be **cipher_byte**.
 - (d) If **cipher_byte** is equal to the i^{th} byte of the input ciphertext, then update **known_plaintext** = **known_plaintext** + **current_byte**.
 - (e) Go to next iteration of Step 3.

The above algorithm uses the fact that first byte is independent of all other bytes, so it can be directly found.

Once the first byte is found, it can be kept constant and second byte can be found using the same method, this continues for the rest of the bytes.

The number of computations performed in the above algorithm will be $128 \cdot 8$ or 2^{10} .

The code for the above algorithm can be found in `decrypt_password.py`.

The password given was: `ktirlqhtlqijmmhqmgkplijngrluiqlq`

The plaintext solved is: `lhlgmjmkmgqlompmoltmglllqlmlgmh`

Directly adding the plaintext as the result, didn't advance us to the next level.

The ciphertext didn't have multiple plaintexts corresponding to it as well.

We thought that the input is an 8-byte value which is actually represented as 16 characters, and we tried printing the value of each byte which was represented by the plaintext:

The plaintext solved was: `lhlgmjmkmgqlompmoltmglllqlmlgmh`

The byte values were:

[98, 97, 116, 117, 113, 107, 105, 122, 121, 110, 113, 99, 107, 103, 97, 114]

Surprisingly, all byte values lie within the range of ascii values of 'a' and 'z'.

Finally, we tried the string generated by getting the ascii values of each of the above bytes, and this was the final answer.

Final Result was: `batuqkizynqckgar`

5.1.2 Solving for A and E

A and E comprise the key of this encryption algorithm.

Using lemma 1, we know that A is a lower triangular matrix.

Let the input 8x1 vector be as follows

$$\begin{bmatrix} i_1 & i_2 & i_3 & i_4 & i_5 & i_6 & i_7 & i_8 \end{bmatrix}^T$$

Let the E 8x1 vector be as follows

$$\begin{bmatrix} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 \end{bmatrix}^T$$

Let the A 8x8 matrix be as follows

$$\begin{bmatrix} a_{(1,1)} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{(2,1)} & a_{(2,2)} & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{(3,1)} & a_{(3,2)} & a_{(3,3)} & 0 & 0 & 0 & 0 & 0 \\ a_{(4,1)} & a_{(4,2)} & a_{(4,3)} & a_{(4,4)} & 0 & 0 & 0 & 0 \\ a_{(5,1)} & a_{(5,2)} & a_{(5,3)} & a_{(5,4)} & a_{(5,5)} & 0 & 0 & 0 \\ a_{(6,1)} & a_{(6,2)} & a_{(6,3)} & a_{(6,4)} & a_{(6,5)} & a_{(6,6)} & 0 & 0 \\ a_{(7,1)} & a_{(7,2)} & a_{(7,3)} & a_{(7,4)} & a_{(7,5)} & a_{(7,6)} & a_{(7,7)} & 0 \\ a_{(8,1)} & a_{(8,2)} & a_{(8,3)} & a_{(8,4)} & a_{(8,5)} & a_{(8,6)} & a_{(8,7)} & a_{(8,8)} \end{bmatrix}$$

The output of the first exponentiation will be

$$\begin{bmatrix} i_1^{e_1} & i_2^{e_2} & i_3^{e_3} & i_4^{e_4} & i_5^{e_5} & i_6^{e_6} & i_7^{e_7} & i_8^{e_8} \end{bmatrix}^T$$

The output of the first linear transform will be

$$\begin{bmatrix} (a_{(1,1)} \cdot i_1^{e_1}) \\ (a_{(2,1)} \cdot i_1^{e_1}) + (a_{(2,2)} \cdot i_2^{e_2}) \\ (a_{(3,1)} \cdot i_1^{e_1}) + (a_{(3,2)} \cdot i_2^{e_2}) + (a_{(3,3)} \cdot i_3^{e_3}) \\ (a_{(4,1)} \cdot i_1^{e_1}) + (a_{(4,2)} \cdot i_2^{e_2}) + (a_{(4,3)} \cdot i_3^{e_3}) + (a_{(4,4)} \cdot i_4^{e_4}) \\ (a_{(5,1)} \cdot i_1^{e_1}) + (a_{(5,2)} \cdot i_2^{e_2}) + (a_{(5,3)} \cdot i_3^{e_3}) + (a_{(5,4)} \cdot i_4^{e_4}) + (a_{(5,5)} \cdot i_5^{e_5}) \\ (a_{(6,1)} \cdot i_1^{e_1}) + (a_{(6,2)} \cdot i_2^{e_2}) + (a_{(6,3)} \cdot i_3^{e_3}) + (a_{(6,4)} \cdot i_4^{e_4}) + (a_{(6,5)} \cdot i_5^{e_5}) + (a_{(6,6)} \cdot i_6^{e_6}) \\ (a_{(7,1)} \cdot i_1^{e_1}) + (a_{(7,2)} \cdot i_2^{e_2}) + (a_{(7,3)} \cdot i_3^{e_3}) + (a_{(7,4)} \cdot i_4^{e_4}) + (a_{(7,5)} \cdot i_5^{e_5}) + (a_{(7,6)} \cdot i_6^{e_6}) + (a_{(7,7)} \cdot i_7^{e_7}) \\ (a_{(8,1)} \cdot i_1^{e_1}) + (a_{(8,2)} \cdot i_2^{e_2}) + (a_{(8,3)} \cdot i_3^{e_3}) + (a_{(8,4)} \cdot i_4^{e_4}) + (a_{(8,5)} \cdot i_5^{e_5}) + (a_{(8,6)} \cdot i_6^{e_6}) + (a_{(8,7)} \cdot i_7^{e_7}) + (a_{(8,8)} \cdot i_8^{e_8}) \end{bmatrix}$$

Observe that if the input 8×1 vector has only one non zero block i_k , then the output blocks upto index $k - 1$ will be zero. k_{th} block of output is $(a_{(k,k)} \cdot (a_{(k,k)} \cdot i_k^{e_k})^{e_k})^{e_k}$. The k^{th} block of the output will depend only on i_k, e_k and $a_{(k,k)}[(k,k)^{th}$ element in A].

So, for each $k \in \{1, 2, 3, 4, 5, 6, 7, 8\}$ multiple inputs are considered where only the k^{th} block is non-zero and takes all possible values in $[0, 127]$. For each such input, k^{th} block of the output is considered. Using such pairs, we calculate possible values of e_k and $a_{(k,k)}$ by iterating through all possible values of e_k and $a_{(k,k)}$ and considering those which provide correct input to output transform.

Now the task is do minimize this set of possible values of e_k and $a_{(k,k)}$ to get unique values. Also we need to find non-diagonal elements of A .

Consider the k^{th} block of output when only the j^{th} block of input is non-zero ($k > j$). k_{th} block of output is $(a_{(k,j)} \cdot (a_{(j,j)} \cdot i_j^{e_j})^{e_j} + a_{(k,k)} \cdot (a_{(k,j)} \cdot i_j^{e_j})^{e_k})^{e_k}$. The k^{th} block of the output will depend on only $e_j, e_k, a_{(k,k)}, a_{(j,j)}, a_{(k,j)}$.

Because we know a reduced set of possible values of $e_j, e_k, a_{(k,k)}, a_{(j,j)}$, we can get possible values of $a_{(k,j)}$ by considering plaintexts where j^{th} block of input is non-zero and considering k^{th} block of the corresponding ciphertext and iterating over all possible values of $a_{(k,j)}$ and considering only those which provide correct input to output transform. Also, we remove values of e_j 's, e_k 's, $a_{(j,j)}$'s and $a_{(k,k)}$'s which have no such possible value of $a_{(k,j)}$, from the set of possible values of e_k 's and $a_{(k,k)}$'s. By doing such calculations for all $j \in \{1, 2, 3, 4, 5, 6, 7\}$ and $k \in [j + 1, 8]$, we get unique values all non-diagonal elements of A . We also get unique values of diagonal elements of $A(a_{(k,k)}$'s) and e_k 's.

We obtain E vector as

$$[85 \ 52 \ 38 \ 72 \ 116 \ 38 \ 66 \ 50]^T$$

We obtain A matrix as

$$\begin{bmatrix} 100 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 122 & 56 & 0 & 0 & 0 & 0 & 0 & 0 \\ 6 & 121 & 40 & 0 & 0 & 0 & 0 & 0 \\ 10 & 97 & 77 & 50 & 0 & 0 & 0 & 0 \\ 58 & 14 & 78 & 10 & 16 & 0 & 0 & 0 \\ 9 & 76 & 114 & 116 & 92 & 87 & 0 & 0 \\ 104 & 30 & 98 & 92 & 104 & 44 & 14 & 0 \\ 13 & 91 & 54 & 58 & 113 & 17 & 37 & 103 \end{bmatrix}$$

The code for getting input and output pairs can be found in the script `getInputOutputPairs.py`.

The code for this part can be found in the script `getKeySandDecrypt.py`.

Once we have E and A , getting the password was just reduced to bruteforcing over all possible inputs and finding the input corresponding to the encrypted password. The input was then converted to ASCII format and the result was `batuqkizynqckgar`.

6 Appendix

This section explains each of the things used in between the solutions without proper explanation.

6.1 Index of Coincidence

The **Index of Coincidence** is a measure of how similar a frequency distribution is to the uniform distribution.

$$I.C. = \frac{\sum_{i=A}^{i=Z} f_i(f_i - 1)}{N(N - 1)}$$

where f_i is the count of letter i (where $i = A, B, \dots, Z$) in the ciphertext, and N is the total number of letters in the ciphertext.

Important facts about the *Index of Coincidence*:

- The *Index of Coincidence* of valid English text is about 0.066.
- The *Index of Coincidence* for uniform distribution of English text is about 0.038.
- The *Index of Coincidence* remains the same for the ciphertext and plaintext if cipher is **Mono-alphabetic** (i.e. Substitution Cipher).
- The *Index of Coincidence* of ciphertext is closer to uniform distribution if cipher is **Poly-alphabetic** (such as Vigenere Cipher).

We can get an approximate idea of what kind of cipher is used to generate the ciphertext by using the *Index of Coincidence*.

6.2 Chi-squared Statistic

The **Chi-squared Statistic** is a measure of how similar two categorical probability distributions are. If the two distributions are identical, the chi-squared statistic is 0, if the distributions are very different, some higher number will result. The formula for the chi-squared statistic is:

$$\chi^2(C, E) = \sum_{i=A}^{i=Z} \frac{(C_i - E_i)^2}{E_i}$$

where C_A is the count (not the probability) of letter A , and E_A is the expected count of letter A .

Important facts about the *Chi-squared Statistic*:

- If the *Chi-squared Statistic* of a ciphertext against *uniform distribution* is very low (~ 50 or less), then it is highly probable that the cipher is *Poly-alphabetic*.
- If the *Chi-squared Statistic* of a ciphertext against *valid English text* is high and the cipher is *Mono-alphabetic*, then it can be solved by trying keys and lowering it.

We can get an approximate idea of whether the cipher is *Poly-alphabetic* or not by using *Chi-squared Statistic*.

6.3 Vigenere Cipher

The Vigenere Cipher is a polyalphabetic substitution cipher.

Suppose, the length of the encryption key is k , then the string formed by picking out each letter with a multiple of k letters in between them will be a *Caesar Cipher*.

Since each such string is a *Caesar Cipher*, the *Index of Coincidence* of this string will be closer to that of valid English text rather than closer to uniform distribution.

Using the above principle, we can crack the *Vigenere Cipher*.