CIS 3362 Homework #2: Substitution Cipher, Vigenere

Part A: Code Break Questions

1) Decode the following message, which was encrypted using the substitution cipher. Make sure to discuss all the steps you took, the key you arrived at, and the decoded message.

```
TEDNDNWAGCODCNTLEPFFMQKDQKLAXMTAXMLCWVTDQQACHPFTDHMNDBAGFXEDXMPV
CDYMPQXEPSMNTGXMQTNFAAJOACDTAQLPHVGNBDTEAQMAOTEMNMHMNNPKMNXDNLFA
NDQKBEMCMTEMVCDYMBPNHPWZMBEPTDLPQXADNTEMOAFFABDQKDOWAGPCMTEMODCN
TTACMPXTEDNHMNNPKMMHPDFHMPTXHPCDQAPTLNGLOMXGBDTETEMOAFFABDQKLAXM
BACXZFGMEPDCMXKGDTPCVFPWMCPQXDBDFFNQPDFHPDFWAGPKDOTLPCX
```

Solution

First, the letter frequencies can be found using the online Cryptool posted on the website.

The most frequent letters are:

```
M    10.3%
D    10%
P    8.4%
A    7.4%
T    7.4%
```

Generally, the most common letter in English is E, therefore M may map to E. Additionally, the Cryptotool produces several trigrams: WAG QKD DQK HMN PQX TEM KDO. It's reasonable to assume that a subset of these correspond to the common trigrams of THE, AND, and ING. With the guess that M -> E, TEM is a likely candidate for THE. AND and ING must have a matching central letter, of which the only valid pairing is DQK and PQX. A is the second-most frequent letter in English, therefore it's reasonable to wager that D corresponds to A and thus, DQK -> AND and PQX -> ING. Plugging these in results in the following text:

```
THA-A------A--T-HI--ENDAND--GET-GE----TANN---I-TA-E-A----GHAGEI-
-A-EINGHI-E-T-GENT--------AT-N-I-----ATH-NE--THE-E-E--IDE-GA----
-AND-HE-ETHE--A-E-I--I--E-HITA-ING-A-THE------ANDA----I-ETHE-A--
TT--EIGTHA--E--IDEE-IA--EITG-I-AN-IT-----EG--ATHTHE------AND--
GE---G---EHIA-EGD-ATI---I-E-INGA-A---NIA--IA----IDA-T-I-G
```

which is noticeably strange for several reasons, such as the text ending in `I-G` rather than `ING`, and other odd strings. P is just below D in frequency; thus, it is possible that DQK actually maps to ING and PQX to AND. This results in:

```
THI-I------I--T-HA--ENGING--DET-DE----TINN---A-TI-E-I----DHIDEA-
-I-EANDHA-E-T-DENT--------IT-N-A-----ITH-NE--THE-E-E--AGE-DI----
-ING-HE-ETHE--I-E-A--A--E-HATI-AND-I-THE------INGI----A-ETHE-I--
TT--EADTHI--E--AGEE-AI--EATD-A-IN-AT-----ED--ITHTHE------ING--
DE---D---EHAI-EDG-ITA---A-E-ANDI-I---NAI--AI----AGI-T-A-D
```

which looks more reasonable. At this point, some more possibilities can be explored. Repeating characters are always valuable; a notable one is in the string -HA--ENGING. Taking common doubles into account and looking at the string will likely bring the word CHALLENGING to mind. Plugging this guess in leads to:

```
THI-I------I--TCHALLENGINGC-DET-DEC---TINN---ALTI-E-I---LDHIDEA-
-I-EANDHA-E-T-DENT-L------IT-NCA-----ITH-NE--THE-E-E--AGE-DI-CL-
-ING-HE-ETHE--I-E-A--A--E-HATICAND-I-THE--LL--INGI----A-ETHE-I--
TT--EADTHI--E--AGEE-AIL-EATD-A-IN-ATC--C-ED--ITHTHE--LL--INGC-
DE---D-L-EHAI-EDG-ITA--LA-E-ANDI-ILL-NAIL-AIL---AGI-TCA-D
```

Other likely candidates can be noticed, such as the first two words being THIS IS.

```
THISIS-----I-STCHALLENGINGC-DET-DEC---TINN---ALTI-ESI---LDHIDEA-
-I-EANDHA-EST-DENTSL------IT-NCA---S-ITH-NE--THESE-ESSAGESDISCL-
SING-HE-ETHE--I-E-AS-A--E-HATICAND-ISTHE--LL--INGI----A-ETHE-I-
STT--EADTHIS-ESSAGEE-AIL-EATD-A-IN-ATCS-C-ED--ITHTHE--LL--INGC-
DE---D-L-EHAI-EDG-ITA--LA-E-ANDI-ILLSNAIL-AIL---AGI-TCA-D
```

A sort of snowballing effect occurs as things stand out. THESE -ESSAGES DISCL-SING is one example, or ST-DENT and -EAD THIS -ESSAGE. Eventually, through picking out more and more substituted letters piece by piece, you come to the answer:

```
This is your first challenging code to decrypt. In normal times,
I would hide a prize and have students look for it on campus with
one of these messages disclosing where the prize was. Maybe what
I can do is the following: if you are the first to read this
message, email me at dmarino at cs.ucf.edu with the following
codeword: BLUEHAIREDGUITARPLAYER and I will snail mail you a gift
card.
```

2) Decode the following ciphertext that was encoded using the Vigenere cipher with a keyword from this wordlist:

https://www.ef.com/wwen/english-resources/english-vocabulary/top-1000-words/

To help you automate your task, I will guarantee that the substring "mate" will appear somewhere in the plaintext.

Here is the ciphertext:

```
vbmqsexciboriijxuwpaiygifzurtmwomisdnyglqaxiazmzrrwhwslampvlbbat
glemoemxzcohdhftvqnsexflvgogjpghwjbkqaolwixevgbrsxqhgwuhcyyfwtnp
rmwnpxgfmxrmpvnkwmflrdmlvrvmtiawmpmchmuxfhqqjrlgqiqmwtdzwraynmwa
jjubblrkbrsx
```

Solution

I created a program, **hw2_2.c**, to take the possible keywords as input and output potential plaintexts. Only plaintexts containing the word “mate” were displayed, allowing a visual scan to determine which keyword and which plaintext was correct—though, there was only one valid result in the end, and it explains why this was such a simple task! The keyword was "investment"

```
Normally, Vigenere is challenging to break, but since I gave you
both a list of possible keywords and something to look for in the
plaintext you should be able to automate this, get the key, and
then decrypt it so I don’t need to give you a ton of ciphertext.
```

3) Decode the following message, which was encrypted using the Vigenere cipher. Make sure to discuss all the steps you took, the key you arrived at, and the decoded message.

```
nywtlizzhbsenhehpgbyqwgrwelyorjcclbrkgafjxsebxkdlntyslxvzwzysocy
rslsycwpuwbrxlogxjinfmxzaffgywbwtgccwpxmaekooyytmynreqpqysumllvs
ubnwcdxxfvvhzxvexwrfkqmpdnffcwkrmpexrafjkrtmlxtokzpqhzxxysielyso
cvhycvurlpnbmjfjafjwcyxswelqttxxvmuwkrtcogttignpfxnhlivamysjlxtv
isuejxygmswlocqpnmxvjkocllpxfidswnjonhthzuagtyukrafjkysonaxcuakl
ydibzemwbnfyxmaieoocsdbxllfolbrpnaxeekdcwtlxtpzrlbneqtlwfgitnzol
lszohlypxmhqrclqzcytkixmsywocvmmffhpdlnmtgboocwpshnxialbfwfmaing
ybxthtlikvpbseqhkowgyrmtmikssdlk
```

Solution

Decrypting this cipher is a bit more difficult, since no possible keywords are given. Thus, you must first determine the key length. This can be accomplished by calculating the index of coincidence for multiple key lengths, which I automated through the use of **hw2_3a.c**. (As an aside, I actually coded this as a TA last year!)

Key	IC
1	0.040701
2	0.042329
3	0.040304
4	0.041632
5	0.050965
6	0.041930
7	0.039773
8	0.042090
9	0.039163
10	0.066401
11	0.042778
12	0.040385
13	0.040211
14	0.040067
15	0.051478
16	0.039655
17	0.042001
18	0.039557
19	0.038988

I made the maximum length rather high just in case, but it seems the key length with the IC closest to the English average of 0.067 is key length 10 with an IC of 0.066. Thus, we now know the key length to a high degree of likelihood.

The next step is creating a shifted key. First, treat the first bin as though it has been shifted by 'A' or 0. Then treat the second bin as though it has been shifted 0, 1, 2,…, 25 times. For each possible shift, determine the MIC between the first bin and the shifted second bin. The shift which results in an MIC closest to 0.067 (or rather, the highest MIC) is the relative shift between the first and second bin. For example, if the optimal relative shift of the second bin is 9, then the second letter of the shifted key is J. Repeat this step by comparing the first and third bin, first and fourth bin, etc. This results in a shifted key of AJ…

Of course, all that is simply theory. For the application of it, I created the program **hw2_3b.c**. It produces 26 possible keys from the optimal shifted key, then prompts the user to enter the one they desire (in case the user needs to make changes to account for strange letter frequencies). Miraculously (in my opinion), one of the shifted keys was `FLUTTERSHY`. As an aside, a couple of years ago, the key was travmtwelve and the letter frequencies made most possibilities look like gibberish, including travytaexvq. If this problem came to that point, guesses would have to be made as to which portions of the selected potential keyword are correct by identifying trends in the prototype plaintext due to the repetitive nature of the Vigenere cipher. Fortunately, that doesn't need to happen here.

Entering the keyword `FLUTTERSHY` allowed the program to properly decrypt the plaintext:

```
In case I hadn't told you all, my daughter is in to the My Little
Pony series and if you are reading this, you might have discovered
that. What is funny is that she and I like to do jumbles, and she
jumbled the keyword to break this code for me and I couldn't get
it for the life of me. So I decided to run a program to see if any
rearrangement of the letters she gave me mapped to a dictionary
word and they didn't. So I gave up. And then when she showed me
the answer, I realized it was so obvious, so I wanted to make sure
a regular dictionary attack, where you tried all the words in a
set, didn't work for this problem.
```

Part B: Written Questions Similar to Quiz/Exam Questions

4) Prove that encrypting a plaintext with two successive affine cipher keys is no more secure than encrypting a plaintext with a single set of affine cipher keys.

Solution

Let any two valid affine ciphers be $f(x) = (ax+b) \bmod 26$ and $g(x) = (cx+d) \bmod 26$. If we compose the two, we get:

$$g(f(x)) = (c(ax+b) + d) \bmod 26$$
$$= (acx + (bc+d)) \bmod 26$$

Since gcd(a, 26) = 1 and gcd(c, 26) = 1, it follows that gcd(ac, 26) = 1. Thus, the coefficient in front of x is a valid value for the original key a in the cipher. Similarly, bc+d is a valid value mod 26. Thus, given a, b, c and d from any two affine functions, we can compose the two functions and come up with an equivalent pair a' = ac mod 26 and b' = (bc+d) mod 26 that form a single affine key that is equivalent to the given function composition.

5) Find $67^{-1} \bmod 148$.

Solution

148 = 2*67 + 14
67 = 4*14 + 11
14 = 1*11 + 3
11 = 3*3 + 2
3 = 1*2 + 1

1 = 3 – 1*2
1 = 3 – 1(11 – 3*3) = – 1*11 + 4*3
1 = – 1*11 + 4(14 – 1*11) = 4*14 – 5*11
1 = 4*14 – 5(67 – 4*14) = – 5*67 + 24*14
1 = – 5*67 + 24(148 – 2*67) = 24*148 – 53*67

$67^{-1} \bmod 148 = (-53) \bmod 148 =$ **95**

6) For an alphabet of size 85, a set of affine encryption keys is a = 46, b = 22. (Thus the encryption function is f(x) = (46x + 22) % 85.) Determine the corresponding set of decryption keys.

Solution

The decryption function is d(x) = c(x – b) % 85, where $c = 46^{-1} \bmod 85$ and b = 22.

85 = 1*46 + 39
46 = 1*39 + 7
39 = 5*7 + 4
7 = 1*4 + 3
4 = 1*3 + 1

1 = 4 – 1*3
1 = 4 – 1(7 – 1*4) = – 1*7 + 2*4
1 = – 1*7 + 2(39 – 5*7) = 2*39 – 11*7
1 = 2*39 – 11(46 – 1*39) = – 11*46 + 13*39
1 = – 11*46 + 13(85 – 1*46) = 13*85 – 24*46

$c = 46^{-1} \bmod 85 = (-24) \bmod 85 = 61$

Now, we can do the work of inverting the function. Swap x and y and solve for y:

$x \equiv (46y+22) \bmod 85$
$(x - 22) \equiv 46y \bmod 85$
$61(x - 22) \equiv 61(46y) \bmod 85$
$y \equiv (61x - 1342) \bmod 85$
$y \equiv (61x + 18) \bmod 85$

It follows that **$f^{-1}(x) = (61x + 18) \% 85$.** The corresponding decryption keys are **a = 61** and **b = 18.**

7) A set of letters consists of 10 As, 25 Bs, 50 Cs, 5 Ds, 10 Es, and 60 Fs. What is the index of coincidence of the set? **Leave your answer as a fraction in lowest terms.**

Solution

First, the size *n* of the set is 10+25+50+5+10+60 = 160. Applying the formula for the index of coincidence results in:

$$\frac{10 * 9 + 25 * 24 + 50 * 49 + 5 * 4 + 10 * 9 + 60 * 59}{160 * 159}$$

$$= \frac{90 + 600 + 2450 + 20 + 90 + 3540}{160 * 159}$$

$$= \frac{6790}{160 * 159}$$

$$= \mathbf{\frac{679}{2544}}$$

8) The set of letters S consists of 10 As, 20 Bs, 45 Cs, 35 Ds, and 40 Es. The set of letters T consists of 30 As, 45 Bs, 25 Cs, 40 Ds and 60 Es. What is the mutual index of coincidence between sets S and T? **Leave your answer as a fraction in lowest terms.**

Solution

The size of the first set $n = 150$ and the size of the second set $m = 200$. Applying the formula for the mutual index of coincidence results in:

$$\frac{10 * 30 + 20 * 45 + 45 * 25 + 35 * 40 + 40 * 60}{150 * 200}$$

$$= \frac{300 + 900 + 1125 + 1400 + 2400}{150 * 200}$$

$$= \frac{6125}{150 * 200}$$

$$= \boldsymbol{\frac{49}{240}}$$