**PLAYFAIR CRYPTANALYSIS**

Our preliminary step is to perform individual letter
frequency and digraphic counts.  The former because high
frequency ciphertext letters follow closely the high
frequency letters they represent and will be located in
the upper rows; similarly, low frequency letters follow
their plain counterparts (UVWXYZ) and may be located at
the last row of the square.  A digraph count is useful
because cipher digraphs follow closely the frequency of
their plaintext digraphs. i.e. TH = HM. The frequency of
HM must be high for a normal length message. Also
tetragraphs may be tested THAT, TION, THIS for
corresponding their frequencies in the square.

All the authors agree that a probable word is need for
entry into the Playfair. Due to its inherent
characteristics, Playfair cipher words will follow the
same pattern as their plaintext equivalents; they carry
their pattern into the cipher.

Given:   Tip "er one day entere"     Hampian. 10/1952

EU  SM  FV  DO  VC  PB  FC  GX  DZ  SQ  DY  BA  AQ  OB
ZD  AC  OC  ZD  ZC  UQ  HA  FK  MH  KC  WD  QC  MH  DZ
BF  NT  BP  OF  HA  SI  KE  QA  KA  NH  EC  WN  HT  CX
SU  HZ  CS  RF  QS  CX  DB  SF  SI  KE  FP  (106)

We set up a combined frequency tally with letters to the
right and left of the reference letter shown:

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

This particular message has no significant repeats.

Cipher  GX  DZ  SQ  DY  BA  AQ  OB  ZD  AC
Plain   ..  ER  ON  ED  AY  EN  TE  RE  ..

Note the first and last pair reversal.

It is necessary to take each set of these pair
equalities and establish the position of the four
letters with respect to each other. They must conform to
the above three rules for row, column, and rectangle.

The six different sets of pairs of know equalities are
set up:

   1          2          3         4        5
er = DZ    on = SQ    ed = DY   ay = BA   en = AQ
------     -------    ------    -------   -------
E D R Z    O S N Q    E D Y     Y A B     E A N Q
D          S          D         A         A
R   E D    N   O S    Y         B         N   E A
Z   Z R    Q   Q N                        Q   Q N

   6
te = OB
-------
T O E B
O
E   T O
B   B E

The three possible relations of the letters are labeled
Vertical (v), Horizontal (h), Diagonal (d).  Our object
is to combine the letters in each of the set of pairs.

Combine 1 and 3:  E R D Z Y

      1/v - 3/v        1/h - 3/h        1/d - 3/h
      ---------        ---------        ---------
          E            E D Y R Z          E D Y
          D                               Z R
          Y
          R
          Z

Combine 2 and 5: O N S Q E A

      2/h - 5/d        2/d - 5/h         2/d - 5/d
      ---------        ---------         ---------
       O S N Q          E A N Q             S O
           A E              S O             N Q
                                            A E

Note that all the equalities hold for all letters.

Set number 6 combines only with the last combination: T
E O B N S Q A

 2/d - 5/d - 6/v                2/d - 5/d - 6/d
----------------                ---------------
         T                          S O T
       S O                          N Q
       A E                          A E B
         B
       N Q

which we now combine with 4:

                 2/d - 5/d - 6/d - 4/h
                 ---------------------
                     S T O
                   Y A E B           (rearranged and
                     N   Q            equalities hold)

only one combination of 1 and 3 will combine with the
above: S T O Y A B E D N Q Z R

             1/d - 2/d - 3/h - 4/h - 5/d - 6/d
             ---------------------------------
                     S T O
                   Y A E B D
                     N   Q
                         Z R

Arranged in a 5 X 5 square:

                    . . S T O
                    D Y A B E
                    . . . . .
                    . . N . Q
                    R . . . Z

We see that O is in the keyword, the sequence NPQ
exists, the letters S T Y are in the keyword, and three
of the letters U V W X  are in needed to fill the bottom
row.

                   ----------
                    . . S T O| C
                    D Y A B E|
                    . . . . .|
                    . . N P Q|
                    R . . . Z| U V W X

With the exception of F G H I K L M which must in order
fill up the 3rd and 4th rows, the enciphering square is
found as:

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

Our plaintext message starts off: YOUNG RECRUIT DRIVER
ONE DAY ENTERED STORE ROOM ....

  Written by Alex Biryukov (Weizmann Institute of Science, Rehovot, Israel) in 2001 for a course taught there entitled Methods of Cryptanalysis

**Lecture 3**
"Cryptanalysis of the Classical Ciphers"

A quick look forward for those, who want some reading before the lecture.
Here are the [lecture notes](http://www.wisdom.weizmann.ac.il/~albi/cryptanalysis/lecture3.ps.gz)  (ps, gzipped) written by Ilya Safro.
(Print with 600 or 1200 dpi to get better quality:  lpr -P12laser11 lecture3.ps)
The 'after the lecture' notes are written in *light green italic.*

We will concentrate on the cryptanalysis of the classic schemes that we have described.
(see [LANAKI's course](http://www.fortunecity.com/skyscraper/coding/379/lesson1.htm), lectures 1-4, 10-12, or the [Army Field Manual](http://www.und.nodak.edu/org/crypto/crypto/army.field.manual/), here is its
[table of contents](http://www.wisdom.weizmann.ac.il/~albi/cryptanalysis/TOC.ps)). See also [extended lecture notes for lecture 1](http://www.wisdom.weizmann.ac.il/~albi/cryptanalysis/lecture1.ps) (sections 1.1, 1.2) for
a classification of cryptanalytic attacks, and a sketch on methods of cryptanalysis.

We will try to cover the following attack methods
*[we used D.Stinson's "Cryptography: Theory and Practice" book, pp.31-34, for the first two topics]:*

1. Frequency analysis, Index of Coincidence *[Chapter 2 of the Army Field Manual]*
2. Kasiski's method (for example, for Carroll's Vigenere)
3. Anagramming (for arbitrary transposition ciphers)
4. Probable word method *(*[*Rosette stone*](http://www.wisdom.weizmann.ac.il/~albi/cryptanalysis/) *is an interesting historic example)*
5. Vowel - consonants splitting *[see Chapter 4 of the Army Field manual]*
6. Decimation
7. Improbable word (for multi-letteral ciphers, *this is the way you solve puzzle 3 of Hw1)*

Meanwhile enjoy the following story (taken from LANAKI's course lecture 17,
historic part of which is in turn taken from [Khan's book](http://www.wisdom.weizmann.ac.il/~albi/cryptanalysis/recread.htm).) Interestingly, here
is the [same story](http://www.history.navy.mil/faqs/faq60-2.htm) from a totally different angle.

**DIGRAPHIC CIPHERS: PLAYFAIR**

Perhaps the most famous cipher of 1943 involved the
**future president of U.S.**, **J. F. Kennedy,** Jr. [KAHN]
On 2 August 1943, Australian Coastwatcher Lieutenant
Arthur Reginald Evans of the Royal Australian Naval
Volunteer Reserve saw a pinpoint of flame on the dark
waters of Blackett Strait from his jungle ridge on
[Kolombangara Island](http://www.wisdom.weizmann.ac.il/~albi/cryptanalysis/map.jpg), one of the Solomons. He did not
know that the Japanese destroyer Amagiri had rammed and
sliced in half an American patrol boat PT-109, under
the command of Lieutenant John F. Kennedy, United States
Naval Reserve.  Evans received the following message at
0930 on the morning of the 2 of August 1943:

      29gps

     KXJEY  UREBE  ZWEHE  WRYTU  HEYFS
     KREHE  GOYFI  WTTTU  OLKSY  CAJPO
     BOTEI  ZONTX  BYBWT  GONEY  CUZWR
     GDSON  SXBOU  YWRHE  BAAHY  USEDQ

                                  /0930/2

Translation:

     PT BOAT ONE OWE NINE LOST IN ACTION IN BLACKETT
     STRAIT TWO MILES SW MERESU COVE X CREW OF TWELVE
     X REQUEST ANY INFORMATION.

The coastwatchers regularly used the Playfair system.
Evans deciphered it with the key ROYAL NEW ZEALAND NAVY
and learned of Kennedy's fate. Evans reported back to
the coastwatcher near Munda, call sign PWD, that Object
still floating between Merusu and Gizo, and at 1:12 pm,
Evans was told by Coastwatcher KEN on Guadalcanal that
there was a possibility of survivors landing either on
Vangavanga or near islands.  That is what Kennedy and
his crew had done. They had swum to Plum Pudding Island
on the Southeastern tip of Gizo Island.

Several messages passed between PWD, KEN and GSE
(Evans). The Japanese made no attempt to capture Kennedy
even though they had access to the various messages. The
importance to the crew was missed even though many P-40's
could have been spotted in the Search and Rescue (SAR) attempt.
Maybe the Japanese didn't want to waste the time or men
because the exact location of the crew was not
specified. A Japanese barge chugged past Kennedy's
hideout. On 09:20 a.m. on Saturday morning 7 August 1943,
two natives found the sailors, who had moved to Gross
Island, and had reported the find to Evans. He wrote a brief
message: *Eleven survivors PT boat on Gross Is X Have*
*sent food and letter advising senior come here without*
*delay X Warn aviation of canoes crossing Ferguson RE.*
The square Evans used was based on the key PHYSICAL
EXAMINATION :
               P H Y S I
               C A L E X
               M N T O B
               D F G K Q
               R U V W Z

The encipherment did not split the doubled letters (*Gross* and
*crossing*) as is the rule:

     XELWA  OHWUW  YZMWI  HOMNE  OBTFW
     MSSPI  AJLUO  EAONG  OOFCM  FEXTT
     CWCFZ  YIPTF  EOBHM  WEMOC  SAWCZ
     SNYNW  MGXEL  HEZCU  FNZYL  NSBTB
     DANFK  OPEWM  SSHBK  GCWFV  EKMUE

**A message of this length alone suffices for the solution of**
**Playfair.**There were four more messages in the same key,
including one of 335 letters, beginning:

XYAWO  GAOOA  GPEMO  HPQCW  IPNLG  RPIXL
TXLOA  NNYCS  YXBOY  MNBIN  YOBTY  QYNAI ...,

for

*Lieut. Kennedy considers it advisable that he pilot PT*
*boat tonight X ...*

These five messages detailed the rescue arrangements, which
offered the Japanese a chance to not only to get the crew (and
change all history!) but also the force coming out to save it.
**All of the messages could have been solved within an hour by**
**even a moderately experienced cryptanalyst.**Yet some ten hours
later, at 10:00 p.m. Kennedy and his crew was rescued.

Digraphic substitution refers to the use of pairs of
letters to substitute for other pairs of letters. The
Playfair system was originated by the noted British
scientist, Sir Charles Wheatstone (1802 - 1875) but, as
far as known, it was not employed for military  or
diplomatic use during his lifetime. About 1890 it was
adopted for use by the British Foreign Office on the
recommendation of Lord Lyon Playfair (1818-1898) and
thereafter by mistake identified with its sponsor.

**Encipherment**

The Playfair is based on a 25 letter alphabet (omit J)
set up in a 5 X 5 square.  A keyword is written in
horizontally into the top rows of the square and the
remaining letters follow in regular order.  So for the
key = LOGARITHM, we have:
                L O G A R
               I T H M B
               C D E F K
               N P Q S U
               V W X Y Z

In preparation for encipherment, the plaintext is
separated into pairs. Doubled letters such as SS or NN
are separated by a null.

For example, "COME QUICKLY WE NEED HELP"  we have

       CO ME QU IC KL YW EN EX ED HE LP

There are three rules governing encipherment:

1.   When the two letters of a plain text pair are in
     the same column of the square, each is enciphered
     by the letter directly below it in that column. The
     letter at the bottom is enciphered by the letter at
     the top of the same column.

                  Plain       Cipher
                   OP           TW
                   IC           CN
                   EX           QG

2.   When the two letters of a plain text pair are in
     the same row of the square, each is enciphered by
     the letter directly to its right in that row.  The
     letter at the extreme right of the row is enciph-
     ered by the letter at the extreme left of the same
     row.
                       Plain       Cipher
                      YW           ZX
                      ED           FE
                      QU           SN

3.   When two letters are located in different rows and
     columns, they are enciphered by the two letters
     which form a rectangle with them, beginning with
     the letter in the SAME ROW with the first letter of
     the plaintext pair. (This occurs about 2/3 of the
     time.)

                     Plain       Cipher
                      CO           DL
                      ME           HF
                      KL           CR
                      LP           ON

Decipherment, when the keyword is known, is accomplished
by using the rules in reverse.

**Identification Of The Playfair**

The following features apply to the Playfair:

1. It is a substitution cipher.

2. The cipher message contains an even number of
   letters.

3. A frequency count will show no more than 25 letters.
   (The letter J is not found.)

4. If long repeats occur, they will be at regular (even)
   intervals.  In most cases, repeated sequences will be
   an even number of letters.

5. Many reversals of digraphs.

**Peculiarities**

1. No plaintext letter can be represented in the cipher
   by itself.

2. Any given letter can be represented by 5 other
   letters.

3. Any given letter can represent 5 other letters.

4. Any given letter cannot represent a letter that it
   combines with diagonally.

5. It is twice as probable that the two letters of any
   pair are at the corners of a rectangle, than as in
   the same row or column.

6. When a cipher letter has once been identified as a
   substitute for a plaintext letter, their is a 20%
   chance that it represents the same plaintext letter
   in each other appearance.

The goal of recovery of the 5 X 5 square and various
techniques for accomplishing this are the focus for
solving the Playfair.   Colonel Parker Hitt describes
Lieutenant Frank Moorman's approach to solving the
Playfair which addresses the keyword recovery logically.
[HITT].  Other writers [ELCY], [BOW2], [FRE4], and
[MAST] do an admirable job of discussing the process.
However, W. M. Bowers Volume I on Digraphic Substitution
presents the easiest protocol for students. [BOWE]

**PLAYFAIR CRYPTANALYSIS**

Our preliminary step is to perform individual letter
frequency and digraphic counts.  The former because high
frequency ciphertext letters follow closely the high
frequency letters they represent and will be located in
the upper rows; similarly, low frequency letters follow
their plain counterparts (UVWXYZ) and may be located at
the last row of the square.  A digraph count is useful
because cipher digraphs follow closely the frequency of
their plaintext digraphs. i.e. TH = HM. The frequency of
HM must be high for a normal length message. Also
tetragraphs may be tested THAT, TION, THIS for
corresponding their frequencies in the square.

All the authors agree that a probable word is need for
entry into the Playfair. Due to its inherent
characteristics, Playfair cipher words will follow the
same pattern as their plaintext equivalents; they carry
their pattern into the cipher.

Given:   Tip "er one day entere"     Hampian. 10/1952

EU  SM  FV  DO  VC  PB  FC  GX  DZ  SQ  DY  BA  AQ  OB
ZD  AC  OC  ZD  ZC  UQ  HA  FK  MH  KC  WD  QC  MH  DZ
BF  NT  BP  OF  HA  SI  KE  QA  KA  NH  EC  WN  HT  CX
SU  HZ  CS  RF  QS  CX  DB  SF  SI  KE  FP  (106)

We set up a combined frequency tally with letters to the
right and left of the reference letter shown:

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

This particular message has no significant repeats.

Cipher  GX  DZ  SQ  DY  BA  AQ  OB  ZD  AC
Plain   ..  ER  ON  ED  AY  EN  TE  RE  ..

Note the first and last pair reversal.

It is necessary to take each set of these pair
equalities and establish the position of the four
letters with respect to each other. They must conform to
the above three rules for row, column, and rectangle.

The six different sets of pairs of know equalities are
set up:

   1          2          3         4        5
er = DZ    on = SQ    ed = DY   ay = BA   en = AQ
------     -------    ------    -------   -------
E D R Z    O S N Q    E D Y     Y A B     E A N Q
D          S          D         A         A
R   E D    N   O S    Y         B         N   E A
Z   Z R    Q   Q N                        Q   Q N

   6
te = OB
-------
T O E B
O
E   T O
B   B E

The three possible relations of the letters are labeled
Vertical (v), Horizontal (h), Diagonal (d).  Our object
is to combine the letters in each of the set of pairs.

Combine 1 and 3:  E R D Z Y

      1/v - 3/v        1/h - 3/h        1/d - 3/h
      ---------        ---------        ---------
          E            E D Y R Z          E D Y
          D                               Z R
          Y
          R
          Z

Combine 2 and 5: O N S Q E A

      2/h - 5/d        2/d - 5/h         2/d - 5/d
      ---------        ---------         ---------
       O S N Q          E A N Q             S O
           A E              S O             N Q
                                            A E

Note that all the equalities hold for all letters.

Set number 6 combines only with the last combination: T
E O B N S Q A

 2/d - 5/d - 6/v                2/d - 5/d - 6/d
----------------                ---------------
         T                          S O T
       S O                          N Q
       A E                          A E B
         B
       N Q

which we now combine with 4:

                 2/d - 5/d - 6/d - 4/h
                 ---------------------
                     S T O
                   Y A E B           (rearranged and
                     N   Q            equalities hold)

only one combination of 1 and 3 will combine with the
above: S T O Y A B E D N Q Z R

             1/d - 2/d - 3/h - 4/h - 5/d - 6/d
             ---------------------------------
                     S T O
                   Y A E B D
                     N   Q
                         Z R

Arranged in a 5 X 5 square:

                    . . S T O
                    D Y A B E
                    . . . . .
                    . . N . Q
                    R . . . Z

We see that O is in the keyword, the sequence NPQ
exists, the letters S T Y are in the keyword, and three
of the letters U V W X  are in needed to fill the bottom
row.

                   ----------
                    . . S T O| C
                    D Y A B E|
                    . . . . .|
                    . . N P Q|
                    R . . . Z| U V W X

With the exception of F G H I K L M which must in order
fill up the 3rd and 4th rows, the enciphering square is
found as:

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

Our plaintext message starts off: YOUNG RECRUIT DRIVER
ONE DAY ENTERED STORE ROOM ....

**SERIATED PLAYFAIR**

Perhaps the best known variation of the Playfair system,
and one which adds greatly to its security, is called
the Seriated Playfair.

The plain text is written horizontally in two line
periodic groups as shown below in period six

       C O M E Q U    E N E E D H    M E D I A T
       I C K L Y W   (X)E L P I M    E L Y T O M

The vertical pairs are formed and enciphered by the
regular Playfair rules. Based on the keyword LOGARITHM,
the above message is enciphered:

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

we take the ciphertext off horizontally by the same
route by which the plain text was written in for
encipherment:

NLBCS  PCDFG  XZQQC  DCMGC  GQTBH  CFTRH  FGWHG  B.

**Solution of Seriated Playfair:**

We assume a period of 4 - 10 which fits most of the
cases encountered.  Of prime importance is determination
of the period. We test the various periods and eliminate
any test where we find a vertical pair consisting of two
appearances of the same letter.

If the message enciphered above is tested this way, in
all periods from 4 - 10, it will be found that period 6
is correct. All others will show a doubled vertical
pair.

Charles A. Leonard [PLAf] detailed a method to determine
impossible periods mathematically:

              S2
           -------    = Q & R
           S2 - S1

 where: S2 - S1 = Period, Q = quotient, R = remainder

 Substituting known S values in this formula and solving
for Q and R, a doubled vertical pair will occur in
period S2 - S1 in the following cases:

     1.  When Q is an odd number and R is greater than
         zero;
     2.  When Q is an even number and R is zero.

Cipher letter position numbers in our message are:

 A   B   C   D   E   F   G   H   I   K   L    etc.
     3   4   8       9  10  25           2
    24   7  16      27  19  30
    36  15          31  21  34
        17              32
        20              35
        26

Period  Letter  S2 - S1     Q   R   Result
  4       F     31 - 27     7   3   Eliminated-Case 1
  5       C     20 - 15     4   0   Case 2
  6       C     26 - 20     4   2   possible
  7       H     34 - 30             Eliminate-last gp
  8       D     16 - 8      2   0   Case 2
  9       C     26 - 17     2   8   possible
          G     19 - 10     2   1   possible
          H     34 - 25     3   7   Case 1
  10      C     17 - 7      1   7   Case 1

When a periodic group S2 - S1 does not occur in message
the last group is inspected. If it is shorter than the
regular groups of the period being tested, a double
vertical pair may show at S2- S1 value equal to the
length of this final group. If so, eliminate.

The mono and digraphic frequency counts are made.
Plaintext high frequency digraphs and tetragraphs do not
carry their identity over into the cipher and are not
recognizable. Entry must be made with a probable word.
Patterns do carry over to the two line groups and will
repeat.

The placing of the probable word is important. Given a
cipher text slice with period 6 found using the Leonard
procedure:

   HKILVP   PBVBAA   BHRPOU   TBITFE   UCEVZK
   RNFTZU   HZWVFR   UDTKBD   UIBYNS   EXBZAR

and the probable phrase "is destined to", the word
destined could be in any of the following positions when
enciphered in period 6:

DESTIN  .DESTI   ..DEST   ...DES    ....DE
ED....  NED...   INED..   TINED.    STINED

The DE = ED reversal in all arrangements is noted and
found in the cipher text portion:

            BHRPOU   TBITFE   UCEVZK
            UDTKBD   UIBYNS   EXBZAR
                     .desti
                     ned..

adding the additional information:

            BHRPOU   TBITFE   UCEVZK
            UDTKBD   UIBYNS   EXBZAR
                 .   sdesti
                 i   nedto.

we develop several equations:

                   ed = IB
 -I = UD, sn = TU, de = BI, ST = TY, to =FN, I- =ES

these translate to the following equalities:

   1          2          3         4        5
SN = TU    DE = BI    ST = TY   TO = FN   I- = ES
-------    -------    ------    -------   -------
S T N U    D B E I    S T Y     T F O N   I E - S
T          B          T         F         E
N   S T    E   D B    Y         O   T F   -   I E
U   U N    I   I E              N   N O   S   S -

   6         7
-I = UD    ED = IB
-------    -------
- U I D    E I D B
U          I
I   - U    D   E I
D   D I    B   B D

After some work (and with some assumptions to be tested
we develop a tentative square for the system:

              1/d-2/d -3/h-4/v- 5/h -6/h
              --------------------------
                      -
                    O U N
                      I E
                      D B
                    F S T Y

check:
TO=FN+   + = yes
SN=TU+
ST=TY+                        letters left: A C E G H K
I-=ES -=t  IT =ES                           L M P Q R V
DE=BI+                                      W X Z
ED=IB+
-I=UD+

from here we need to expand on the cipher text or choose
another probable word.