# PLAYFAIR CRYPTANALYSIS OF A CHALLENGE CIPHER

BURTON ROSENBERG
UNIVERSITY OF MIAMI

## Introduction

A Playfair cipher is a digraph substitution, where each letter pair of the plaintext is replaced by a letter pair in the ciphertext. The chart of substitutions is summarized by the placement of letters in a table, called the *Playfair Square*, constructed in some customary manner from a keyword. The table is used to transform letter pair $(a, b)$ with letter pair $(c, d)$ by first locating the letters $a$ and $b$ in the table, noting whether the two letters share a row, share a column, or share neither.

In the case where $a$ and $b$ share neither row nor column, a square is formed with two other letters $(c, d)$, where $c$ is in the same row as $a$ and the same column as $b$; and $d$ is in the same row as $b$ and the column as $a$. The substitution is then to replace the letter by the other letter in the same row: $c$ for $a$ and $d$ for $b$, written $(a, b) \mapsto (c, d)$.

In the case where $a$ and $b$ are in the same row, the substitution is to letters $c$ and $d$, which are one to the right, wrapping around if $a$ or $b$ is the right-most letter in a row, to the leftmost letter in the same row.

Finally, the case where $a$ and $b$ are in the same column, the substitution is to letters $c$ and $d$, which are one below, wrapping around if $a$ or $b$ is the bottom-most letter in a column to the top-most letter of the same column.

It is not possible to encode a double letter, such as $aa$, and such digraphs need to be broken up, customarily by inserting the letter $x$, i.e., $axa$. What happens for messages discussing matters involving the text $xxx$ is not customarily specified.

## Some easy observations.

That the first sort, of the square, substitutes by following rows, and that of the other sorts, substitutes are made to the left and down, are the typical conventions but obviously have no significance to the strength of the cipher.

- Note that the rows can be cyclically permuted, as well as the columns can be cyclically permuted, without changing the encipherment.
- Note that if $(a, b) \mapsto (c, d)$ then $(b, a) \mapsto (d, c)$.
- While substitutions of the first case are involutions, those of the other cases are not.

*Date*: created: September 8, 2004; revised: January 27, 2018.

```
FIRST CASE:

    a * c
    *   *     a square is formed
    d * b

SECOND CASE:

    c a * d b   c and d in a common row

THIRD CASE:

    c
    a
    *     c and d in a common column
    d
    b
```

**Figure 1:** Three substitution cases.

## The obstruction theorems

Given a pair of known or proposed substitutions, they might be combined to give information on the arrangement of the substitutions. In effect, substitutions obstruct the possible arrangements of other substitutions.

**Theorem 0.1** (Obstruction theorem IA)**.** Let $a, b, c, d, e$ and $f$ be variables representing distinct letters. If $(a, b) \mapsto (c, d)$ and $(a, e) \mapsto (f, b)$ are both substitutions, then $(a, b) \mapsto (c, d)$ is arranged in a column and $(a, e) \mapsto (f, b)$ is arranged in a square. This theorem also holds when $(a, e) \mapsto (f, b)$ is replaced by any of:

$$\begin{aligned}
&(e, a) \mapsto (b, f), \quad (b, e) \mapsto (f, a), \\
&(e, b) \mapsto (a, f), \quad (c, e) \mapsto (f, d), \\
&(e, c) \mapsto (d, f), \quad (d, e) \mapsto (f, c), \text{ or } (e, d) \mapsto (c, f).
\end{aligned}$$

**Theorem 0.2** (Obstruction theorem IB)**.** Let $a, b, c, d, e$ and $f$ be variables representing distinct letters. If $(a, b) \mapsto (c, d)$ and $(a, e) \mapsto (b, f)$ are both substitutions then $(a, b) \mapsto (c, d)$ is arranged in a row and $(a, e) \mapsto (b, f)$ is arranged in a square. The theorem also holds when $(a, e) \mapsto (b, f)$ is replaced by any of,

$$\begin{aligned}
&(a, e) \mapsto (f, b), \quad (b, e) \mapsto (a, f), \\
&(e, b) \mapsto (f, a), \quad (c, e) \mapsto (d, f), \\
&(e, c) \mapsto (f, d), \quad (d, e) \mapsto (c, f), \text{ or } (e, d) \mapsto (f, c).
\end{aligned}$$

**Proof:** Assume in order to establish a contradiction that $(a, b) \mapsto (c, d)$ is arranged in a square. Then $a$ and $b$ are in distinct columns and rows. Therefore $(a, e) \mapsto (f, b)$ cannot be the type of substitution all in a column or a row. If arranged in a square $(a, e) \mapsto (f, b)$ would have $a$ and $b$ in the same column, a contradiction. So $(a, b) \mapsto (c, d)$ cannot be arranged in a square.

If both $(a, b) \mapsto (c, d)$ and $(a, e) \mapsto (f, b)$ are arranged in a row or column, they are both in a single row or both in a single column. Then there would be six distinct letters in that row or column, which is impossible. So $(a, e) \mapsto (f, b)$ must be arranged in a square.

Having established that $(a, e) \mapsto (f, b)$ is arranged in a square, $a$ and $b$ are in distinct rows, so $(a, b) \mapsto (c, d)$ cannot be arranged in a row. On the other hand, we can arrange $(a, b) \mapsto (c, d)$ in a column consistent with $(a, e) \mapsto (f, b)$ being arranged in a square.

This shows the proof for the first claim of IA. The argument repeats, with sometimes minor changes, to establish all the remaining claims.

**Theorem 0.3** (Obstruction theorem II)**.** Let $a, b, c$ and $d$ be variables representing distinct letters. Then $(a, b) \mapsto (c, d)$ and $(c, d) \mapsto (a, b)$ are both substitutions if and only if $(a, b) \mapsto (c, d)$ is arranged in square. The theorem also holds if $(c, d) \mapsto (a, b)$ is replaced by $(d, c) \mapsto (b, a)$.

**Theorem 0.4** (Obstruction theorem III)**.** If the substitution $(a, b) \mapsto (c, d)$ contains only three distinct letters, then either $a = d$ or $b = c$ and the arrangement is either in a row or on a column, in which the letters appear consecutively.

**Theorem 0.5** (Obstruction theorem IV)**.** Let $a, b, c, d, e$ and $f$ be variables representing distinct letters. If $(a, b) \mapsto (c, d)$ and $(a, e) \mapsto (f, c)$ are both substitutions then either $(a, b) \mapsto (c, d)$ is arranged in a square and $(a, e) \mapsto (f, c)$ is arranged in a row or $(a, e) \mapsto (f, c)$ is arranged in a square and $(a, b) \mapsto (c, d)$ is arranged in a column.

## Application to a challenge cipher

We look at an example given by Jim Gillogly,

> This message was received by an intercept station in Scotland. The frequency and format indicate that it is a most urgent message from one of our agents who landed a week ago in Norway. His controllers have been unable to read it. Although it clearly uses his backup cipher, the Playfair, the keys assigned to him do not work. We cannot reach him before his normal scheduled transmission in two weeks, so we urgently request that you attempt to decrypt this and let us know the contents. In case it helps, he is carrying materials to assist a previously dropped team in their work regarding the Norsk Hydro facility at Rjukan. His recognition code should appear in the message: It is "beware ice weasels." If he is operating under duress, he should include the phrase "red penguin frenzy." He will use "STOP" between sentences and "END" at the end.

See `http://www.pbs.org/wgbh/nova/decoding/faceoff.html.` This is the ciphertext,

```
VY TE SY ED LU TE RV LF
NV UH DW AR DL CF FB SD
EW NP XK IC FT RE OL KA
LZ YL SL TO BK EV LY AR
MK RB OD NA LD YP LA ET
OL QA DF HS FZ WN AI DS
MU RU OL HR YL LO TW FY
LD IC VL US VS SF ZY LU
NF FX LK TG BC DO BF AL
EW RP FY WL HU LD AR LI
TF LA BF FZ CY FU UF BG
```

We first discover the placement of the crib. With luck, only one among the recognition code or the distress signal will have any potential of matching and only in one particular location — else we will have to proceed under several tentative suppositions as to the true match. We must consider several cases: that the crib is broken on either odd or even letters; that the crib is in the middle of the text surrounded by `STOP`; that the crib is at the beginning of the text; that the crib is at the end of the text followed by either `END` or `ENDX`.

There are some methods for finding the crib. Since there is no sequence `xy zs` appearing twice in the cipher text, where the small letters are variables, then `STOP` is broken as `S TO P`, and so we can look for a pair `xy` separated by 10 digrams. For instance,

```
ET OL QA DF HS FZ WN AI DS MU RU OL HR
?s to pr ed pe ng ui nf re nz ys to p?
```

We get lucky and this is the only possibility. This gives us a collection of substitution pairs to begin our reconstruction of the Playfair square. The Obstruction Theorems give the following information: `FZ->NG` is arranged as a square; `MU->NZ` is arranged in a column; and `AI->NF` is arranged in a row. Fitting this together, and placing `F` in the upper left hand corner of the square (which we can do since a cyclic permutation of rows or columns does not affect the cipher),

```
F I * N A
      M
      *
G * * Z
    * U
```

The Obstruction Theorems also identify `DF->ED` as arranged as a row or column, and given the partially filled square, it must be a column with `D` preceding `F` in its column position. This determines that there is no row between the `M` and `Z`. We also use substitution `WN->UI` to place `W`,

```
F I * N A
      M
G   * Z
E W * U
D
```

Consider `RU->YS`. There isn't room in the row or column containing `U` for this substitution to be arranged as a row or column, so it is arranged as a square. This means that `Y` is in the same column as `U`, in the only open row. `R` is in the same row as `Y`. Considering `QA->PR`, if this is arranged in a square then `R` is in the same row as `A`, not the same row as `Y`. So `QA->PR` is not arranged in a square, it is arranged as a column, and so `R` is in the same column as `A` We also use substitution `HS->PE` to place `H`,

```
F I * N A
H   * M P
G   * Z Q
E W * U S
D   * Y R
```

The obstruction theorems identify `OL-TO` as arranged in a row or column, and considering the space left in the partially filled square, it must be a arranged in a column. The pair `ET-?S` aligns `T` with `S`. Recognizing the word `FINAL`, we place it in the rightmost column,

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

At this point we might notice the alphabetic spiral and attempt to complete the pattern. Because filling the table always uses an ascending sequence of letters, and only `K` remains unplaced between `H` and `M`, and only `X` between `W` and `Z`, they are placed,

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

A similar reasoning concludes that `V` must be in the final column, and therefore one of `B` or `C` is in the final column. Of the four possible placements, we recognize the word `victory`, and this completes the square,

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

## Conclusion

This note showed the cryptanalysis of a Playfair cipher. It used the method of a crib to get a possible sampling of the digraph substitutions. To those were applied Obstruction Theorems, which sorted out the type of a substitution — whether arranged in a square, or a column or row.

Various other logical arguments were made to further place letters, especially once it was recognized that the keyword was placed as a spiral with the remaining letters not used in the keyword placed consecutively in the spiral pattern.

It remains to consider how this process can be automated. In any case, it is probably possible to break a Playfair by brute force, as the 25! placements of letters has complexity approximated by,

$$\begin{aligned}
\log_2(25!) &= \sum_{i=1}^{25} \log_2(i) \\
&\leq 9 \log_2(32) + 8 \log_2(16) + 4 \log_2(8) + 2 \log_2(4) + \log_2(2) \\
&= 45 + 32 + 12 + 4 + 1 = 94 \text{ bits.}
\end{aligned}$$