CIS 3362 Homework #3: Playfair, Hill Solutions
by Seth Jones and Arup Guha

1) (15 pts) By hand, using the Playfair Cipher, encrypt the plaintext "SAMMATTHEWSWENTUNDERGROUND" with the keyword "THREEHOLEPUNCH" and the padding character "X".

Solution

Here is the Playfair square, putting in the unique letters in the keyword as they appear in order, followed by the rest of the alphabet:

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

Here are the encrypted digrams:

SA -> IG MX ->XR MA -> ID TX -> RV TH -> HR EW -> HY
SW -> KZ EN -> NF TU -> RL ND -> UF ER -> OE GR -> DO
OU -> RC ND -> UF

The final ciphertext is: `IGXR IDRV HRHY KZNF RLUF OEDO RCUF`

2) (15 pts) You have matching plaintext and ciphertext for a Hill cipher with a 2 x 2 key. In particular, the ciphertext "GEOG" maps to the plaintext "CAKE". What are the possible encryption and decryption keys consistent with this information?

Solution

We'll show two possible approaches to this problem. Either can be used to obtain the encryption keys or the decryption keys. In particular, we'll

(a) Use the technique of setting up equations on paper and solve them to find possible decryption keys

and

(b) Use the technique of brute force with a program to find the possible encryption keys.

Of course, once we know an encryption key, we can use the formula presented in class to find the corresponding decryption key. (Similarly, the same formula works to uncover the encryption key given the decryption key since the two matrices are inverses of one another.)

First, let $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be the decryption key. Using the given information under mod, we have the two following equations:

$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 6 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ $\quad\quad$ $\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 14 \\ 6 \end{pmatrix} = \begin{pmatrix} 10 \\ 4 \end{pmatrix}$

GE → CA $\quad\quad$ OG → KE

These are equations under mod 26. First let's expand the matrix multiplication on the left-hand side concering the equations with a and b:

$6a + 4b \equiv 2 \ (mod\ 26)$
$14a + 6b \equiv 10 \ (mod\ 26)$

Note that if some integer d is such that ax + by = c (mod n) and d|a and d|b and d|c and d|n, then we can deduce that $\frac{a}{d}x + \frac{b}{d}y \equiv \frac{c}{d} \ (mod\ \frac{n}{d})$. We prove the result here:

By definition of mod, we have:

$ax + by = nz + c$, for some integer z.

Since each of a, b, c and n are divisibe by d, we can divide this equation through by d, resulting in each portion in parentheses being integers:

$(\frac{a}{d})x + (\frac{b}{d})y = (\frac{n}{d})z + \frac{c}{d}$

Now, take this equation under mod $\frac{n}{d}$ and we have

$\frac{a}{d}x + \frac{b}{d}y \equiv \frac{c}{d} \ (mod\ \frac{n}{d})$, as desired. Thus, we can rewrite both equations under mod 13:

$3a + 2b \equiv 1 \ (mod\ 13)$
$7a + 3b \equiv 5 \ (mod\ 13)$

Multiply the top equation by 3 and the bottom equation by 2:

$9a + 6b \equiv 3 \ (mod\ 13)$
$14a + 6b \equiv 10 \ (mod\ 13)$

Subtract the top equation from the bottom to get:

$5a \equiv 7 \ (mod\ 13)$

Using the Extended Euclidean Algorithm, we have:

$13 = 2 \times 5 + 3$
$5 = 1 \times 3 + 2$
$3 = 1 \times 2 + 1$

$3 - 2 = 1$
$3 - (5 - 3) = 1$
$2 \times 3 - 1 \times 5 = 1$
$2(13 - 2 \times 5) - 1 \times 5 = 1$
$2 \times 13 - 5 \times 5 = 1$
It follows that $5^{-1} \equiv -5 \equiv 8 \pmod{13}$. Multiplying through by 8 we get

$8(5a) \equiv 8(7) \pmod{13}$
$a \equiv 56 \equiv 4 \pmod{13}$

Substitute this value back into $3a + 2b \equiv 1 \pmod{13}$, to get $3(4) + 2b \equiv 1 \pmod{13}$, so $2b \equiv -11 \equiv 2 \pmod{13}$. It follows that $b \equiv 1 \pmod{13}$.

This gives us four possible solutions for a and b mod 26 since these solutions are mod 13: (4, 1), (4, 14), (17, 1) and (17, 14).

Now, let's do the same work for c and d:

$6c + 4d \equiv 0 \pmod{26}$
$14c + 6d \equiv 4 \pmod{26}$, divide by 2 to get

$3c + 2d \equiv 0 \pmod{13}$
$7c + 3d \equiv 2 \pmod{13}$, multiply equations by 3 and 2 and subtract:

$9c + 6d \equiv 0 \pmod{13}$
$14c + 6d \equiv 4 \pmod{13}$

$8(5c) \equiv 8(4) \pmod{13}$
$c \equiv 32 \equiv 6 \pmod{13}$

Substituting in the top equation we get $3(6) + 2d \equiv 0 \pmod{13}$, so $2d \equiv -18 \pmod{13}$ and $d \equiv -9 \equiv 4 \pmod{13}$

This means that the ordered pair (c, d) mod 16 could be any one of these four: (6, 4), (6, 17), (19, 4) and (19, 17).

Thus, there area total of 16 solutions for (a, b, c, d), but some of these will yield matrices with determinants that aren't relatively prime to 26. If we manually test the corresponding determinants for all 16 options, we find that 6 of the matrices are actual candidates for the real decryption matrix:

$$\begin{pmatrix} 4 & 1 \\ 19 & 4 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 19 & 17 \end{pmatrix}, \begin{pmatrix} 17 & 1 \\ 6 & 17 \end{pmatrix}, \begin{pmatrix} 17 & 1 \\ 19 & 4 \end{pmatrix}, \begin{pmatrix} 17 & 14 \\ 6 & 17 \end{pmatrix}, \begin{pmatrix} 17 & 14 \\ 19 & 17 \end{pmatrix}$$

Alternate idea where a program is written to find the potential encryption keys. Set up the following equations:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \end{pmatrix} \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 10 \\ 4 \end{pmatrix} = \begin{pmatrix} 14 \\ 6 \end{pmatrix}$$

CA → GE KE → OG

Using the first equation, it's clear that $2a \equiv 6\ (mod\ 26)$ and $2c \equiv 4\ (mod\ 26)$, meaning that $a \equiv 3\ (mod\ 13)$ and $c \equiv 2\ (mod\ 13)$. We can simply brute force all possible values for b and d (code is attached as h3q2.py) and check if the corresponding determinant is relatively prime with 26 and just output every ordered quadruplet that satisfies both equations. Here is the output of the program:

```
3 9 2 3
3 22 2 3
3 9 15 16
3 22 15 3
16 9 15 3
16 9 15 16
```

Rewriting as matrices we have:

$$\begin{pmatrix} 3 & 9 \\ 2 & 3 \end{pmatrix}, \begin{pmatrix} 3 & 22 \\ 2 & 3 \end{pmatrix}, \begin{pmatrix} 3 & 9 \\ 15 & 16 \end{pmatrix}, \begin{pmatrix} 3 & 22 \\ 15 & 3 \end{pmatrix}, \begin{pmatrix} 16 & 9 \\ 15 & 3 \end{pmatrix}, \begin{pmatrix} 16 & 9 \\ 15 & 16 \end{pmatrix}$$

3) (30 pts) Playfair Decrypt This ciphertext was produced via the Playfair cipher. Since it might be difficult for students to completely decrypt this on their own, eventually some matching plaintext and ciphertext will be revealed. (Probably 4 or 5 pairs.) Note: The padding character is q.

```
tikqdifmfrstfodebrhzhizfziyshpktmzbsegizemabkhzimzmfabnizfoi
kyidzkemnsshzveqiydzytspfrstpbtyhimzmfabniycsbbeosteslhuziyi
bkoctsiqbdzkpyyzyizkdkmttvsvrvtmsnwczduhhbdrytspletsavleodyz
knfzbeisfaeldrcdshzikmcmhksrrdvsmtmttibuzytzwczduxelabzpmtmr
xvgostiqzwptzenwvsmtvrwdytbuzyfzdrektzkcanyzmtmktzkckcbpgsbz
cdsytzzunwgdxfofbgtybepdqsmreztpiqdhpdyzmttbslhizfzitzcfdpmz
mnzihkzcbuztspnkxatyizoshkgawfldktye
```

Solution

To start this problem, I started to do analysis on the ciphertext, I looked at the diagram representation of the ciphertext vs the popularity of digrams in generalized English text. I started by mapping the most popular diagrams, MT -> TH and ZI -> HE as per https://cs.wellesley.edu/~fturbak/codman/letterfreq.html. For this part it's mostly guess and check,

try mapping a popular diagram to one similarly represented by our digram frequency chart. Once I would find a mapping I would want to try I would make a square of the letters, similar to the notes, and start to try to make inferences as to what letters would be in the keyword based off of how the squares were formed. For example, if I saw that a square would lead me to get a letter out of order, I would know 1 or both of those letters would be used in the keyword, assuming that the digram guessing was correct.

Once the Hints came out, this process became a lot easier.
LE TM ED OON EM was the first few digrams given. Adding the padding char we get LE TM ED OQ ON EM, and I was able to inference that the next few letters were probably going to be ORE. “Let me do one more”
TI KQ DI FM FR ST FO
VVVV
LE TM ED OQ ON EM OR
I immediately notice that FO -> OR, which gives a huge hint, it means its either horizontal or vertical, because F->O when mapped with O. Horizontal also would imply that all of the letters were used in the keyword. My first assumption is that it was horizontal due to it being kind of awkward for OR to be vertical since they are so close in the alphabet. This would imply an ROF in the keyword somewhere (When I saw this I instantly thought it would be first row, and it would be PROF like professor, this ended up not being true). This is confirmed when we see FR map to ON, this would mean that the row would have to be NROF. Since we can see FM -> OQ, we can also assume that MQ will be in the same columns as OF. We also see a mapping of TM ->KQ, since we know mq are next to eachother, we can also assume now that this is a row case, where T and K are in the same row as M and Q. Since these letters are so close to each other in the natural alphabet, and hard to find a word that makes any sense out of this ordering, I can only guess that they are in the same row and unused from the keyword. Due to the low use of the last letters of the alphabet, I also took an educated guess that this would not be the last row. Since T mapped to K I could also assume that the row wrapped around from T to K.
I also knew the NROF would probably not be the top row because it looks like it may be a word with some letters taken out, meaning it would be row 2 or 3. This would mean that the table is as follows so far:

N	R	O	F	
K		M	Q	T

This gives some huge hints so far, if we assume that our row placement for row 4 is correct, we know that only 1 letter out of UVWXYZ is used, because there is 6 letters after T and only 5 spaces.
Now we continue to look for combinations that contain the letters we have on our board so we can try to build up some sort of full row or columns.
Since ST -> EM we can also, and we see that S is used due to our mapping going immediately from Q to T in our box, that would mean that the box would look like this:

		S		E
N	R	O	F	

K		M	Q	T

Since ED -> DI we know that they have to be in a row or column, If it was column, that would mean that D is the last letter in the box, the bottom right, so that didn't make sense. So that meant, D would wrap around, and I would be after D, leaving us with:

D	I	S		E
N	R	O	F	
K		M	Q	T

To double check what im doing is making sense, using cryptool, with my letters put in, I hit the decrypt button and double check that everything looks normal, and also that our given words are correct. At this point the beginning looks like "-ETMEDOQONEMORE". So I believe im on a decent track, because now given other hints, we do have MORE as the last word, meaning that earlier assumption about the finished phrase was correct, and the E naturally appeared there helps me to believe that E is in the correct spot as well.
Our current plaintext is:

-ETMEDOQONEMORE----------TQ--------ST--------QO--RD--RS------TDKOII----------ME--OI------S----------S-EM-I----------ME----------Q------Q-DO--------IN------ME----NS----O---D-O------EI--D-----MDNN----Q-QQ----------Q-------EM----------Q----------
INDT---------Q-T----------EI----------OR---------K---I----K-------QQ--------E---N-----KO----------D-------S---------TQ--

Doing some playing around, I put L where it would go naturally in the alphabet and press decrypt and get:
LETMEDOQONEMORE----------TQ--------ST--------QO--RD--RS------TDKOII----L-----ME--OI------SL----------S-EM-I----------ME-L----------LQ------QLDO--------IN----TIME--TINS-----O---D-O--KI--EI--DL----MDNN----QLQQ----------QL------EM-L----------LQ----------INDT--------LQLT----------EI----------OR---------K---I----K-------QQ--------EL--N-----KO----------D-------S-------KITQ—

This unlocks the L in the beginning of the plaintext which we know is correct, and it unlocks a word "TIME" later on in the plaintext which gives me a lot of confidence that im on the right path.

Our remaining unused letters are ABCGHUPVWXYZ, and we can see because theres no letters between L and M that P is used, and we know that atleast one of the last 6 letters are used.

At this point because I knew there were 5 out of last 6 letters used, I did a version of brute force. I filled in the last row with all the combinations (only 5) and found the plaintext that made the most sense

LETMEDOQONEMORE----------WE--TQ--------ST--------QO--RD--RS--VE--TDKOII--YIL-
UE--ME--OI------SL------EV----S-EM-I----EV------ME-L--------VE--U-LQLYIWILQLDO--------
INT---TIME--TINS-----O---D-O--KI--EI--DL----MDNNIUW-QLQQ----T-----UKY-------QL--
WU--EM-L------OUWILQLIUST-------INDT--------LQLT-------------EIT---KD---UOR---------
KW--I----K---EU--QQ--------EL--N-----KO-----------D----T--SW----XOKITQTY

seemingly this unlocks the ending “XOKITTY”, indicating to me that this is correct.

D	I	S		E
N	R	O	F	
K	L	M	Q	T
U	V	W	X	Y

LETMEDOQONEMORE----------WE--TQ--------ST--------QO--RD--RS--VE--TDKOII--YIL-
UE--ME--OI------SL------EV----S-EM-I----EV------ME-L--------VE--U-LQLYIWILQLDO--------
INT---TIME--TINS-----O---D-O--KI--EI--DL----MDNNIUW-QLQQ----T-----UKY-------QL--
WU--EM-L------OUWILQLIUST-------INDT--------LQLT-------------EIT---KD---UOR---------
KW--I----K---EU--QQ--------EL--N-----KO-----------D----T--SW----XOKITQTY

Since we know P And Z are used for sure I try plugging these in to see if I can find anything else. It was at this point that I also realized that I had weird “ “ characters that were copied into cryptool from me copying straight from problem, after fixing this and then adding p and z to our blank spaces we get (**I think this had a big effect, I think the spaces that changed the segments made about 50% of characters to be corrupted**):

LETMEDOQONEMOREP------FOREWE--TQTO----ERST----RETOQO--
RDFORSTUDENTSTOD--RYPTEVENT-ISONEM---T--TOQO--RD------SW-YIM--REVE----
MEPL--NTEXTEVENTU-LQLYIWILQLDO--NE----INT-ISTIME--TINSTE-DOF--DI--ITIN---
-RETL----IONIWILQLQLE--ET-E--NEYWIT--FELQLOWU--
EMPLOYEQEYOUWILQLIUST---ETOFINDT-E----TELQLT-E----------EW-ENYOU--
QPRO---T--SEMPLOYEQEPL--SETELQ--IM--FORE-E--ESTOKORE------E-ISD----TERSW--
--XOKITQTY

At this point I couldn’t figure out what word could be made with the Z, I tried a bunch of different combinations of the last 5 letters to see what made sense, eventually seeing that certain letters being added helped it make slightly more sense. Turns out, there is no other letters, Z is the final letter, the rest is alphabetical.

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

Here is the final message with spaces inserted between words, padding characters removed and adding some punctuation:

LET ME DO ONE MORE PRIZE BEFORE WE GET TO CIPHERS THAT ARE TOO HARD FOR STUDENTS TO DECRYPT. EVEN THIS ONE MIGHT BE TOO HARD, WHICH IS WHY I MAY REVEAL SOME PLAINTEXT EVENTUALLY. I WILL DO MONEY AGAIN THIS TIME, BUT INSTEAD OF HIDING IT IN A SECRET LOCATION, I WILL LEAVE THE MONEY WITH A FELLOW UCF EMPLOYEE. YOU WILL JUST HAVE TO FIND THEM AND TELL THEM A MAGIC PHRASE WHEN YOU APPROACH THIS EMPLOYEE. PLEASE TELL HIM BEFORE HE GOES TO KOREA TO HAVE HIS DAUGHTERS WATCH XOKITTY.

4) (40 pts) This ciphertext was encrypted via the Hill cipher using a 3 x 3 key. Note: The intended approach to break this one might take a little bit of compute time.

```
kcppxxtlafgktqiqhhgistvgmxfmkknffomkpkhjcgpjsuitqzgykohrwtvj
xomjgtwgvrfagkolfrtqitqvgkofajtxxurkykoyompeeecblfrlplwurywq
ywiekjctb
```

Solution

Here are lists that I first looked at:

https://word-lists.com/word-lists/100-most-common-5-letter-words/
https://word-lists.com/word-lists/100-most-common-6-letter-words/

To start this problem, I knew it was going to involve some amount of brute force, and my goal was to kind of lean in a direction of cutting stuff out early so I can limit this time. To start this, I would try to look at each individual block of 3 and see if I could create a mapping to make that segment = “the”. I figured the popularity of the word “the” I could maybe get lucky and the word would land in a perfect block. So, every possible mapping to “the” I would decrypt with that matrix, and see if I could find a couple other “the”s in the plaintext, if I could, I would print the message. This ran for a while and then didn’t find anything, I figured that there were either too many results for me to properly search, or maybe “the” didn’t pop up as a perfect block at all.

I started to add other filters, I took the top 100 most common 5 and 6 letter words and added them to a text file, and used these as another filter. I chose 5 or 6 letter words because the chance of coincidence would be lower. Some of these words contained “the” and we were getting a lot of false positives just based on our other criteria already being “the” (ie: their) and random chance, so I removed these from the lists.

I also set a portion of the code to instant filter out certain letter frequencies went way above a regular rate for english, these letters were J, Q, X, and Z. If they showed up more than a few times in a single row then we cut the row, this was the biggest drop in workload.

I also added another filter that dropped certain matrices based on the even and odd combinations of the rows, I found that certain row combinations could never produce a proper determinant and cut them early. I also check determinant relatively prime with 26.

Due to the number of results I had, I sorted by most matches to see if I could find any patterns. I got really lucky, member matched with “faculty member” and “remember”.

```
lfr [[22, 10, 13], [10, 17, 18], [15, 7, 16]] ['please',
'member', 'member']
```

Here is the resulting plaintext:

To obtain your gift, please find faculty member Kyle Dencker. Remember he will only give you the gift if you tell him the message from the previous question.

It was actually the only result that matched 3 words, there were some similar matrices that produced similar columns such as:

2 lfr [[9, 23, 13], [10, 17, 18], [15, 7, 16]] ['please', 'member']

toobtainyburtifgpleasefiadfacultymemoerxyledenckrrrememberhejilyonlygiveyouthetifgifyougell hizthemessatefromthepreviousquestioaxx

This solution solves it in about 30-60 seconds, but i got very lucky with the amount of words that matched, I think.