Assignment #2

CPEN 442

September 28, 2015

Department of Electrical and Computer Engineering
University of British Columbia
Vancouver, Canada

I. Problem #1

1. 1. ———————
   MD5 Hash = f0878b056247191f7fcfa6642c2d58de [1].

   2. I have added spaces to the recovered plain text to divide the words.

      THEN YOUVE HEARD MORE THAN I CAN SPEAK TO COMMA ANSWERED THE GAFFER DOT I KNOW NOTHING ABOUT JOOLS DOT MR DOT BILBO IS FREE WITH HIS MONEY COMMA AND THERE SEEMS NO LACK OF IT BUT I KNOW OF NO TUNNEL MAKING DOT I SAW MR DOT BILBO WHEN HE CAME BACK COMMA A MATTER OF SIXTY YEARS AGO COMMA WHEN I WAS A LAD DOT ID NOT LONG COME PRENTICE TO OLD HOLMAN DOT HIM BEING MY DADS COUSIN DOT COMMA BUT HE HAD ME UP AT BAGEND HELPING HIM TO KEEP FOLKS FROM TRAMPLING AND TRAPESSING ALL OVER THE GARDEN WHILE THE SALE WAS ON DOT AND IN THE MIDDLE OF IT ALL MR

   3. This was a monoalphabetic cipher.

   4. The key for this cipher is shown in Table I.

   5. **Approach** My first approach was to brute force the 26 combinations for deciphering a Caeser cipher. After looking through the the combinations I found none of them made sense. So I proceeded to do a frequency analysis using a Java program I developed myself[2]; seen in Figure 1. The analysis showed similar characteristics to the English alphabet so this suggested a monoalphabetic cipher[3].

      Next I proceeded to determine the common substrings of length two to five along with their frequency count. I started with the top occurring two and three letter substrings; shown in Figure 2 and Figure 3 respectively. I found *'JN'* and *'GJ'* in the top three occurring for

Fig. 1: Frequency Analysis - Problem 1

```
Q : 45  [10.47% ] #############################################
N : 42  [ 9.77% ] ##########################################
L : 39  [ 9.07% ] #######################################
G : 35  [ 8.14% ] ###################################
W : 32  [ 7.44% ] ################################
P : 28  [ 6.51% ] ############################
B : 27  [ 6.28% ] ###########################
R : 24  [ 5.58% ] ########################
J : 21  [ 4.88% ] #####################
M : 19  [ 4.42% ] ###################
X : 18  [ 4.19% ] ##################
H : 17  [ 3.95% ] #################
C : 12  [ 2.79% ] ############
F : 11  [ 2.56% ] ###########
K : 10  [ 2.33% ] ##########
O : 10  [ 2.33% ] ##########
U : 9   [ 2.09% ] #########
A : 8   [ 1.86% ] ########
T : 7   [ 1.63% ] #######
Z : 7   [ 1.63% ] #######
I : 5   [ 1.16% ] #####
S : 2   [ 0.47% ] ##
V : 1   [ 0.23% ] #
Y : 1   [ 0.23% ] #
D : 0   [ 0.00% ]
E : 0   [ 0.00% ]
```

Fig. 2: Common Two Letter Substrings - Problem 1

String	Count
QG	12
JN	12
GJ	11
RQ	9
BW	8

Fig. 3: Common Three Letter Substrings - Problem 1

String	Count
RQG	9
GJN	7
BWF	6
CQP	6
QPP	5

TABLE I: Monoalphabetic Cipher Key

Cipher Text	Letter
A	K
B	I
C	C
D	-
E	-
F	G
G	T
H	R
I	Y
J	H
K	B
L	A
M	L
N	E
O	W
P	M
Q	O
R	D
S	V
T	P
U	F
V	J
W	N
X	S
Y	X
Z	U

digraphs and *'GJN'* as the second most occuring three letter string. Since the two digraphs occurred more often *'GJN'*, this suggested that the pairs were common in the English language and not just because of the fact that *'GJN'* was common. After looking at common digraphs, I tried substituting *'THE'* for *'GJN'*.

Next I tried to find the word *AND* because it was a common word and *AN* is a common digraph. I looked for a three letter substring that also had a two letter prefix that was high occurring. One option was *RQG*; however, this would suggest that *'A'* did not have a high frequency according to my earlier analysis so this was unlikely. *'BWF'* was the next option but again 'B' did not have a high occurrence in the ciphertext. The next option that was plausible was *'LWR'*. The frequency of *'L'* in the ciphertext was similar to that of *'A'* in English.

Since I had $N \rightarrow E$, I tried to find common pairs of letters that began with *'E'*. This process involved trying a few combinations and looking at the output text to see if any new words formed or if there were any series of letters that seemed unlikely to make a word. I tried looking for *'ER'*; this involved finding pairs in the ciphertext that began with *'N'* and ended with a letter that had a high frequency since *'R'* is common in English. This led me to try mapping *'H'* to *'R'*; I saw the word *'HEARD'* and *'THERE'* in part of the decrypted ciphertext which led me to believe this could be correct.

The next unsolved, common digraph from the ciphertext beginning with *'N'* was *'NX'*. I tried pairing *'X'* with *'S'* to make *'ES* which was the next common two letter sequence begging with *'E'*. I did not see many new words except *'SEE'* but it did not look like this added letter caused any problems decrypted text in terms of forming words.

At this point, I could not find any new patterns with the substrings. So, I tried checking to see which characters in English had a high occurrence and were not yet paired in my key. First I tried assigning *'O'* to *'Q'* and *'I'* to *'P'*. But then I got the sequence *'HEARD IORE THAN'*. There are two words here, *'HEARD'* and *'THAN'*. This left *'IORE'* in th middle which did not make sense. Swapping the two pairs did not reveal any words either. So I backtracked and tried *'I'* for *'B'*. After doing this, I got a phrase *'AND IN THE'* towards the end of my excerpt, so I kept this pairing.

Reading the first part of the decrypted text I found *'HEARD _ORE THAN I'*. I tried using the word *'MORE'* in that place which led be substitute *'M'* for *'P'*. Next I substituted *'L'* for *'M'* since they were both the next unused most frequent letter in the English alphabet and ciphertext respectively. This revealed the word *'MIDDLE'* towards the end of the text. For the same reason as the previous pair, I inserted *'C'* for *'C'*. This revealed a repeated five letter sequence which was *'COMMA'*. I suspected that this was used to replace the punctuation of the same name. This also explained why the sequence *'DOT'* appeared many times as well so I assumed this meant a period in the sentence.

Looking at the common three letter sequences, I found *'BWF'* to match with *'IN_'*; I then tried to form *'ING'*. The word *'NOTHING'* became visible so I moved on. From this point forward, I would look at the decrypted text and try to make out partial words. This led me to solve for the remaining letters. For example, from *'PQWNI'* and *'MONE_'*, I tried *'Y'* for *'I'*. With fewer letters remaining, it became more manageable to substitute the next frequent letter in the English alphabet for the next frequent letter in the ciphertext. After each substitution, I would look at the decrypted text to see if this solution made sense in terms of forming new words and not creating any patterns that were not words.

When it came down to three letters, *'J'*, *'Q'*, and *'Z'*, I searched online[4] for the text to find that *'V'* mapped to *'J'* as I was unaware of the word *'JOOL'*. This excerpt was from *The Fellowship of the Ring*[4]. It is unknown if *'D'* maps to *'Q'* and *'E'* maps to *'Z'* or if *'D'* maps to *'Z'* and *'E'* maps to *'Q'* because the letters *'D'* and *'E'* did not appear in the ciphertext.

II. Problem #2

1. 1. I have added spaces to the recovered plain text to divide the words.

    LET ME KNOW X WHEN THEYRE BACK DOT LUPIN NODXDED X DOT WITH A WAVE TO THE OTHERS COMMA KINGSLEY WALKED AWAY INTO THE DARKNESS TOWARD THE GATE DOT HARRY THOUGHT HE HEARD THE FAINTEST POP AS KINGSLEY DISAPXPARATED IUST BEYOND THE BURXROWS BOUNDARIES DOT MR DOT AND MRS DOT WEASLEY CAME RACING DOWN THE BACKSTEPS COMXMA GINXNY BEHIND THEM DOT BOTH PARENTS HUGXGED RON BEFORE TURNING TO LUPIN AND TONKS DOT X THANK YOU COMXMA SAID MRS DOT WEASLEY COMMA FOR OUR SONS DOT DONT BE SILLY COMXMA MOLXLY COMXMA SAID TONKS AT ONCE DOT HOWS GEORGE ASKED LUPIN DOT WHATS WRONG WITH X HIM PIPED UP RON DOT HES LOST BUT THE X END OF MRS DOT WEASLEYS X SENTENCE WAS DROWNED IN A GENERAL OUTCRY DOT A THESTRAL HAD IUST SOARED IN TO SIGHT AND LANDED A FEW FEXET FROM THEM DOT BILL AND FLEUR SLID FROM ITS BACK COMXMA WINDSWEPT BUT UNHURT DOT BILL THANK GOD COMMA THANK GOD MRS DOT WEASLEY RAN FORWARD COMMA BUT X THE HUG BILXL BESTOWED UPON HER WAS PERFUNCTORY DOT LOXOKING DIRECTLY AT HIS FATHER COMXMA HE SAID COMXMA MADEYES DEAD X DOT NOBODY SPOKE COMXMA NOBODY MOVED X DOT HARRY FELT AS THOUGH SOMETHING INSIDE HIM WAS FALLING COMXMA FALXLING THROUGH THE EARTH COMMA LEAVING HIM FOREVER DOT WE SAW IT COMMA SAID BILXL FLEUR NODXDED COMMA TEAR TRACKS GLITTERING ON HER CHEEKS IN THE LIGHT FROM THE KITCHEN WINDOW DOT IT HAPXPENED IUST AFTER WE BROKE OUT OF THE CIRCLE MADEYE AND DUNG WERE CLOSE BY US COMMA THEY WERE HEADING NORTH TOO DOT VOLDEMORT HE CAN FLY WENT STRAIGHT FOR THEM DOT DUNG PANICKED COMMA I HEARD HIM CRY OUT COMMA MADEYE TRIED TO STOP HIM COMXMA BUT HE DISAPPARATED X DOT VOLDEMORTS CURSE HIT MADEYE FULL IN THE FACE COMXMA HE FELXL BACKWARD OFXF HIS BROXOM AND THERE WAS NOTHING WE COULD X DO COMXMA NOTHING COMMA WE HAD HALF A DOZEN OF THEM ON OUR OWN TAIL BILLS VOICE BROKE DOT OF COURSE YOU COULDNT HAVE DONE ANYTHING COMMA SAID LUPIN DOT THEY ALL STOOD LOOKING AT EACH OTHER DOT HARRY COULD NOT QUITE COMPREHEND IT DOT MADEYE DEAD IT COULD NOT BE MADEYE COMXMA SO TOUGH COMMA SO BRAVE COMMA THE CONSUMXMATE SURVIVOR AT LAST IT SEEMED TO DAWN ON EVERYONE COMXMA THOUGH NOBODY SAID IT COMXMA THAT THERE WAS NO POINT OF WAITING IN THE YARD ANYMORE COMXMA AND IN SILENCE THEY FOLXLOWED MR DOT AND MRS DOT WEASLEY BACK IN TO THE BURROW COMMA AND IN TO THE LIVING ROOM COMXMA WHERE FRED AND GEORGE WERE LAUGHING TO GET HER DOT WHATS WRONG SAID FRED COMMA SCANNING THEIR FACES AS THEY ENTERED COMMA WHATS HAPPENED WHOS MADEYE COMMA SAID MR DOT WEASLEY COMXMA DEAD DOT X THE TWINS GRINS TURNED TO GRIMACES OF SHOCK DOT NOBODY SEXEMED TO KNOW WHAT TO DO DOT X TONKS WAS CRYING SILENTLY INTO A HANDKERCHIEF SHE HAD BEXEN CLOSE TO MADEYE COMMA HARXRY KNEW COMXMA HIS FAVORITE AND HIS PROTGAT THE MINISTRY OF MAGIC DOT HAGRID COMMA WHO HAD SAT DOWN ON THE FLOXOR IN THE CORNER WHERE HE HAD MOST SPACE COMXMA WAS DABBING AT HIS EYES WITH X HIS TABLE CLOTHSIZED HANDKERCHIEF DOT BILL WALKED OVER TO THE SIDE BOARD AND PULXLED OUT A BOTTLE OF X FIREWHISKY AND SOME GLASXSES DOT HERE COMXMA HE SAID COMXMA AND WITH A WAVE OF HIS WAND COMMA EH SENT TWELVE FULXL GLASXSES X SOARING THROUGH THE ROXOM TO EACH OF THEM COMXMA HOLDING THE THIRTEENTH ALOFT DOT MAD EYE DOT MADEYE COMMA THEY ALL SAID COMMA AND X DRANK DOT MADEYE COMXMA ECHOED HAGRID COMXMA A LITTLE LATE COMMA WITH A HICCUP DOT THE FIRE WHISKY SEARED HARRYS THROAT X

2. This was a Playfair cipher.

3. The key I found was:

Z	F	I	H	K
G	L	X	N	O
P	V	A	Y	U
S	C	M	B	E
D	T	R	Q	W

4. **Approach** I first did a frequency analysis on the ciphertext seen in Figure 4. This distribution did not look as similar to the English alphabet as the distribution from Problem #1. This was especially

Fig. 4: Frequency Analysis - Problem 2

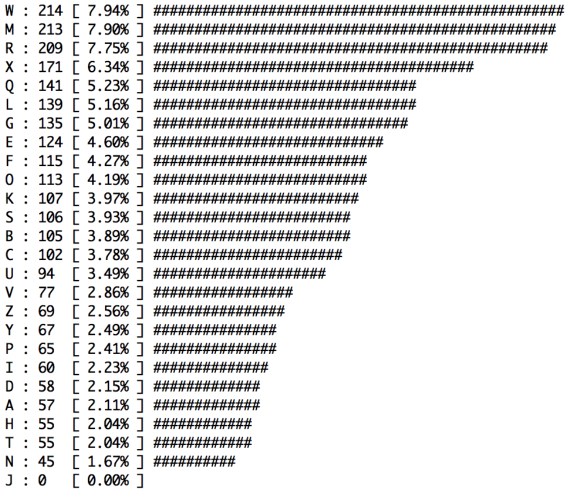

Fig. 5: Common Four Letter Substrings - Problem 2

String	Count
RARM	26
ELRA	25
XERM	24
TSXE	9
MPZR	8

evident for the letters that occurred less often such as *'P', 'I', 'D', 'A', 'H', 'T',* and *'N'*. In the ciphertext, they all had a similar frequency whereas in the distribution for common letters in English, there is a larger variation. The lack of *'J'*s in a ciphertext of almost 3000 characters, suggested a Playfair cipher so next I tried looking at the letters in pairs. My main thought process throughout this exercise was determining which pairs of letters encoded another pair of letters.

Based on Problem #1, I first tried to identify which pairs encoded 'COMMA'. I looked at the possible partitions of the word 'COMMA' and came up with '_C OM', 'CO MM', 'OM MA', 'MM A_'; where '_' represents some unknown letters. Now for a Playfair cipher, I need to insert an 'X' for repeating letters. The two combinations that match this case then become 'CO MX' and 'MX MA'. Next, I looked at the common four letter sequences (Figure 5 displays the top five) to determine which ones match the pattern that makes up the text 'COMMA'. I found the following matches:

- RA RM → MX MA
- EL RA → CO MX
- XE RM → OM MA

Breaking up into pairs I got:

Fig. 6: Possible RM to MA / RA to MX Orientations

_	X	A	M	R
_	_	_	_	_
X	_	_	_	_
A	_	_	_	_
M	_	_	_	_
R	_	_	_	_

Fig. 7: Possible Rectangle XE to OM Orientation

X	O	_	M	E
M	E	_	X	O
_	_	_	_	_
E	M	_	O	X
O	X	_	E	M

- RA → MX
- RM → MA
- EL → CO
- XE → OM

Next, I tried to find a possible orientation for these letters. In order for 'RM' to encode 'MA', this suggests they are either in the same row or same column since 'M' appears on both sides of the mapping. I used this information along with pair 'RA' → 'MX' to determine that 'X' was also in the same row or column. The possible combinations are shown in the Figure 6.

I tried looking at the possible rectangle orientations to map 'XE' to 'OM' and came up with four scenarios; Figure 7. I often used this technique of drawing the four rectangle orientations to determine which combinations fit with the key I had so far.

From this, I see that 'X' and 'M' are not in the same row and so 'RMAX' should appear in a column. Using the encodings I found from 'COMMA', I got a possible partial key shown in Figure 8.

This is what I used as a starting point. Next, I looked at high occurring digraphs (Figure 9). I considered 'TH' and 'HE'. Also, based on the last problem, 'DOT'

Fig. 8: Possible Key from COMMA

_	_	_	_	_
_	L	X	O	_
_	_	A	_	_
_	C	M	E	_
_	_	R	_	_

Fig. 9: Common Two Letter Substrings - Problem 2

String	Count
RM	57
QF	55
WG	38
XE	31
EL	29

Fig. 10: Possible Key After Looking at Digraphs

_	_	_	_	_
G	L	X	O	_
_	_	A	_	_
_	C	M	E	_
D	T	R	W	_

could be broken into ‘DO’ and ‘OT’. From the key I had so far, ‘ER’, which is a common digraph in English, appeared as well. So I tried to find a common occurring digraph that began with ‘M’. After trying a few combinations I ended up with the key shown in Figure 10.

This key revealed words such as ‘LET ME’ at the beginning of the ciphertext. The two digraphs that I looked at earlier, ‘TH’ and ‘HE’ were still not in the key so I tried adding these two in. I suspected ‘QF’ would match to ‘TH’ since it was an unused frequent pair in the ciphertext. I tried matching ‘KB’, the next frequent pair, with ‘HE’ (Figure 11).

At this point, the decrypted text I had so far did not form many words. I also found an occurrence of ‘QW’ which suggests that there may be something wrong with my key. I still needed ‘AN’, another frequent English digraph, so I tried ‘HX’ to ‘AN’. This revealed the word ‘WHEN’ in the text.

I had ‘OC RC WO OG RO QK BO’ encode ‘LE TM E_ XN WX WH EN’ in the beginning of my text. I could not think of a word that ended in ‘NW’ so I either had ‘OG’ or ‘RO’ mapped incorrectly. I attempted to fill in the blank in the phrase ‘LET ME ____ WHEN’ and looked at the letters in that part ignoring the X’s,

Fig. 11: Possible Key After Looking at QF and KB

_	_	_	_	_
G	L	X	O	_
_	F	A	K	H
_	C	M	E	B
D	T	R	W	Q

Fig. 12: Possible Key After Trying for KNOW

_	F	_	H	K
G	L	X	N	O
_	_	A	_	_
_	C	M	B	E
D	T	R	Q	W

Fig. 13: Cipher Key Found

Z	F	I	H	K
G	L	X	N	O
P	V	A	Y	U
S	C	M	B	E
D	T	R	Q	W

‘OGNW’. After some time, I looked through some common four letter words[3] and it made sense that a possible word was ‘KNOW’. So I tried rearranging the key I had in order to set ‘OG’ to encode ‘NO’. The new key is shown in in Figure 12.

At this point I started looking through the decrypted text looking for any patterns I could fill in. The plain text began with ‘LET ME KNOW X WH EN TH __ RE’, so I tried common words beginning with ‘TH’ and found ‘THEY’ to be a possible fit; this meant that ‘BU’ could encode ‘EY’. Next, I came across ‘TO THE OTHER_ COMMA’ and I tried replacing the blank with an ‘S’. This meant ‘CM’ $\rightarrow$ ‘SC’.

I found that ‘HX’ $\rightarrow$ ‘IN’ from the text ‘WALKED AWAY _N TO THE’. I came across ‘HARRY’ and determined that this was from the *Harry Potter* book series. Having read the books, I found that ‘ZH’ $\rightarrow$ ‘KI’ in order to form the name *Kingsley*. I had two more spaces in my key and found ‘PM’ $\rightarrow$ ‘AS’ based on the name *Weasley*. And the last letter remaining was a ‘V’. The key that I found to solve this cipher is shown in Figure 13. Since there is not ‘J’ in the key, some words such as ‘JUST’ will show as ‘IUST’ in the recovered plain text.

III. Problem #3

$x =$ “lzl1u”
$y =$ “p500a”
$CRC(x) == CRC(y) ==$ 0x399d9dec

When $x =$ “lzl1u” and $y =$ “p500a”, then we have the case such that $CRC(x) == CRC(y) ==$ 0x399d9dec. I found this by implementing a program[2] (crc-collisions/assn2-3.cpp), that iterated through a sequence of alphanumeric strings where letters were only lowercase. On each iteration, I utilized the Boost library[5] to perform the CRC computations. I then checked to see if the result was previously calculated and if it was then I found a value for x and y. Otherwise, I would store the result in a map and continue to the next string. Finding the values took 140.507 seconds (2 minutes 20.507 seconds). After finding the two results, I later verified my answers using pycrc.

IV. Problem #4

$x =$ "f0878b056247191f7fcfa6642c2d58de"
$y =$ "08apxe8"
$CRC(x) == CRC(y)$ == 0x78211f19

The md5 hash value for my student number was f0878b056247191f7fcfa6642c2d58de[1]. I used a similar approach to Problem 3 and iterated through a sequence of alphanumeric strings to check the CRC value of each one. However, I did not store each CRC computation result since I only needed to compare the values against the CRC value of the md5 hash of my student number. The CRC value for my md5 hash was 0x78211f19. After 1199.19 seconds (19 minutes 59 seconds), I found that when $y =$ "08apxe8", there is a collision. I verified this answer using pycrc as well.

References

[1] S. Walker. (2015). *md5 Hash Generator* [Online]. Available http://www.miraclesalad.com/webtools/md5.php

[2] K. Wou. (2015, Sept. 22). *ciphertext - GitHub Repository* [Online]. Available https://github.com/kwou/ciphertext

[3] S. Bryce. (n.d.). [Online]. Available http://scottbryce.com/cryptograms/stats.htm

[4] J. R. R. Tolkien. (2012). *The Fellowship of the Ring: Being the First Part of The Lord of the Rings* [Online]. Boston: Houghton Mifflin Harcourt. Available https://books.google.ca/books?id=aWZzLPhY4o0C

[5] D. Walker (2001). *CRC Library* [Online]. Available http://www.boost.org/doc/libs/1_59_0/libs/crc/