

## Decipherment of Playfair cipher

The Playfair cipher is a digraphic substitution cipher. It is among the hardest cipher types we have encountered, because it does a good job of disguising the letter frequency patterns of English. The patterns are still there, and can still be used.

Recall that in Playfair letters are encoded in pairs, and that a 5-by-5 key square is used. The one below is a boring square, but illustrates the idea once again:

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

There are three rules.

- 1) If the two letters are in the same row, they are each encoded using the letter to their right. (AB goes to BC). If there is no such letter, because you are at the end of the row, then use the first letter instead (AE goes to BA)
- 2) If the two letters are in the same column, they are each encoded by the letter below (AF goes to FL). If no such letter, back to top (AV -> FA).
- 3) If the two letters are in different row and column, they form the corner of a rectangle and are encoded by the other corners, using the letters in the same row. (AG goes to BF, AU goes to EQ)

We'll use a running example from Helen Fouche Gaines' **Cryptanalysis**.

```
HR KY LD ZX NQ EO ND EC TC TI AD CT AK RH LB GT SN AN
UN ON DR HX PE BN ZC DT KV EQ HD AO HR DU RF TQ OB DE
QD HR KY YA HZ HB BU KZ EQ XG TI BI KY RI CQ HR CE CO
SX RM BC TH CG QD RK NQ IT DC WT FV UB YA GU HE CZ NU
LB IQ YK FV UB IQ WD QB UN KM DE TD KA HR NU OU
```

As usual, we need a plain frequency count

| A | B  | C  | D  | E | F | G | H  | I | J | K  | L | M | N  | O | P | Q  | R  | S | T  | U  | V | W | X | Y | Z |
|---|----|----|----|---|---|---|----|---|---|----|---|---|----|---|---|----|----|---|----|----|---|---|---|---|---|
| 7 | 11 | 11 | 14 | 9 | 2 | 4 | 12 | 7 | 0 | 10 | 3 | 2 | 11 | 6 | 2 | 11 | 11 | 2 | 11 | 10 | 3 | 2 | 4 | 6 | 5 |

It is also usual that the frequency count is more informative when sorted so as to group the letters by frequency group. We are sure that ETAOIN SHRDLU will be mainly in the first 12 by frequency.

### ***High frequency***

|    |    |    |    |    |    |    |    |    |    |   |
|----|----|----|----|----|----|----|----|----|----|---|
| D  | H  | B  | C  | N  | Q  | R  | T  | K  | U  | E |
| 14 | 12 | 11 | 11 | 11 | 11 | 11 | 11 | 10 | 10 | 9 |

### ***Medium Frequency***

|   |   |   |   |   |
|---|---|---|---|---|
| A | I | O | Y | Z |
| 7 | 7 | 6 | 6 | 5 |

### ***Low Frequency***

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| G | X | L | V | F | M | P | S | W | J |
| 4 | 4 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 0 |

The high frequency group contains all the letters with greater than average frequency. These are likely to be either E and T or related in some way to E and T

The first useful fact specifically for cryptanalysis of Playfair is:

**Each plaintext letter is translated by no more than 5 different ciphertext letters.**

This helps because it gives us the connection between the frequencies of ciphertext digraphs and the (known) frequencies of plaintext letters. The most frequent plaintext digraphs are very likely to have E T A O I N in them, so the most frequent ciphertext digraphs are likely to be translations of these plaintext digraphs. The top few plaintext digraphs, in order of likely frequency are TH, IN, ER, RE, AN, HE, AR, EN, TI, TE, AT, ON, HA, OU, IT.

In our cryptogram, HR is the most common with a frequency of 5, KY is present with frequency of 3. This would be more useful if we had more text.

The second useful fact is that

**Letters which are in the same row of the keysquare as a common letter will tend to be high frequency in the ciphertext.**

This is because Playfair will almost always translate a letter by one of its rowmates.

The next useful fact is:

**If Playfair translates the pair  $xy$  as  $pq$  it also translates  $yx$  as  $qp$ .**

This greatly increases the value of ER,RE, because the ciphertext for RE will be the reverse of that for ER, and both digraphs will be frequent. TI and IT, might be helpful too, though IT actually isn't that frequent.

In this cryptogram, KY (3) and YK (1) are both present, as are UN (2) and NU (2). We might tentatively suppose TH->HR. Because there isn't much text, this is a guess.

The next useful fact is a little non-obvious.

**Guesses like TH -> HR, where the ciphertext and the plaintext share a letter, are particularly useful.**

When Playfair uses the diagonal rule, the ciphertext letters are drawn from the same rows as the plaintext letters, and these rows are different. The diagonal rule can't apply to TH -> HR, because H would have to be in the same row as T, and in the same row as R, but H and R need to be in different rows. So we can't be using the diagonal rule. If so, THR must be present, in that order, in either a row or a column. It could wrap around

.THR.

R. .TH

But either way, this gives us a clue about what the square is. Actually, we don't have to worry about this right now, because we have some freedom in how we lay out the square. This is because many re-orderings of the square produce the same transformation.

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

Move leftmost column to right

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

Move bottom row to top

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

So, actually, we can just write THR in the middle of our guessed square and be done with it. The square may come out a bit scrambled, but it will do the right thing. We need to consider both the column possibility and the row possibility, because Playfair is not quite symmetrical: the ciphertext letters are drawn from the same row as the plaintext letters more often than they are drawn from the same column.

```

. . . . .      . . . . .
. . . . .      . T . . .
. T H R .      . H . . .
. . . . .      . R . . .
. . . . .      . . . . .

```

OK, now lets look at the contexts in which HR occurs

```

HR KY LD ZX NQ EO ND EC TC TI AD CT AK RH LB GT SN AN
UN ON DR HX PE BN ZC DT KV EQ HD AO HR DU RF TQ OB DE
QD HR KY YA HZ HB BU KZ EQ XG TI BI KY RI CQ HR CE CO
SX RM BC TH CG QD RK NQ IT DC WT FV UB YA GU HE CZ NU
LB IQ YK FV UB IQ WD QB UN KM DE TD KA HR NU OU

```

Notice that HR KY is repeated, and that there are 3 other tetragraphs with HR in.

**Question:** In Playfair, are we certain to see a repeated tetragraph if a four letter word occurs two times in the message? How about if it occurs three times?

Probably HR KY is *this, that* or *they*. On the assumption that it is *this* we have a four-letter equation. IS -> KY

```

I
K           I . K
.   I K . S Y   . .
S           Y . S
Y

```

We don't know exactly where the fifth letter is in the vertical and horizontal arrangements, and the diagonal pattern stands can have any number of rows and columns replacing the dots.

At this point there is not really enough to pin down details of the square. In practice, you need to guess a keyword. For this message, if we are lucky, we know that the word **condemnation** is likely to appear. It will be split up into bigrams one of two ways

CO ND EM NA TI ON

Or

.C ON DE MN AT IO N.

We now need to scan these possibilities across the message. We can rule out the first one because we have letters that map to themselves. (I'm cheating by skipping to the one that is going to work

CO ND EM NA TI ON  
EO ND EC TC TI AD

However, this one is a candidate, actually rather a good one, because it has several three-letter equations

.C ON DE MN AT IO N.

EO ND EC TC TI AD CT

Specifically, ON -> ND and DE -> EC give us a lot to work with, because they share D.

We must have

OND

The next fact is sometimes useful if there is a keyword.

**It is likely that the keyword will use many of the most frequent letters and at least a couple of vowels.**

If so, the frequent letters and vowels will tend to cluster in the top rows.

