Fall 2019
Chris Christensen
MAT/CSC 483
Section 11

Cryptanalysis of the columnar transposition cipher

We will assume that the columnar transposition uses a rectangular array that was completely filled.

Here is the ciphertext:

```
ASAIR  ITFNM  IMTKL  SOIEE  M
```

The "key" to cryptanalyzing the ciphertext is to determine the number of columns; i.e., the length of the keyword. There are 21 letters in the ciphertext. Because we know that the message completely fills the rectangle, this suggests either a 3×7 or a 7×3 array.

We arrange the ciphertext in columns.

Either

A	I	T	M	T	S	E
S	R	F	I	K	O	E
A	I	N	M	L	I	M

or

A	F	L
S	N	S
A	M	O
I	I	I
R	M	E
I	T	E
T	K	M

.

The solution is by anagramming (making a word or portion(s) of word(s) by rearranging letters) a row.

The 7×3 arrangement seems unlikely because it has a string `TKM` with no vowels that is unlikely. Also, the `III` is unlikely. So, let us try the 3×7 arrangement. Notice that there are 7! = 5040 arrangements of the columns. We would like to not have to try all of them!

A I T M T S E
S R F I K O E
A I N M L I M

In the first row, MATE seems to leap out. This leaves ITS. Perhaps, a slightly wrong guess – ESTIMAT- seems to be a possibility.

Let us rearrange the columns.

E S T I M A T
E O K R I S F
M I L I M A N

Not quite, but there are two Ts in the first row. Let us swap those columns.

E S T I M A T
E O F R I S K
M I N I M A L

This works. Notice that because we have multiple rows that are permuted the same way, we can use multiple anagramming for cryptanalysis.

It is often worthwhile to write the ciphertext in columns, cut out the columns, and rearrange the columns to do the anagramming.

Determining the dimension of the rectangle

Frequencies can help to determine the dimensions of the rectangle. In English approximately 40% of plaintext consists of vowels. Therefore, for the correct dimension, each row of the rectangle should be approximately 40% vowels. Consider our choice between 3×7 and 7×3.

For a 3×7 rectangle, each row should contain approximately 2.8 vowels. Let us note the difference between this estimate and the actual count:

							Number of vowels	Difference
A	I	T	M	T	S	E	3	0.2
S	R	F	I	K	O	E	3	0.2
A	I	N	M	L	I	M	3	0.2

The sum of the differences is 0.6.

For a 7×3 rectangle:

			Number of vowels	Difference
A	F	L	1	0.2
S	N	S	0	1.2
A	M	O	2	0.8
I	I	I	3	1.8
R	M	E	1	0.2
I	T	E	2	0.8
T	K	M	0	1.2

The sum of the differences is 6.2. It appears that the 3×7 rectangle is more likely.

Using digraph frequencies to arrange the columns

Digraph frequencies can be used to help in the cryptanalysis in place of just looking for reasonable pairings of the columns. For example, consider our ciphertext above `ASAIR ITFNM IMTKL SOIEE M`. Again, we'll assume that a 3×7 rectangle is appropriate.

A I T M T S E
S R F I K O E
A I N M L I M

We will pair the first column with each of the other columns on the right and consider how likely it is that such digraphs will occur in English. The frequencies we will use come from Sinkov. Recall that there are $26 \times 26 = 676$ digraph frequencies.

```
AI   311    AT  1019    AM   182    AT  1019    AS   648    AE    13
SR     9    SF     8    SI   390    SK    30    SO   234    SE   595
AI   311    AN  1216    AM   182    AL   681    AI   311    AM   182
     631        2243         754        1730        1193         790
```

The most likely pairing is

```
AT
SF
AN
```

Oops! We know that this is not the correct pairing, but the second most likely pairing is correct. (During cryptanalysis, we don't always get the correct result on the first try.)

Once we have a pairing, we could then continue using digraph frequencies to select columns to add on the left and on the right. Etc.