TCSS 581 Cryptology

Homework 1

Tianyi Li
Student ID: 1827924

Tianyi Li
Student #:1827924

1.1 Decrypt the ciphertext provided at the end of the section on mono-alphabetic substitution ciphers. (Hint: the first word encrypted in the given cipher text is "cryptographic".)

JGRMQOYGHMVBJWRWQFPWHGFFDQGFPFZRKBEEBJIZQQOCIBZKLFAFGQVFZFWWE
OGWOPFGFHWOLPHLRLOLFDMFGQWBLWBWQOLKFWBYLBLYLFSFLJGRMQBOLWJVFPF
WQVHQWFFPQOQVFPQOCFPOGFWFJIGFQVHLHLROQVFGWJVFPFOLFHGQVQVFILE
OGQILHQFQGIQVVOSFAFGBWQVHQWIJVWJVFPFWHGFIWIHZZRQGBABHZQOCGFHX

- In this question, all underlined lowercase letters indicated plaintext and uppercase letters are ciphertext.

- Since the instructor has given the hint that the first word is 'cryptographic'. We can rewrite the whole cipher.

cryptographicWyWtFPWarFFDtrFPFZyKiEEicIZttoCIiZKLFAFrthFZFWWE
orWoPFrFaWoLPaLyLoLFDpFrtWiLWiWtoLKFWigLiLgLFSFLcryptioLWchFPF
WthatWFFPtothFPtoCFPorFWFcIrFthaLaLyothFrWchFPFoLFarththFILE
ortILatFtrIthhoSFAFriWthatWIchWchFPFWarFIWIaZZytriAiaZtoCrFaX

- Following the steps indicated from the textbook, we can start by listing the frequency distribution of characters in the ciphertext, and get ride off the ones has been figured out.

Letters	Frequencies	Letters*	Frequencies*
F	37	e	12.7
W	21	n	6.7
L	17	s	6.3
P	10	d	4.3
I	9	l	4.0
Z	7	u	2.8
E	4	w	2.4
K	3	m	2.4
C	3	f	2.2
A	3	b	1.5
S	2	v	1.0
D	2	k	0.8
X	1	x	0.2
N	0	j	0.2
T	0	z	0.1
U	0	q	0.1

*The frequency distribution of individual letters in the English language is known (Textbook, figure 1.3, page 11).

- We can now starting mapping letters corresponding to the frequencies.

  Swap 'F' to 'e':

cryptographicWyWtePWareeDtrePeZyKiEEicIZttoCIiZKLeAertheZeWWE
orWoPereaWoLPaLyLoLeDpertWiLWiWtoLKeWigLiLgLeSeLcryptioLWchePe
WthatWeePtothePtoCePoreWecIrethaLaLyotherWchePeoLearththeILE
ortILatetrIthhoSeAeriWthatWIchWchePeWareIWIaZZytriAiaZtoCreaX

We see 'cryptioL' in the last bit on the second line, which is probably 'cryption':

cryptographicWyWtePWareeDtrePeZyKiEEicIZttoCIiZKneAertheZeWWE
orWoPereaWonPanynoneDpertWinWiWtonKeWigningneSencryptionWchePe
WthatWeePtothePtoCePoreWecIrethananyotherWchePeonearththeInE
ortInatetrIthhoSeAeriWthatWIchWchePeWareIWIaZZytriAiaZtoCreaX

This helps a lot. Then we can see 'KeWigning', which can be 'designing':

cryptographic systePsareeDtrePeZydiEEicIZttoCIiZdneAertheZessE
orsoPereasonPanynoneDpertsinsistondesigningneSencryptionschePe
sthatseePtothePtoCePoresecIrethananyotherschePeonearththeInE
ortInatetrIthhoSeAeristhatsIchschePesareIsIaZZytriAiaZtoCreaX

We see 'systeP' as the word right after 'cryptographic', which could be 'system':

cryptographic systemsareeDtremeZydiEEicIZttoCIiZdneAertheZessE
orsomereasonmanynoneDpertsinsistondesigningneSencryptionscheme
sthatseemtothemtoCemoresecIrethananyotherschemeonearththeInE
ortInatetrIthhoSeAeristhatsIchschemesareIsIaZZytriAiaZtoCreaX

There is 'eDtreme' on the first half of the first line and it could be ''extreme''; and 'neAertheZess' can be 'nevertheless':

cryptographic systemsareextremelydiEEicIlttoCIildneverthelessE
orsomereasonmanynonexpertsinsistondesigningneSencryptionscheme
sthatseemtothemtoCemoresecIrethananyotherschemeonearththeInE
ortInatetrIthhoSeveristhatsIchschemesareIsIallytrivialtoCreaX

The 'diEEicIlt' on the second half of the first line can be 'difficult':

cryptographic systemsareextremelydifficulttoCuildneverthelessf
orsomereasonmanynonexpertsinsistondesigningneSencryptionscheme
sthatseemtothemtoCemoresecurethananyotherschemeonearththeInf
ortunatetruthhoSeveristhatsuchschemesareusuallytrivialtoCreaX

We are left with:

Letters	Frequencies	Letters*	Frequencies*
C	3	w	2.4
S	2	b	1.5
X	1	k	0.8

- Therefore, the answer is: "cryptographic systems are extremely difficult to build nevertheless for some reason many non experts insist on designing new encryption schemes that seem to them to be more secure than any other scheme on earth the unfortunate truth however is that such schemes are usually trivial to break"

1.2 Provide a formal definition of the Gen, Enc, and Dec algorithms for both the mono-alphabetic substitution and Vigenére ciphers.

- The basic distinction to the 3 algorithms are below:
  - Gen – key generation
  - Enc – input encryption key, plaintext; output ciphertext
  - Dec – input decryption key, ciphertext; output plaintext

- In mono-alphabetic substitution cipher:
  - Let m is the plaintext, c is the ciphertext.
  - Plaintext space = ciphertext space = $\{0, 1, \ldots, 25\}^{|m|}$.
  - Key space = 1-to-1 mapping of {0, 1, ….., 25} = 26!
  - Let π be the permutations of {0, 1, ….., 25}.
  - Gen():

    for i = 25 to 1:

    pick j uniformly at random from {0, 1, ….., i}

    swap π[i] and π[j]

    return π
  - Enc($k = \pi$, $m = m_1 m_2 \ldots\ldots m_l$):

    return $c = \pi(m_1)\, \pi(m_2) \ldots\ldots \pi(m_l)$
  - Dec($k = \pi$, $c = c_1, c_2, \ldots\ldots, c_l$):

    for $i = 1$ to l:

    for $j = 0$ to 25:

    if $\pi[j] == c_i$:

    $m_i = \pi[j]$

    return $m = m_1 m_2 \ldots\ldots m_l$
  - Therefore, Gen() = π , *a permutation of* {0, 1, ….., 25}.

$$\text{Enc}(k = \pi, m = m_1 m_2 \ldots\ldots m_l) = \pi(m_1)\, \pi(m_2) \ldots\ldots \pi(m_l)$$
$$= c_1, c_2, \ldots\ldots, c_l = c$$
$$\text{Dec}(k = \pi, c = c_1, c_2, \ldots\ldots, c_l) = \pi^{-1}(c_0)\, \pi^{-1}(c_1) \ldots\ldots \pi^{-1}(c_l)$$
$$= m_1 m_2 \ldots\ldots m_l = m$$

- In Vigenére cipher:
  - Let m is the plaintext, c is the ciphertext,
  - And $c = c_1, c_2, \ldots\ldots, c_l$ can be divided into t parts where each part can be viewed as having been encrypted using a shift cipher.
  - Plaintext space = $\{0, 1, \ldots\ldots, 25\}^{|m|}$, m = length of the message.
  - Let key $k = k_1\, k_2 \ldots\ldots\, k_t$, where each k_i is a letter of the alphabet.
  - Key space = $\{0, 1, \ldots\ldots, 25\}^{|k|=t}$
  - Gen():

    return uniform key $k = k_1\, k_2 \ldots\ldots\, k_t$, where $k_i \in \{0, 1, \ldots\ldots, 25\}$
  - Enc($k = k_1\, k_2 \ldots\ldots\, k_t$, $m = m_1\, m_2 \ldots\ldots\, m_l$):

    $j = 1$

    for i = 1 to I:

    $c_i = m_i + k_j$, #shifts m_i forward by k_j positions

    $j \mathrel{+}= 1$

    $j = j \bmod t$

    return $(m_1 + k_1, \ldots\ldots, m_t + k_t, m_{t+1} + k_1, \ldots\ldots, m_{2t} + k_t, \ldots\ldots) \bmod 26$
  - Dec($k = k_1\, k_2 \ldots\ldots\, k_t$, $c = c_1, c_2, \ldots\ldots, c_l$):

    $j = 1$

    for $i = 1$ to l:

    $m_i = c_i - k_j$, #shifts c_i backward by k_j positions

    $j \mathrel{+}= 1$

    $j = j \bmod t$

    return $(c_1 - k_1, \ldots\ldots, c_t - k_t, c_{t+1} - k_1, \ldots\ldots, c_{2t} - k_t, \ldots\ldots) \bmod 26$
  - Therefore, Gen() = uniform key $k = k_1\, k_2 \ldots\ldots\, k_t$, where $k_i \in \{0, 1, \ldots\ldots, 25\}$

    Enc($k = k_1\, k_2 \ldots\ldots\, k_t$, $m = m_1\, m_2 \ldots\ldots\, m_l$)

    $= (m_1 + k_1, \ldots\ldots, m_t + k_t, m_{t+1} + k_1, \ldots\ldots, m_{2t} + k_t, \ldots\ldots) \bmod 26$

    $= c_1, c_2, \ldots\ldots, c_l = c$

    Dec_π ($k = k_1\, k_2 \ldots\ldots\, k_t$, $c = c_1, c_2, \ldots\ldots, c_l$)

    $= \pi^{-1}(c_0)\, \pi^{-1}(c_1) \ldots\ldots\, \pi^{-1}(c_l)$

    $= (c_1 - k_1, \ldots\ldots, c_t - k_t, c_{t+1} - k_1, \ldots\ldots, c_{2t} - k_t, \ldots\ldots) \bmod 26$

    $= m_1\, m_2 \ldots\ldots\, m_l = m$

1.3 Consider an improved version of the Vigenére cipher, where instead of using multiple shift ciphers, multiple mono-alphabetic substitution ciphers are used. That is, the key consists of *t* random permutations of the alphabet, and the plaintext characters in positions *i, t + i, 2t + i*, and so on are encrypted using the i^{th} permutation. Show how to break this version of the cipher

- From to the textbook, the authors introduced the Kasiski's method for determining *t* works for this cipher.
    - Identify repeated patterns of length 2 or 3 in the ciphertext, to find out the certain bigrams or trigrams that appear frequently in the plaintext (e.g. 'th' and 'the').
    - Then use those repeated fragments in the ciphertext and compile a list of the distances that separate the repetitions.
    - Thus, the overall the distances between these occurrences are then used to be multiples of the keyword length of the keyword.
- The Kasiski's method is still good for the scenario proposed in this question.
- However, there should be a frequency table listed for each of the *t* keys, and carry out an attack like on the mono-alphabetic cipher, before the second time of the attack. The rest would be the same until the t^{th} time of attack.
- Hence, with sufficient long plaintext, this method would work.

1.4 In an attempt to prevent Kasiski's attack on the Vigenére cipher, the following modification has been proposed. Given the period *t* of the cipher, the plaintext is broken up into blocks of size *t*. Within each block, the Vigenére cipher works by encrypting the i^{th} character with the i^{th} key (using a basic cipher). Letting the key be $k_1, ..., k_t$, this means the i^{th} character in each block is encrypted by adding k_i to it, modulo 26. The proposed modification is to encrypt the i^{th} character in the j^{th} block by adding k_{i+j} modulo 26.

(a) Show that decryption can be carried out.
(b) Describe the effect of the above modification on Kasiski's attack.
(c) Devise an alternate way to determine the period for this scheme.

a) In this scenario, Kasiski's method would still work for determining *t* works for this cipher.

- Kasiski's method:
  - Identify repeated patterns of length 2 or 3 in the ciphertext, to find out the certain bigrams or trigrams that appear frequently in the plaintext (e.g. 'th' and 'the').
  - Then use those repeated fragments in the ciphertext and compile a list of the distances that separate the repetitions.
  - Thus, the overall the of length 2 or 3 in the ciphertext distances between these occurrences are then used to be multiples of the keyword length of the keyword.
- The repeated patterns in the ciphertext does not change in this case, therefore the attack would be carried out as the same to the t^{th} time of attack.

b) The only affect is that the key has letter scrambled, such that 'CAKE' is now 'EKCA'. Thus, a rearrangement of the letters to the key is needed to obtain the original key. This means the key is half-encrypted as well (e.g. letters are scrambled, but not swapped).

c) I am not sure what this question is asking.

1.5 Show that the shift, substitution, and Vigenére ciphers are all trivial to break using a known-plaintext attack. How much known plaintext is needed to completely recover the key for each of the ciphers?

- In shift cipher:
  - Assume we are work with English, therefore there are 26 letters in the alphabets.
  - Given a single plaintext character m_i and ciphertext character c_i,
  - From textbook, the key is simply $k = (c - m)$ mod 26,
  - Say if we know one letter from the plaintext, then we can find the key to that letter by $k = (c_i - m_i)$ mod 26,
  - Since k is consist throughout the whole message m, then k is known for every $m_i \in m$.
  - Therefore, one single plaintext character would suffice to recover the key of the cipher.

- In mono-alphabetic cipher:
  - Assume we are work with English, therefore there are 26 letters in the alphabets.
  - Given a plaintext character m_i, and the corresponding ciphertext character c_i,
  - Let m is the plaintext, c is the ciphertext.
  - Plaintext space = ciphertext space = $\{0, 1, ..., 25\}^{|m|}$.
  - Key space = 1-to-1 mapping of {0, 1, ….., 25} = 26!
  - Let π be the permutations of {0, 1, ….., 25}.
  - Then $\pi(m_i) = c_i$
  - Because each ciphered letter is corresponding to a plaintext character without any patterns, unlike the shift ciphers
  - Then, in order to fully decipher and find the key (26 different letters), we need to obtain 25 different plaintext letters of the alphabet.
  - Since all letters are 1-to-1 mapping to the ciphered letters, knowing the 25 input plaintexts would yield a definite last input.
  - Then a full π permutation (26 letters) can be deduced.
  - Therefore, to fully determine the key, a plaintext containing 25 distinct letters of the alphabet should be given to recover the full key of the cipher.
  - However, in a normal English setting, we need to wait and obtain for way more than 25 letters in order to have 25 distinct letters to occur.

- In Vigenére cipher:
  - Assume we are work with English, therefore there are 26 letters in the alphabets.

- Given a single plaintext character m_i and ciphertext character c_i,
- And c = c_1, c_2, ……,c_l can be divided into *t* parts where each part can be viewed as having been encrypted using a shift cipher.
- Plaintext space = {0, 1, ….., 25}$^{|m|}$, m = length of the message.
- Let key $k = k_1 k_2 \ldots\ldots k_t$, where each k_i is a letter of the alphabet.
- Key space = {0, 1, ….., 25}$^{|k|=t}$
- Similar method would be used as the shift cipher
- From textbook, the parts of the key is simply $k_i = (c_i - m_i) \bmod 26$,
- Say if we know one letter from the plaintext, then we can find the key to that letter by $k = (c_i - m_i) \bmod 26$,
- If this process is repeated for *t* times, with *t* consecutive letters, then the whole key $k = k_1 k_2 \ldots\ldots k_t$ will be obtained.
- Then *k* will be known for every $m_i \in m$.
- Therefore, *the* encryption of *t* consecutive characters of plaintext is needed to recover the key of the cipher.

1.6 Show that the shift, substitution, and Vigenére ciphers are all trivial to break using a chosen-plaintext attack. How much plaintext must be encrypted in order for the adversary to completely recover the key? Compare to the previous question.

- The solutions are the same for the shift and Vigenére ciphers as the previous question (1.5).

- In shift cipher (answer from Q1.5):
  - Assume we are work with English, therefore there are 26 letters in the alphabets.
  - Given a single plaintext character m_i and ciphertext character c_i,
  - From textbook, the key is simply $k = (c - m) \bmod 26$,
  - Say if we know one letter from the plaintext, then we can find the key to that letter by $k = (c_i - m_i) \bmod 26$,
  - Since k is consist throughout the whole message m, then k is known for every $m_i \in m$.
  - Therefore, one single plaintext character would suffice to recover the key of the cipher.

- In Vigenére cipher (answer from Q1.5):
  - Assume we are work with English, therefore there are 26 letters in the alphabets.
  - Given a single plaintext character m_i and ciphertext character c_i,
  - And $c = c_1, c_2, \ldots\ldots, c_l$ can be divided into t parts where each part can be viewed as having been encrypted using a shift cipher.
  - Plaintext space = $\{0, 1, \ldots\ldots, 25\}^{|m|}$, m = length of the message.
  - Let key $k = k_1 k_2 \ldots\ldots k_t$, where each k_i is a letter of the alphabet.
  - Key space = $\{0, 1, \ldots\ldots, 25\}^{|k|=t}$
  - Similar method would be used as the shift cipher
  - From textbook, the parts of the key is simply $k_i = (c_i - m_i) \bmod 26$,
  - Say if we know one letter from the plaintext, then we can find the key to that letter by $k = (c_i - m_i) \bmod 26$,
  - If this process is repeated for t times, with t consecutive letters, then the whole key $k = k_1 k_2 \ldots\ldots k_t$ will be obtained.
  - Then k will be known for every $m_i \in m$.
  - Therefore, *the* encryption of t consecutive characters of plaintext is needed to recover the key of the cipher.

- In mono-alphabetic cipher, using chosen-plaintext attack for an encryption, we can select a set of plaintext that contains 25 distinct letters of the alphabet and follow the rest of steps we used in previous exercise (1.5) to recover the full set of key.

- In mono-alphabetic cipher:
  - Assume we are work with English, therefore there are 26 letters in the alphabets.
  - Given a plaintext character m_i, and the corresponding ciphertext character c_i,
  - Let m is the plaintext, c is the ciphertext.
  - Plaintext space = ciphertext space = $\{0, 1, ..., 25\}^{|m|}$.
  - Key space = 1-to-1 mapping of {0, 1, ….., 25} = 26!
  - Let π be the permutations of {0, 1, ….., 25}.
  - Then $\pi(m_i) = c_i$
  - Because each ciphered letter is corresponding to a plaintext character without any patterns, unlike the shift ciphers
  - Then, in order to fully decipher and find the key (26 different letters), we need to obtain 25 different plaintext letters of the alphabet.
  - Since all letters are 1-to-1 mapping to the ciphered letters, knowing the 25 input plaintexts would yield a definite last input.
  - Then a full π permutation (26 letters) can be deduced.
  - Therefore, to fully determine the key, a plaintext containing 25 distinct letters of the alphabet should be given to recover the full key of the cipher.
- Thus, comparing to the previous exercise, we can select a set of plaintext that contains 25 distinct letters, instead of waiting for way more than 25 letters in order to have 25 distinct letters to occur. Hence, much less of plaintext will be required.