Methods for the Reconstruction of Primary Alphabets

William F. Friedman and Elizabeth S. Friedman

Publication No. 21

RIVERBANK LABORATORIES
DEPARTMENT OF CIPHERS
RIVERBANK
GENEVA, ILL.
1918

INTRODUCTORY NOTE

It is not our intention in this brochure to describe any newly-discovered methods of cipher solution, or indeed, to make a detailed analysis of even any one system. We do not claim any remarkable achievement in putting forth the few principles herein described. They are meant rather as a stimulant to the more advanced student of deciphering. Therefore no attempt has been made to make any exhaustive analysis of different systems, or of varying methods of using the same system. The methods here given are issued primarily as an outline or suggestion to the cipher student who is more or less familiar with complicated systems, and who therefore will be quick to see the application of the present principles to any variations of known methods. For him who wishes to go farther into the subject, these suggestions will be found to yield a wealth of possibilities for research, which would need volumes to describe.

KEY-WORD ALPHABETS

In Riverbank Publication No. 15 a method was shown for reconstructing a Primary Alphabet from any one of the Secondary Alphabets. In that monograph only Key-Word Alphabets were considered. It is our purpose in this pamphlet to deal not only with Key-Word Alphabets, but with Arbitrarily- and Random-mixed Alphabets as well.

Let us consider the first of the examples at the end of Publication 15. We are given the deciphering alphabet:

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
QMTAZSCUXLIWPYNEBDRFVGHJKO
```

You will note that XY of the upper alphabet represents JK of the lower. Each being a group of two infrequent consonants, we shall assume as a starting point that the letters constituting the two pairs are sequent in the Primary Alphabet. Then, taking the values by twos as they result from the preceding groups, we should have, with XY equalling JK, JK equalling LI and so on:

```
XY    TV    PQ
JK    FG    EB
LI    SC    ZM
WX    RT    OP
HJ    DF    NE
UL    AS    YZ
VW    QR    KO
GH    BD    IN
CU    MA    XY
```

Having reached the starting point, we conclude this part of the operation and begin to build the alphabet proper from the pairs thus obtained. Beginning with the first pair, XY, which has Y for its second letter, we follow it with that pair which has Y for its first letter, which is YZ; then YZ is followed by that group which has Z for its first letter, which is ZM. This process is continued until all the pairs have been used. Thus:

```
XYZMASCULINEBDFGHJKOPQRTVWX
```

The reconstruction of the Primary Alphabet from a single secondary alphabet, as in the foregoing example, is possible only where the system from which the deciphering or secondary alphabet came is a Primary Alphabet System in which the two components are identical. (See Riverbank Publication No. 18, Table V, 4a.) It is possible, of course, to have a Primary Alphabet System in which the components are not identical, in which case it is impossible to recover the Primary Alphabets from a single one of the secondaries.

However, with the preceding method in mind, let us consider a case in which we shall deal with two deciphering or secondary alphabets.

```
  I   ABCDEFGHIJKLMNOPQRSTUVWXYZ
 II   ZWABOCFYGHIKJLMENPDRQSTUVX
III   PUKEVYWTORDQABCSFGZXHIJLMN
```

In the preceeding example twenty-six pairs resulted. In this example it will be found that fifty-two pairs will result.

Choosing a pair of sequent letters in Alphabet I, follow it by its equivalent pair in Alphabet II: then finding the second pair in Alphabet III, which letters are not recorded, take their equivalents in Alphabet I. Repeat this cycle until the starting point is reached.

```
1) VW
       ST (1
   PH
       EY
   DF
       BC
   NO
       LM
3) XY
       UV (3
   BE
       WO
   GI
       FG
   QR
       NP
5) ZA
       XZ (5
   TS
       RD
```

Here the alternate groups have been placed to the right, forming two columns of pairs. It will be found that the pairs in the left column comprise the letters of one Primary Alphabet, those on the right the other Primary Alphabet. As when dealing with a single alphabet, take as a starting point any group and follow it by the pair which has for its second letter that which was the first in the former pair. The column on the left, then, will unite in this manner:

```
123456, etc.
VWXYZALPHBETSCDFGIJKMNOQRU
```

The right column forms the remaining Primary Alphabet, taking its pairs in the same order and positions.

```
123456, etc.
STUVXZKEYWORDABCFGHIJLMNPQ
```

Having recoverd the Primary Alphabets, whose key words are KEYWORDALPH(A)BETS, it is ascertained that the secondary alphabets used in their construction were the second and ninth deciphering alphabets — that is, of the twenty-six possible deciphering alphabets to be used, one even-numbered and one odd-numbered alphabet brought the results shown. If two even-numbered or two odd-numbered alphabets had been taken, the cycle would have been concluded with twenty-six pairs instead of fifty-two. Let us note the results when two odd-numbered alphabets are used. The following are the first and seventeenth deciphering alphabets:

```
  I   ABCDEFGHIJKLMNOPQRSTUVWXYZ
 II   KOBCRFGWHIJELMNYPQADSTUVXZ
III   FJPQLSTIUVXGZKEHYWNMORDABC
```

Here the cycle is completed after twenty-six pairs. Hence it is not possible, as in the preceding example, to follow pair by pair with the last and first letters of the succeeding groups comprising the indicators. But nevertheless, when dealing with Keyword Alphabets, it is a fairly simple matter to build the Primary Alphabet. In this case, for instance, starting with VW in the left column, one would naturally look for a pair to follow it, the letters of which would most likely be found at the end of the alphabet, such as XY, XZ, or YZ. Following down the left column, the fifth pair from VW is seen to be XY. Searching then for a Z, which would very probably follow Y, it is seen to be the first letter of the fifth group from XY. This gives a clue, and it is found that by taking every fifth pair the Primary Alphabet is completed. Exactly as before, the other Primary Alphabet will result from the column on the right, by taking pairs in the same order — namely every fifth group.

Now it is rarely, if ever, possible to tell with which of the deciphering alphabets one is dealing, until the Primary Alphabets are known. Hence, the secondary alphabets being derived from messages deciphered, it may happen that the first two tried would not at once yield results. Another trial should be made in such a case, with a different pair of alphabets. Any two alphabets whose interval is 13, such as alphabets 9 and 22, 2 and 15, and the like, will be found incapable of yielding to the foregoing method of reconstruction. However, this contingent need hardly be considered, for it would be a rare case indeed where not more than two alphabets could be found whose interval was other than 13.

VW	TU
GI	GH
LP	EY
OQ	NP
SC	AB
XY	VX
JK	IJ
HB	WO
RU	QS
DF	CF
ZA	ZK
MN	LM
ET	RD
VW	

ARBITRARILY-MIXED ALPHABETS

We shall first discuss the reconstruction of a single arbitrarily-mixed alphabet, i.e., a Primary Alphabet System where the components are identical. By an arbitrarily-mixed alphabet, is meant one which is made up according to some pre-arranged plan, and yet which presents the appearance of being mixed at random[1].

Such an alphabet would be, for example, using the key word DEMOCRA(C)Y:

```
3 4 5 6 2 7 1 8
D E M O C R A Y
B F G H I J K L
N P Q S T U V W
X Z
```

```
AKVCITDBNXEFPZMGQOHSRJUYLW
```

Or taking the columns after the manner of an alternate vertical transposition cipher:

```
DBNXZPFEMGQSHOCITUJRAKVWLY
```

Once aware of any such system of forming an alphabet, it is comparatively easy to rebuild the generating rectangle. Take, for instance, the first of the two alphabets above. It is advisable here, as with key-word alphabets, to make the attack upon the XYZ part of the alphabet. Note here that X and Z are found four intervals apart, and that the third letters preceding are D and E respectively, themselves four intervals apart. This would lead one to place them in a possible rectangle thus:

```
D E
B F
N P
X Z
```

If this is correct, it at once shows that DE is a part of the key word, and that C, O and Y are also in the key word. Returning to the alphabet, we look for Q, R, or S to follow P, as this seems to be the simplest method of procedure. The three letters after Z in the alphabet are MGQ, which, if they are made the adjoining column on the right, will bring G and Q in their correct alphabetical sequence. Following MGQ one finds OHS which will place H after G, and S after Q, signifying that R is in the key word. Now we have

```
D E M O
B F G H
N P Q S
X Z
```

At once we guess that the key word is DEMOCRACY or DEMOCRATIC, and a trial quickly proves the former.

In the second alphabet derived from this rectangle, or any alphabet of the same nature, the juxtaposition of such letters as XZ, QS, TU, and VW, should reveal at once the alternate vertical method of transposition.

To proceed, then, let us take the following deciphering alphabet, derived from a Primary Alphabet System in which the two components are identical.

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
NSUFABHCPEDOGRJWTYXKIMLZQV
```

[1] See Riverbank Publication No. 17, pages 22 and 23; also Gioppi, *La Crittografia*, page 54, and De Viaris, *L'art de Dechiffrer*, page 126.

In the case of an Arbitrarily-Mixed Alphabet, the method of formation precludes the possibility of dealing directly with pairs of letters. But let us begin with any letter in the direct alphabet, and form a sequence, such as A - N - R - Y, etc. By writing this sequence of equivalents on two strips of cross-section paper, it is made a simple matter to shift the strips until the correct pairs are found. With the sequence started above written on two strips and set in the first possible position, that is, shifted one space from identity of equivalents, they appear thus:

```
ANRYQTKDFBSXZVMGHCUIPWLOJEA
 ANRYQTKDFBSXZVMGHCUIPWLOJE
```

Now if this is the correct position, we should be able to obtain from here a sequence of equivalents which will build into a generating rectangle. Beginning then, we have NAEJOLWPIUC, etc., which from an inspection for symmetry in a possible rectangle, seems very improbable. Hence we try the next position. Here we get RAJLPUHMZSF, etc., which is likewise unpromising. Of course, it is not always necessary to record the sequence of equivalents resulting from any position. Usually a simple rapid inspection will reveal the possibilities of construction.

Finally we have the strips in this position:

```
ANRYQTKDFBSXZVMGHCUIPWLOJEANRYQTKDFBS
           ANRYQTKDFBSXZVMGHCUIPWLOJE
```

Here the resulting sequence is XAGQIFOZNHTPBJVRCKWSEMYUDLX. Immediately it becomes apparent that V, W, Y, and X are four intervals apart. If we assume, then, that these are the final letters in columns of four letters each, we have:

```
P R U S
B C D E
J K L M
V W X Y
```

Searching for Z, we find that the letters preceding it adjoin those above in alphabetical sequence, and also indicate that N is in the key word. Thus:

```
P R U S I
B C D E F
J K L M O
V W X Y Z
```

Now looking for G which may follow F, we find AGQ, which is checked for the succeeding position on the right by having Q follow O, since P has already been placed in the key word. Similarly NHT checks as the remaining column, with H to succeed G, and T following Q, and R and S placed in the first line of the rectangle. The rectangle is now complete, the key word being PRUS(S)IAN.

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

In a Primary Alphabet System where the components are not identical, not one, but two alphabets are to be reconstructed. As was shown in the case of Key-Word Alphabets, it will be found that the number of letters in the sequence of equivalents resulting from any two deciphering alphabets will vary — when two even-numbered or two odd-numbered alphabets are used, the sequence will end with 26 letters; when one even- and one odd-numbered alphabet are used, the sequence will yield 52 letters. But whereas in the case of Key-Word Alphabets it was comparatively easy to recover the Primary Alphabet with only 26 pairs, it is a much more difficult matter here. Therefore it is always best, if the first two alphabets tried result in only a 26-letter sequence, to discard them and try others until a sequence of 52 letters is procured.

Here are two deciphering alphabets, derived presumably from such a system:

```
  I   ABCDEFGHIJKLMNOPQRSTUVWXYZ
 II   DXEQWSCUNVBHMAGPFOYKIRZJTL
III   ZITNHBDXFEGAYQKCJSVPMULOWR
```

Beginning with `A` of Alphabet I, follow it by its equivalent `D` in Alphabet II, then finding `D` in Alphabet III take its equivalent in Alphabet I, which is `G`; follow `G` by `C` in Alphabet II, and continue the cycle until the starting point is reached. This process is the same as that with Key-Word Alphabets, except that here one can deal with only single letters and not with pairs. The sequence of equivalents should be written, as for a single alphabet, on strips of cross-section paper, which may be shifted at will.

In the present case, since two alphabets will result from the sequence, we designate the first, third, fifth, etc., letters as *a*, and they will constitute one Primary Alphabet, and the remaining, or *b* letters, will be found to comprise the other Primary Alphabet. In shifting the strips to find the sequence of pairs from which the generating rectangle may be built, *a* letters must be paired with *a* letters, and *b*'s with *b*'s.

```
a A
  D
a G
  C
a P
  P
a T
  K
a O
  G
a K
  B
a F
  S
a RA
  OD b
a XG
  JC b
a QP
  FP b
a IT
  NK b
a DO
  QG b
a NK
  AB b
a LF
  HS b
a ER
  WO b
a YX
  TJ b
a CQ
  EF b
a JI
  VN b
a SD
  YQ b
a MN
  MA b
a UL
  IH b
a BE
  XW b
a HY
  UT b
a VC
  RE b
a ZJ
  LV b
a WS
  ZY b
a AM
  DM b
a GU
  CI b
a PB
  PX b
a TH
```

Having shifted the strips space by space, making a careful trial each time for the generating rectangle, we soon have the strips in the position here shown. As a result from the *a* letters, we have the sequence `RAMNKZJITHYXGULFWSDOVCQPBER`. The juxtaposition of such letters as `MN`, `YX`, and `QP`, seem to indicate the alternate vertical form of transposition. But upon trying to place the first few letters in the columns of a rectangle, it is seen that `RAM` and `NKZJ` certainly cannot be adjacent columns; letting them rest for the moment, then, we pass on to the letters preceding and following `YX`. Placing these in columns so as to alphabetize `X` and `Y`, we have:

```
U T
G H
X Y
```

Now, if we go to `Z` and take the adjacent letters, the next column on the right will be `ZKN` or `ZKI`. The former is very unlikely, for it would signify that both `J` and `I` are in the key word, which is not probable; but if `ZJI` is made the next column of letters, it will bring `J` in the position following `H`, and `I` in the key word. Then we note that by placing `W` in its natural position before `X`, we also have `F` before `G`, and `L` in the key. We now have:

```
L U T I
F G H J
W X Y Z
```

From this point on, it is a very simple matter to build the remaining part of the rectangle, by seeking to fit alphabetical sequences together. The key word is found to be `REVOLUTI(O)N`:

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

Now without moving the strips, the sequence of *b* letters is `ODMABLVNKUTJCIHSZYQGREFPXWO`. By the same process as before the generating rectangle is quickly recovered, and the key word is found to be `AMERIC(A)N`.

Being now in possession of the Primary Alphabets, it is ascertained that the deciphering alphabets used in this instance were the fifth and sixteenth, with AMERICAN the key word for the outer or text alphabet, and REVOLUTION for the inner or cipher alphabet. In other words, to have a sequence of 52 equivalents the interval between the two deciphering alphabets must be odd. If the interval is even, as for instance, if the deciphering alphabets were the fourth and tenth, or the fifth and thirteenth, the sequence would yield 26 letters only, as stated before. It is not impossible to build the generating rectangles and Primary Alphabets from a sequence of 26, but it is a process which takes time and patience.

Let us examine the long sequence on page 6 and above. Note that AM is found 28 places removed from RA, MN 28 places from AM, etc. In other words, there is always a symmetry of position, or definite interval, between pairs; and once the number of places between succeeding pairs is ascertained, the alphabet may then be built mathematically. Therefore, if in any given case it should be found impossible to discover two deciphering alphabets which will results in a sequence of 52 letters, the Primary Alphabets may be obtained from the sequence of 26 letters, if a series of 25 tests is made for each possible position of the strips — that is, with the strips set for the first possible position, the pairs at first one, then two, then three intervals, and on to twenty-five, are tested for the generating rectangle. This necessitates approximately 25×25 trials; hence it is advisable to use alphabets which will yield the 52-letter sequence, if possible.

In the foregoing paragraphs we have dealt with only one form of Arbitrarily-Mixed Alphabet, namely, that in which the system of formation was a key word generating rectangle. There are, of course, many methods by which an alphabet may be built up, but whatever the method, its very use will enable it in most cases to be discovered.

```
a TA
  JR b
a BO
  UP b
a MC
  GT b
a HE
  SB b
a XN
  KM b
a VW
  YH b
a ZI
  DX b
a JR
  LV b
a UP
  FZ b
a GT
  QJ b
a SB
  AU b
a KM
  OG b
a YH
  CS b
a DX
  EK b
a LV
  NY b
a FZ
  WD b
a QJ
  IL b
a AU
  RF b
a OG
  PQ b
a CS
  TA b
a EK
  BO b
a NY
  MC b
a WD
  HE b
a IL
  XN b
a RF
  VW b
a PQ
  ZI b
```

RANDOM-MIXED ALPHABETS

The process of recovering Random-Mixed Alphabets is very much the same as that used for Arbitrarily-Mixed Alphabets. But whereas it is always possible to make certain the recovery of an arbitrarily-mixed Primary Alphabet by reconstructing the generating rectangle, or whatever the system of formation, it is not possible to check a random-mixed alphabet in the same way. Here the proof must be found in the solution of cipher text by means of the Primary Alphabets obtained. Let us consider the following alphabets:

```
  I    ABCDEFGHIJKLMNOPQRSTUVWXYZ
 II    RUTEBWQSXLONGMPZIVAJFYHKCD
III    IJPCTNFGHDAEUBRVWXQZLKMSOY
```

From these deciphering alphabets results the sequence of 52 equivalents here shown, beginning with A of Alphabet I and continuing the process as described on page 2. Now if the two strips on which the sequence is recorded are shifted to any position, such as for example the one here shown, it will be noted that the alphabet made up from the *a* letters is exactly the same as that formed by the*b* letters — in other words, the Primary Alphabet System from which the foregoing deciphering alphabets were derived, is one in which the components are identical.

The next problem is to ascertain if the alphabet resulting from this position is the Primary Alphabet: Let us suppose that the following is a portion of the message from which the deciphering alphabets were obtained:

```
   Key:  G E N E R A L
Cipher:  H C J N Z Z P
  Text:  H U N S A R E
```

Now placing the alphabet resulting from the position of the sequence as shown here on two strips, set H of the lower or cipher alphabet to H of the upper, and find the value of the key letter G. It is G. Then resetting the strips so that C of the lower equals U of the upper, it is noted that the key letter, or in this case E, again points to G. Similarly, when J is set to N, N is below G, when N equals S, E is below G. Since each key letter points always to the same letter G, this signifies that G is the first letter of the original Primary Alphabet.

TAUPQJRFZILVWDXNYHEKMCSBOGT

$\mathtt{YHEKM\underline{C}SBOGTAUPQJRFZILVWDXNYH\underline{E}KMCS}$

Now returning to the sequence of equivalents, the strip on the right is shifted until it has its beginning opposite G, keeping in mind that *a* letters must be opposite *a* letters. The alphabet then reads GRDMAZNSPLHOJWKTFXCUIYBQVE. This may or may not be the original Primary Alphabet. But it is an alphabet which will solve any message enciphered by means of that Primary Alphabet, for there must necessarily be a symmetry of position in any alphabet thus derived, which makes it exactly as efficacious as the original itself.

So it is, also, with any system of two components which are random-mixed alphabets, even though not identical. To illustrate, here are the Primary Alphabets of such a system:

```
A -- SZNGDKWFJEOYTCUXBVLQMRAHPI
B -- WKAZHRMPBJQNTVCXDLIOFESGUY
```

The seventh and twelfth deciphering alphabets derived from these Primaries are as follows:

```
  I    ABCDEFGHIJKLMNOPQRSTUVWXYZ
 II    ANJOXUVPTGRYSKCZDIBWLFMEQH
III    VHGKOETQFPBWRSJAIMYZCNXDUL
```

If there is derived from these deciphering alphabets a sequence of equivalents beginning AAPZTWL, etc., and the strips upon which the sequence is recorded are shifted, two of the possible resulting alphabets may be as follows:

```
From a letters:    NAOQKIJWLBYDPUXMGCRSVHETZFN
From b letters:    KACDRTGMYNQOZLESVJIBFPXWHUK
```

Now let us test these alphabets upon some cipher text which was enciphered by means of the original Primary Alphabets.

```
   Key:  A M E R I C A N
  Text:  C O M P A N Y D
Cipher:  X D X Z C D V X
```

Taking the pair of alphabets shown immediately preceding and setting them so that C of the lower equals X of the upper, it will be found that A, or the key letter for such encipherment, equals S, if the message has been enciphered by the Vigenere method. Resetting them so that O equals D, the encipherment of the second letter, the key letter M in this case, again equals S. Such is found to be the case with each succeeding encipherment. This would indicate that S was the first letter of the Primary Alphabet in which the text letters were found. Therefore in any succeeding cipher text, if the key letter is set to S as in the alphabets used in this portion of text, these alphabets will be found to solve the text exactly as easily as the original Primaries, whose actual recovery then becomes unnecessary.

It may be said in conclusion that these methods for recovering alphabets, although here given as applying to Primary Alphabet Systems alone, may be utilized in many other forms in systems of cipher. The real student of the science will be quick to see their application in manifold ways.