First homework Joosep Parts 221963IVCM

Tasks

Task 1. Assume that the Affine cipher is implemented in $\mathbb{Z}_{97}$, not in $\mathbb{Z}_{26}$. (Imagine that we just extended alphabet, added a set of special symbols. But the first 26 letters stay the same as in English alphabet.)

1. For this modified Affine cipher over $\mathbb{Z}_{97}$, the encryption and decryption functions remain structurally the same but adapt to the new modulo.

$$\text{Encryption: } Enc(m) = a \cdot m + b(\mod 97)$$
$$\text{Decryption: } Dec(c) = a^{-1}(c - b)(\mod 97)$$

2. To calculate the total number of keys, we need to consider the options for both a and b in the key $k = (a, b)$. Since 97 is a prime number, we use Euler's totient function to find the number of choices for a:

$$\phi(97) = 97 - 1 = 96$$

For the additive part b, we have 97 choices, ranging from 0 to 96 inclusive:

$$\text{Choices for } b = 97$$

Therefore, the total number of possible keys is:

$$\text{Total number of keys} = 96 \times 97 = 9312$$

3. We read English alphabet and get numerical equivalents for B, U, H are 1, 20, and 7. We do equations to solve for a and b. To find the encryption and decryption keys, solve the following equations to find a and b.

$$\begin{aligned} 81 &= a \cdot 1 + b \mod 97, \\ 71 &= a \cdot 20 + b \mod 97, \\ 37 &= a \cdot 7 + b \mod 97. \end{aligned}$$

Solve for a and b:
First, eliminate b by substituting it from the first equation into the third:

$$\begin{aligned} 6a + 81 &\equiv 37 \mod 97, \\ 6a &\equiv -44 \mod 97, \\ 6a &\equiv 53 \mod 97. \end{aligned}$$

Now, find the multiplicative inverse of 53 mod 97, which is $a^{-1} = 66$.

$$\begin{aligned} a &= 66 \cdot 53 \mod 97, \\ a &= 25 \mod 97. \end{aligned}$$

Finally, solve for b using the first equation:

$$\begin{aligned} b &= 81 - 25 \mod 97, \\ b &= 56 \mod 97. \end{aligned}$$

So, $(a, b) = (25, 56) \mod 97$.
Decrypting the Cipher goes with decryption, $Dec = (a^{-1}, b) = (66, 56)$.
Decrypt $c' = [59, 62, 90]$ using $m' = a^{-1}(c' - b) \mod 97$:

$$\begin{aligned} m'_1 &= 66 \cdot (59 - 56) \mod 97 = 66 \cdot 3 \mod 97 = 198 \mod 97 = 4, \\ m'_2 &= 66 \cdot (62 - 56) \mod 97 = 66 \cdot 6 \mod 97 = 396 \mod 97 = 8, \\ m'_3 &= 66 \cdot (90 - 56) \mod 97 = 66 \cdot 34 \mod 97 = 2244 \mod 97 = 13. \end{aligned}$$

So, $m' = [4, 8, 13]$, which corresponds to the airport code "EIN".

Task 2. Look carefully at the following frequency diagrams.

1. Well just like we did in class, by following the patterns and deducting possibility we could come to a 'certain' conclusion by analyzing the English alphabet frequency.

   (a) Figure 1: Diagram is identical to Diagram 3 this seems to correspond to a normal distribution, so I would say it represents plaintext . Distribution of Z, Q and J is >0.2 percent and letters such as E, T and A are most highest frequency.

   (b) Figure 2: Diagram 2. This looks like plain text distribution however, in addition to shifting the frequencies have also changed places. So it affine cipher.

   (c) Figure 3: Diagram 3 this seems to correspond to a normal distribution, so I would say it represents permutation as we already have one plaintext identified and permutation could look exactly like plaintext because it does not change frequency. Distribution of Z, Q and J is >0.2 percent and letters such as E, T and A are most highest frequency.

   (d) Figure 4: Diagram 4 - Vigenère cipher because the keyword causes letters to repeat in the ciphertext, which makes frequency more uniform, meaning, it's likely that the key was shorter than the message and letters start to repeat.

   (e) Figure 5: Diagram 5 this is shift cipher locations of frequencies have changed locations but the order is same.

2.

Let the plaintext alphabet be represented as $\mathcal{A} = \{A, B, \ldots, Z\}$.
Let the ciphertext alphabet be represented as $\mathcal{B} = \{A, B, \ldots, Z\}$.
Let the most frequent letter in the plaintext be 'E'.
Let the most frequent letter in the ciphertext be 'T'.

According to Figure 1 or 3, in the plaintext, the most frequent letter is 'E'.
According to Figure 5, in the ciphertext, the most frequent letter is 'T'.

If we consider 'A' to be 0 and 'Z' to be 25, then:
$E = 4, \quad T = 19$

The shift from 'E' to 'T' can be represented as:
Shift = $T - E$
Shift = $19 - 4$
Shift = 15

Therefore, the shift cipher has shifted the letters by 15 places.

Task 3. Assuming that the rate of English language is 1.8, find unicity distance of affine cipher.
The unicity distance U for an affine cipher can be calculated using Shannon's formula:

$$U = \frac{\log_2 M}{R - 1}$$

For further references see http://practicalcryptography.com/cryptanalysis/text-characterisation/statistics/unicity-distance

Where:

- U is the unicity distance
- R is the rate of the natural language (given as 1.8 for English)
- M is the size of the key space

For an affine cipher over an alphabet of size N, the key space size M is $N \times \phi(N)$, given that a (the multiplicative key) must be chosen such that $\gcd(a, N) = 1$.
If we're dealing with the English alphabet, $N = 26$, and Euler's totient function $\phi(26) = 12$.
Therefore, $M = 26 \times 12 = 312$.
Plugging these into Shannon's formula, we get:

$$U = \frac{\log_2 312}{1.8 - 1} = \frac{\log_2 312}{0.8} \approx \frac{8.29}{0.8} \approx 10.36$$

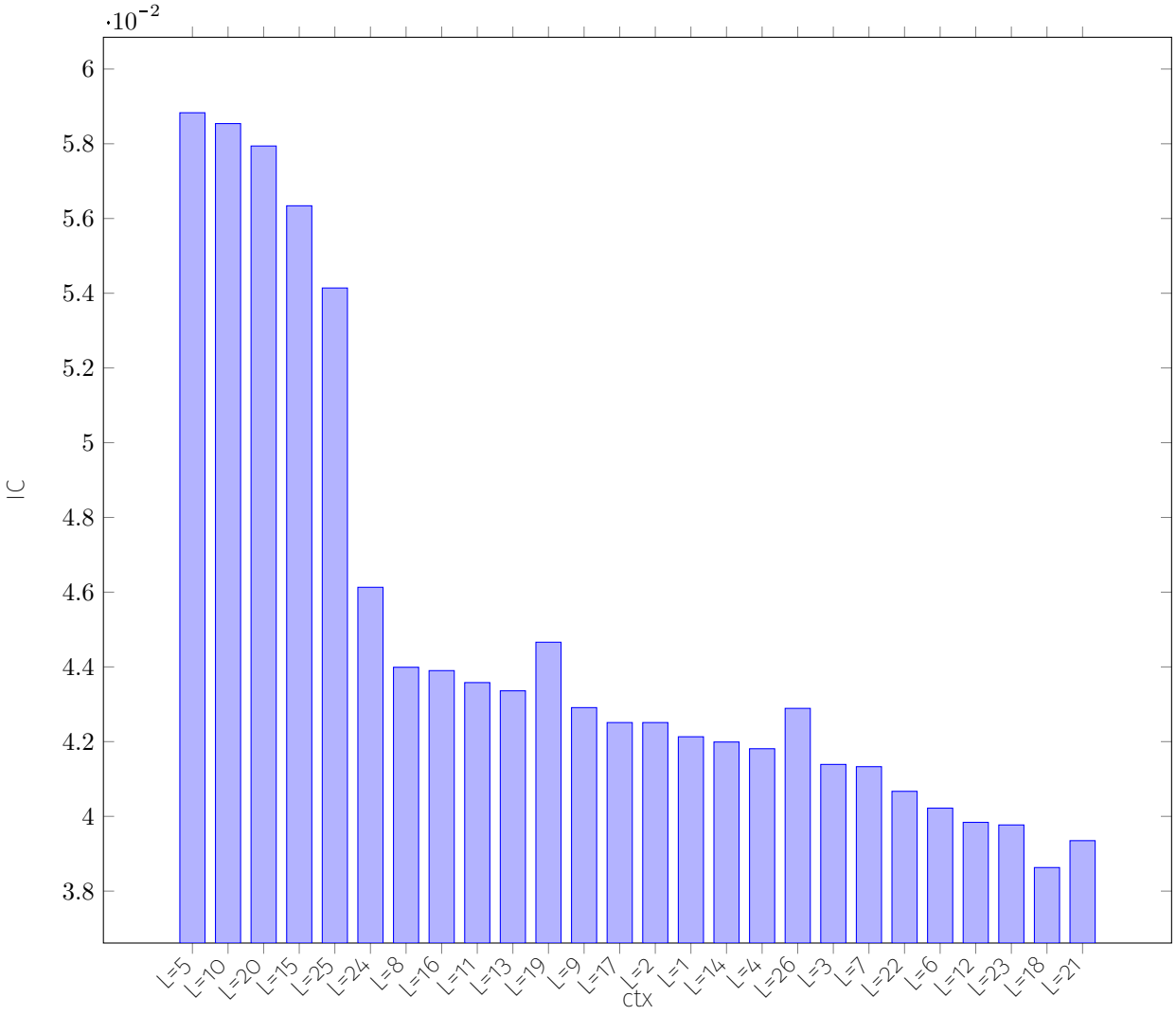
So the unicity distance U for an affine cipher with these parameters is approximately 10.36 characters.

Task 4. Suppose you intercepted the following ciphertext $c = 00010010 \quad 00000111 \quad 11101010$. You know that a 3-letter word was encrypted using one-time pad (to convert letters to binary strings ASCII table was used). Can you bruteforce keys and learn the message that was encrypted?

No, bruteforcing keys in the case of a one-time pad is not possible. Because for me the key is truly random and it has been used one. Even if I brute force and get some answer, I don't know if the message unencrypted is correct. Suppose 00010010 is ASCII code for letter 'A'. Even if I bruteforce with key 00010010 I would get back 'A', with 00010011 I would get something different etc. I could get up to 256 different plaintexts eventually, but for none of which I know to be true.

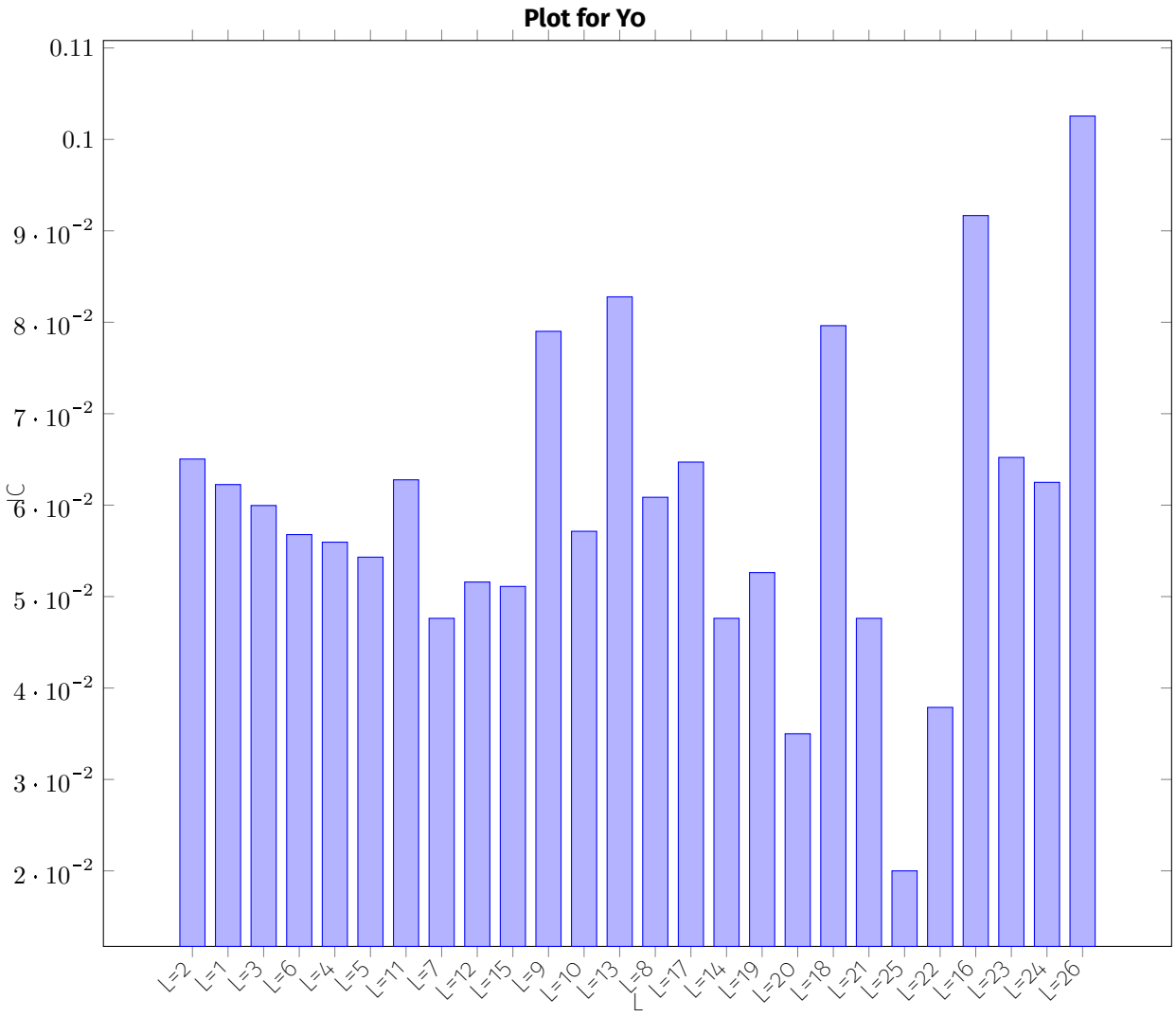
Task 5. You have intercepted the following ciphertext encrypted using Vigenere cipher. You have a crypto-analyst friend who can help you break the cipher, but he asked you to **find key length**.

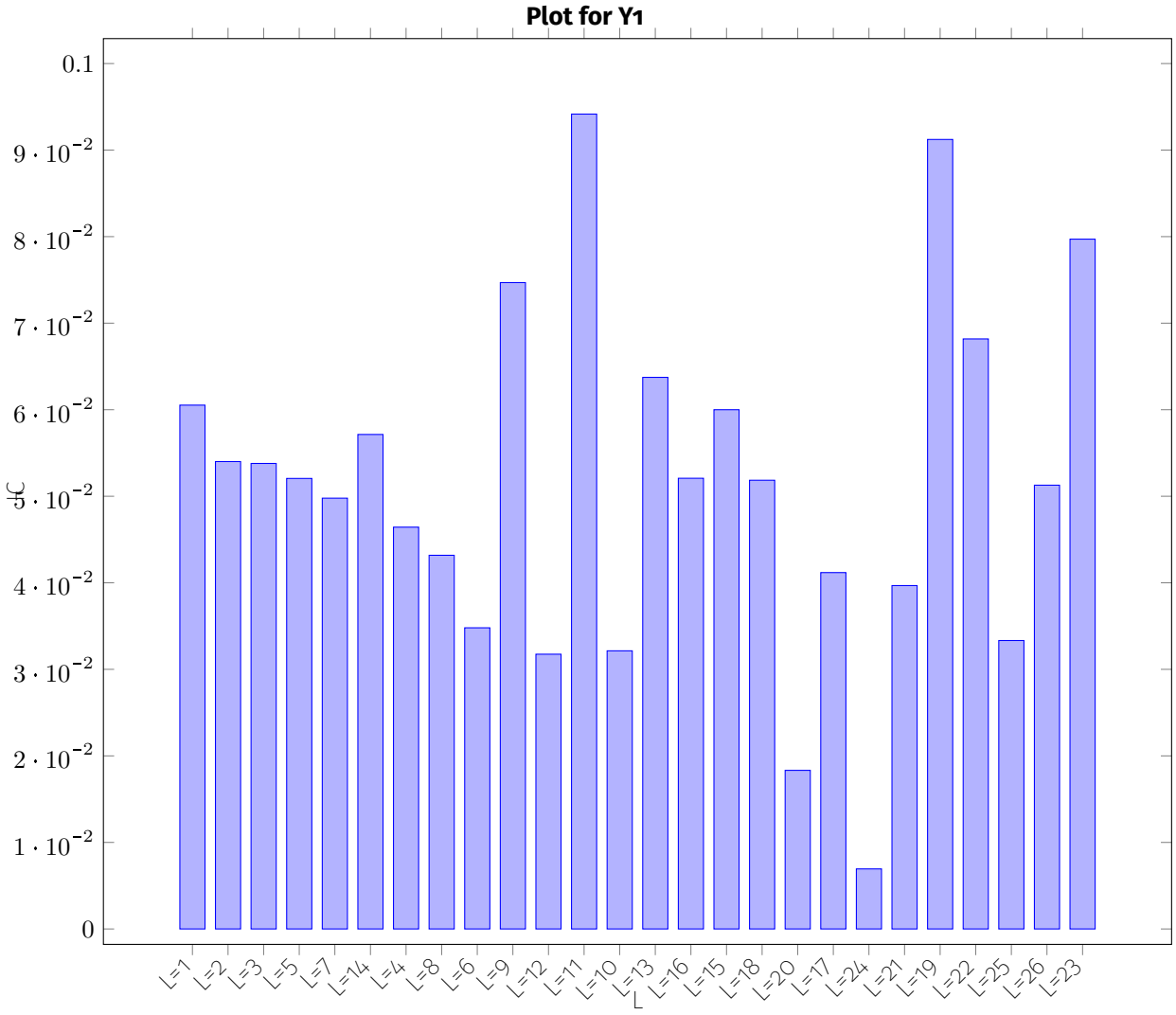
I use https://www.dcode.fr/index-coincidence to find the key length for ctx. Index of Coincidence on dCode.fr [online website], retrieved on 2023-09-24, https://www.dcode.fr/index-coincidence
I get the following results for ctx:

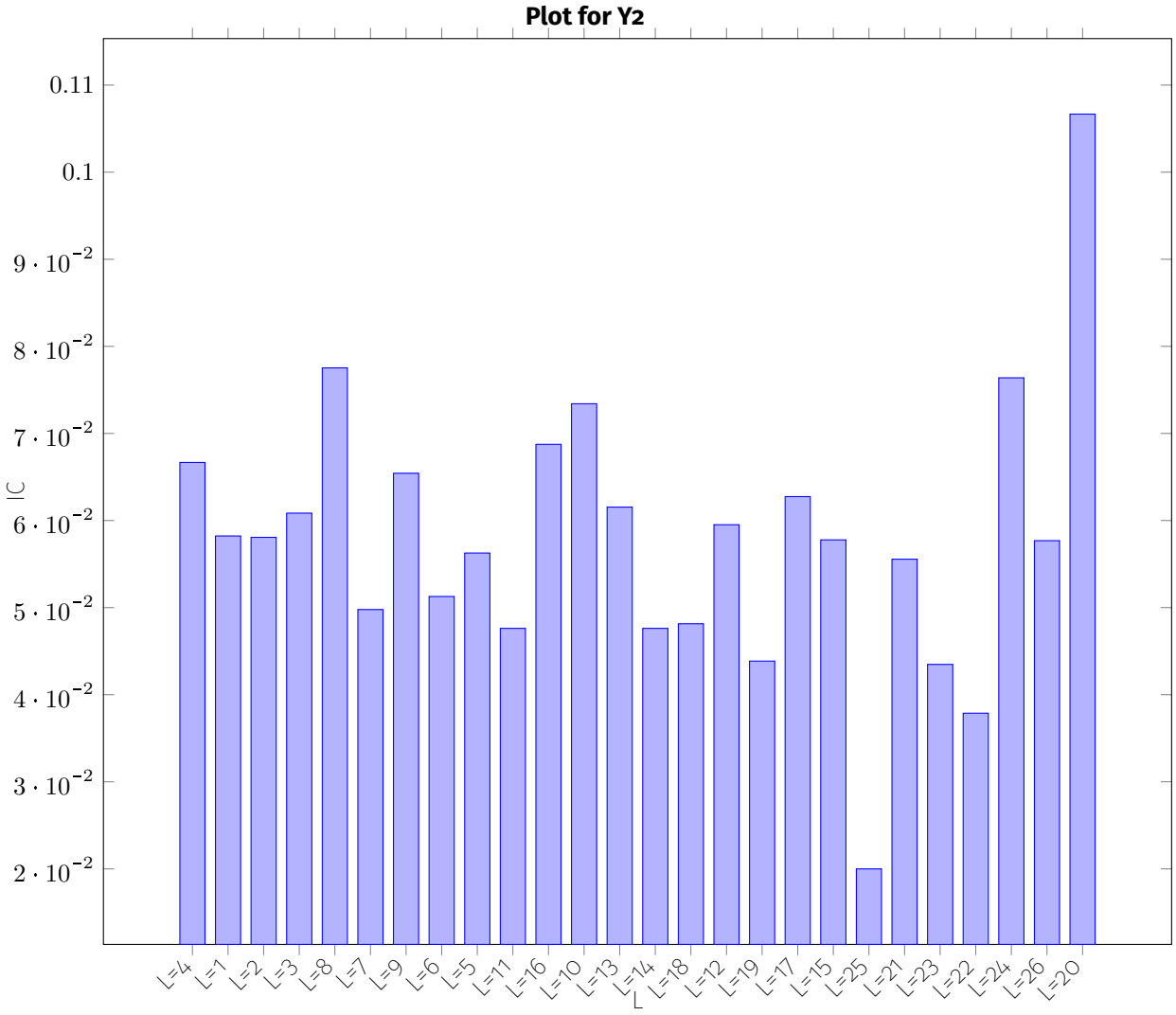


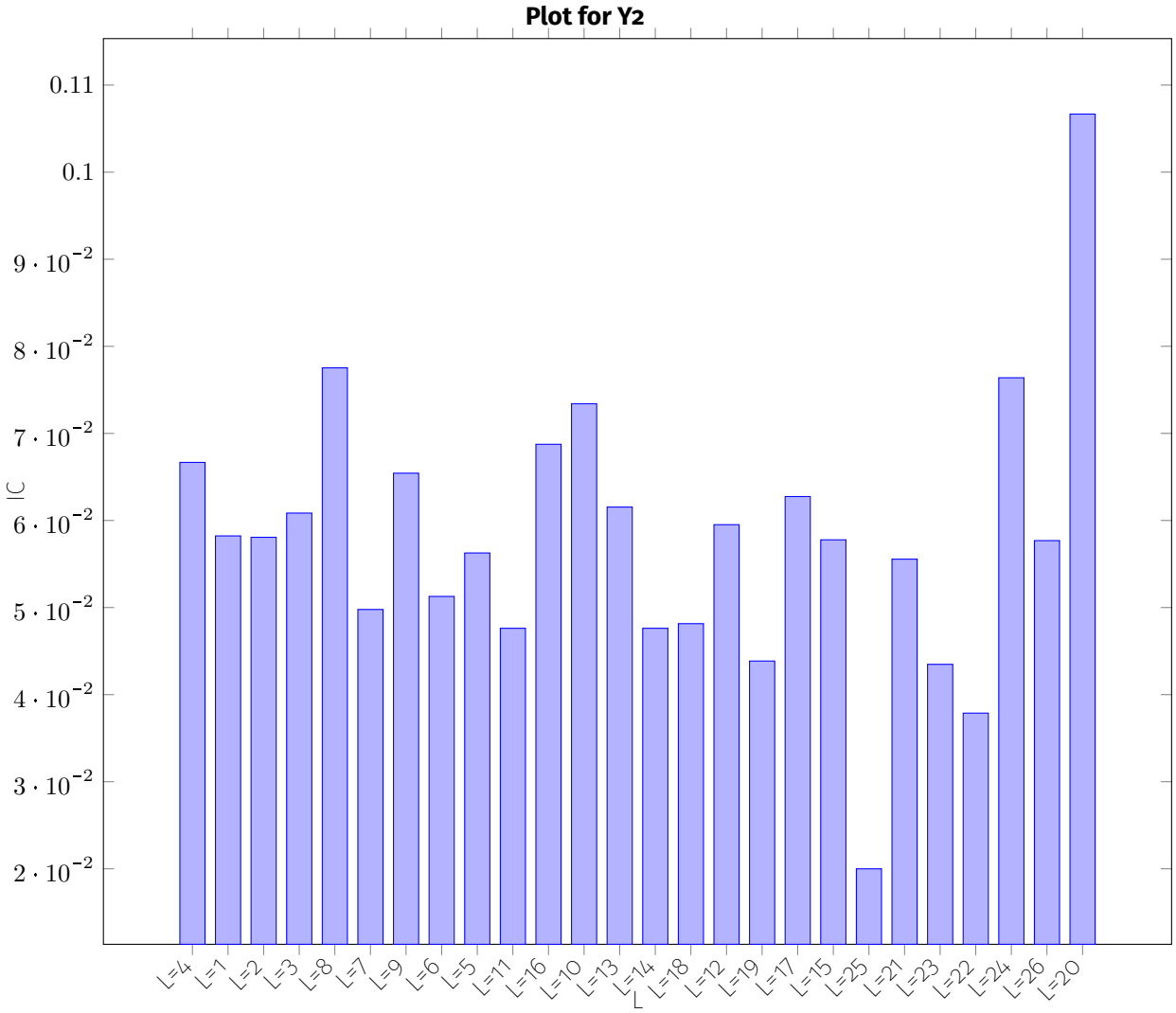
1) So it could be either lenght of 5, 10, 20, 15 or 25. Length of 5 being the highest.
2) If I exclude the very low end of number 0.03 and the high end 0.05 we can see that the majority of letters on average the IC for each letter is slightly above 0.04 and the average is $ctxIC = 0.04213$, this leads me to believe that we can assume this Vigenere chipher is with keyword length of $l = 5$ with somewhat good confidence. (see index of confidence table)
Now after taking Every 5th letter from ctx we get: FZAQFXWPNKFIKEELMEEGDBTNPAXAHPPLVKITZTUEYBOGYOKXMNTZIMFMPGOAAXXIDDRKIKVDPQQHNRGDOSFMATGVTHMPKIXEAKRACLOEMAMHJXJUENGT
Now find the key by shifting cesar ciphers to reveal something meaningfu. We should see something in english. But that only means it would work if the key is infact an english word. So I might not be able to tell if I have found the key or not unless I try evey key on the chiphertext to reveal that ctx^l is something meaningful. Anyway, the answer is key lenght is 5 because statistically

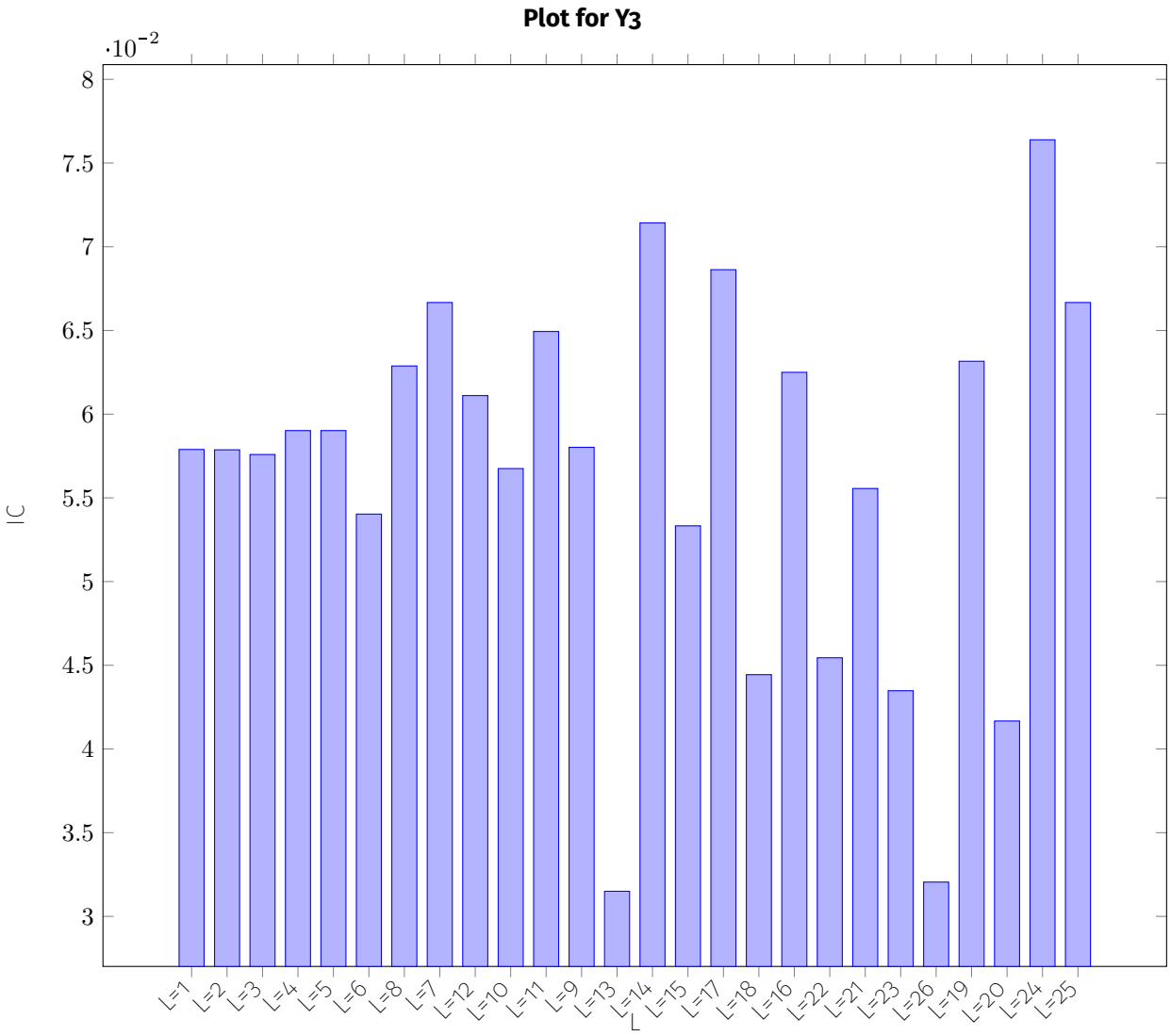
Bonus Task As for the bonus, I don't think encrypting it twice would add exta security. But the added strenght only comes from the fact that it's encrypted twice. If the offender expects chiphger text to be encrypted once but it's twice then there would be added benefit. But encrypting a message using a Vigenere cipher with a key of length 3 and then encrypting the result with a different key of length 5 would be the same as encrypting it once with 15-character key lenght. Because for the keys 3,5 LCM is 15 and similar key can be formed. But eventually, when analyzed any keylenght will leak some statistical probability. So it's not a bad idea to encrypt it twice but the benefit is marginal. Safest way would to be increase the key size in general in proportion to the text size to avoid leakage.


Plot for Yo
0.11
0.1
$9 \cdot 10^{-2}$
$8 \cdot 10^{-2}$
$7 \cdot 10^{-2}$
$6 \cdot 10^{-2}$
$5 \cdot 10^{-2}$
$4 \cdot 10^{-2}$
$3 \cdot 10^{-2}$
$2 \cdot 10^{-2}$
$\frac{1}{C}$
L=2
L=1
L=3
L=6
L=4
L=5
L=11
L=7
L=12
L=15
L=9
L=10
L=13
L=8
L=17
L=14
L=19
L=20
L=18
L=21
L=25
L=22
L=16
L=23
L=24
L=26
L


Plot for Y1
0.1
$9 \cdot 10^{-2}$
$8 \cdot 10^{-2}$
$7 \cdot 10^{-2}$
$6 \cdot 10^{-2}$
$5 \cdot 10^{-2}$
$4 \cdot 10^{-2}$
$3 \cdot 10^{-2}$
$2 \cdot 10^{-2}$
$1 \cdot 10^{-2}$
0
$\mathcal{L}$
L=1
L=2
L=3
L=5
L=7
L=14
L=4
L=8
L=6
L=9
L=12
L=11
L=10
L=13
L=16
L=15
L=18
L=20
L=17
L=24
L=21
L=19
L=22
L=25
L=26
L=23


Plot for Y2
IC
0.11
0.1
$9 \cdot 10^{-2}$
$8 \cdot 10^{-2}$
$7 \cdot 10^{-2}$
$6 \cdot 10^{-2}$
$5 \cdot 10^{-2}$
$4 \cdot 10^{-2}$
$3 \cdot 10^{-2}$
$2 \cdot 10^{-2}$
L=4
L=1
L=2
L=3
L=8
L=7
L=9
L=6
L=5
L=11
L=16
L=10
L=13
L=14
L=18
L=12
L=19
L=17
L=15
L=25
L=21
L=23
L=22
L=24
L=26
L=20
L


Plot for Y2
IC
0.11
0.1
$9 \cdot 10^{-2}$
$8 \cdot 10^{-2}$
$7 \cdot 10^{-2}$
$6 \cdot 10^{-2}$
$5 \cdot 10^{-2}$
$4 \cdot 10^{-2}$
$3 \cdot 10^{-2}$
$2 \cdot 10^{-2}$
L=4
L=1
L=2
L=3
L=8
L=7
L=9
L=6
L=5
L=11
L=16
L=10
L=13
L=14
L=18
L=12
L=19
L=17
L=15
L=25
L=21
L=23
L=22
L=24
L=26
L=20
L


Plot for Y3
$\cdot 10^{-2}$
8
7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
IC
L=1
L=2
L=3
L=4
L=5
L=6
L=8
L=7
L=12
L=10
L=11
L=9
L=13
L=14
L=15
L=17
L=18
L=16
L=22
L=21
L=23
L=26
L=19
L=20
L=24
L=25
L

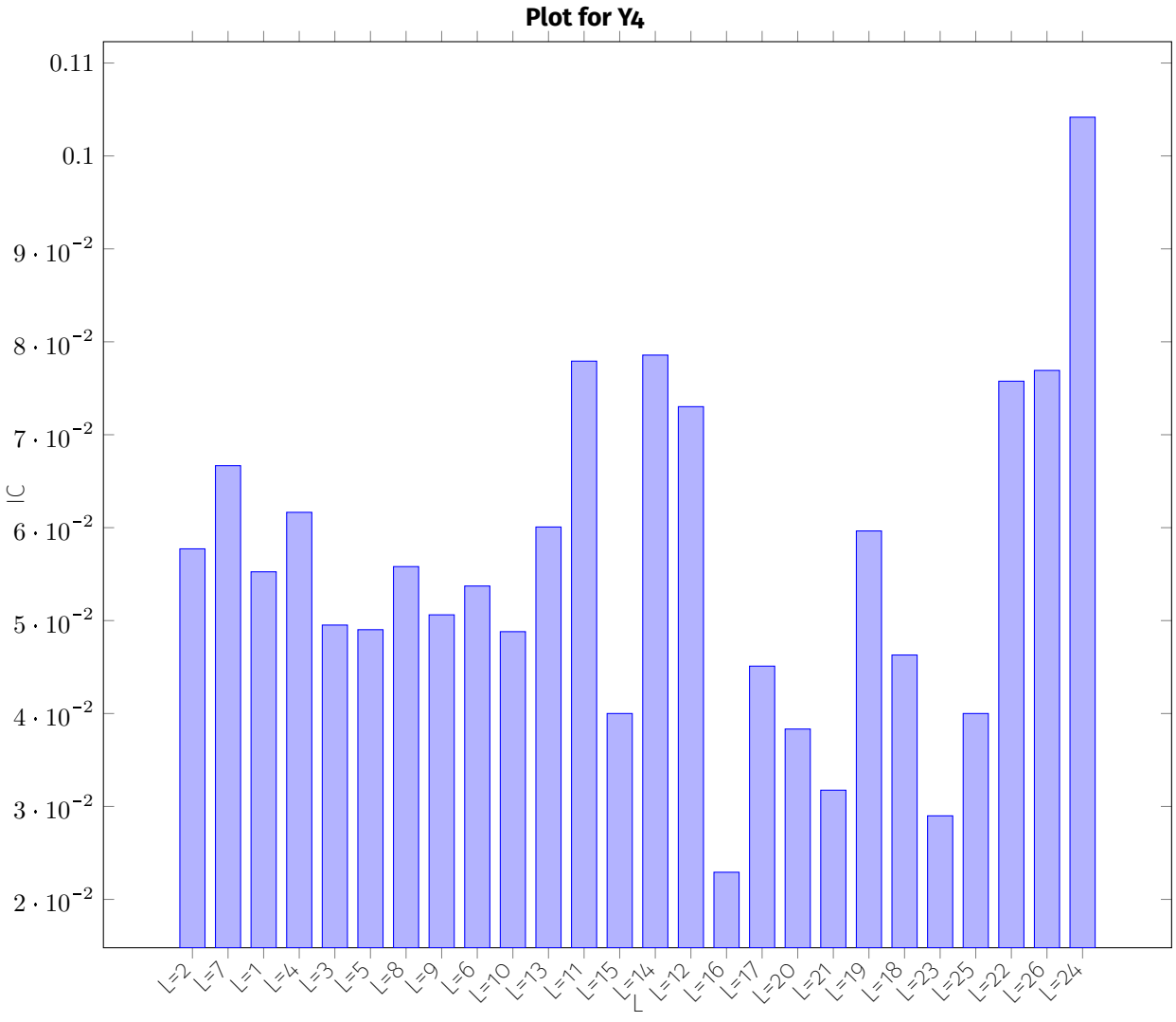

Plot for Y4
0.11
0.1
$9 \cdot 10^{-2}$
$8 \cdot 10^{-2}$
$7 \cdot 10^{-2}$
$6 \cdot 10^{-2}$
$5 \cdot 10^{-2}$
$4 \cdot 10^{-2}$
$3 \cdot 10^{-2}$
$2 \cdot 10^{-2}$
IC
L=2
L=7
L=1
L=4
L=3
L=5
L=8
L=9
L=6
L=10
L=13
L=11
L=15
L=14
L=12
L=16
L=17
L=20
L=21
L=19
L=18
L=23
L=25
L=22
L=26
L=24
L

Table 1: Index of Coincidence for Different Key Lengths

N	IC
2	0.056706
3	0.050608
4	0.047559
5	0.04573
6	0.04451
7	0.043639
8	0.042986
9	0.042478
10	0.042133
11	0.041739
12	0.041461
13	0.041227
14	0.041026
15	0.040852
16	0.040699
17	0.040565
18	0.040445
19	0.040338
20	0.040242
21	0.040155
22	0.040076
23	0.040003
24	0.039937
25	0.039876