Homework #1

Due date: **27/10/2023**
Notes:

- Your work (code + written answers) must be submitted through SUcourse+.
- Winzip your programs and add a readme.txt document (**if necessary**) to explain the programs and how to use them.
- Name your winzip file as **"cs411_507_hw01_yourname.zip"**

---

1. (**15 pts**) Consider the shift cipher. Show that the ciphertext "NLPDLC" can be decrypted into a meaningful word. Find out this word and the corresponding encryption key.
Word: CAESAR
Key: 11
Method: I used the Brute-Force shifting algorithm for this question. Then I tried adding each value in range(26) to each of the chars of the string "NLPDLC" and analyzed the possible answers. Please check the code for the application of the Brute-Force algorithm.

---

2. (**20 pts**) Consider the ciphertext generated by Affine Cipher over Z_{26}. As a hint, you are told that the most frequent letter in the plaintext is 'T'. Find the plaintext, the encryption and decryption keys. Show your work.

"J gjg mxa czjq ayr arpa. J ulpa cxlmg ayerr ylmgerg rqrwrm hzdp ax gx ja hexmn."

Plaintext: I did not fail the test. I just found three hundred eleven ways to do it wrong.
Encryption and Decryption Keys: Encryption key = (11, 25), Decryption key = (19, 19).
Method: First, I found the most frequent letter in the ciphertext. It was the letters "a" and "r" with each of their number of occurrences being 8. So, I had to analyze these 2 cases where a*value(t)+b = value(a) or a*value(t)+b = value(r). Then, I found the possible a values for the formula "ax+b=y" by finding the numbers that are coprime to 26. After that, for each of the a values, I found the corresponding b value by using the formula where x = value(t) = 19. After finding every possible a and b duos, I found the corresponding gamma and theta values. Then, I obtained each of the possible candidate plaintexts and one of them was actually meaningful. Therefore, there were no need to check the Case R since Case A gave the result. Please check the code for a more detailed step-by-step methodology.

---

3. **(15 pts)** Assume that you design a new affine cipher where you encrypt two letters at a time, where your alphabet is

{'A':0, 'B':1, 'C':2, 'D':3, 'E':4, 'F':5, 'G':6, 'H':7, 'I':8, 'J':9, 'K':10, 'L':11, 'M':12, 'N':13, 'O':14, 'P':15, 'Q':16, 'R':17, 'S':18, 'T':19, 'U':20, 'V':21, 'W':22, 'X':23, 'Y':24, 'Z':25, '.':26, ' ':27, ',':28, '!':29}.

In other words, you group your plaintext message in bigrams (i.e., two-character words) and encrypt each bigram of the plaintext separately using this affine cipher. For example, if the first two letters of a plaintext is "TH" then it will be encoded as follows

TH ⇒ 19 × 30 + 7 = 577.

If the number of letters in the plaintext is not a multiple of two, you pad it with the letter "X" at the end. Determine the modulus and the size of the key space.

The Modulus: Since the number of possible bigrams is 30*30 = 900, the modulus should be 900.
Key Space Size: The size of the key space can be found with the formula:
phi(900) * 900 = 216000 (which is calculated using the code).

---

4. (**15 pts**) Is the affine cipher defined in question (3) secure against the letter frequency analysis?
Answer: Although it is more complex and somewhat better than the classical version of the affine cipher, it is still possible to make a frequency analysis on this approach. For example, according to the research of Oxford College of Emory University, the most common bigrams in English are: "th", "he", "in", "en", "nt" and so forth. Therefore, using this information and a slightly modified version of the affine cipher algorithm, it is possible to crack this type of ciphering. For more detailed information: https://mathcenter.oxford.emory.edu/site/math125/englishLetterFreqs/

---

5. **(20 pts**) Consider the following ciphertext that is encrypted with the affine cipher defined in question (3):

"ZHOFC.BNZCLRZ WNJ.XGI.WMBDV.MEJ!GGYKGDZ ERGMWNJ.KDGD RSW"

Find the key and decrypt the ciphertext.

(**Hint 1:** The plaintext is a sentence that ends with a dot.)
(**Hint 2:** The length of the plaintext (plen) is not a multiple of 2; plen = 2k+1 for an integer k.)

Answer: Since the plain text ends with a dot and the length of it is an odd number, the only thing that I can conclude is that the ".X" bigram corresponds to "SW".

6. (15 pts) If we select a different shift amount for every letter in the plaintext uniformly randomly, the shift cipher becomes a one-time-pad with perfect security. Suppose p_α is the probability of the plaintext letter α from the Turkish alphabet, where α ∈ {A, B, C, Ç, …, Z}. Suppose also that p_β is the probability of the ciphertext letter β, where β ∈ {A, B, C, Ç, …, Z}. Demonstrate that $p_\beta = 1/29$ for every β ∈ {A, B, C, Ç, …, Z} independent of the values of p_α.

Answer: This one-time-pad shift cipher encryption can be shown as:

$$(x + s) \bmod 29 = y$$

where x represents a plaintext letter and x ∈ {0, 1, 2, … , 27, 28},
s represents the shifting amount which is uniformly randomly selected,
y represents a ciphertext letter.

Since the shifting amount is uniformly randomly selected, the probability of getting any p_β value is 1/29 independent of the values of p_α:

$p_\beta = (1/29*p_A) + (1/29*p_B) + (1/29*p_C) \ldots + (1/29*p_Z) = 1/29*(p_A + p_B + p_C + \ldots + p_Z) = 1/29$

---

7. BONUS (20 pts) The following was encrypted using the Vigènere cipher:

"FNZ FFZZMLQQZVO GAXXH PZ UPU QXGIHU UY NWJXR AHBDLPOMK YOUPZM, VOZAYCD. J TGQH B XUIJJZM ARS XOAH, BZJ D JP AT GLWUTB LO EVDWF AL GRHUI. OKPGMC L NME IRU NKGLFHK DQ UTK JUEQX JI UTK PQJHKMVF, KKO L MABZ WIQ YOLDWE GLUFRZ OFMBZV BE ZCHZ AVZQ JZ YKUJZM. D OPHK OKF NRPH TWE, D OPHK NRNQ VZRQXK, RKPY UIH MABZV ZAA FQPI YJPFFOHHT IOOKPGZ FQPIOIJ XTE. D OPHK NRNQ MMHBF JZHEE JJQF NE HHO, FNJXHT O'QH MATB FFMYZG QQXCDQE ZJ KBHK ADJFN DQ UTKH, BFF LMRN ARY KBNOO ROQ'Y CHBDZ KUJLKN WIQS. CHSQ ZCHZ TGQH CDUPJIF ZCH TAAK IPD EJX, FMZ DW, JF CDOM PU TRV SUJG. JF'Y ALSEZ-MDUQ YJXQ, FNZB LZUR KPI ZJ PBWK DW IQXZ. L XMTO WP FXVYFX OI HVDUKH, BXEJVIM, O NKBXR NHU ALA ISAS CHSQ. GIG ZQZ D NOAC OKBF O VP PZRT JPUTB WP M MMDWQEVUE, NAO LU'E G HRTF VMHDUUPV HDGQHZMXY, WIMZ'N ZIMZ DW JE. VMHDUUPV BDK OKF PKVG UTGO OJQ ZCHSQ, KQHSK YOROQ UQHS FNZP TBKVNT AL NXDT HPUOUTB OJRK DQ UTK KDTF, UA VVON KDTEOJQBFK ADJFN DQ UTKDU XAXF, WIQOM WSGZC, WIQOM VUDABJMQ GIG UTKDU TOOZQDQ, ZCDU U QIRX U YCDMX LVOM AT OKF SXJXOP GIG LUYN WIAYZ VUATZV BZJ RHFB UQHS FNZP; UTUPJI U'S XROHOIFFP OI PZ TKVUU FNVW JF'Y GROS HZHO ZUOKJZM WXU M MMDWQEVUE. MTY L TTGGO OAZ RHFB LMRN PKNSBUX, WXU EOHSMK HZFBGYZ L TTGGO CQ NVSQK OI PZ FKVUT, U YCDMX YOHFB ST VPGR DQ FYUOLPZ. O GRWQ ZCH TFOXNZ XKVYFE OI VQDOIJ, UTK WOVQ YFB - UTGO'V BXR DW JE. OO'V OAZ V PBFZZU PR OIWFXRZFU AX GRHUI, DW'T XUQLOS CDWI ATZ'V JZYDGF, IOOK PZK'N VUASVFI."

Attack it and find the key length and the key. Note that only the letter characters are encrypted.

Key Length: The key length should be 5. To find the key length, I analyzed the number of coincidences by shifting the ciphertext to right and counting the matchings between the original ciphertext and the ciphertexts that are shifted to right. After finding the number of coincidences for each right shift amount, I had the array: [42, 33, 33, 46, ***66***, 51, 42, 26, 42, ***81***, 31, 43, 39, 45, ***70***, 31, 37, 43, 35, ***62***, …]. Analyzing this array, I found out that the distance between each of the significantly big numbers is 5, i.e. the distance between 66 and 81 is 5, 70 and 81 is again 5 and 81 and 70 is 5 too. Therefore, I concluded that statistically the length of the key should be 5.

The Key: MGVDB (12, 6, 21, 3, 1). To find the key, I have created 5 sub-ciphertext dictionaries (since it is the most possible key length) in which the number of appearances of each letter inside of each sub-ciphertext is stored. I created this to analyze the most common words in each of the sub-ciphertext and I mapped the most common letter in each ciphertext to the letter "E" at first and checked whether this shifting amount holds for the top 3 most common letters of the each sub-ciphertext. For the first 4 letter of the key, it directly held. However, 5th letter of the key required a second check in which I mapped the most common letter to the letter "T" and then calculated the shifting amount. By creating these dictionaries and applying the frequency analysis (in which I checked whether such mapping is logical or not for each of the sub-ciphertext), I concluded that the key is "MGVDB".

The Paragraph:

THE CENTRIPETAL FORCE ON OUR PLANET IS STILL FEARFULLY STRONG ALYOSHA I HAVE A LONGING FOR LIFE AND I GO ON LIVING IN SPITE OF LOGIC THOUGH I MAY NOT BELIEVE IN THE ORDER OF THE UNIVERSE YET I LOVE THE STICKY LITTLE LEAVES AS THEY OPEN IN SPRING I LOVE THE BLUE SKY I LOVE SOME PEOPLE WHOM ONE LOVES YOU KNOW SOMETIMES WITHOUT KNOWING WHY I LOVE SOME GREAT DEEDS DONE BY MEN THOUGH IVE LONG CEASED PERHAPS TO HAVE FAITH IN THEM YET FROM OLD HABIT ONES HEART PRIZES THEM HERE THEY HAVE BROUGHT THE SOUP FOR YOU EAT IT IT WILL DO YOU GOOD ITS FIRST RATE SOUP THEY KNOW HOW TO MAKE IT HERE I WANT TO TRAVEL IN EUROPE ALYOSHA I SHALL SET OFF FROM HERE AND YET I KNOW THAT I AM ONLY GOING TO A GRAVEYARD BUT IT'S A MOST PRECIOUS GRAVEYARD THAT'S WHAT IT IS PRECIOUS ARE THE DEAD THAT LIE THERE EVERY STONE OVER THEM SPEAKS OF SUCH BURNING LIFE IN THE PAST OF SUCH PASSIONATE FAITH IN THEIR WORK THEIR TRUTH THEIR STRUGGLE AND THEIR SCIENCE THAT I KNOW I SHALL FALL ON THE GROUND AND KISS THOSE STONES AND WEEP OVER THEM THOUGH IM CONVINCED IN MY HEART THAT ITS LONG BEEN NOTHING BUT A GRAVEYARD AND I SHALL NOT WEEP FROM DESPAIR BUT SIMPLY BECAUSE I SHALL BE HAPPY IN MY TEARS I SHALL STEEP MY SOUL IN EMOTION I LOVE THE STICKY LEAVES IN SPRING THE BLUE SKY THAT'S ALL IT IS ITS NOT A MATTER OF INTELLECT OR LOGIC ITS LOVING WITH ONES INSIDE WITH ONES STOMACH