~~CONFIDENTIAL~~
~~Modified Handling Authorized~~

NATIONAL SECURITY AGENCY

MILITARY CRYPTANALYTICS
Part II

INTERIM EDITION
(First Section)

By
WILLIAM F. FRIEDMAN
and
LAMBROS D CALLIMAHOS

National Security Agency
Washington 25, D. C.

December 1955

record taken from WFF's home

~~CONFIDENTIAL~~
~~Modified Handling Authorized~~

NATIONAL SECURITY AGENCY

MILITARY CRYPTANALYTICS
Part II

INTERIM EDITION
(First Section)

By
WILLIAM F. FRIEDMAN
and
LAMBROS D. CALLIMAHOS

NSATL S-70,022

National Security Agency
Washington 25, D C.

December 1955

A likely impossibility is always preferable to an unconvincing possibility

—Aristotle

Preface

This text represents an extensive expansion and revision, both in scope and content, of the earlier work entitled "Military Cryptanalysis, Part II by William F Friedman This expansion and revision, as indicated in the preface of the first volume of this present series, was necessitated by the considerable advancement made in the art since the publication of the original version

The first 9 chapters of this text are being published in this interim edition to make it readily available for use in cryptologic training programs of the National Security Agency and of the Service Cryptologic Agencies The last 5 chapters and the appendices will be issued in an interim version in the near future, after which the entire text will be formally published in one volume

—L D C

TABLE OF CONTENTS

MILITARY CRYPTANALYTICS, PART II

Periodic Polyalphabetic Substitution Systems

(BLANK)

CHAPTER I

INTRODUCTORY REMARKS

1. General.--a. This text constitutes the second in the series of six basic texts on the science of cryptanalytics as applied to military communications in general, without regard to the particular Service involved (i.e., army, navy, or air force traffic).

b. It is assumed that the reader has studied Military Cryptanalytics, Part I, and is familiar with the cryptologic concepts, principles, and techniques of solution of the various cryptosystems treated in that text, this background is a necessary prerequisite to the understanding of the principles expounded in the present text.

c. It is taken for granted that the student has acquired a background of generalized cryptologic terminology from the study of the first text and its accompanying glossary. The new terms which appear in this text are usually defined upon their first occurrence; these terms, as well as others which are necessary for cross-reference, are included in the glossary to this volume (Appendix 1).

d. As has been already indicated, each text has its accompanying course of problems in cryptanalysis, so that the student may have the opportunity of applying the principles learned to practical examples, and in so doing develop skill in the analysis of the types of cryptosystems treated in this text. The problems which pertain to this text constitute Appendix 10.

e. As was the case in the preceding text, this present volume is written from the standpoint that the reader has had a minimum of mathematical background, not beyond elementary algebra; the authors have endeavored to enhance this background gradually and progressively, to enable the student to be better versed in the mathematical techniques and applications in the art of cryptanalysis. As before, footnotes are used to give additional general information about the subject being treated, or to amplify mathematical principles in details which may be beyond the average reader; therefore certain footnotes may be passed over by the student. The next text, Military Cryptanalytics, Part III, will contain a comprehensive treatment of the fundamentals of cryptomathematics, and will of necessity recapitulate the points of mathematical observations made in the first two texts.

f. The next paragraph will bridge the gap between the basic concept of the systems treated in the previous text and the principles to be introduced in this present volume.

2. The essential difference between monoalphabetic and polyalphabetic substitution.--a. In the substitution methods thus far discussed it has been pointed out that their basic feature is that of monoalphabeticity. From the cryptanalytic standpoint, neither the nature of the cipher symbols, nor their method of production is an essential feature, although these may be differentiating characteristics from the cryptographic standpoint. It is true that in the cases of monoalphabetic substitution with variants and in syllabary squares and code charts, there is a departure, more or less considerable, from strict monoalphabeticity. In some of the cases of variant systems indeed, there may be available two or more wholly independent sets of equivalents, which, moreover, may even be arranged in the form of completely separate alphabets. Thus, while a loose terminology might permit one to designate such systems as polyalphabetic, it is better to reserve this nomenclature for those cases wherein polyalphabeticity is the essence of the method, specifically introduced with the purpose of imparting a positional variation in the substitutive equivalents for plaintext letters, in accordance with some rule directly or indirectly connected with the absolute positions the plaintext letters occupy in the message. This point calls for amplification.

b. In monoalphabetic substitution with variants the object of having different or multiple equivalents is to suppress, so far as possible by simple methods, the characteristic frequencies of the individual letters occurring in plain text. As has been noted, it is by means of these characteristic frequencies that the cipher equivalents can usually be identified. In these systems the varying equivalents for plaintext letters are subject to the free choice and caprice of the enciphering clerk; if he is careful and conscientious in the work, he will really make use of all the different equivalents afforded by the system; but if he is slip-shod and hurried in his work he will use the same equivalents repeatedly rather than take pains and time to refer to the charts, tables, or diagrams to find the variants. Moreover, and this is a crucial point, even if the individual enciphering clerks are extremely careful, when many of them employ the same system it is entirely impossible to insure a complete diversity in the encipherments[1] produced by two or more clerks working at different message centers. The result is inevitably to produce plenty of repetitions and near-repetitions or isologous sequences in the texts emanating from several stations, and when texts such as these are all available for study they are open to solution, by a comparison of their similarities and differences.

---

[1] It must be noticed however, that a complete diversity of enciphering is sometimes not necessarily an optimum desideratum from the standpoint of cryptosecurity, a complete diversity of encipherments, in the case of isologs, would lay bare all the elements of a variant system See in this connection the example given in subpar 62b in Military Cryptanalytics, Part I

c. In true polyalphabetic systems, on the other hand, there is established a rather definite procedure which automatically determines the shifts or changes in equivalents or in the manner in which they are introduced, so that these changes are beyond the momentary whim or choice of the enciphering clerk. When the method of shifting or changing the equivalents is scientifically sound and sufficiently complex, the research necessary to establish the values of the cipher characters is much more prolonged and difficult than is the case even in complicated monoalphabetic substitution with variants, as will later be seen. These are the objects of true polyalphabetic substitution systems. The number of such systems is quite large, and it will be possible to describe in detail the cryptanalysis of only a few of the more common or typical examples of methods encountered in practical military communications.

d. The three methods, (1) single-equivalent monoalphabetic substitution, (2) monoalphabetic substitution with variants, and (3) true polyalphabetic substitution, show the following relationships as regards the equivalency between plaintext and ciphertext units:

A. In method (1), there is a set of 26 symbols, a plaintext letter is always represented by one and only one of these symbols; conversely, a symbol always represents the same plaintext letter. The equivalence between the plaintext and the cipher letters is constant in both encipherment and decipherment.

B. In method (2), there is a set of n symbols, where n may be any number greater than 26 and often is a multiple of that number; a plaintext letter may be represented by 1, 2, 3, . . . different symbols; conversely, a symbol always represents the same plaintext letter, the same as is the case in method (1). The equivalence between the plaintext

and the cipher letters is variable in encipherment but constant in decipherment.[2]

C. In method (3), there is as in the first method, a set of 26 symbols; a plaintext letter may be represented by 1, 2, 3, . . . 26 different symbols; conversely, a symbol may represent 1, 2, 3, . . . 26 different plaintext letters, depending upon the system and the specific key. The equivalence between the plaintext and the cipher letters is variable in both encipherment and decipherment.

---

[2] As has been pointed out in the previous text, there is a monoalphabetic method in which the inverse result obtains, the correspondence being constant in encipherment but variable in decipherment, this is a method not found in the usual books on cryptography but in an essay on that subject by Edgar Allan Poe entitled, in some editions of his works, A few words on secret writing and in other editions Cryptography The method is to draw up an enciphering alphabet such as the following (using Poe's example)

Plain....	**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**
Cipher...	**S**	**U**	**A**	**V**	**I**	**T**	**E**	**R**	**I**	**N**	**M**	**O**	**D**	**O**	**F**	**O**	**R**	**T**	**I**	**T**	**E**	**R**	**I**	**N**	**R**	**E**

In such an alphabet because of repetitions in the cipher component, the plaintext equivalents are subject to a considerable degree of variability as will be seen in the deciphering alphabet

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

This type of variability gives rise to ambiguities in decipherment A cipher group such as $\overline{\text{TIE}}_{\text{C}}$ would yield such plaintext sequences as REG, FIG, TEU, REU, etc , which could be read only by context No system of such a character would be practical for serious usage For a further discussion of this type of cipher alphabet see Friedman William F , Edgar Allan Poe, Cryptographer, Signal Corps Bulletins Nos 97 (July-Sept) and 98 (Oct -Dec), 1937

3. Example of polyalphabetic substitution.--a. A simple example may be used to illustrate what is meant by true polyalphabetic substitution. Suppose that two correspondents agree upon a numerical key, for example, 74030274, each digit of which means that the plaintext letter to which the digit applies as a key number is to be replaced by the letter that stands a corresponding number of places to the right of it in the normal alphabet. For example, if R_p is to be enciphered by key number 7 it is to be replaced by Y_c. The numerical key is written over the letters of the plain text, letter for letter, and is repeated until the whole text is covered.[3] Let the message be REINFORCEMENTS BEING RUSHED. The encipherment of this message is shown in Fig. 1, below. For convenience in counting forward (to the right) to find cipher equivalents, a normal alphabet is given at the top of the figure. To decipher such a crypto-

Normal alphabet:	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Key:	74030 27474 03027 47403 02747
Plain text:	REINF ORCEM ENTSB EINGR USHED
Cipher text:	YIIQF QYGLQ EQTUI IPRGU UUOIK

Figure 1.

gram, the clerk writes the numerical key over the cipher letters and then counts backward (i.e., to the left) in the normal alphabet as many places as indicated by the key number standing over each letter.

b. Instead of writing the key over and over again in order to cover the plain text completely, the text may be written in sets of letters corresponding in length to the length of the key. Thus the text may be written underneath a single appearance of the key in successive short horizontal lines leaving space between the lines for the insertion of cipher equivalents, as shown in Fig. 2. Instead of enciphering the letters by individual repeated countings, two strips bearing normal alphabets may

7 4 0 3 0 2 7 4
R E I N F O R C
E M E N T S B E
I N G R U S H E
D

Figure 2.

[3] This system being described is known in cryptologic literature as the Gronsfeld cipher

be juxtaposed in the proper relative positions to encipher a whole column of letters at one setting of the strips. Thus for the first column, with the key number 7, the strips are juxtaposed so that the first letter in the column, viz., R_p (which is to be represented by the seventh letter to the right of it, and is therefore to be enciphered by Y_c of the lower strip) is directly above Y_c, as follows:

Plain: ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher: ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ

The equivalents for the rest of the letters of the first column may now be written under their respective plaintext letters, reference being made to the enciphering alphabet to see what the cipher letters should be: $E_p = L_c$; $I_p = P_c$, and $D_p = K_c$. For the second column, the two strips are juxtaposed as follows:

Plain: ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher: ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ

The cipher equivalents for the second column are: $E_p = I_c$; $M_p = Q_c$; and $N_p = R_c$. The process is continued in this manner until all the columns have been enciphered as shown in the diagram below:

7 4 0 3 0 2 7 4
R E I N F O R C
Y I I Q F Q Y C

E M E N T S B E
L Q E Q T U I I

I N G R U S H E
P R G U U U O I

D
K

The cipher text is then transcribed in five-letter groups for transmission, viz., YIIQF QYGLQ EQTUI IPRGU UUOIK. This systematized procedure has the merit of being faster, less laborious, and less liable to error than the method shown in subpar. 3a.

4. Primary classification of polyalphabetic systems.--a. A primary classification of polyalphabetic systems into two rather distinct types may be made: (1) periodic systems and (2) aperiodic systems. When the enciphering process involves a cryptographic treatment which is repetitive in character, and which results in the production of cyclic phenomena in the cryptographic text, the system is termed periodic. When the enciphering process is not of the type described in the foregoing general terms, the system is termed aperiodic. The substitution in both cases involves the use of two or more cipher alphabets.

b. The cyclic phenomena inherent in a periodic system may be exhibited externally, in which case they are said to be patent, or they may not be exhibited externally, and must be uncovered by a preliminary step in the analysis in which case they are said to be latent. The periodicity may be quite definite in nature, and therefore determinable with mathematical exactitude allowing for no variability, in which case the periodicity is said to be fixed. In other instances the periodicity is more or less flexible in character and even though it may be determinable mathematically allowance must be made for a degree of variability subject to limits controlled by the specific system under investigation. The periodicity is in this case said to be flexible, or variable within limits.

5. Primary classification of periodic systems.--a. Periodic polyalphabetic substitution systems may primarily be classified into two kinds:

(1) Those in which only a few of a whole set of cipher alphabets are used in enciphering individual messages, these alphabets being employed repeatedly in a fixed sequence throughout each message. Because it is usual to employ a secret word, phrase, or number as a key to determine the number identity, and sequence with which the cipher alphabets are employed, and this key is used over and over again in encipherment, this method is often called the repeating-key system, or the repeating-alphabet system. In this text the designation "repeating-key system" will be used.[4]

(2) Those in which all the cipher alphabets comprising the complete set for the system are employed one after the other successively in the encipherment of a message, and when the last alphabet of the series has been used, the encipherer begins over again with the first alphabet. This is commonly referred to as a progressive-alphabet system because the cipher alphabets are used in progression.

6. Sequence of study of polyalphabetic systems.--a. In the studies to be followed in connection with polyalphabetic systems, the order in which the work will proceed conforms very closely to the classifications made in pars. 4 and 5. Periodic polyalphabetic substitution ciphers will come first, because they are, as a rule, the simpler and because a thorough understanding of the principles of their analysis is prerequisite to a comprehension of how aperiodic systems are solved. But in the final analysis the solution of examples of both types rests upon the conversion or reduction of polyalphabeticity into monoalphabeticity. If this is possible, solution can always be achieved, granted there are sufficient data in the final monoalphabetic distributions to permit of solution by recourse to the ordinary principles of frequency.

---

[4] French terminology calls this the "double-key method", but there is no logic in such nomenclature

b. First in the order of study of periodic systems will come the analysis of repeating-key systems. Some of the more simple varieties will be discussed in detail, with examples. Subsequently, ciphers of progressive alphabet systems will be discussed. There will then follow a treatment of polyalphabetic bipartite systems, monome-dinome systems with cyclic additives, and periodic digraphic systems. Aperiodic systems will be treated in detail in Military Cryptanalytics, Part III.

CHAPTER II

THEORY OF REPEATING-KEY SYSTEMS

7. Classification of cipher alphabets upon the basis of their derivation.--a. The substitution processes in polyalphabetic methods involve the use of a plurality of cipher alphabets. The latter may be derived by various schemes, the exact nature of which determines the principal characteristics of the cipher alphabets and plays a very important role in the preparation and solution of polyalphabetic cryptograms. For these reasons it is advisable, before proceeding to a discussion of the principles and methods of analysis, to point out these various types of cipher alphabets, show how they are produced, and how the method of their production or derivation may be made to yield important clues and shortcuts in analysis.

b. A primary classification of cipher alphabets for polyalphabetic substitution may be made into the two following types:

(1) Independent or unrelated cipher alphabets.
(2) Derived or interrelated cipher alphabets.

c. Independent cipher alphabets may be disposed of in a very few words. They are merely separate and distinct alphabets showing no relationship to one another in any way. They may be compiled by the various methods discussed in par. 39 of Military Cryptanalytics, Part I. The solution of cryptograms written by means of such alphabets is rendered more difficult by reason of the absence of any relationship between the equivalents of one cipher alphabet and those of any of the other alphabets of the same cryptogram. On the other hand, from the point of view of practicability in their production and their handling in encrypting and decrypting, they present some difficulties which make them less favored by cryptographers than interrelated cipher alphabets.

d. Derived or interrelated alphabets, as their name indicates, are most commonly produced by the interaction of two primary components, which when juxtaposed at the various points of coincidence can be made to yield secondary alphabets.

8. Primary components and secondary alphabets.--Two basic, slidable sequences or components of n characters each will yield n secondary alphabets. The components may be classified according to various schemes. For cryptanalytic purposes the following classification will be found useful:

Case I. The primary components are both normal sequences.

a. The sequences proceed in the same direction. (The secondary alphabets are direct standard alphabets.) (Pars. 20-22.)

b. The sequences proceed in opposite directions. (The secondary alphabets are reversed standard alphabets; they are also reciprocal cipher alphabets.) (Subpars. 20i, 21g.)

Case II. The primary components are not both normal sequences.

a. The plain component is normal, the cipher component is a mixed sequence. (The secondary alphabets are mixed alphabets.) (Pars. 26-35.)

b. The plain component is a mixed sequence, the cipher component is normal. (The secondary alphabets are mixed alphabets.) (Par. 36.)

c. Both components are mixed sequences.

1. Components are identical mixed sequences.

(a) Sequences proceed in the same direction. (The secondary alphabets are mixed alphabets.) (Par. 41.)

(b) Sequences proceed in opposite directions. (The secondary alphabets are reciprocal mixed alphabets.) (Par. 55.)

2. Components are different mixed sequences. (The secondary alphabets are mixed alphabets.) (Par. 56.)

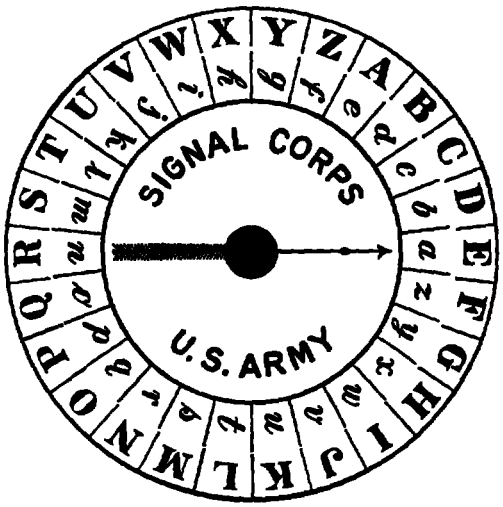
9. The use of key words to indicate number, identity, and sequence of cipher alphabets employed.--a. If reference is made to the two settings of alphabet strips in subpar. 3b, it will be noted that in the first setting, $A_p = H_c$, and in the second setting, $A_p = E_c$. If the eight settings of the strips are studied it will be found that the letters which A_p represents successively are H, E, A, D, A, C, H, and E, giving the word HEADACHE. These settings, when first presented in the foregoing description, correspond merely to the numerical key 74030274, but this numerical key is also expressible in terms of letters, which when put together properly spell a word. This is only another way of showing that key words may be employed in this type of substitution as in those previously described. Key words of various lengths and composition may be used, consisting of single words, long phrases or sentences. In general, the longer the key the greater is the degree of cryptographic security.

b. The number of elements in the key--that is, that number of letters or figures composing it--determines the number of alphabets to be employed. The identity of each element of the key, the specific letter or figure it happens to be, determines specifically which of a set of cipher alphabets pertaining to the whole system will be used. And the specific sequence or relative order of the elements of the key determines specifically the sequence with which the cipher alphabets are employed within the encipherment. The total number of cipher alphabets pertaining to or composing the system may be limited or unlimited. When they are produced as a result of the sliding of two basic or primary alphabets against each other, the number is limited to 26 in the English alphabet.

c. A brief notation for indicating or designating a specific key letter is to suffix the subscript "k" to it, just as the subscripts "p" and "c" are suffixed to letters to indicate letters of the plain text or cipher text, respectively. When the key letter occurs in an equation, it can be enclosed within parentheses to avoid ambiguity. Thus $B_p(D_k) = E_c$ means that the plaintext letter B when enciphered by the key letter D (in a certain alphabet system) yields the cipher letter E.

10. Cipher disks.--a. In subpar. 3b, it was noted that the separate alphabets employed in the encipherment are produced by the use of only two strips of paper bearing the normal alphabet. Such strips are often referred to as sliding alphabets, because they can be shifted or slid against each other in any one of 26 points of contact or coincidence. Exactly the same results, so far as cipher equivalents are concerned, can be obtained by the use of other devices. First, there are the so-called cipher wheels or cipher disks in which an alphabet is written on the periphery of a rotating disk, the circumference of which is divided into 26 equal segments, and this disk is made to revolve concentrically upon a similar but slightly larger fixed disk. Fig. 3 shows the now obsolete U.S. Army Cipher Disk, which is of this simple type. Here the alphabetic sequences are printed on glossy celluloid, are permanent, and admit of no variation. The use of unglazed celluloid upon which blank segments appear would permit of writing letters and erasing them as often as desirable. Thus, quick and easy change of alphabets would be possible.

b. The cipher alphabets produced by the cipher disk shown in the figure are merely reversed standard alphabets, the same as are produced by the use of sliding strips of paper, and by the use of certain tables which are discussed below. The method of employing the disk needs no discussion. It may serve in monoalphabetic or polyalphabetic substitution with a key word or key number.



To encipher a message the key letter or the first letter of the key word or phrase is set opposite 'a' Let us assume it to be E The cipher letters to be written are those opposite the text letter when a on the circle is set opposite 'E' on the card For example send powder' would be written MARBPQIBAN" To use a key word or phrase each letter is used in turn to encipher one letter only When the last letter of the key word is used repeat until all letters of the message are enciphered Numbers when enciphered with the disk must be spelled out

Figure 3.

11. Square tables.--a. Tables known in the literature of cryptography under various names, such as "Vigenère Square", "Vigenere Table", "Square Table", "Pythagorean Table", etc., are often employed in polyalphabetic substitution. All the results produced by their use can be duplicated by the employment of sliding alphabets or revolving disks. The modern form of the Vigenère Square is shown in Fig. 4. Such a square may be used in various ways, differing from one another in minor details. The most common method is to consider the top line of the table as containing the plaintext letters, the first column at the left as containing the key letters. Then each successive horizontal line contains the cipher equivalents for the plaintext sequence ABC...Z enciphered by the key letter which stands at its left in the first column. Thus, the cipher alphabet corresponding to key letter D is the sequence of letters in the fourth horizontal line under the plaintext line, where $A_p = D_c$, $B_p = E_c$, etc. It will be easy to remember, in using such a table, that the equivalent of a given plaintext letter, T_p, for example, enciphered by a given key letter, O_k, lies at the intersection of the vertical column headed by T, and the horizontal row begun by O. In this case $T_p(O_k) = H_c$. The same result will be found on referring to sliding, direct standard alphabets.

Plain text

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

Figure 4. The Vigenere Square

b. Minor modifications of the Vigenere Square are encountered. If the top line is made a reversed normal sequence, leaving the interior of the table unchanged, or if the successive horizontal rows are made to contain the reversed normal sequence, leaving the top row (plain text) unchanged, then the results given by using the table are the same as those given by using the cipher disk shown in Fig. 3. Again, the same general results can be obtained by using a set of alphabets in tabular form known under the names of Porta's Table and Napoleon's Table, which is shown in Fig. 5.

AB	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
CD	A	B	C	D	E	F	G	H	I	J	K	L	M
	O	P	Q	R	S	T	U	V	W	X	Y	Z	N
EF	A	B	C	D	E	F	G	H	I	J	K	L	M
	P	Q	R	S	T	U	V	W	X	Y	Z	N	O
GH	A	B	C	D	E	F	G	H	I	J	K	L	M
	Q	R	S	T	U	V	W	X	Y	Z	N	O	P
IJ	A	B	C	D	E	F	G	H	I	J	K	L	M
	R	S	T	U	V	W	X	Y	Z	N	O	P	Q
KL	A	B	C	D	E	F	G	H	I	J	K	L	M
	S	T	U	V	W	X	Y	Z	N	O	P	Q	R
MN	A	B	C	D	E	F	G	H	I	J	K	L	M
	T	U	V	W	X	Y	Z	N	O	P	Q	R	S
OP	A	B	C	D	E	F	G	H	I	J	K	L	M
	U	V	W	X	Y	Z	N	O	P	Q	R	S	T
QR	A	B	C	D	E	F	G	H	I	J	K	L	M
	V	W	X	Y	Z	N	O	P	Q	R	S	T	U
ST	A	B	C	D	E	F	G	H	I	J	K	L	M
	W	X	Y	Z	N	O	P	Q	R	S	T	U	V
UV	A	B	C	D	E	F	G	H	I	J	K	L	M
	X	Y	Z	N	O	P	Q	R	S	T	U	V	W
WX	A	B	C	D	E	F	G	H	I	J	K	L	M
	Y	Z	N	O	P	Q	R	S	T	U	V	W	X
YZ	A	B	C	D	E	F	G	H	I	J	K	L	M
	Z	N	O	P	Q	R	S	T	U	V	W	X	Y

Figure 5.

In this table the alphabets are all reciprocal, for example, $G_p(W_k) = R_c$, $R_p(W_k) = G_c$. Reciprocal alphabets when arranged in this form are sometimes called complementary alphabets. Note that in each alphabet either of two letters may serve as key letter indifferently: $G_p(W_k)$ or $G_p(X_k) = R_c$.

c. Another modification of the basic table, and one that employs numbers instead of letters as cipher equivalents is shown in Fig. 6. Since many more than 26 different equivalents are available (100 pairs of digits from 00 to 99), it is possible to insert many plaintext elements in the top line of the table in addition to the 26 letters. For example, one could have the 10 digits, a few common double-letter combinations, such as DD, LL, RR, and SS; a few of the most frequently used digraphs, such as TH, ER, IN, or even such common syllables as ENT, ING, and ION. The table shown in Fig. 6 was used by the Italian army in World War I, and was known as the "Cifrario militare tascabile" (Pocket military cipher).

*	a b c d e f g h ı	*	j k l m n o p q r	*	s t u v w x y z	*	0 1 2 3 4 5 6 7 8 9	*
a	10 11 12 13 14 15 16 17 18	a	19 20 21 22 23 24 25 26 27	a	28 29 30 31 32 33 34 35	a	36 37 38 39 40 41 42 43 44 45	a
b	11 12 13 14 15 16 17 18 19	b	20 21 22 23 24 25 26 27 28	b	29 30 31 32 33 34 35 36	b	37 38 39 40 41 42 43 44 45 10	b
c	12 13 14 15 16 17 18 19 20	c	21 22 23 24 25 26 27 28 29	c	30 31 32 33 34 35 36 37	c	38 39 40 41 42 43 44 45 10 11	c
d	13 14 15 16 17 18 19 20 21	d	22 23 24 25 26 27 28 29 30	d	31 32 33 34 35 36 37 38	d	39 40 41 42 43 44 45 10 11 12	d
e	14 15 16 17 18 19 20 21 22	e	23 24 25 26 27 28 29 30 31	e	32 33 34 35 36 37 38 39	e	40 41 42 43 44 45 10 11 12 13	e
f	15 16 17 18 19 20 21 22 23	f	24 25 26 27 28 29 30 31 32	f	33 34 35 36 37 38 39 40	f	41 42 43 44 45 10 11 12 13 14	f
g	16 17 18 19 20 21 22 23 24	g	25 26 27 28 29 30 31 32 33	g	34 35 36 37 38 39 40 41	g	42 43 44 45 10 11 12 13 14 15	g
h	17 18 19 20 21 22 23 24 25	h	26 27 28 29 30 31 32 33 34	h	35 36 37 38 39 40 41 42	h	43 44 45 10 11 12 13 14 15 16	h
ı	18 19 20 21 22 23 24 25 26	i	27 28 29 30 31 32 33 34 35	ı	36 37 38 39 40 41 42 43	ı	44 45 10 11 12 13 14 15 16 17	ı
j	19 20 21 22 23 24 25 26 27	j	28 29 30 31 32 33 34 35 36	j	37 38 39 40 41 42 43 44	j	45 10 11 12 13 14 15 16 17 18	j
k	20 21 22 23 24 25 26 27 28	k	29 30 31 32 33 34 35 36 37	k	38 39 40 41 42 43 44 45	k	10 11 12 13 14 15 16 17 18 19	k
l	21 22 23 24 25 26 27 28 29	l	30 31 32 33 34 35 36 37 38	l	39 40 41 42 43 44 45 10	l	11 12 13 14 15 16 17 18 19 20	l
m	22 23 24 25 26 27 28 29 30	m	31 32 33 34 35 36 37 38 39	m	40 41 42 43 44 45 10 11	m	12 13 14 15 16 17 18 19 20 21	m
n	23 24 25 26 27 28 29 30 31	n	32 33 34 35 36 37 38 39 40	n	41 42 43 44 45 10 11 12	n	13 14 15 16 17 18 19 20 21 22	n
o	24 25 26 27 28 29 30 31 32	n	33 34 35 36 37 38 39 40 41	o	42 43 44 45 10 11 12 13	o	14 15 16 17 18 19 20 21 22 23	o
p	25 26 27 28 29 30 31 32 33	p	34 35 36 37 38 39 40 41 42	p	43 44 45 10 11 12 13 14	p	15 16 17 18 19 20 21 22 23 24	p
q	26 27 28 29 30 31 32 33 34	q	35 36 37 38 39 40 41 42 43	q	44 45 10 11 12 13 14 15	q	16 17 18 19 20 21 22 23 24 25	q
r	27 28 29 30 31 32 33 34 35	r	36 37 38 39 40 41 42 43 44	r	45 10 11 12 13 14 15 16	r	17 18 19 20 21 22 23 24 25 26	r
s	28 29 30 31 32 33 34 35 36	s	37 38 39 40 41 42 43 44 45	s	10 11 12 13 14 15 16 17	s	18 19 20 21 22 23 24 25 26 27	s
t	29 30 31 32 33 34 35 36 37	t	38 39 40 41 42 43 44 45 10	t	11 12 13 14 15 16 17 18	t	19 20 21 22 23 24 25 26 27 28	t
u	30 31 32 33 34 35 36 37 38	u	39 40 41 42 43 44 45 10 11	u	12 13 14 15 16 17 18 19	u	20 21 22 23 24 25 26 27 28 29	u
v	31 32 33 34 35 36 37 38 39	v	40 41 42 43 44 45 10 11 12	v	13 14 15 16 17 18 19 20	v	21 22 23 24 25 26 27 28 29 30	v
w	32 33 34 35 36 37 38 39 40	w	41 42 43 44 45 10 11 12 13	w	14 15 16 17 18 19 20 21	w	22 23 24 25 26 27 28 29 30 31	w
x	33 34 35 36 37 38 39 40 41	x	42 43 44 45 10 11 12 13 14	x	15 16 17 18 19 20 21 22	x	23 24 25 26 27 28 29 30 31 32	x
y	34 35 36 37 38 39 40 41 42	y	43 44 45 10 11 12 13 14 15	y	16 17 18 19 20 21 22 23	y	24 25 26 27 28 29 30 31 32 33	y
z	35 36 37 38 39 40 41 42 43	z	44 45 10 11 12 13 14 15 16	z	17 18 19 20 21 22 23 24	z	25 26 27 28 29 30 31 32 33 34	z
*	a b c d e f g h ı	*	j k l m n o p q r	*	s t u v w x y z	*	0 1 2 3 4 5 6 7 8 9	*

Figure 6.

12. Square tables employing mixed alphabets.--a. In the tables thus far shown the alphabets have been direct or reversed standard sequences, but just as mixed sequences may be written upon sliding strips and revolving disks, so can mixed alphabets appear in tabular form. The table shown in Fig. 7, based upon the keyword sequence derived from the word LEAVENWORTH, is an example that is equivalent to the use of a strip

Plain text

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

Cipher

Figure 7.

bearing that keyword sequence sliding against another strip bearing the normal alphabet. The usual method of using such a table is the same as that in the preceding cases. The only difference is that the key letters must now be sought in a mixed sequence, whereas in the preceding tables they were located in direct or reversed standard sequences. Example, using Fig. 7: $C_p(S_k) = X_c$.

b. Fig. 8 illustrates a case in which a mixed alphabet is sliding against itself. The usual method of employing such a table is exactly the same as that explained before. The only difference is that both the plaintext letters and the key letters must be looked for in mixed sequences. Example, using Fig. 8: $U_p(R_k) = V_c$.

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

Figure 8.

c. In employing sliding alphabets it is usual to set the key letter as located in the cipher component opposite the letter A as located in the plain component or the key letter as located in the cipher component may be set opposite the initial letter of the plain component. In all examples preceding that in Fig. 8, the key letter has been A. In Fig. 8, since the plain component is also a mixed sequence and its initial letter is Q, the sliding alphabets are set against each other so that the given key letter in the cipher component is opposite Q in the plain component. Thus, to duplicate the results given by the use of Fig. 8 in finding the value of $U_p(R_k)$, it is necessary to set the sliding strips in the following relative positions:

Plain: QUESTIONABLYCDFGHJKMPRVWXZ
Cipher: QUESTIONABLYCDFGHJKMPRVWXZQUESTIONABLYCDFGHJKMPRVWXZ

Here it is seen that $U_p(R_k) = V_c$, which is identical with the result obtained from the use of the table. There are other ways of using the table, however, each having a correspondingly modified method of employing sliding strips in order to obtain identical results; these ways and methods will be discussed in the next paragraph.

d. In addition to cryptographic schemes in which there are cipher squares composed of slides of a basic sequence to produce the various alphabets, it is of course possible to have n unrelated, random-mixed alphabets (here n can be any number) used cyclically to encipher the letters of a message. Such a scheme cannot be reduced to two components, and therefore would require the alphabets written out in a matrix or in tabular form. For instance a system might incorporate 1,000 different mixed alphabets, numbered from 1-1,000, then alphabet 1 might be used to encipher the first letter of each message, alphabet 2 the second letters, etc. There are also cryptographic schemes in which certain alphabets out of a total are selected for enciphering a given message, the selection being governed by an indicator, or the date, or a similar convention. These systems will be discussed in greater detail in Chapter XI.

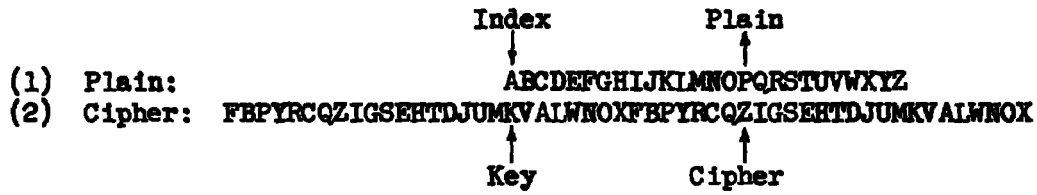
13. Further remarks on primary components.--a. In preceding paragraphs it has been shown that the equivalents obtainable from the use of square tables may be duplicated by the use of revolving cipher disks or of sliding primary components It was also stated that there are various ways of employing such tables, disks, and sliding components. Cryptographically the results may be quite diverse from different methods of using such paraphernalia, since the specific equivalents obtained from one method may be altogether different from those obtained from another method. But from the cryptanalytic point of view the diversity referred to is of little significance; only in one or two cases does the specific method of employing these cryptographic instrumentalities have an important bearing upon the procedure in cryptanalysis. However, it is advisable that the student learn something about these different methods before proceeding with further work.

b. There are, not two, but four letters involved in every case of finding equivalents by means of sliding primary components; furthermore, the determination of an equivalent for a given plaintext letter is representable by two equations involving four elements, usually letters. Three of these letters are by this time well-known to and understood by the student, viz., θ_k, θ_p, and θ_c. The fourth element or letter has been passed over without much comment, but cryptographically it is just as important a factor as the other three. Its function may best be indicated by noting what happens when two primary components are juxtaposed, for the purpose of finding equivalents. Suppose these components are the following sequences:

(1) 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
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Now suppose one is merely asked to find the equivalent of P_p when the key letter is K. Without further specification, the cipher equivalent cannot be stated; for it is necessary to know not only which K will be used as the key letter, the one in the component labeled (1) or the one in the component labeled (2), but also what letter the K_k will be set against, in

order to juxtapose the two components. Most of the time, in the preceding text, these two factors have been tacitly assumed to be fixed and well understood: the K_k is sought in the mixed, or cipher component, and this K is set against A in the normal, or plain component. Thus:



With this setting, $P_p = Z_c$.

c. The letter A in this case may be termed the index letter, symbolized A_i. The index letter constitutes the fourth element involved in the two equations applicable to the finding of equivalents by sliding components. The four elements are therefore these:

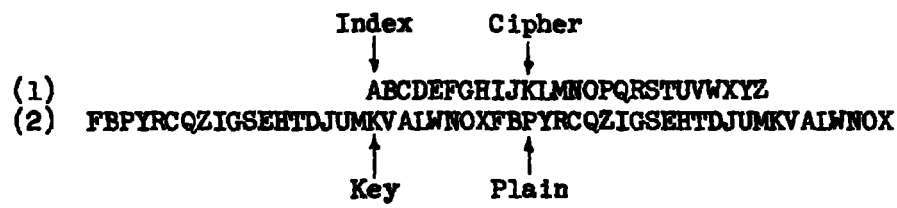
(1) The key letter, θ_k
(2) The index letter, θ_i
(3) The plaintext letter, θ_p
(4) The cipher letter, θ_c

The index letter is commonly the initial letter of the component; but this, too, is only a convention. It might be any letter of the sequence constituting the component, as agreed upon by the correspondents. However, in the subsequent discussion it will be assumed that the index letter is the initial letter of the component in which it is located, unless otherwise stated.

d. In the foregoing case the enciphering equations are as follows:

$$(I) \quad K_k = A_i; \; P_p = Z_c$$

But there is nothing about the use of sliding components which excludes other methods of finding equivalents than that shown above. For instance, despite the labeling of the two components as shown above, there is nothing to prevent one from seeking the plaintext letter in the component labeled (2), that is, the cipher component, and taking as its cipher equivalent the letter opposite it in the other component labeled (1). Thus:



Thus:

$$(II) \quad K_k = A_i; \; P_p = K_c$$

e. Since equations (I) and (II) yield different resultants, even with the same index, key, and plaintext letters, it is obvious that an accurate formula to cover a specific pair of enciphering equations must include data showing in what component each of the four letters comprising the equations is located. Thus, equations (I) and (II) should read:

(I) K_k in component (2) = A_i in component (1); P_p in component (1) = Z_c in component (2).
(II) K_k in component (2) = A_i in component (1); P_p in component (2) = K_c in component (1).

For the sake of brevity, the following notations will be used:

(1) $K_{k/2} = A_{i/1}$; $P_{p/1} = Z_{c/2}$

(2) $K_{k/2} = A_{i/1}$; $P_{p/2} = K_{c/1}$

f. Employing two sliding components and the four letters entering into an enciphering equation, there are, in all, twelve different resultants possible for the same set of components and the same set of four basic elements. These twelve differences in resultants arise from a set of twelve different enciphering conditions,[1] as set forth below (the notation adopted in subpar. e is used):

(1) $\theta_{k/2} = \theta_{i/1}$; $\theta_{p/1} = \theta_{c/2}$

(2) $\theta_{k/2} = \theta_{i/1}$; $\theta_{p/2} = \theta_{c/1}$

(3) $\theta_{k/1} = \theta_{i/2}$; $\theta_{p/1} = \theta_{c/2}$

(4) $\theta_{k/1} = \theta_{i/2}$; $\theta_{p/2} = \theta_{c/1}$

(5) $\theta_{k/2} = \theta_{p/1}$; $\theta_{i/1} = \theta_{c/2}$

(6) $\theta_{k/2} = \theta_{c/1}$; $\theta_{i/1} = \theta_{p/2}$

(7) $\theta_{k/2} = \theta_{p/1}$, $\theta_{i/2} = \theta_{c/1}$

(8) $\theta_{k/2} = \theta_{c/1}$; $\theta_{i/2} = \theta_{p/1}$

(9) $\theta_{k/1} = \theta_{p/2}$; $\theta_{i/1} = \theta_{c/2}$

(10) $\theta_{k/1} = \theta_{c/2}$; $\theta_{i/1} = \theta_{p/2}$

(11) $\theta_{k/1} = \theta_{p/2}$; $\theta_{i/2} = \theta_{c/1}$

(12) $\theta_{k/1} = \theta_{c/2}$; $\theta_{i/2} = \theta_{p/1}$

g. The twelve resultants obtainable from juxtaposing sliding components as indicated under the preceding subparagraph may also be obtained either from one square table, in which case twelve different methods of finding equivalents must be applied, or from twelve different square tables, in which case one standard method of finding equivalents will serve all purposes.

[1] Equations (1) and (2) are the most widely used and are referred to in cryptographic literature as the Vigenère type of encipherment, (5) and (6) are the equations of the Beaufort type, and (9) and (10) are the Delastelle type

h. If but one table such as that shown below in Fig. 9 is employed, the various methods of finding equivalents are difficult to keep in mind.

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

Figure 9.

For example:

(1) For enciphering equations $\theta_k/2 = \theta_i/1$; $\theta_p/1 = \theta_c/2$:

Locate θ_p in top sequence; locate θ_k in first column; θ_c is the letter within the square at intersection of the two lines thus determined.

Thus:

$$K_{k/2} = A_{i/1}; \; P_{p/1} = Z_{c/2}$$

(2) For enciphering equations $\theta_{k/2} = \theta_{i/1}$; $\theta_{p/2} = \theta_{c/1}$:

Locate θ_k in first column; follow line to right to θ_p; proceed up this column; θ_c is letter at top.

Thus:

$$K_{k/2} = A_{i/1};\ P_{p/2} = K_{c/1}$$

(3) For enciphering equations $\theta_{k/1} = \theta_{i/2}$; $\theta_{p/1} = \theta_{c/2}$:

Locate θ_k in top sequence and proceed down column to θ_i; locate θ_p in top sequence; θ_c is letter at other corner of rectangle thus formed.

Thus:

$$K_{k/1} = A_{i/2};\ P_{p/1} = X_{c/2}$$

Only three different methods have been shown and the student no doubt already has encountered difficulty in keeping them segregated in his mind. It would obviously be very confusing to try to remember all twelve methods, but fortunately this is not necessary. If one standard or fixed method of finding equivalents is followed with several different tables, this difficulty disappears. Suppose that the following method is adopted: Arrange the square so that the plaintext letter may be sought in a separate sequence, arranged alphabetically, above the square and so that the key letter may be sought in a separate sequence, also arranged alphabetically, to the left of the square; look for the plaintext letter in the top row; locate the key letter in the 1st column to the left; find the letter standing within the square at the intersection of the vertical and horizontal lines thus determined. Then twelve squares, equivalent to the twelve different conditions listed in subpar. f, can readily be constructed. However, to avoid confusing the student with a multiplicity of unnecessary details which have no direct bearing upon basic principles, one and only one standard method of finding equivalents by means of sliding components will be selected from among the twelve available, as set forth in the preceding subparagraphs. Unless otherwise stated, this method will be the one denoted by the first of the formulas listed in subpar. f, viz.:

$$\theta_{k/2} = \theta_{i/1};\ \theta_{p/1} = \theta_{c/2}$$

Calling the plain component "1" and the cipher component "2", this will mean that the key letter on the cipher component will be set opposite the index, which will be the first letter of the plain component; the plaintext letter to be enciphered will then be sought on the plain component and its equivalent will be the letter opposite it on the cipher component.

CHAPTER III

THEORY OF SOLUTION OF REPEATING-KEY SYSTEMS

14. The three steps in the analysis of repeating-key systems.--a. The method of enciphering according to the principle of the repeating key has been illustrated in subpar. 3b. The analysis of a cryptogram produced by a periodic polyalphabetic system, regardless of the kind of cipher alphabets employed, or their method of production, resolves itself into three distinct and successive steps:

(1) Determination of the length of the repeating key, which is the same as the determination of the exact number of alphabets involved in the cryptogram.

(2) Allocation or distribution of the letters of the cipher text into the respective cipher alphabets to which they belong. This is the step which reduces the polyalphabetic text to monoalphabetic terms.

(3) Analysis of the individual monoalphabetic distributions to determine plaintext values of the cipher letters in each distribution or alphabet.

b. The foregoing steps will be treated in the order in which mentioned. The first step may be described briefly as that of determining the period. The second step may be described briefly as that of reduction to monoalphabetic terms. The third step may be designated as identification of cipher-text values.

15. First step: finding the length of the period.--a. The determination of the period, that is, the length of the key or the number of cipher alphabets involved in a cryptogram enciphered by the repeating-key method is, as a rule, a relatively simple matter. The cryptogram itself usually manifests externally certain phenomena which are the direct result of the use of a repeating key. The principles involved are, however, so fundamental in cryptanalysis that their elucidation warrants a somewhat detailed treatment. This will be done in connection with a short example of encipherment, shown in Fig. 10.

Message

THE ARTILLERY BATTALION MARCHING IN THE REAR OF THE ADVANCE GUARD KEEPS ITS COMBAT TRAIN WITH IT INSOFAR AS PRACTICABLE.

(Key: BLUE, using direct standard alphabets)

Cipher Alphabets

Plain.....	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
(1)....	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	A
(2)....	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
(3)....	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
(4)....	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D

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

Figure 10.

Cryptogram:

U S Y E S E C P M P L C C L N X B W C S O X U V D S C R H T

H X I P L I B C I J U S Y E E G U R D P A Y B C X O F P J W

J E M G P X V E U E L E J Y Q M U S C X J Y M S G L L E T A

L E D E C G B M F I

b. Regardless of what system is used, identical plaintext letters enciphered by the same cipher alphabet[1] must yield identical cipher letters. Referring to Fig. 10, such a condition is brought about every time that identical plaintext letters happen to be enciphered with the same key letter (i.e., every time identical plaintext letters fall into the same column in the encipherment).[2] Now since the number of columns or positions with respect to the key is very limited (except in the case of very long key words), and since the repetition of letters is an inevitable condition in plain text, it follows that there will be in a message of fair length many cases where identical plaintext letters must fall into the same column. They will thus be enciphered by the same cipher alphabet, resulting, therefore, in the production of many identical letters in the cipher text and these will represent identical letters in the plain text. When identical plaintext polygraphs fall into identical columns the result is the formation of identical ciphertext polygraphs, that is, repetitions of groups of 2, 3, 4, . . . letters are exhibited in the cryptogram. Repetitions of this type will hereafter be called causal repetitions, because they are produced by a definite, traceable cause, viz., the encipherment of identical letters by the same cipher alphabets.

c. It will also happen, however, that different plaintext letters falling in different columns will, by mere accident, produce identical cipher letters. Note, for example, in Fig. 10 that in the first column (under B_k), R_p becomes S_c and that in the second column (under L_k), H_p also becomes S_c. The production of an identical ciphertext letter in these two cases (that is, a repetition where the plaintext letters are different alphabets) is merely fortuitous. It is, in everyday language, "a mere coincidence", or "an accident." For this reason repetitions of this type will hereafter be called accidental repetitions.

d. A consideration of the phenomenon pointed out in subpar. c makes it obvious that in polyalphabetic ciphers it is important that the cryptanalyst be able to tell whether the repetitions he finds in a specific

[1] It is to be understood of course that cipher alphabets with single equivalents are meant in this case

[2] The frequency with which this condition may be expected to occur can be definitely calculated A discussion of this point will be treated in Military Cryptanalytics, Part III

case are causal or accidental in their origin, that is, whether they represent actual encipherments of identical plaintext letters by identical keying elements, or mere coincidences brought about purely fortuitously.

e. Now accidental repetitions will, of course, happen fairly frequently with individual letters, but less frequently with digraphs, because in this case the same kind of an "accident" must take place twice in succession. Intuitively one feels that the chances that such a purely fortuitous coincidence will happen two times in succession must be much less than that it will happen every once in a while in the case of single letters. Similarly, intuition makes one feel that the chances of such accidents happening in the case of three or more consecutive letters are still less than in the case of digraphs, decreasing very rapidly as the repetition increases in length.

f. The phenomena of cryptographic repetition may, fortunately, be dealt with statistically, thus taking the matter outside the realm of intuition and putting it on a firm mathematical or objective basis. Moreover, often the statistical analysis will tell the cryptanalyst when he has arranged or rearranged his text properly, that is, when he is approaching or has reached monoalphabeticity in his efforts to reduce polyalphabetic text to its simplest terms. By means of the binomial distribution,[3] it is possible to compute tables of the expected number of digraphs, trigraphs, and other polygraphs occurring exactly 0, 1, 2, 3, . . . x times in a sample of random text of a given size; then the repetitive phenomena in a cryptogram under study may be compared with the phenomena expected by pure chance (i.e., in samples of random text of the same size as the cryptogram) as a means of evaluating whether or not the observed repetitions and their number are significant. If the observed repetitive phenomena are no more than would normally be expected by chance, then these phenomena cannot be used as a basis for cryptanalytic attack, if however these repetitions are highly unlikely to have occurred by chance, then they are open to interpretation and exploitation. The tables derived from the binomial distribution are given in subpar. g, below.[4]

---

[3] This distribution, as well as the Poisson exponential distribution (which is an approximation to the binomial) will be treated in Military Cryptanalytics Part III

[4] The tables illustrated here have been computed using the formula for the number of comparisons as $\frac{(N-t+1)(N-t)}{2}$ where N is the number of letters in the sample size and t is the length of the polygraph

Strictly speaking, the formula $\frac{(N-2t+2)(N-2t+1)}{2}$ should be used to discount overlapping repetitions such as the "repeated tetragraph' in the sequence ABCABCA, however in most statistical computations especially where analytical machine techniques are employed the scoring is almost invariably predicated upon the first formula The two formulas are practically equivalent, except for small values of N when the second formula is the more precise one for the number of comparisons

g. Fig. 11a is a table of the expected number of digraphs appearing exactly 2, 3, 4, . . . 10 times in samples of random text of sizes 100 to 1,000 letters, by hundreds. Fig. 11b is a table of the expected number of trigraphs appearing exactly 2, 3, and 4 times by chance in these sample sizes, and Figs. 11c and d contain the data for tetragraphs and pentagraphs, respectively. As an illustration of the use of these tables, from Fig. 11a

No. of Letters	Expected number of digraphs occurring exactly x times E(2)	E(3)	E(4)	E(5)	E(6)	E(7)	E(8)	E(9)	E(10)
100	6.21	0.298	0.011						
200	21.8	2.12	0.154	0.009					
300	42.5	6.23	0.683	0.060	0.004				
400	65.3	12.8	1.87	0.220	0.022	0.002			
500	88.1	21.6	3.97	0.582	0.071	0.008			
600	110	32.3	7.11	1.25	0.184	0.023	0.003		
700	129	44.3	11.4	2.35	0.403	0.059	0.008	0.001	
800	145	57.1	16.8	3.96	0.777	0.130	0.019	0.003	
900	158	70.1	23.2	6.16	1.36	0.257	0.043	0.006	0.001
1000	169	83.0	30.6	9.03	2.21	0.466	0.085	0.014	0.002

Figure 11a.

No. of letters	Exp. no. of trigraphs E(2)	E(3)	E(4)
100	0.269	0.001	
200	1.10	0.004	
300	2.48	0.014	
400	4.40	0.033	
500	6.85	0.064	
600	9.81	0.111	0.001
700	13.3	0.175	0.002
800	17.3	0.261	0.003
900	21.8	0.371	0.005
1000	26.8	0.505	0.008

Figure 11b.

No. of letters	Tetragraphs E(2)	E(3)
100	0.010	
200	0.043	
300	0.096	
400	0.171	
500	0.270	
600	0.389	
700	0.530	
800	0.693	
900	0.877	
1000	1.08	0.001

Figure 11c.

No. of letters	Penta. E(2)
100	
200	0.002
300	0.004
400	0.007
500	0.011
600	0.015
700	0.021
800	0.027
900	0.034
1000	0.042

Figure 11d.

we observe that in a sample of 300 letters of random text we may expect about 43 (rounded off to the nearest integer) digraphs to occur twice, 6 digraphs to occur three times, and about 1 digraph to occur four times. The meaning of the decimal fractions in that row under E(4), E(5), and E(6) may be interpreted as follows: the entry 0.683 under E(4) means that in 100 samples of 300 random letters each, about 68 of them will have a digraph occurring 4 times within the sample, the entry 0.060 under E(5)

means that in 100 samples of 300 random letters each, 6 of them will have a digraph occurring 5 times; and the entry 0.004 under E(6) means that in 1,000 samples of 300 random letters, 4 of them may be expected to have a digraph occurring 6 times. From Fig. 11b, we note that a sample of 300 random letters may be expected to contain 2 or 3 trigraphs occurring twice; and in 1,000 of such samples, 14 may be expected to contain a trigraph occurring three times. From Fig. 11c, we note that in 1,000 samples of 300 random letters, 96 of them may be expected to contain a repeated tetragraph, while the chance of a tetragraph occurring three times in these 1,000 samples is so small as to be practically non-existent, note that, under E(3) of the last row of this Figure, if we had 1,000 samples of 1,000 letters each, only 1 of them may be expected to contain a threefold occurrence of a tetragraph. From Fig. 11d, we see that if we had 1,000 samples of 300 random letters, only 4 of them may be expected to contain a pentagraphic repetition (i.e., a pentagraph occurring twice), and that in these 1,000 samples there is, in unmathematical but nevertheless precise language, not a ghost of a chance that a pentagraph will occur three times.

h. The foregoing tables may also be used to determine the cumulative values of digraphs and polygraphs expected to appear x or more times in samples of random text. Using Fig. 11a as an example, in a 300-letter sample of random text, if the entries under E(2) to E(6) are added together, the sum (49.477) indicates that about 49 digraphs may be expected to occur at least twice (i.e., two or more times); if the values E(3) to E(6) are added together, the sum (6.977) shows that 7 digraphs may be expected to occur three or more times; if the entries under E(4) to E(6) are added together, the sum (0.747) shows that in 100 such samples of random text, about 75 of them will contain a digraph occuring at least four times; and if the entries under E(5) and E(6) are added, their sum (0.065) shows that in 1,000 300-letter samples of random text, 65 of these may be expected to contain a digraph occurring five or more times.

i. As an illustration of the application of the foregoing discussion, it is indicated that if a cryptanalyst were to have at hand only the cryptogram of Fig. 10, with the repetitions underlined as below, a statistical study of the number and length of the repetitions within the message would tell him that while some of the digraphic repetitions may be accidental, the chances that they all are accidental are small. In the case of the tetragraphic repetition he would realize that the chances of its being accidental are very small indeed.

U S Y E S E C P M P L C C L N X B W C S O X U V D S C R H T

H X I P L I B C I J U S Y E E G U R D P A Y B C X O F P J W

J E M G P X V E U E L E J Y Q M U S C X J Y M S G L L E T A

L E D E C G B M F I

j. A consideration of the facts therefore leads to but one conclusion, viz., that the repetitions exhibited by the cryptogram under investigation are not accidental but are causal in their origin; and the cause is in this case not difficult to find: repetitions in the plain text were actually enciphered by identical alphabets. In order for this to occur, it was necessary that the tetragraph USYE, for example, fall both times in exactly the same relative position with respect to the key. Note, for example, that USYE in Fig. 10 represents in both cases the plaintext polygraph THEA. The first time it occurred it fell in positions 1-2-3-4 with respect to the key, the second time it occurred it happened to fall in the very same relative positions, although it might just as well have happened to fall in any of the other three possible relative positions with respect to the key, viz., 2-3-4-1, 3-4-1-2, or 4-1-2-3.

k. Lest the student be misled, however, a few more words are necessary on this subject. In the preceding subparagraph the word "happened" was used; this word correctly expresses the idea in mind, because the insertion or deletion of a single plaintext letter between the two occurrences would have thrown the second occurrence one letter forward or backward, respectively, and thus caused the polygraph to be enciphered by a sequence of alphabets such as can no longer produce the cipher polygraph USYE from the plaintext polygraph THEA. On the other hand, the insertion or deletion of this one letter might bring the letters of some other polygraph into similar columns so that some other repetition would be exhibited in case the USYE repetition had thus been suppressed.

l. The encipherment of similar letters by similar cipher alphabets is therefore the cause of the production of repetitions in the cipher text in the case of repeating-key ciphers. What principles can be derived from this fact, and how can they be employed in the solution of cryptograms of this type?

m. If a count is made of the number of letters from and including the first USYE to, but not including, the second occurrence of USYE, a total of 40 letters is found to intervene between the two occurrences. This number, 40, must, of course, be an exact multiple of the length of the key.[4] Having the plain text before one, it is easily seen that it is the 10th multiple; that is, the 4-letter key has repeated itself 10 times between the first and the second occurrence of USYE. It follows, therefore, that if the length of the key were not known, the number 40 could safely be taken to be an exact multiple of the length of the key; in other words, one of the factors of the number 40 would be equal to the length of the key. The word "safely" is used in the preceding sentence to mean that the interval 40 applies to a repetition of 4 letters and it has been shown that the chances are small that this repetition is accidental. The factors of 40 are 2, 4, 5, 8, 10, and 20. So far as this single repetition of USYE is concerned, if the length of the key were not known, all that could be said about the latter would be

[4] Barring that is cases such as those mentioned in subpar 16b and in footnote 6 on p 32

that it is equal to one of these factors. The repetition by itself gives no further indications. How can the exact factor be selected from among a list of several possible factors?

n. Let the intervals between all the repetitions in the cryptogram be listed. They are as follows:

Repetition	Interval	Factors
1st USYE to 2d USYE	40	2, 4, 5, 8, 10, 20
1st BC to 2d BC	16	2, 4, 8
1st CX to 2d CX	25	5
1st EC to 2d EC	88	2, 4, 11, 22, 44
1st LE to 2d LE	16	2, 4, 8
2d LE to 3d LE	4	2
1st LE to 3d LE	20	2, 4, 5, 10
1st JY to 2d JY	8	2, 4
1st PL to 2d PL	24	2, 3, 4, 6, 8, 10, 12
1st SC to 2d SC	52	2, 4, 13, 26
(1st SY to 2d SY, already included in USYE.)		
(1st US to 2d US, already included in USYE.)		
2d US to 3d US	36	2, 3, 4, 6, 9, 18
1st US to 3 US	76	2, 4, 19, 38
(1st YE to 2d YE, already included in USYE.)		

o. Are all these repetitions causal repetition? Since, from Fig. 11c, we find that the expected number of tetragraphs appearing twice (i.e., a tetragraphic repetition) in 100 letters of random text is 0.01, this decimal fraction means that in 100 such cases only 1 may be expected to contain a repeated tetragraph; thus it is a 99-to-1 chance of the USYE repetitions occurring accidentally. From Fig. 11a, the expected number of digraphs occurring 3 times is 0.298; so the chances against the threefold occurrence of the two digraphs LE and US are quite high. We expect only 6.5 digraphs to occur at least 2 times in 100 letters of random text, but in our sample we have 10 digraphs appearing two or more times. The chances are very great, therefore, that the majority of these repetitions are causal, so that it is not astonishing that the intervals betweeen all the various repetitions, except in one case, contain the factors 2 and 4.

p. This means that if the cipher is written out in either 2 columns or 4 columns, all these repetitions (except the CX repetition) would fall into the same columns. From this it follows that the length of the key is either 2 or 4, the latter, on practical grounds, being more probable than the former. Doubts concerning the matter of choosing between a 2-letter and a 4-letter key will be dissolved when the cipher text is distributed into its component uniliteral frequency distributions.

q. The repeated digraph CX in the foregoing message is an accidental repetition, as will be apparent by referring to Fig. 10. Had the message been longer there would have been more such accidental repetitions, but, on the other hand, there would be a proportionately greater number of causal repetitions. This is because the phenomenon of repetition in plain text is so all-pervading.

r. Sometimes it happens that the cryptanalyst quickly notes a repetition of a polygraph of four or more letters, the interval between the first and second occurrences of which has only two factors, of which one is a relatively small number, the other a relatively high prime number.[5] He may therefore assume at once that the length of the key is equal to the smaller factor without searching for additional recurrences upon which to corroborate his assumption. Suppose, for example, that in a relatively short cryptogram the interval between the first and second occurrences of a polygraph of five letters happens to be a number such as 203, the factors of which are 7 and 29. Evidently the number of alphabets may at once be assumed to be 7, unless one is dealing with messages exchanged among correspondents known to use long keys. In the latter case one could assume the number of alphabets to be 29.

s. The foregoing method of determining the period in a polyalphabetic cipher is commonly referred to in the literature as "factoring the intervals between repetitions", or more often it is simply called "factoring." Because the latter is an apt term and is brief, it will be employed hereafter in this text to designate the process.

t. As an aid in the determination of possible periods in cases under study, there is given at the end of this chapter a table of the factors of all numbers from 1 to 400 inclusive (cf. pp. 41 - 44).

16. General remarks on factoring.--a. The statement made in par. 4 with respect to the cyclic phenomena said to be exhibited in cryptograms of the periodic type now becomes clear. The use of a short repeating key produces a periodicity of recurrences or repetitions collectively termed "cyclic phenomena", an analysis of which leads to a determination of the length of the period or cycle, and this gives the length of the key. Save for a rare exception mentioned below, only in the case of relatively short cryptograms enciphered by a relatively long key does factoring fail to lead to the correct determination of the number of cipher alphabets in a repeating-key cipher; and of course, the fact that a cryptogram contains repetitions whose factors show constancy is in itself an indication and test of its periodic nature. It also follows that if the cryptogram is not a repeating-key cipher, then factoring will show no definite results, and conversely the fact that it does not yield definite results at once indicates that the cryptogram is not a periodic, repeating-key cipher.

---

[5] A prime number is defined as one which is exactly divisible only by itself and 1

b. There are two main cases in which factoring leads to no definite results. One is in the case of monoalphabetic substitution ciphers. Here recurrences are very plentiful as a rule, and the intervals separating these recurrences may be factored, but the factors will show no constancy, there will be several factors common to many or most of the recurrences. This in itself is an indication of a monoalphabetic (monographic) substitution cipher, if the very fact of the presence of many recurrences fails to impress itself upon the inexperienced cryptanalyst. The other case in which the process of factoring is nonsignificant involves certain types of nonperiodic, polyalphabetic ciphers. In certain of these ciphers, recurrences of digraphs, trigraphs, and even longer polygraphs may be plentiful in a long message, but the intervals between such recurrences bear no definite multiple relation to the length of the key, such as in the case of the true periodic, repeating-key cipher, in which the alphabets change with successive letters and repeat themselves over and over again.[6]

c. Factoring is not the only method of determining the length of the period of a periodic, polyalphabetic substitution cipher, although it is by far the most common and easily applied. At this point it will merely be stated that when the message under study is relatively short in comparison with the length of the key, so that there are only a few cycles of cipher text and no long repetitions affording a basis for factoring, there are several other methods available. However, it being deemed inadvisable to interject the data concerning those other methods at this point, they will be explained subsequently. It is desirable at this juncture merely to indicate that methods other than factoring do exist and are used in practical work.

d. Fundamentally, the factoring process is merely a more-or-less simple mathematical method of studying the phenomena of periodicity in cryptograms. It will usually enable the cryptanalyst to ascertain definitely whether or not a given cryptogram is periodic in nature, and if so, the length of the period, stated in terms of the cryptographic unit involved. By the latter statement is meant that the factoring process may be applied not only in analyzing the periodicity manifested by cryptograms in which the plaintext units subjected to cryptographic treatment are monographic in nature (i.e., are single letters) but also in studying the periodicity exhibited by those occasional cryptograms wherein the plaintext units are digraphic, trigraphic, or n-graphic in character. The student should bear this point in mind when he comes to the study of substitution systems of the latter sort.

---

[6] One further case which might be mentioned is that of periodic ciphers in which the key word or phrase contains repeated polygraphic segments, such as in NATIONALORGANIZATION A five-letter word enciphered in alphabets 2-6 will have the same ciphertext equivalents as the same word enciphered in alphabets 16-20, thus giving rise to a causal repetition, but the interval between the occurrences will not reflect the length of the true period Such repetitions are referred to as being the result of the sub-cycles in the total key, these phenomena are often encountered in the study of certain machine cipher systems

17. Second step: distributing the cipher text into the component monoalphabets.--a. After the number of cipher alphabets involved in the cryptogram has been ascertained, the next step is to rewrite the message in groups corresponding to the length of the key, or in columnar fashion, whichever is more convenient, and this automatically divides up the text so that the letters belonging to the same cipher alphabet occupy similar positions in the groups, or, if the columnar method is used, fall in the same column. The letters are thus allocated or distributed into the respective cipher alphabets to which they belong. This reduces the polyalphabetic text to monoalphabetic terms.

b. Then separate uniliteral frequency distributions for the thus isolated individual alphabets are compiled. For example, in the case of the cipher in subpar. 15i, having determined that four alphabets are involved, and having rewritten the message in four columns, a frequency distribution is made of the letters in the first column, another is made of the letters in the second column, and so on for the four columns. Each of the resulting distributions is therefore a monoalphabetic frequency distribution. If these distributions do not give the characteristic irregular crest-and-trough appearance of monoalphabetic frequency distributions, including the expected number of blanks (Λ), and if the ϕ_o of these distributions do not meet the range of the expected value of ϕ_p (or do not yield I.C.'s in the close vicinity of the expected value of 1.73), then the analysis which led to the hypothesis as regards the number of alphabets involved is fallacious. In fact, the ϕ or I.C. of these individual distributions may be considered to be an index of the correctness of the factoring process; for theoretically, and practically, the individual distributions constructed upon the correct hypothesis will tend to conform more closely to the expected ϕ or I.C. of a monoalphabetic frequency distribution than will the distributions constructed upon an incorrect hypothesis. These considerations will be discussed in the next paragraph.

18. Statistical proof of the monoalphabeticity of the distributions. --a. The student is already familiar with the monographic ϕ test for determining the relative monoalphabeticity of a distribution; this test was discussed in detail in par. 27 of Military Cryptanalytics, Part I. The formulas for monographic ϕ_p and ϕ_r were stated as $\phi_p = .0667N(N-1)$, and $\phi_r = .0385N(N-1)$, where N is the total number of elements in the distribution. The ϕ_o was calculated by the formula $\phi_o = \Sigma f(f-1)$, where f is the frequency of each element of the distribution. The I.C. was defined as the ratio of ϕ_o to ϕ_r; the monographic I.C. of English plain text was given as 1.73, as compared with the I.C. of 1.00 for random text.[7]

b. The ϕ test may be applied to the distributions of periodic polyalphabetic ciphers to confirm the monoalphabeticity of the distributions (made on an hypothesis of n alphabets) and thereby confirm the length of the period; this test is particularly applicable in difficult cases, as for

[7] A more convenient formula for the monographic I C is $\frac{26\Sigma f(f-1)}{N(N-1)}$, which is equivalent to $\frac{\phi_o}{\phi_r}$

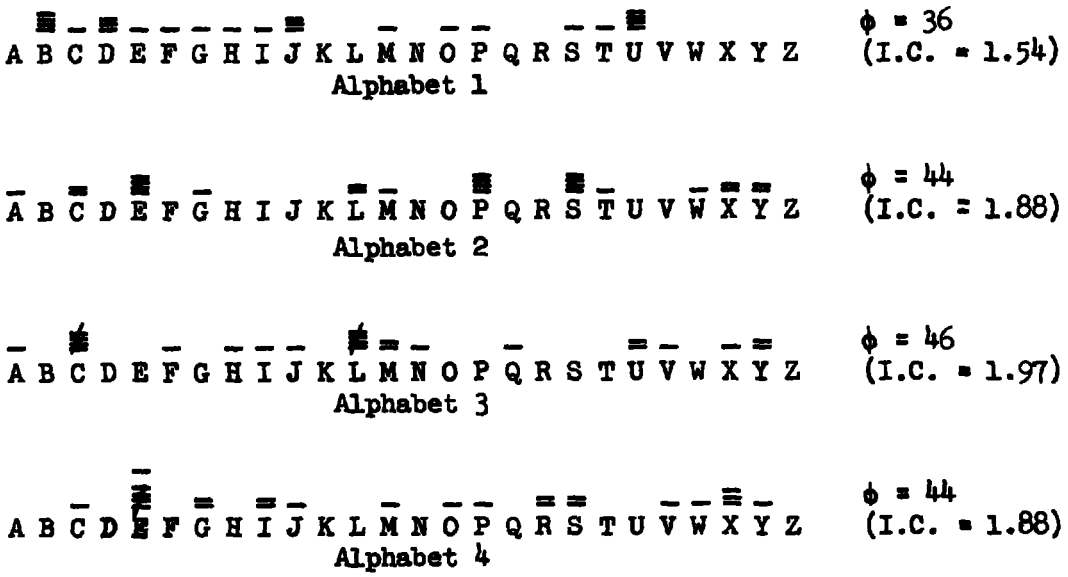
instance where there are insufficient polygraphic repetitions in a short text, or where the factoring resolves itself into two or more periods. If the correct period is assumed, then the ϕ test applied to each of the alphabets should approximate fairly closely and consistently the value for ϕ_p; and, conversely, if an incorrect period is assumed, the ϕ_o should approximate the value of ϕ_r more than it does ϕ_p. It is to be remarked that small deviations for the expected values are usual, and indeed is the normal situation, but that large deviations are rare. The degree of deviation that may be expected may be determined by statistical means, under the concept of the standard deviation of ϕ; however, this topic is reserved for treatment in Military Cryptanalytics, Part III.

c. For reference purposes, there is appended in Fig. 12, below, a table of the expected values of ϕ_p and ϕ_r for sample sizes (N) from 11 to 100, inclusive. Since the I.C. is an expression of monoalphabeticity in terms of a ratio, the evaluation of distributions will be expressed in terms of the I.C. in the future, unless the ϕ values are more convenient in specific cases.

N	ϕ_p	ϕ_r	N	ϕ_p	ϕ_r	N	ϕ_p	ϕ_r	N	ϕ_p	ϕ_r	N	ϕ_p	ϕ_r
11	7.34	4.23	29	54	31	47	144	83	65	277	160	83	454	262
12	8.80	5.08	30	58	33	48	150	87	66	286	165	84	465	268
13	10.4	6.00	31	62	36	49	157	90	67	295	170	85	476	275
14	12.1	7.00	32	66	38	50	163	94	68	304	175	86	488	281
15	14.0	8.08	33	70	41	51	170	98	69	313	180	87	499	288
16	16.0	9.23	34	75	43	52	177	102	70	322	186	88	511	294
17	18.1	10.5	35	79	46	53	184	106	71	331	191	89	522	301
18	20.4	11.8	36	84	48	54	191	110	72	341	197	90	534	308
19	22.8	13.2	37	89	51	55	198	114	73	351	202	91	546	315
20	25.3	14.6	38	94	54	56	205	118	74	360	208	92	558	322
21	28.0	16.2	39	99	57	57	213	123	75	370	213	93	571	329
22	30.8	17.8	40	104	60	58	221	127	76	380	219	94	583	336
23	33 8	19.5	41	109	63	59	228	132	77	390	225	95	596	343
24	36.8	21.2	42	115	66	60	236	136	78	401	231	96	608	351
25	40.0	23.1	43	120	69	61	244	141	79	411	237	97	621	358
26	43.4	25.0	44	126	73	62	252	145	80	422	243	98	634	366
27	46.8	27.0	45	132	76	63	261	150	81	432	249	99	647	373
28	50.4	29.1	46	138	80	64	269	155	82	443	255	100	660	381

Figure 12.

d. The uniliteral distributions for the four columns of cipher text in Fig. 10 are made, and are shown below together with the ϕ and I.C. values for these distributions:



It can be seen that these distributions are clearly monoalphabetic, since the ϕ values are in the vicinity of ϕ_p (which, for this size sample, is 40) rather than close to ϕ_r (in this case, 23). It is pointed out that any other assumed periods, which are not a multiple of 4, will not yield monoalphabetic distributions. But if in this case a period of 8 were assumed, these 8 distributions would be monoalphabetic, however it would be noticed that the distributions for alphabets 1 and 5 would be quite similar, and likewise alphabets 2 and 6, 3 and 7, and 4 and 8--therefore these 8 distributions could then be combined into the four distributions of the correct hypothesis.[8]

e. A further application of the ϕ test should be considered at this time, that is, the cases where the length of the period is large as compared with the size of the sample of cipher text.[9] Although a single ϕ test for small values of N would rarely give a reliable result, it is nevertheless possible to apply this test for small values of N, provided it is possible to obtain the average for a number of such tests. Thus it is usually possible to determine the period of a polyalphabetic cipher, where the number of alphabets is large and the number of letters per distribution is small, even though there are no long repetitions.

---

[8] Cf the discussion on the χ (chi) test in par 37

[9] The discussion which follows and the example illustrated are based on certain parts of the pioneer work of Dr Solomon Kullback in his treatise Statistical Methods in Cryptanalysis (Revised Edition), Washington 1938

(1) Let us consider the following cryptogram which is known to be enciphered by periodic polyalphabetic substitution, where the number of alphabets is between 40 and 50:

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

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

If we assume a period of 50, the cryptogram would be written out on this width, as illustrated below; the ϕ_o value for each column is obtained and is included in the diagram:

	1	1	2	2	3	3	4	4	5
5	0	5	0	5	0	5	0	5	0

HSKUSPMFHDUJJIXMSPTPOIPCIWKZVUYPPNEUSAIGBOOGAOPGPR
HBOUCSHPVGHQXZSACKRKVBGHMVSFRYYTKHKVWZXVLIJHWARLKF
IJSLTMHKAHQTUVTXSMECFCSKTGOOYBXZVLIJRYACDWEJMSCAFP
IEAXOKAQDWEXPYPQHDNOJIXNZJGNUDOARFUERJOYBDOKEIKDUV
TDVEVLETDOAFROUNYNBDVQOBEGGSHQHXOPUZCOCUKKZLTPHKRT
CCOASBZUGBUBBUNOVTPOVMIZDEPQFVKZ
40222020202000002002620002200022002020002020000002

We note that there are 32 columns wherein N = 6, and 18 columns wherein N = 5; we also note that the ϕ values range between ϕ and 6. In the diagram below, keeping the data from the two categories of N separate, the column labelled "ϕ" is the observed value of ϕ; the column labelled "x"

N = 6		
ϕ	x	ϕx
0	17	0
2	13	26
4	1	4
6	1	6
	32	36

N = 5		
ϕ	x	ϕx
0	13	0
2	5	10
4	0	0
6	0	0
	18	10

is the number of times the particular value of ϕ occurred; and the column labelled "ϕx" is the product of the preceding two columns (given as a means of arriving at the average value of ϕ). The average value of ϕ (symbolized by $\overline{\phi}$, which is read as "ϕ bar") where N = 6 is $\frac{36}{32}$, or 1.13; the $\overline{\phi}$ where N = 5 is $\frac{10}{18}$, or 0.56; in other words, the average value of ϕ is derived by adding up all the ϕ_o values for a given column length, and then dividing by the number of columns of that length. The values of $\overline{\phi}$ are compared with

ϕ_p and ϕ_r below, and, since $\bar{\phi}$ more closely approximates ϕ_r than it does ϕ_p, the conclusion is reached that the period of the cryptogram is not 50.[10]

	N = 6	N = 5
$\bar{\phi}$	1.13	0.56
ϕ_p	2.00	1.33
ϕ_r	1.15	0.77

(2) Since we have discarded 50 as a possible period, we write the cryptogram on a width of 49 and examine its statistics; if this width fails, then we write the cryptogram on widths of 48, 47, ...etc. The results of assumptions of widths from 49 down to 44 are shown in the diagram below:

	49 alphabets		48 alphabets		47 alphabets
	N = 6	N = 5	N = 6	N = 5	N = 6
$\bar{\phi}$	1.57	0.67	1.33	0.33	0.85
ϕ_p	2.00	1.33	2.00	1.33	2.00
ϕ_r	1.15	0.77	1.15	0.77	1.15

	46 alphabets		45 alphabets		44 alphabets	
	N = 7	N = 6	N = 7	N = 6	N = 7	N = 6
$\bar{\phi}$	1.00	1.15	1.50	0.69	1.22	1.23
ϕ_p	2.80	2.00	2.80	2.00	2.80	2.00
ϕ_r	1.66	1.15	1.66	1.15	1.66	1.15

These assumed periods are discarded one by one because $\bar{\phi}$ did not come up to plaintext expectations.

[10] Kullback ibid, p 42, carries the analysis further by considering the sigmages of the deviations, and expresses these in terms of probability statements (assuming for convenience that this sigmage is normally distributed) He shows that, for N = 6, only 15% of monoalphabetic text would yield a value of $\bar{\phi}$ as small as or smaller than 1 13, whereas 52% of random text would yield a value of $\bar{\phi}$ as large as or larger than 1 13, for N = 5, only 35% of monoalphabetic text would yield a value of $\bar{\phi}$ as small as or smaller than 0 56, whereas 75% of random text would yield a value of $\bar{\phi}$ as large as or larger than that observed This is more compactly illustrated in the diagram below·

	N = 6	N = 5
$\bar{\phi}$, N letters of monoalphabetic text	$P(\bar{\phi} \leq 1\ 13) = .15$	$P(\bar{\phi} \leq 0\ 56) = .35$
$\bar{\phi}$, N letters of random text	$P(\bar{\phi} \geq 1\ 13) = .52$	$P(\bar{\phi} \geq 0\ 56) = .75$

(3) Now testing for a period of 43, we write out the cryptogram on this width and take the ϕ_o of the columns, as is illustrated below:

```
      1   1   2   2   3   3   4  4
  5   0   5   0   5   0   5   0  3
HSKUSPMFHDUJJIXMSPTPOIPCIWKZVUYPPNEUSAIGBOO
GAOPGPRHBOUCSHPVGHQXZSACKRKVBGHMVSFRYYTKHKV
WZXVLIJHWARLKFIJSLTMHKAHQTUVTXSMECFCSKTGOOY
BXZVLIJRYACDWEJMSCAFPIEAXOKAQDWEXPYPQHDNOJI
XNZJGNUDOARFUERJOYBDOKEIKDUVTDVEVLETDOAFROU
NYNBDVQOBEGGSHQHXOPUZCOCUKKZLTPHKRTCCOASBZU
GBUBBUNOVTPOVMIZDEPQFVKZ
20244424264024246040444620482204204222424 62
```

The data from this hypothesis are tabulated in the following diagram:

N = 7		
ϕ	x	ϕx
0	4	0
2	6	12
4	11	44
6	3	18
	24	74

N = 6		
ϕ	x	ϕx
0	3	0
2	9	18
4	4	16
6	1	6
8	1	8
14	1	14
	19	62

$\bar{\phi} = \frac{74}{24} = 3.08$ $\bar{\phi} = \frac{62}{19} = 3.26$

The comparison of the $\bar{\phi}$ values with the values for ϕ_p and ϕ_r is illustrated in the following diagram:

	N = 7	N = 6
$\bar{\phi}$	3.08	3.26
ϕ_p	2.80	2.00
ϕ_r	1.66	1.15

There seems to be no doubt that the period of the cryptogram is 43.

(4) In this example, it is noted that ϕ counts were made of the columns of the write-outs of the cryptogram on the various widths, in this case, ϕ's are more convenient to use than their translation in terms of the decimal values of I.C.'s. If however we had much <u>longer</u> columns, the I.C. values might give a quicker portrayal of the relative goodness of the columns. Note that we could have used I.C.'s in expressing the <u>final</u> result

of each assumption of a period, as follows:

	50 alphabets		49 alphabets		48 alphabets		47 alphabets
	N = 6	N = 5	N = 6	N = 5	N = 6	N = 5	N = 6
I.C.	0.98	0.73	1.37	0.87	1.16	0.43	0.74

	46 alphabets		45 alphabets		44 alphabets		43 alphabets	
	N = 7	N = 6	N = 7	N = 6	N = 7	N = 6	N = 7	N = 6
I.C.	0.60	1.00	0.90	0.60	0.73	1.07	1.86	2.83

It will be observed that the data may be evaluated much more rapidly in the case of I.C.'s, since we need to keep only one invariant index (1.73) in mind in comparing our results with the expected value of the I.C. for English plain text.[11]

---

[11] In the calculations of this problem Kullback treated the statistics for the long and short columns separately, this was done primarily to make easier the determination of the sigmages of the deviations However, it is possible to derive a single statistic (either ϕ or I C) for each hypothesis of a period, eliminating the necessity of looking at two sets of data for each assumption of a period

For instance in the hypothesis of 50 alphabets, the total number of comparisons is $32\left(\frac{6x5}{2}\right)+18\left(\frac{5x4}{2}\right)=660$ thus the ϕ values (which by definition are twice the expected number of coincidences) are the following

$$\phi_p = 2(660)(.0667) = 88$$
$$\phi_r = 2(660)(.0385) = 51$$

The observed ϕ is the sum of the ϕ values for the long and short columns, so that $\phi_o = 36 + 10 = 46$ The I C is thus $\frac{46}{51} = 0.90$

In the hypothesis of 43 alphabets, the number of comparisons is $24\left(\frac{7x6}{2}\right)+19\left(\frac{6x5}{2}\right)$, thus

$$\phi_p = \frac{2(789)}{15} = 105$$
$$\phi_r = \frac{2(789)}{26} = 61$$
$$\phi_o = 74 + 62 = 136$$

The I C of $\frac{136}{61} = 2.23$ points to this hypothesis as the correct period (Note that .0667 is approximately equal to $\frac{1}{15}$ so that computation is facilitated by using $\frac{1}{15}$ and $\frac{1}{26}$ instead of .0667 and .0385 for the plaintext and random constants respectively)

Having established the period as 43 solution of this example is predicated on the assumption that standard alphabets are involved, this procedure will be taken up in par 21 in the next chapter

19. Third step: solving the monoalphabetic distributions.--The difficulty experienced in analyzing the individual or isolated frequency distributions depends mostly upon the type of cipher alphabets that is used. It is apparent that mixed alphabets may be used just as easily as standard alphabets, and, of course, the cipher letters themselves give no indication as to which is the case. However, just as it was found that in the case of monoalphabetic substitution ciphers, a uniliteral frequency distribution gives clear indications as to whether the cipher alphabet is a standard or a mixed alphabet, by the relative positions and extensions of the crests and troughs in the table, so it is found that in the case of repeating-key ciphers, uniliteral frequency distributions for the isolated or individual alphabets will also give clear indications as to whether these alphabets are standard alphabets or mixed alphabets. Only two or three such frequency distributions are necessary for this determination;[12] if they appear to be standard alphabets, similar distributions can be made for the rest of the alphabets; but if they appear to be mixed alphabets, then it is best to compile triliteral frequency distributions for all the alphabets. The analysis of the values of the cipher letters in each table proceeds along the same lines as in the case of monoalphabetic ciphers. The analysis is more difficult only because of the reduced size of the tables, but if the message be very long, then each frequency distribution will contain a sufficient number of elements to enable a speedy solution to be achieved.

[12] In certain mixed-alphabet periodic ciphers, it is possible that a distribution for one alphabet might reflect the phenomena expected for a standard alphabet, for instance, if the plain- and cipher components are identically-mixed sequences running in the same direction, and if for one alphabet $A_p = A_c$, then the distribution for that alphabet will be the normal uniliteral distribution It is for this reason that we must make distributions for at least two alphabets to determine whether or not a polyalphabetic cipher is composed of standard alphabets

Table of Factors

Numbers 1-400 inclusive

1	
2	
3	
4	2
5	
6	2 3
7	
8	2 4
9	3
10	2 5
11	
12	2 3 4 6
13	
14	2 7
15	3 5
16	2 4 8
17	
18	2 3 6 9
19	
20	2 4 5 10
21	3 7
22	2 11
23	
24	2 3 4 6 8 12
25	5
26	2 13
27	3 9
28	2 4 7 14
29	
30	2 3 5 6 10 15
31	
32	2 4 8 16
33	3 11
34	2 17
35	5 7
36	2 3 4 6 9 12 18
37	
38	2 19
39	3 13
40	2 4 5 8 10 20
41	
42	2 3 6 7 14 21
43	
44	2 4 11 22
45	3 5 9 15
46	2 23
47	
48	2 3 4 6 8 12 16 24
49	7
50	2 5 10 25
51	3 17
52	2 4 13 26
53	
54	2 3 6 9 18 27
55	5 11
56	2 4 7 8 14 28
57	3 19
58	2 29
59	
60	2 3 4 5 6 10 12 15 20 30
61	
62	2 31
63	3 7 9 21
64	2 4 8 16 32
65	5 13
66	2 3 6 11 33
67	
68	2 4 17 34
69	3 23
70	2 5 7 10 14 35
71	
72	2 3 4 6 8 9 12 18 24 36
73	
74	2 37
75	3 5 15 25
76	2 4 19 38
77	7 11
78	2 3 6 13 26 39
79	
80	2 4 5 8 10 16 20 40
81	3 9 27
82	2 41
83	
84	2 3 4 6 7 12 14 21 28 42
85	5 17
86	2 43
87	3 29
88	2 4 8 11 22 44
89	
90	2 3 5 6 9 10 15 18 30 45
91	7 13
92	2 4 23 46
93	3 31
94	2 47
95	5 19
96	2 3 4 6 8 12 16 24 32 48
97	
98	2 7 14 49
99	3 9 11 33
100	2 4 5 10 20 25 50

101	
102	2 3 6 17 34 51
103	
104	2 4 8 13 26 52
105	3 5 7 15 21 35
106	2 53
107	
108	2 3 4 6 9 12 18 27 36 54
109	
110	2 5 10 11 22 55
111	3 37
112	2 4 7 8 14 16 28 56
113	
114	2 3 6 19 38 57
115	5 23
116	2 4 29 58
117	3 9 13 39
118	2 59
119	7 17
120	2 3 4 5 6 8 10 12 15 20 24 30 40 60
121	11
122	2 61
123	3 41
124	2 4 31 62
125	5 25
126	2 3 6 7 9 14 18 21 42 63
127	
128	2 4 8 16 32 64
129	3 43
130	2 5 10 13 26 65
131	
132	2 3 4 6 11 12 22 33 44 66
133	7 19
134	2 67
135	3 5 9 15 27 45
136	2 4 8 17 34 68
137	
138	2 3 6 23 46 69
139	
140	2 4 5 7 10 14 20 28 35 70
141	3 47
142	2 71
143	11 13
144	2 3 4 6 8 9 12 16 18 24 36 48 72
145	5 29
146	2 73
147	3 7 21 49
148	2 4 37 74
149	
150	2 3 5 6 10 15 25 30 50 75
151	
152	2 4 8 19 38 76
153	3 9 17 51
154	2 7 11 14 22 77
155	5 31
156	2 3 4 6 12 13 26 39 52 78
157	
158	2 79
159	3 53
160	2 4 5 8 10 16 20 32 40 80
161	7 23
162	2 3 6 9 18 27 54 81
163	
164	2 4 41 82
165	3 5 11 15 33 55
166	2 83
167	
168	2 3 4 6 7 8 12 14 21 24 28 42 56 84
169	13
170	2 5 10 17 34 85
171	3 9 19 57
172	2 4 43 86
173	
174	2 3 6 29 58 87
175	5 7 25 35
176	2 4 8 11 16 22 44 88
177	3 59
178	2 89
179	
180	2 3 4 5 6 9 10 12 15 18 20 30 36 45 60 90
181	
182	2 7 13 14 26 91
183	3 61
184	2 4 8 23 46 92
185	5 37
186	2 3 6 31 62 93
187	11 17
188	2 4 47 94
189	3 7 9 21 27 63
190	2 5 10 19 38 95
191	
192	2 3 4 6 8 12 16 24 32 48 64 96
193	
194	2 97
195	3 5 13 15 39 65
196	2 4 7 14 28 49 98
197	
198	2 3 6 9 11 18 22 33 66 99
199	
200	2 4 5 8 10 20 25 40 50 100

201	3 67
202	2 101
203	7 29
204	2 3 4 6 12 17 34 51 68 102
205	5 41
206	2 103
207	3 9 23 69
208	2 4 8 13 16 26 52 104
209	11 19
210	2 3 5 6 7 10 14 15 21 30 35 42 70 105
211	
212	2 4 53 106
213	3 71
214	2 107
215	5 43
216	2 3 4 6 8 9 12 18 24 27 36 54 72 108
217	7 31
218	2 109
219	3 73
220	2 4 5 10 11 20 22 44 55 110
221	13 17
222	2 3 6 37 74 111
223	
224	2 4 7 8 14 16 28 32 56 112
225	3 5 9 15 25 45 75
226	2 113
227	
228	2 3 4 6 12 19 38 57 76 114
229	
230	2 5 10 23 46 115
231	3 7 11 21 33 77
232	2 4 8 29 58 116
233	
234	2 3 6 9 13 18 26 39 78 117
235	5 47
236	2 4 59 118
237	3 79
238	2 7 14 17 34 119
239	
240	2 3 4 5 6 8 10 12 15 16 20 24 30 40 48 60 80 120
241	
242	2 11 22 121
243	3 9 27 81
244	2 4 61 122
245	5 7 35 49
246	2 3 6 41 82 123
247	13 19
248	2 4 8 31 62 124
249	3 83
250	2 5 10 25 50 125
251	
252	2 3 4 6 7 9 12 14 18 21 28 36 42 63 84 126
253	11 23
254	2 127
255	3 5 15 17 51 85
256	2 4 8 16 32 64 128
257	
258	2 3 6 43 86 129
259	7 37
260	2 4 5 10 13 20 26 52 65 130
261	3 9 29 87
262	2 131
263	
264	2 3 4 6 8 11 12 22 24 33 44 66 88 132
265	5 53
266	2 7 14 19 38 133
267	3 89
268	2 4 67 134
269	
270	2 3 5 6 9 10 15 18 27 30 45 54 90 135
271	
272	2 4 8 16 17 34 68 136
273	3 7 13 21 39 91
274	2 137
275	5 11 25 55
276	2 3 4 6 12 23 46 69 92 138
277	
278	2 139
279	3 9 31 93
280	2 4 5 7 8 10 14 20 28 35 40 56 70 140
281	
282	2 3 47 94 141
283	
284	2 4 71 142
285	3 5 15 19 57 95
286	2 11 13 22 26 143
287	7 41
288	2 3 4 6 8 9 12 16 18 24 32 36 48 72 96 144
289	17
290	2 5 10 29 58 145
291	3 97
292	2 4 73 146
293	
294	2 3 6 7 14 21 42 49 98 147
295	5 59
296	2 4 8 37 74 148
297	3 9 11 27 33 99
298	2 149
299	13 23
300	2 3 4 5 6 10 12 15 20 25 30 50 60 75 100 150

301	7 43
302	2 151
303	3 101
304	2 4 8 16 19 38 76 152
305	5 61
306	2 3 6 9 17 18 34 51 102 153
307	
308	2 4 7 11 14 22 28 44 77 154
309	3 103
310	2 5 10 31 62 155
311	
312	2 3 4 6 8 12 13 24 26 39 52 78 104
313	
314	2 157
315	3 5 7 9 15 21 35 45 63 105
316	2 4 79 158
317	
318	2 3 6 53 106 159
319	11 29
320	2 4 5 8 10 16 20 32 40 64 80 160
321	3 107
322	2 7 14 23 46 161
323	17 19
324	2 3 4 6 9 12 18 27 36 54 81 108 162
325	5 13 25 65
326	2 163
327	3 109
328	2 4 8 41 82 164
329	7 47
330	2 3 5 6 10 11 13 22 30 33 55 66 110 165
331	
332	2 4 83 166
333	3 9 37 111
334	2 167
335	5 67
336	2 3 4 6 7 8 12 14 16 21 24 28 42 48 56 84 112 168
337	
338	2 13 26 169
339	3 113
340	2 4 5 10 17 20 34 68 84 170
341	11 31
342	2 3 6 9 18 19 38 57 114 171
343	7 49
344	2 4 8 43 86 172
345	3 5 15 23 69 115
346	2 173
347	
348	2 3 4 6 12 29 58 87 116 174
349	
350	2 5 7 10 14 25 35 50 70 175
351	3 9 13 27 39 117
352	2 4 8 11 16 22 32 44 88 176
353	
354	2 3 6 59 118 177
355	5 71
356	2 4 89 178
357	3 7 17 21 51 119
358	2 179
359	
360	2 3 4 5 6 8 9 10 12 15 18 20 24 30 36 40 45 60 72 90 120 180
361	19
362	2 181
363	3 11 33 121
364	2 4 7 13 14 26 28 52 91 182
365	5 73
366	2 3 6 61 122 183
367	
368	2 4 8 16 23 46 92 184
369	3 9 41 123
370	2 5 10 37 74 185
371	7 53
372	2 3 4 6 12 31 62 93 124 186
373	
374	2 11 17 22 34 187
375	3 5 15 25 75 125
376	2 4 8 47 94 188
377	13 29
378	2 3 6 7 9 14 18 21 27 42 54 63 126
379	
380	2 4 5 10 19 20 38 76 95 190
381	3 127
382	2 191
383	
384	2 3 4 6 8 12 16 24 32 48 64 96 128
385	5 7 11 35 55 77
386	2 193
387	3 9 43 129
388	2 4 97 194
389	
390	2 3 5 6 10 13 15 26 30 39 65 78 130 195
391	17 23
392	2 4 7 8 14 28 49 56 98 196
393	3 131
394	2 197
395	5 79
396	2 3 4 6 9 11 12 18 22 33 36 44 66 99 132 198
397	
398	2 199
399	3 7 19 21 57 133
400	2 4 5 8 10 16 20 25 40 50 80 100 200

CHAPTER IV

REPEATING-KEY SYSTEMS WITH STANDARD CIPHER ALPHABETS

20. Solution by applying principles of frequency.--a. In the light of the foregoing principles, let the following cryptogram be studied:

Message

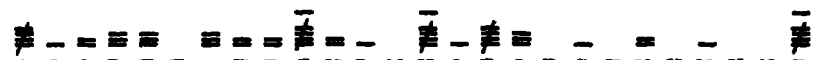
	5				10				15				20				25
A.	A	U	K	H	Y	J	A	M	K	I	Z	Y	M	W	M	J	M	I	G	X	N	F	M	L	X
B.	E	T	I	M	I	Z	H	B	H	R	A	Y	M	Z	M	I	L	V	M	E	J	K	U	T	G
C.	D	P	V	X	K	Q	U	K	H	Q	L	H	V	R	M	J	A	Z	N	G	G	Z	V	X	E
D.	N	L	U	F	M	P	Z	J	N	V	C	H	U	A	S	H	K	Q	G	K	I	P	L	W	P
E.	A	J	Z	X	I	G	U	M	T	V	D	P	T	E	J	E	C	M	Y	S	Q	Y	B	A	V
F.	A	L	A	H	Y	P	O	I	X	W	P	V	N	Y	E	E	Y	X	E	E	U	D	P	X	R
G.	B	V	Z	V	I	Z	I	I	V	O	S	P	T	E	G	K	U	B	B	R	Q	L	L	X	P
H.	W	F	Q	G	K	N	L	L	L	E	P	T	I	K	W	D	J	Z	X	I	G	O	I	O	I
J.	Z	L	A	M	V	K	F	M	W	F	N	P	L	Z	I	O	V	V	F	M	Z	K	T	X	G
K.	N	L	M	D	F	A	A	E	X	I	J	L	U	F	M	P	Z	J	N	V	C	A	I	G	I
L.	U	A	W	P	R	N	V	I	W	E	J	K	Z	A	S	Z	L	A	F	M	H	S			

All repetitions of trigraphs and longer polygraphs are underlined, these repetitions are tabulated in the diagram below, together with their locations, intervals, and factors.

Repetition	Location	Interval	Factors
LUFMPZJNVC	D2, K12	160	2, 4, 5, 8, 10, 16, 20, 32, 40, 80
JZXIG	E2, H17	90	2, 3, 5, 6, 9, 10, 15, 18, 30, 45
EJK	B20, L10	215	5, 43
PTE	E12, G12	50	2, 5, 10, 25
QGK	D18, H3	85	5, 17
UKH	A2, C7	55	5, 11
ZLA	J1, L16	65	5, 13

b. The factor 5 appears in all cases; it is practically certain that the number of alphabets is five, and the I.C.'s should confirm this hypothesis. Since the text already appears in groups of five letters, it is unnecessary to rewrite the message. The next step is to make a uniliteral frequency distribution for Alphabet 1 to see if it can be determined whether or not standard alphabets are involved. It is as follows:

Alphabet 1



I.C. = 1.44

c. Although the indications for fitting the distribution to the normal are not very clear cut, yet if one takes into consideration the small amount of data, the assumption of a direct standard alphabet with $W_c = A_p$, is worth further test. Accordingly a similar distribution is made for Alphabet 2.

Alphabet 2

I.C. = 1.47

d. There is every indication of a direct standard alphabet, with $H_c = \overline{A}_p$. Let similar distributions be made for the last three alphabets. They are as follows:

Alphabet 3

I.C. = 1.71

Alphabet 4

I.C. = 1.36

Alphabet 5

I.C. = 1.91

e. After but little experiment it is found that the distributions can best be made to fit the normal when the following values are assumed:

Alphabet 1.....$A_p = W_c$
Alphabet 2.....$A_p = H_c$
Alphabet 3.....$A_p = I_c$
Alphabet 4.....$A_p = T_c$
Alphabet 5.....$A_p = E_c$

f. Note the key word given by the successive equivalents of A_p: WHITE. And also note what may appear to be a discordant note in the values of three of the I.C.'s above.[1] Nevertheless, the real proof of the cryptanalytic

[1] As has been remarked in subpar 17b small deviations from the expected values are usual and in fact may be anticipated, whereas large deviations are rare In the case of the I C 's at hand the value for the third alphabet (1 71) almost coincides with the expected 1 73 but the values for the 1st 2d and 4th alphabets seem too low while the value for the 5th (1 91) is "on the high side", i e , a positive deviation instead of a negative deviation For the benefit of the mathematical reader these deviations from the expected plain can be proven to be in the nature of only about 1σ for samples of these sizes (55 and 54 tallies), so that the deviations observed are not really significant after all For the statistically curious the formulas for the standard deviation of ϕ and I C for English plain text are given below (where N is the sample size)

$$\sigma(\phi) = \sqrt{(.0048)N^3 + (.1101)N^2 + (.1149)N}$$

$$\sigma(I.C.) = \frac{.026}{(N-1)\sqrt{N}} \sqrt{(.0048)N^2 + (.1101)N - .1149}$$

Derivation of these formulas will be left for the extensive treatment of cryptomathematics in Military Cryptanalytics, Part III It might be noted that cryptanalysts are usually much more deeply concerned with the deviation of an observed ϕ or I C from random rather from an estimated or expected plain This is of course especially true in situations wherein the value of the ϕ_p is unknown (such as would be in the case of a 10x10 bipartite matrix of unknown composition or in the case of a polyalphabetic encipherment of an unknown code), in such situations only the deviations from random could be measured The formulas for the standard deviation of ϕ and I C for 26 letter random text are as follows

$$\sigma(\phi) = .0272\sqrt{N(N-1)}$$

$$\sigma(I.C.) = \frac{.00711}{\sqrt{N(N-1)}}$$

Since sigmage is defined as the difference between the observed and the expected number divided by the standard deviation it may be shown that the I C of Alphabet 1 in the example is $\frac{1.44 - 1.00}{.013} = 3.38\sigma$ over random, for this type of distribution (which follows the χ^2 distribution) this amounts to less than 1 chance in 300 of being produced at random

pudding (i.e., the correctness of the analysis) is, of course, to test the values of the solved alphabets on the cryptogram. The five complete cipher alphabets are as follows:

Plain text

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

g. Applying these values to the first few groups of our message, the following is found:

	1	2	3	4	5		1	2	3	4	5		1	2	3	4	5		1	2	3	4	5		1	2	3	4	5	
Cipher...	A	U	K	H	Y		J	A	M	K	I		Z	Y	M	W	M		J	M	I	G	X		N	F	M	L	X	. . .
Plain....	E	N	C	O	U		N	T	E	R	E		D	R	E	D	I		N	F	A	N	T		R	Y	E	S	T	. . .

h. Intelligible text at once results, and the solution can now be completed very quickly. The complete message is as follows:

ENCOUNTERED RED INFANTRY ESTIMATED AT ONE REGIMENT AND MACHINE GUN COMPANY IN TRUCKS NEAR EMMITSBURG. AM HOLDING MIDDLE CREEK NEAR HILL 543 SOUTHWEST OF FAIRPLAY. WHEN FORCED BACK WILL CONTINUE DELAYING REDS AT MARSH CREEK. HAVE DESTROYED BRIDGES ON MIDDLE CREEK BETWEEN EMMITSBURG-TANEYTOWN ROAD AND RHODES MILL.

i. In the foregoing example (which is typical of the system erroneously attributed, in cryptographic literature, to the French cryptographer Vigenère, although to do him justice, he made no claim of having "invented" it), direct standard alphabets were used, but it is obvious that reversed standard alphabets may be used and the solution accomplished in the same manner. In fact, the cipher disk once used by the United States Army for a number of years yields exactly this type of cipher, which is also known in the literature as the Beaufort Cipher, and by other names. In fitting the isolated frequency distributions to the normal, the direction of "reading" the crests and troughs is merely reversed.[2]

21. Solution by completing the plain-component sequence.--a. There is another method of solving this type of cipher, which is worthwhile explaining, because the underlying principles will be found useful in many cases. It is a modification of the method of solution by completing the plain-component sequence, already explained in Military Cryptanalytics, Part I.

b. After all, the individual alphabets of a cipher such as the one just solved are merely direct standard alphabets. It has been seen that monoalphabetic ciphers in which standard cipher alphabets are employed may

[2] If standard alphabets are employed wherein a letter (usually J) is omitted, this omission must be taken into consideration in fitting the distributions to the normal, or in applying the method of completing the plain-component sequence treated in par 21

be solved almost mechanically by completing the plain-component sequence. The plain text reappears on only one generatrix and this generatrix is the same for the whole message. It is easy to pick this generatrix out of all the other generatrices because it is the only one which yields intelligible text. Is it not apparent that if the same process is applied to the cipher letters of the individual alphabets of the cipher just solved that the plaintext equivalents of these letters must all reappear on one and the same generatrix? But how will the generatrix which actually contains the plaintext letters be distinguishable from the other generatrices, since these plaintext letters are not consecutive letters in the plain text but only letters separated from one another by a constant interval? The answer is simple. The plaintext generatrix should be distinguishable from the others because it will show more and a better assortment of high-frequency letters, and can thus be selected by the eye from the whole set of generatrices. If this is done with all the alphabets in the cryptogram, it will merely be necessary to assemble the letters of the thus selected generatrices in proper order, and the result should be consecutive letters forming intelligible text.

c. An example will serve to make the process clear. Let the same message be used as before. Factoring showed that it involves five alphabets. Let the first ten cipher letters in each alphabet be set down in a horizontal line and, under the assumption that direct standard alphabets are involved, let the normal alphabet sequences be completed.[3] Thus:

---

[3] If reversed standard alphabets are assumed it would first be necessary to convert the cipher letters of each isolated alphabet into their normal plain-component equivalents, and then to proceed as in the case of direct standard alphabets

	Alphabet 1	Alphabet 2	Alphabet 3	Alphabet 4	Alphabet 5
1	AJZJNEZAIJ	UAYMFTHYLK	KMMIMIBMVU	HKWGLMHZMT	YIMXXIRMEG
2	BKAKOFABJK	VBZNGUIZML	LNNJNJCNWV	ILXHMNIANU	ZJNYYJSNFH
3	CLBLPGBCKL	WCAOHVJANM	MOOKOKDOXW	JMYINOJBOV	AKOZZKTOGI
4	DMCMQHCDLM	XDBPIWKBON	NPPLPLEPYX	KNZJOPKCPW	BLPAALUPHJ
5	ENDNRIDEMN	YECQJXLCPO	OQQMQMFQZY	LOAKPQLDQX	CMQBBMVQIK
6	FOEOSJEFNO	ZFDRKYMDQP	PRRNRNGRAZ	MPBLQRMERY	DNRCCNWRJL
7	GPFPTKFGOP	AGESLZNERQ	QSSOSOHSBA	NQCMRSNFSZ	EOSDDOXSKM
8	HQGQULGHPQ	BHFTMAOFSR	RTTPTPITCB	ORDNSTOGTA	FPTEEPYTLN
9	IRHRVMHIQR	CIGUNBPGTS	SUUQUQJUDC	PSEOTUPHUB	GQUFFQZUMO
10	JSISWNIJRS	DJHVOCQHUT	TVVRVRKVED	QTFPUVQIVC	HRVGGRAVNP
11	KTJTXOJKST	EKIWPDRIVU	UWWSWSLWFE	RUGQVWRJWD	ISWHHSBWOQ
12	LUKUYPKLTU	FLJXQESJWV	VXXTXTMXGF	SVHRWXSKXE	JTXIITCXPR
13	MVLVZQLMUV	GMKYRFTKXW	WYYUYUNYHG	TWISXYTLYF	KUYJJUDYQS
14	NWMWARMNVW	HNLZSGULYX	XZZVZVOZIH	UXJTYZUMZG	LVZKKVEZRT
15	OXNXBSNOWX	IOMATHVMZY	YAAWAWPAJI	VYKUZAVNAH	MWALLWFASU
16	PYOYCTOPXY	JPNBUIWNAZ	ZBBXBXQBKJ	WZLVABWOBI	NXBMMXGBTV
17	QZPZDUPQYZ	KQOCVJXOBA	ACCYCYRCLK	XAMWBCXPCJ	OYCNNYHCUW
18	RAQAEVQRZA	LRPDWKYPCB	BDDZDZSDML	YBNXCDYQDK	PZDOOZIDVX
19	SBRBFWRSAB	MSQEXLZQDC	CEEAEATENM	ACOYDEZREL	QAEPPAJEWY
20	TCSCGXSTBC	NTRFYMARED	DFFBFBUFON	ADPZEFASFM	RBFQQBKFXZ
21	UDTDHYTUCD	OUSGZNBSFE	EGGCGCVGPO	BEQAFGBTGN	SCGRRCLGYA
22	VEUEIZUVDE	PVTHAOCTGF	FHHDHDWHQP	CFRCGHCUHO	TDHSSDMHZB
23	WFVFJAVWEF	QWUIBPDUHG	GIIEIEXIRQ	DGSCHIDVIP	UEITTENIAC
24	XGWGKBWXFG	RXVJCQEVIH	HJJFJFYJSR	EHTDIJEWJQ	VFJUUFOJBD
25	YHXHLCXYGH	SYWKDRFWJI	IKKGKGZKTS	FIUEJKFXKR	WGKVVGPKCE
26	ZIYIMDYZHI	TZXLESGXKJ	JLLHLHALUT	GJVFKLGYLS	XHLWWHQLDF

Figure 13.

d. If the high-frequency generatrices underlined in Fig. 13 are selected and their letters are juxtaposed in columns, the consecutive letters of intelligible plain text immediately present themselves. Thus:

Selected Generatrices:

For Alphabet 1, generatrix 5...E N D N R I D E M N
For Alphabet 2, generatrix 20..N T R F Y M A R E D
For Alphabet 3, generatrix 19..C E E A E A T E N M
For Alphabet 4, generatrix 8...O R D N S T O G T A
For Alphabet 5, generatrix 23..U E I T T E N I A C

Columnar juxtaposition of letters from selected generatrices:

1	2	3	4	5
E	N	C	O	U
N	T	E	R	E
D	R	E	D	I
N	F	A	N	T
R	Y	E	S	T
I	M	A	T	E
D	A	T	O	N
E	R	E	G	I
M	E	N	T	A
N	D	M	A	C

Plain text: ENCOUNTERED RED INFANTRY ESTIMATED AT ONE REGIMENT AND MAC . . .

e. Solution by this method can thus be achieved without the compilation of any frequency tables whatever and is very quickly attained. The inexperienced cryptanalyst may have difficulty at first in selecting the generatrices which contain the most and the best assortment of high-frequency letters, but with increased practice, a high degree of proficiency is attained. After all it is only a matter of experiment, trial, and error to select and assemble the proper generatrices so as to produce intelligible text. The ocular selection of the correct generatrix in each alphabet may be narrowed down to a considerably restricted choice from a comparatively few generatrices, using a short-cut procedure which has much merit and is easy to apply, as will now be demonstrated.

f. Let the generatrices be completed as in Fig. 13, and then let us encircle all the letters J, K, Q, X, and Z in each of the ten columns belonging to Alphabet 1. Now let us cross out all generatrices containing two or more of these low-frequency letters, under the premise that it is unlikely that the correct generatrix will contain more than one of these low-frequency letters.[4] This procedure is extended to the generatrices pertaining to Alphabets 2-5 (cf. Fig. 14). It will be observed that with this procedure there have been eliminated 13, 15, 11, 16, and 16 generatrices from Alphabets 1-5, respectively, thus considerably simplifying the inspection of the remaining generatrices.

g. The selection of the correct generatrix from those remaining may now be facilitated by the use of a rough weighting or "scoring" procedure, in which the eight highest-frequency letters (ETNROAIS) are assigned a weight of 1 and the remaining letters a weight of ϕ.[5] The sum of the

---

[4] This premise can be substantiated statistically By means of the binomial theorem, it may be shown that for 10-letter generatrices an average of 60% (i e 16 generatrices out of 26) of incorrect generatrices will be eliminated while the chance of re ecting the correct case is only 0 8% If the generatrices contained 15 letters an average of 82% (i e 21 out of 26) of incorrect generatrices may be expected to be eliminated with a risk of 1 8% as the chance of rejecting the correct case Thus to avoid excessive risk the threshhold of only 2 letters of the JKQXZ group should be raised when the generatrices contain more than 12-15 letters

[5] In considering the h ghest frequency English plaintext letters in descending freqiency order there is a sharp drop after the S, therefore this seems the obvious place to divide the letters into two classes or categories There are also other more cogent mathematical reasons to substantiate the fact that this 8-18 split is the best possible division of the English plaintext letters into two classes for weighting purposes The scoring system discussed actually involves three weights 1 ϕ, and $-\infty$, the elimination of generatrices on the basis of two or more occurrences of one of the low freqiency letters (JKQXZ) is tantamount to assigning a weight of $-\infty$ to the eliminated generatrices

The set of alphabet str ps prepared for use in connection with the courses in Military Cryptanalytics has been designed with this weighting system in mind The letters ETNROAIS are printed in red the rest of the letters in black with the letters JKQXZ in minuscule type Thus in using these strips one searches for the most redness in the generatrices discounting those generatrices in which two or more of the minuscule letters are present

~~CONFIDENTIAL~~

weights for each generatrix is then recorded at the side of the generatrix; the correct generatrix may be expected to have the highest or near-highest score in that particular alphabet. The result of the generatrix elimination and the summing of the weights is shown in Fig. 14, below:

	Alphabet 1	Alphabet 2	Alphabet 3	Alphabet 4	Alphabet 5
1	~~AJZJNEZAIJ~~	2 UAYMFTHYLK	2 KMMIMIBMVU	~~HKWGLMHZMT~~	~~YIMXXIRMEG~~
2	~~BKAKOFABJK~~	~~VBZNGUIZML~~	~~LNNJNJCNWV~~	5 ILXHMNIANU	~~ZJNYYJSNFH~~
3	0 CLBLPGBCKL	4 WCAOHVJANM	~~MOOKOKDOXW~~	~~JMYINOJBOV~~	~~AKOZZKTOGI~~
4	0 DMCMQHCDLM	~~XDBPIWKBON~~	2 NPPLPLEPYX	~~KNZJOPKGPW~~	2 BLPAALUPHJ
5	7 ENDNRIDEMN	~~YEGQJYLGPO~~	~~OQQMQMFQZY~~	~~LOAKPQLDQX~~	~~GMQBBMVQIK~~
6	7 FOEOSJEFNO	~~ZFDRKYMDQP~~	7 PRRNRNGRAZ	3 MPBLQRMERY	4 DNRCCNWRJL
7	2 GPFFPTKFGOP	~~AGESLZNERQ~~	7 QSSOSOHSBA	~~NQCMRSNFSZ~~	~~EOSDDOXSKM~~
8	~~HQGQULGHPQ~~	5 BHFTMAOFSR	6 RTTPTPITCB	8 ORDNSTOGTA	5 FPTEEPYTLN
9	5 IRHRVMHIQR	4 CIGUNBPGTS	~~SUUQUQJUDC~~	4 PSEOTUPHUB	~~GQUFFQZUMO~~
10	~~JSISWNIJRS~~	~~DJHVOCQHUT~~	4 TVVRVRKVED	~~QTFPUVQIVG~~	4 HRVGGRAVNP
11	~~KTJTKOJKST~~	4 EKIWPDRIVU	3 UWWSWSLWFE	~~RUGQVWRJWD~~	4 ISWHHSBWOQ
12	~~LUKUYPKLTU~~	~~FLJXQESJWV~~	~~VXXTXTMXGP~~	~~SVHRWXSKXE~~	~~JTXIITCXPR~~
13	~~MVLVZQLMUV~~	~~GMCYRFTKXW~~	1 WYYUYUNYHG	3 TWISXYTLYF	~~KUYJJUDYQS~~
14	4 NWMWARMNVW	~~HNLZSGULYX~~	~~XZZVZVOZIH~~	~~UXJTYZUMZG~~	~~LVZKKVEZRT~~
15	~~OXNXBSNOWX~~	4 IOMATHVMZY	5 YAAWAWPAJI	~~VYKUZAVNAH~~	3 MWALLWFASU
16	3 PYOYCTOPXY	~~JPNBUIWNAZ~~	~~ZBBXBXQBKJ~~	3 WZLVABWOBI	~~NXBMMXGBTV~~
17	~~QZPZDUPQYZ~~	~~KQOCVJXOBA~~	2 ACCYCYRCLK	~~XAMWBCXPGJ~~	3 OYCNNYHCUW
18	~~RAQAEVQRZA~~	1 LRPDWKYPCB	~~BDDZDZSDML~~	~~YBNXCDYQDK~~	~~PZDOOZIDVX~~
19	5 SBRBFWRSAB	~~MSQEXLZQDC~~	8 CEEAEATENM	~~ZCOYDEZREL~~	~~QAEPPAJEWY~~
20	4 TCSCGXSTBC	6 NTRFYMARED	2 DFFBFBUFON	4 ADPZEFASFM	~~RBFQQBKFXZ~~
21	2 UDTDHYTUCD	5 OUSGZNBSFE	2 EGGCGCVGPO	4 BEQAFGBTGN	4 SCGRRCLGYA
22	4 VEUEIZUVDE	4 PVTHAOCTGF	0 FHHDHDWHQP	2 CFRCGHCUHO	3 TDHSSDMHZB
23	2 WFVFJAVWEF	1 QWUIBPDUHG	~~GIIEIEXIRQ~~	3 DGSCHIDVIP	8 UEITTENIAC
24	~~XGWGKBWXFG~~	~~RXVJCQEVIH~~	~~HJJFJFYJSR~~	~~EHTDIJEWJQ~~	~~VFJUUFOJBD~~
25	~~YHKHLCXYGH~~	~~SYWKDRFWJI~~	~~IKKGKGZKTS~~	~~FIUEJKFXKR~~	~~WGKVVGPKCE~~
26	~~ZIYIMDYZHI~~	~~TZXLESGXKJ~~	2 JLLHLHALUT	~~GJVFKLGYLS~~	~~XHLWWHQLDF~~

Figure 14.

Note that in this example the correct generatrix in each alphabet is the one with the highest score.[6] This weighting system, crude as it may appear, suffices in cases where the generatrices contain at least 8-10 letters. When the number of letters per generatrix is small, there exist more refined statistical methods for the selection of the correct generatrices, these methods will be treated in par. 34 in the next chapter.

[6] Theoretically the generatrix with the greatest value will be the correct generatrix In actual practice of course the generatrix with the greatest value may not be the correct one but the correct one will certainly be among the three or four generatrices or so with the largest values In any case the test of correctness is whether when juxtaposed, a set of two or three generatrices selected will yield 'good digraphs or trigraphs i e high-frequency digraphs or trigraphs such as occur in normal plain text

h. It has been seen how the key word may be discovered in this type of cryptogram. Usually the key is made up of those letters in the successive alphabets whose equivalents are A_p but other conventions are of course possible. Sometimes a key number is used, such as 8-4-7-1-2, which means merely that A_p is represented by the eighth letter from A (in the normal alphabet) in the first cipher alphabet, by the fourth letter from A in the second cipher alphabet, and so on, as in the classic Gronsfeld cipher. However, the method of solution as illustrated above, being independent of the nature of the key, is the same as before.

22. Solution by the "probable-word method."--a. The common use of plaintext words as key words in cryptograms such as the foregoing makes possible a method of solution that is simple and can be used where the more detailed method of analysis using frequency distributions or by completing the plain-component sequence is of no avail. In the case of a very short message which may show no recurrences and give no indications as to the number of alphabets involved, this modified method will be found most useful.

b. Briefly, the method consists in assuming the presence of a probable word in the message, and referring to the alphabets to find the key letters applicable when this hypothetical word is assumed to be present in various positions in the cipher text. If the assumed word happens to be correct, and is placed in the correct position in the message, the key letters produced by referring to the alphabets will yield the key word. In the following example it is assumed that reversed standard alphabets are known to be used by the enemy.

Message

P G S G G D N R U H V M B G R Y O U U C W M S G L V T Q D O

c. Extraneous circumstances lead to the assumption of the presence of the word REGIMENT. One may assume that this word begins the message. Using sliding normal components, one reversed, the other direct, the key letters are ascertained by noting what the successive equivalents of A_p are. Thus:

Cipher:	P G S G G D N R
Plain text:	R E G I M E N T
"Key"	G K Y O S H A K

The key does not spell any intelligible word. One therefore shifts the assumed word one letter forward and another trial is made.

Cipher:	G S G G D N R U
Plain text:	R E G I M E N T
"Key"	X W M O P R E N

This also yields no intelligible key word. One continues to shift the

assumed word forward one space at a time until the following point is reached:

Cipher:	U H V M B G R Y
Plain text:	R E G I M E N T
"Key"	L L B U N K E R

The key now becomes evident. It is a cyclic permutation of BUNKER HILL. It should be clear that since the key word or key phrase repeats itself during the encipherment of such a message, the plaintext word upon whose assumed presence in the message this test is being based may begin to be enciphered at any point in the key, and continue over into the next repetition of the key if the probable word is longer than the key. When this is the case it is merely necessary to shift the latter part of the sequence of key letters to the first part, as in the case noted: LLBUNKER is permuted cyclically into BUNKER..LL, and thus BUNKER HILL.

d. The examples in subpar. c, above, merely illustrate the theory of "placing" a probable word and recovering the key word. In actual practice, the application of the probable-word method proceeds along slightly different lines of a short-cut manner, as will be described below, using the same message and probable word as stated in the preceding subparagraph.

(1) The cipher text is written in a horizontal line on cross-section paper, and the first five letters or so of the probable word are written columnarwise to the left of the cipher text and one space below it. Assuming first that direct standard alphabets have been used, the successive letters of the cipher are deciphered as R_p, writing the respective key letters (as derived under A_p or the assumed index letter) on the first line just below the cipher text; this assumes that R_p exists at one of the N-8 possible positions (for the word REGIMENT). Then the presence of the letter E_p is assumed in the text (beginning with the second letter of the message), and the successive key letters from these decipherments are inscribed in the second line for N-7 positions. On the third line the process is repeated, assuming that G_p is present in N-6 possible positions beginning with the third letter of the cryptogram, writing the respective decipherments under each Θ_c.

(2) Now if the trigraph $\overline{REG}_p$ exists in the message, then the juxtaposition of $\overline{REG}_p$ at its correct location in the cipher text will yield on a diagonal a plaintext trigraph which is a part of the repeating key, if the key is a plaintext word or phrase. So by examining the possible plaintext trigraphs and extending them to 1, 2, 3, or more places if necessary, all but one will be eliminated by inconsistencies (i.e., implausible plaintext polygraphs), as only one polygraph will keep on yielding valid plain text. If the first trials with direct standard alphabets are not successful, then reversed standard alphabets are tried. It is important to keep in mind that plaintext trigraphs are not necessarily only those which are contained within words; observe the $\overline{LLB}$ trigraph in Fig. 15b, below, which occurs between words at a cyclical repetition of the key phrase BUNKER HILL.

(3) In the following three figures, Fig. 15a is the attempted solution under the premise that the alphabets employed are direct standard, Fig. 15b is the successful trial with reversed standard alphabets,[7] and Fig. 15c is the complete decipherment of the message after the key word has been recovered.

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

Figure 15a.

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

Figure 15b.

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

Figure 15c.

[7] It is interesting to point out a further short cut to this already short-cut method In Fig 15a we derive the first row of key letters (representing R_p) under the cipher, i e YPBBP. . For the second row, we derive C_k under G_c as the first equivalent of E_p, this equivalent is of course under the P_k derived in the first row We now take an alphabet composed of two direct standard sequences and juxtapose them so that P in the upper component is over C in the lower component The rest of the letters in the E_p row (viz the second row) may now be read directly by referring to the direct standard alphabet i e if P_k in component (1) is equivalent to C_k in component (2) then $B_k(1) = O_k(2)$ $M_k(1) = Z_k(2)$ etc In Fig 15b the same procedure is followed still using the direct standard alphabet for finding the equivalents of the key letters in the second and third rows, the reason that direct standard alphabets may still be used is that there has been in effect a conversion into plain-component equivalents The method just described is much faster and less laborious in finding the equivalents for the second and third rows once the initial key letter of each row has been determined

e. Another method for testing cribs where the components are known is illustrated in Fig. 16, below. This method involves the completion of all the generatrices from the cipher text, and searching for the key

Cipher text

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

Figure 16.

by means of a stencil pre-cut to the probable word being tested.

(1) In this example, using the same message and crib as in the preceding subparagraph, the top row of the diagram is the cipher message, the identical row just beneath the cipher consists of the key letters (on the hypothesis of reversed standard alphabets) if the ciphertext letters represent encipherments of A_p, the next row (QHTHH...) consists of the key letters if the ciphertext letters represent encipherments of B_p; and so forth. A stencil or mask is made on cross-section paper of the same size cells as the cross-section paper used to complete the generatrices, with appropriate cells cut out in successive columns to represent the letters of the crib (using for this purpose the reference alphabet to the left of the diagram). This stencil can now be slid along the horizontal axis through successive positions of the diagram, when the correct placement of the crib is reached, the letters of the key (in this case, LLBUNKER) will manifest themselves in the apertures.

(2) If we are to test the crib on the basis of direct standard alphabets (instead of reversed standard as above), the completion diagram of Fig. 16 may still be used, the only change necessary is that the plaintext reference alphabet at the left must be changed to the AZYXW...CDB sequence, to reflect the difference in deriving the key letters in the diagram.

(3) This particular method is very valuable if there are many cribs to be tested, this method also has related applications in other fields of cryptanalysis.

f. Two further important ramifications of the probable-word method should be pointed out at this time; these apply to cases wherein the probable word or crib is considerably shorter than the repeating key, and to cases where the repeating key is composed of a sequence of random letters.

(1) Suppose that in the previous example we were testing the crib $\overline{\text{YOUR}}_p$; at the 5th position of the cipher text we would have recovered the key fragment $\overline{\text{ERHI}}_k$ which appears likely as a part of a key word in English. We would then take this key fragment and slide it along all the remaining positions of the cipher text, at position 15 we would obtain the fragment $\overline{\text{NTTO}}_p$ as possible plain text, and at position 25 we would obtain the fragment $\overline{\text{TWOS}}_p$. Factoring the intervals between the fragments which yielded plain text, we would conclude that the period of the message is 10. The $\overline{\text{NTTO}}_p$ sequence might be preceded by an Ep, which might be expanded into $\overline{\text{MENTTO}}_p$, which would yield the fragments (MO?)$\overline{\text{VEYOUR}}_p$ at position 3 and (FI?)$\overline{\text{VETWOS}}_p$ at position 23. This procedure would be continued until the message is completely solved.

(2) In the foregoing case, $\overline{\text{ERHI}}_k$ was recognized as a possible key fragment because it looked like a plausible sequence of a plaintext key word. If the key had not been a plaintext word, but instead had been, let us say, the arbitrary letters CBNOMRGOWB, then the fragment $\overline{\text{MRGO}}_k$ of this sequence would not have been recognized as part of the key when the crib $\overline{\text{YOUR}}_p$ was tried at the 5th position. The procedure followed in cases where the key is composed of random letters is to assume the crib at position 1, derive the "key", and slide this "key" along the rest of the message to see whether possible plain text results, then the crib is tried at positions 2, 3, 4... in turn, until its placement at the correct position yields decipherments in other parts of the message which are recognized as valid plain text. This technique, although laborious when done by hand, is the basis for solution when analytical machine methods are employed.

g. It has been seen in the probable-word method described in this paragraph that the length of the key is of no particular interest or consequence in the steps taken in effecting the solution. The determination of the length and elements of the key come after the solution rather than before it. In the case illustrated the length of the period is seen to be ten, corresponding to the length of the key (BUNKER HILL).

h. The foregoing method is one of the other methods of determining the length of the key (besides factoring), referred to in subpar. 15c. As will be shown subsequently, the method can also be used as a last resort when known mixed alphabets are employed. This method of solution by searching for a word is contingent upon the following circumstances:

(1) That the word whose presence is assumed actually occurs in the message, is properly spelled, and correctly enciphered.

(2) That the sliding components (or equivalent cipher disks or squares) employed in the search for the assumed word are actually the ones which were employed in the encipherment, or are such as to give identical results as the ones which were actually used.

(3) That the pair of enciphering equations used in the test is actually the pair which was employed in the encipherment; or if a cipher square is used in the test, the method of finding equivalents gives results that correspond with those actually obtained in the encipherment.

i. The foregoing appears to be quite an array of contingencies and the student may think that on this account the method will often fail. But examining these contingencies one by one, it will be seen that successful application of the method may not be at all rare--after the solution of some messages has disclosed what sort of paraphernalia and methods of employing them are favored by the enemy. From the foregoing remark it is to be inferred that the probable-word method has its greatest usefulness not in an initial solution of a system, but only after successful study of enemy communications by more difficult processes of analysis has told its story to the alert cryptanalyst. Although it is commonly attributed to Bazeries, the French cryptanalyst of 1900, the probable-word method is very old in cryptanalysis and goes back several centuries. Its usefulness in practical work may best be indicated by quoting from a competent observer[8]:

> There is another [method] which is to this first method what the geometric method is to analysis in certain sciences and according to the whims of individuals certain cryptanalysts prefer one to the other Certain others, incapable of getting the answer with one of the methods in the solution of a difficult problem conquer it by means of the other with a disconcerting masterly stroke This other method is that of the probable word We may have more or less definite opinions concerning the subject of the cryptogram We may know something about its date and the correspondents, who may have been indiscreet in the subject they have treated On this basis the hypothesis is made that a certain word probably appears in the text
>
> In certain classes of documents, military or diplomatic telegrams, banking and mining affairs, etc , it is not impossible to make very important assumptions about the presence of certain words in the text After a cryptanalyst has worked for a long time with the writings of certain correspondents he gets used to their expressions He gets a whole load of words to try out, then the changes of key and sometimes of system no longer throw into his way the difficulties of an absolutely new study which might require the analytical method '

[8] Givierge M , Cours de Cryptographie Paris 1925, p 30

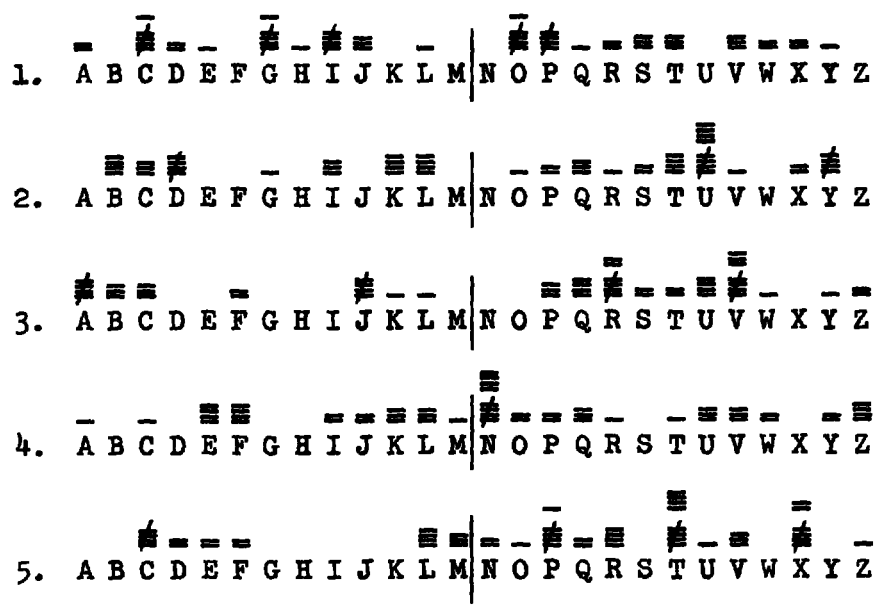
To which I am prompted to add the amusing definition of cryptanalysis attributed to a British wag: "All cryptanalysis is divided into two parts: trance-titution and supposition."

23. The Porta system.--a. The solution of the Porta system, described in subpar. 11b, may properly be treated at this point along with repeating-key systems with standard alphabets, since the enciphering matrix is a known matrix with normal components. The Porta matrix illustrated in Fig. 5 may be visualized as follows:

Plain text

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

b. If the message in par. 20 were enciphered by means of the Porta table, the key word still being WHITE, the distributions for the five alphabets would appear as follows:



Now if a vertical dividing line is drawn between the M and the N of the distributions, each half of the distribution may be used to fit half of the normal frequency distribution (following the Porta rule of encipherment, i.e., each half of the alphabet going to the opposite half). Thus in Alphabet 1 the sequence $\overline{CDEFGHIJ}_C$ may easily be identified as $\overline{NOPQRSTU}_P$; this would fix the key letters as WX, and therefore the $\overline{A...M}_P$ sequence should begin at Y_C. This latter fit may not be ideal, but it is nevertheless plausible. In Alphabets 2, 3, and 5 the $\overline{RST}_P$ sequence may be spotted at $\overline{BCD}_C$, $\overline{ABC}_C$, and $\overline{CDE}_C$, respectively, whereas in Alphabet 4 the trial of E_P as N_C gives a reasonably good matching of that half of the distribution. These assumptions in the first halves of the distributions will of course determine the placements of the letters in the second halves, since, for example in Alphabet 4, if $N_C = E_P$, then $E_C = N_P$; therefore the original assumptions for the first halves will be confirmed or rejected by the goodness of fit of the distributions for the second halves. The keys for these five alphabets are derived as (W,X), (G,H), (I,J), (S,T), and (E,F); from these letters, the repeating key WHITE is obvious.[9]

c. In completing the plain-component sequence in the case of Porta encipherment, the cipher letters of each alphabet are first converted to their /Porta/plain-component equivalents, and then the plain-component sequence is completed from these letters, with a minor modification. This modification consists in completing the converted cipher letters A-M in a downward direction, while the letters N-Z are completed in the opposite (i.e., upward) direction. As an example of this process, let us assume that the message in subpar. 20h has been enciphered by five alphabets in the Porta system, the first forty letters of this encipherment are:

P K T F F C D V I T O B V Z X C V R E E G I V J E T P R K T
O Q C F L P B V P X . . .

The conversion process and plain-component completion of the first three alphabets are shown in the diagram below (employing the procedure of generatrix elimination and weighting as described in subpars. 21f and g):

[9] In some cases the lower half of the Porta alphabets shifts to the right instead of to the left as in the normal form, this possibility must be taken into account in the recovery of the key word

	P C O C G T O P		K D B V I P Q B		T V V R V R C V
1	C P B P T G B C		~~X Q O I V C D O~~	6	G I I E I E P I
3	D O C O S H C D	3	W P N J U D E N		~~H J J F J F O J~~
6	E N D N R I D E		~~V O Z K T E F Z~~		~~I K K G K G N K~~
	~~F Z E Z Q J E F~~	2	U N Y L S F G Y		~~J L L H L H Z L~~
0	G Y F Y P K F G		~~T Z X M R G H X~~	2	K M M I M I Y M
	~~H X G X O L G H~~	3	S Y W A Q H I W		~~L A A J A J X A~~
3	I W H W N M H I		~~R X V B P I J V~~		~~M B B K B K W B~~
	~~J V I V Z A I J~~		~~Q W U G O J K U~~	1	A C C L C L V C
	~~K U J U Y B J K~~	3	P V T D N K L T	0	B D D M D M U D
	~~L T K T X G K L~~	3	O U S E Z L M S	7	C E E A E A T E
2	M S L S W D L M	5	N T R F Y M A R	0	D F F B F B S F
5	A R M R V E M A		~~Z S Q G X A B Q~~	2	E G G C G C R G
	~~B Q A Q U F A B~~	1	Y R P H W B C P	0	F H H D H D Q H

The generatrices with the highest scores are the correct ones.

d. Just as the Vigenère table (consisting of direct standard alphabets) has its complementary table of reversed standard alphabets, a variant of the Porta table might be constructed wherein the lower halves of the sequences run in the opposite direction to the upper half, as is illustrated below:

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

In this case, the method of fitting the distributions to the normal and the method of completing the plain-component sequence must of course be modified to take care of the new situation. Other variations of the Porta idea are possible; these will be treated in subsequent chapters.

e. In applying the probable-word method in Porta, the cryptographic peculiarities of the system greatly facilitate the testing and placing of cribs. As an illustration, let us suppose we have at hand the 40-letter fragment in subpar. 23c (the period being unknown), and let us place under each cipher letter a notation of its class (using "1" to designate a cipher letter from the Group A-M in the normal sequence and "2" to designate a letter from the group N-Z). The cipher text and notations will look as follows:

P	K	T	F	F	C	D	V	I	T	O	B	V	Z	X	C	V	R	E	E
2	1	2	1	1	1	1	2	1	2	2	1	2	2	2	1	2	2	1	1

G	I	V	J	E	T	P	R	K	T	O	Q	C	F	L	P	B	V	P	X...
1	1	2	1	1	2	2	2	1	2	2	2	1	1	1	2	1	2	2	2

Let us suppose that the probable word is INFANTRY, the letters of this word have the class notation of 12112222, but in encipherment the classes would be reversed, viz., 21221111. We now look for the pattern 21221111 in the cipher text, and we find it beginning at the 15th position. The derived key $\binom{\text{E W G I S E W G}}{\text{F X H J T F X H}}$ can easily be recognized as a cyclical repetition of the key word WHITE.

24. The Gronsfeld system. The Gronsfeld system mentioned earlier is identical with a Vigenère system with direct standard alphabets (where only the first 10 alphabets are used), except that a numerical key is involved, the digits of the key indicating how much displacement the plaintext letters should have along the normal sequence; thus only the first ten rows of the Vigenere table are used. The Gronsfeld system is solved just like any Vigenere system, except (when the system is known to be a Gronsfeld) for a minor modification in the use of the probable-word method. The severe limitation of the cipher equivalents possible for a given plaintext letter greatly restricts the placement of trial cribs. For example, in trying the crib $\overline{\text{YOUR}}_p$ in a message, the diagram below (analogous to Fig. 15a)

	T	S	V	H	Y	Q	B	V	Y	I	G	L	M	G	U	X	A	S	R	M	F	K	C	I	A	A	O	V	I	E
Y						9	0		3		0		8				8			2				7		4		2	2		
O								2		7								6			4									0	
U									7		4								3												1
R										4										9											

shows the keys resulting from the only possible Gronsfeld decipherments of Y_p in the first row beneath the cipher; the row just beneath that gives the keys for O_p (where Y_p has been a possibility for the preceding cipher letter); etc. In the example, there are only two places where YOUR_p is a possibility: at the 5th and 14th positions, with the corresponding keys being 0274 and 8639, respectively. If 0274 is slid through the remainder of the cipher text, the appearance of the plaintext fragments MENT (at position 13) and FIVE (at position 21) shows that this key fragment is correct, and that the period is 8. If the solution proved difficult, the plain-component sequences could be completed for the remaining cipher text as shown below, as an aid in

				0	2	7	4					0	2	7	4					0	2	7	4					0	2
T	S	V	H	Y	Q	B	V	Y	I	G	L	M	G	U	X	A	S	R	M	F	K	C	I	A	A	O	V	I	E
S	R	U	G	Y	O	U	R	X	H	F	K	M	E	N	T	Z	R	Q	L	F	I	V	E	Z	Z	N	U	I	X
R	Q	T	F					W	G	E	J					Y	Q	P	K					Y	Y	M	T		
Q	P	S	E					V	F	D	I					X	P	O	J					X	X	L	S		
P	O	R	D					U	E	C	H					W	O	N	I					W	W	K	R		
O	N	Q	C					T	D	B	G					V	N	M	H					V	V	J	Q		
N	M	P	B					S	C	A	F					U	M	L	G					U	U	I	P		
M	L	O	A					R	B	Z	E					T	L	K	F					T	T	H	O		
L	K	N	Z					Q	A	Y	D					S	K	J	E					S	S	G	N		
K	J	M	Y					P	Z	X	C					R	J	I	D					R	R	F	M		

recovering the rest of the text. It must be pointed out that, although reading a Gronsfeld cipher of a lengthy period on ten generatrices alone is quite possible, it may be rather difficult to do so in actual practice unless something concerning the contents or nature of the message plain text is known.

25. Polyalphabetic numerical systems.--a. Periodic number ciphers may be encountered in which the plain component is the normal sequence and the cipher component is what may be regarded as a standard [numerical] sequence. For instance, if the cipher component consisted of the dinomes 01-26 in normal order, this component is in effect an A-Z sequence and analysis would proceed along the lines of any direct standard alphabet cipher. In Fig. 6 we have a numerical Vigenere square consisting of a 36-element "normal" plain component and a cipher component consisting of the dinomes 10-45 in ascending order; this system involves nothing new in techniques of solution, except that in fitting the cipher distributions to the normal (after factoring), allowance has to be made for the beginning and ending points of the A-Z sequence in the 36 elements of the cipher distributions.

b. If periodic numerical ciphers are encountered in which the cipher components are slides of the 00,01...98,99 sequence in normal order, the occurrence in certain alphabets of dinomes within a comparatively narrow range will be an aid to factoring. For example, if the matrix in the illustration below were used for encipherment, the occurrence of the "low

	A	B	C	D	E	F	G	H	.....	V	W	X	Y	Z
1	03	04	05	06	07	08	09	10	.....	24	25	26	27	28
2	41	42	43	44	45	46	47	48		62	63	64	65	66
3	28	29	30	31	32	33	34	35		49	50	51	52	53
4	70	71	72	73	74	75	76	77		91	92	93	94	95
5	32	33	34	35	36	37	38	39	.....	53	54	55	56	57

dinomes" (resulting from encipherments by Alphabet 1) spaced along the cipher text at an interval of 5, and the "high dinomes" (resulting from encipherments by Alphabet 4) likewise spaced along the cipher text at that same interval, would quickly identify the length of the period.

(BLANK)

CHAPTER V

REPEATING-KEY SYSTEMS WITH MIXED CIPHER ALPHABETS, I; DIRECT SYMMETRY OF POSITION

26. Reason for the use of mixed alphabets.--a. It has been seen in the examples considered thus far that the use of several alphabets in the same message does not greatly complicate the analysis of such a cryptogram. There are three reasons why this is so. Firstly, only relatively few alphabets were employed; secondly, these alphabets were employed in a periodic or repeating manner, giving rise to cyclic phenomena in the cryptogram by means of which the number of alphabets could be determined; and, thirdly, the cipher alphabets were known alphabets, by which is meant merely that the sequences of letters in both components of the cipher alphabets were known sequences.

b. In the case of monoalphabetic ciphers it was found that the use of a mixed alphabet delayed the solution to a considerable degree, and it will now be seen that the use of mixed alphabets in polyalphabetic ciphers renders the analysis much more difficult than the use of standard alphabets, but the solution is still fairly easy to achieve.

27. Interrelated mixed alphabets.--a. It was stated in par. 7 that the method of producing the mixed alphabets in a polyalphabetic cipher often affords clues which are of great assistance in the analysis of the cipher alphabets. This is so, of course, only when the cipher alphabets are interrelated secondary alphabets produced by sliding components or their equivalents. Reference is now made to the classification set forth in par. 8, in connection with the types of alphabets which may be employed in polyalphabetic substitution. It will be seen that thus far only Cases Ia and b have been treated. Case IIa will now be discussed.

b. Here one of the components, the plain component, is the normal sequence, while the cipher component is a mixed sequence, the various juxtapositions of the two components yielding mixed alphabets. The mixed component may be a systematically-mixed or a random-mixed sequence. If the 25 successive displacements of the mixed component are recorded in separate lines, a symmetrical cipher square such as that shown in Fig. 17 results therefrom. It is identical in form with the square table shown in Fig. 9.

Plain

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

Figure 17.

c. Such a cipher square may be used in exactly the same manner as the Vigenère square. With the key word BLUE and conforming to the normal enciphering equations ($\theta_{k/2} = \theta_{i/1}$; $\theta_{p/1} = \theta_{c/2}$), the following lines of the square would be used:

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

Figure 18a.

These lines would, of course, yield the following cipher alphabets:

(1)	Plain....	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
	Cipher...	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H
(2)	Plain....	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
	Cipher...	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z
(3)	Plain....	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
	Cipher...	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S
(4)	Plain....	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
	Cipher...	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L

Figure 18b.

28. Principles of direct symmetry of position.--a. It was stated directly above that Fig. 17 is a symmetrical cipher square, by which is meant that the letters in its successive horizontal lines show a symmetry of position with respect to one another. They constitute, in reality, one and only one sequence or series of letters, the sequences being merely displaced successively 1, 2, 3, ... 25 intervals. The symmetry exhibited is obvious and is said to be visible, or direct. This fact can be used to good advantage, as will presently be demonstrated.

b. Consider, for example, the pair of letters G_c and V_c in cipher alphabet (1) of Fig. 18b. The letter V_c is the 15th letter to the right of G_c. In cipher alphabet (2), V_c is also the 15th letter to the right of G_c, as is the case in each of the four cipher alphabets in Fig. 18b, since the relative positions they occupy are the same in each horizontal line in Fig. 18a, that is, in each of the successive recordings of the cipher component as the latter is slid to the right against the plain or normal component. If, therefore, the relative positions occupied by two letters, θ_1 and θ_2, in such a cipher alphabet, C_1, are known, and if the position of θ_1 in another cipher alphabet, C_2, belonging to the same series is known, then θ_2 may at once be placed into its correct position in C_2. Suppose, for example, that as the result of an analysis based upon considerations of frequency, the following values in four cipher alphabets have been tentatively determined:

(1)	Plain....	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
	Cipher...					G										Y					V						
(2)	Plain....	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
	Cipher...					N										G					P						
(3)	Plain....	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
	Cipher...					L										B					I						
(4)	Plain....	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
	Cipher...					W										I					Q						

Figure 19a.

c. The cipher components of these four secondary alphabets may, for convenience, be assembled into a cellular structure, hereinafter called a sequence reconstruction matrix, as shown in Fig. 19b. Regarding the top line of the reconstruction matrix in Fig. 19b as being common to all four secondary cipher alphabets listed in Fig. 19a, the successive lines of the reconstruction matrix may now be termed cipher alphabets, and may be referred to by the numbers at the left.

		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
Plain																											
	1					G										Y					V						
	2					N										G					P						
Cipher	3					L										B					I						
	4					W										I					Q						

Figure 19b.

d. The letter G is common to Alphabets 1 and 2. In Alphabet 2 it is noted that N occupies the 10th position to the left of G, and the letter P occupies the 5th position to the right of G. One may therefore place these letters, N and P, in their proper positions in Alphabet 1, the letter N being placed 10 letters before G, and the letter P, 5 letters after G. Thus:

	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
Plain																										
1						G				P					Y					V	N					

Thus, the values of two new letters in Alphabet 1, viz., $P_c = J_p$, and $N_c = U_p$ have been automatically determined; these values were obtained without any analysis based upon the frequency of P_c and N_c. Likewise, in Alphabet 2, the letters Y and V may be inserted in these positions:

	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
Plain																										
2				V	N										G					P					Y	

This gives the new values $V_c = D_p$ and $Y_c = Y_p$ in Alphabet 2. Alphabets 3 and 4 have a common letter I, which permits of the placement of Q and W in Alphabet 3, and of B and L in Alphabet 4.

e. The new values thus found are of course immediately inserted throughout the cryptogram, thus leading to the assumption of further values in the cipher text. This process, viz., the reconstruction of the primary components, by the application of the principles of direct symmetry of position to the cells of the reconstruction matrix, thus facilitates and hastens solution.

f. It must be clearly understood that before the principles of direct symmetry of position can be applied in cases such as the foregoing, it is necessary that the plain component be a known sequence. Whether it is the normal sequence or not is immaterial, so long as the sequence is known. Obviously, if the sequence is unknown, symmetry, even if present, cannot be detected by the cryptanalyst because he has no base upon which to try out his assumptions for symmetry. In other words, direct symmetry of position

is manifested in the illustrated example because the plain component is a known sequence, and not because it is the normal alphabet. The significance of this point will become apparent later on in connection with the problem discussed in par. 36b.

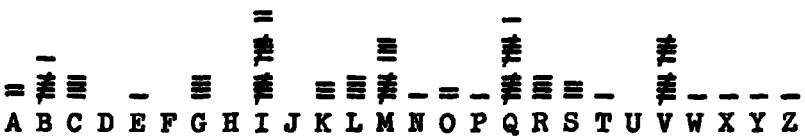
29. Initial steps in the solution of a typical example.--a. In the light of the foregoing principles let a typical message now be studied.

```
                5         10        15        20        25        30
A.  QWBRI VWYCA ISPJL RBZEY QWYEU LWMGW
B.  ICJCI MTZEI MIBKN QWBRI VWYIG BWNBQ
C.  QCGQH IWJKA GEGXN IDMRU VEZYG QIGVN
D.  CTGYO BPDBL VCGXG BKZZG IVXCU NTZAO
E.  BWFEQ QLFCO MTYZT CCBYQ OPDKA GDGIG
F.  VPWMR QIIEW ICGXG BLGQQ VBGRS MYJJY
G.  QVFWY RWNFL GXNFW MCJKX IDDRU OPJQQ
H.  ZRHCN VWDYQ RDGDG BXDBN PXFPU YXNFG
J.  MPJEL SANCD SEZZG IBEYU KDHCA MBJJF
K.  KILCJ MFDZT CTJRD MIYZQ ACJRR SBGZN
L.  QYAHQ VEDCQ LXNCL LVVCS QWBII IVJRN
M.  WNBRI VPJEL TAGDN IRGQP ATYEW CBYZT
N.  EVGQU VPYHL LRZNQ XINBA IKWJQ RDZYF
P.  KWFZL GWFJQ QWJYQ IBWRX
```

b. The principal repetitions of three or more letters have been underlined in the message and the factors (up to 26 only) of the intervals between them are as follows:

QWBRIVWY	45	3, 5, 9, 15
CGXGB	60	2, 3, 4, 5, 6, 10, 12, 15, 20
PJEL	95	5, 19
ZZGI	145	5
BRIV	330	2, 3, 5, 6, 10, 11, 13, 22
BRIV	285	3, 5, 15, 19
KAG	75	3, 5, 15, 25
QRD	165	3, 5, 15
QWB	45	3, 5, 9, 15
QWB	275	5, 11, 25
WIC	130	2, 5, 10, 13, 26
XNF	45	3, 5, 9, 15
YZT	225	3, 5, 15, 25
ZTC	145	5

The factor 5 is common to all of these repetitions, and there seems to be every indication that five alphabets are involved. Since the message already appears in groups of five letters, it is unnecessary in this case to rewrite it in groups corresponding to the length of the key. The uniliteral frequency distribution for Alphabet 1 is as follows:



c. Attempts to fit this distribution to the normal on the basis of a direct or reversed standard alphabet do not give positive results, and it is assumed that mixed alphabets are involved. Individual triliteral frequency distributions are then compiled and are shown below. These tables are similar to those made for single mixed-alphabet ciphers, and are made in the same way except that instead of taking the letters one after the other, the letters which belong to the separate alphabets now must be assembled in separate tables. For example, in Alphabet 1, the trigraph QAC means that A occurs in Alphabet 1; Q, its prefix, occurs in Alphabet 5, and C, its suffix, occurs in Alphabet 2. All confusion may be avoided by placing numbers indicating the alphabets in which they belong above the letters, thus: $\overset{512}{\text{QAC}}$

~~CONFIDENTIAL~~

Alphabet 1

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
QC	GW	NT		TV		AE		AS		UD	UW	IT	UT	QP	NX	-W	LB	LA	LA		IW	NN	QI	UX	QR
PT	OP	TG				AD		WC		FI	QX	II		UP		YW	YW	DE			IW				
	GK	TT				LX		HW		FW	LV	OT				NW	QD	RB			UE				
	OW	WB				LW		ND			LR	SY				QC	QD				LC				
	GL							GV				WC				GI					GP				
	GX							WC				GP				QL					QB				
								XD				AB				RI					NW				
								GB				JF				YV					QE				
								IV				DI				NY					IP				
								NR								SW					UP				
								AK								QW									
								QB																	

Alphabet 2

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
SN	RZ	IJ	IM	GG	MD			MB		IW	QF		WB		BD		ZH	IP	MZ		IX	QB	GN	MJ	
TG	VG	QG	GG	VZ				QG		BZ	BG				OD		IG		CG		QF	VY	BD	QA	
	IE	VG	ID	SZ				QI							VW		LZ		NZ		LV	QY	PF		
	MJ	CB	RG	VD				KL							OJ				MY		IJ	LM	YN		
	SG	IG	KH					MY							MJ				CJ		EG	QB	LN		
	CY	MJ	RZ					XN							VJ				AY			VY			
	IW	AJ													VY							BN			
																						IJ			
																						BF			
																						RN			
																						VD			
																						QB			
																						KF			
																						GF			
																						QJ			

Alphabet 3

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
YH	WR		PB	BY	WE	CQ	RC	IE	CC		IC	WG	WB		SJ						VC	PM	VC	WC	BE
	IK		PK		LC	EX	DC		WK			DR	WF									KJ		WE	TE
	WR		DR		VW	IV			YJ				XF									BR		WI	EY
	CY		WY		XP	TY			CK				XF											TZ	KZ
	WI		XB		WZ	CX			PQ				AC											IZ	TA
	NR		FZ		WJ	DI			PE				XC											TE	EZ
			EC			CX			BJ				IB											BZ	RN
						LQ			TR															PH	DY
						BR			CR																
						DD			VR																
						BZ			PE																
						AD			WY																
						RQ																			
						VQ																			

~~CONFIDENTIAL~~

Alphabet 4

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
ZO	NQ	YA	GG	ZY	NL	MW	AQ	YG	PL	BN		WR	ZQ		FU	GH	BI				GN	FY	GN	ZG	ZG
	DL	JI	GN	YU	NW		YL	GG	JY	JA						GQ	BI						GG	GO	YT
	DN	XU		ZI	NG			BI	JF	DA						JQ	MU						GG	BQ	ZG
	NA	FO		FQ					WQ	JX						GP	GS							DQ	DT
		HN		IW					FQ							GU	DU							EU	YQ
		ND		JL													JD							ZF	GN
		HA		JL													JR							JQ	YT
		LJ		YW													JN								FL
		DQ															BI								
		NL															WX								
		VS																							

Alphabet 5

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
CI			CS		JK	IB	QI	RV	CM		JR		KQ	YB	QA	BQ	MQ	RM	ZC	EL		GI	KI	EQ	
KG			RM		YK	YQ		CM			BV		XI	AB		EQ	RS	CQ	ZC	RV		EI	R-	JQ	
KG						XB		EM			FG		VC	CM		YO			ZE	CN		FM		WR	
CM						ZI		RV			ES		CV			QV				RO		EC			
BI						IV		II			CL		BP			QZ				PY					
						XB		RV			ET		ZQ			YR				YK					
						DB					HL		RW			ZA				QV					
						FM					ZG		DI			HV									
						ZI										CL									
																NX									
																JR									
																JQ									
																YI									

Condensed table of repetitions

1-2-3-4-5-1-2-3
Q W B R I V W Y-2

2-3-4-5-1
C G X G B-2

2-3-4-5
P J E L-2

3-4-5-1
B R I V-3
Z Z G I-2

1-2-3
Q W B-3

2-3-4
X N F-2

3-4-5
Y Z T-2

4-5-1
K A G-2
Z T C-2

5-1-2
Q R D-2
W I C-2

1-2
Q W-5
V P-3
V W-3

2-3
C G-3
C J-3
P J-3
W B-3
W F-3
W Y-3
X N-3

3-4
B R-3
G Q-4
G X-3
J R-3
N F-3
Y Z-3

4-5
R I-3
Y Q-3
Z T-3

5-1
G B-4
I V-3
Q Q-3

d. One now proceeds to analyze each alphabet distribution, in an endeavor to establish identifications of cipher equivalents. First, of course, attempts should be made to separate the vowels from the consonants in each alphabet, using the same test as in the case of a single mixed-alphabet cipher. There seems no doubt that $\overset{2}{W}_c$ and $\overset{5}{Q}_c$ are equivalents of E_p. In the other alphabets the equivalents of E_p are not so clear-cut, but for the moment, let us assume that E_p is the highest θ_c in the particular alphabet, viz., $\overset{1}{I}_c$, $\overset{3}{G}_c$, and $\overset{4}{C}_c$.

e. The letters of greatest frequency in Alphabet 1 are I, M, Q, V, B, G, L, R, S, and C. I_c has already tentatively been assumed to be E_p. If $\overset{2}{W}_c$ and $\overset{5}{Q}_c = E_p$, then one should be able to distinguish the vowels from the consonants among the letters I, M, Q, V, B, G, L, R, S, and C by examining the prefixes of $\overset{2}{W}_c$, and the suffixes of $\overset{5}{Q}_c$. The prefixes and suffixes of these letters, as shown by the triliteral frequency distributions, are these:

Prefixes of $\overset{2}{W}_c$ ($=\overset{2}{E}_p$)	Suffixes of $\overset{5}{Q}_c$ ($=\overset{5}{E}_p$)
Q G K V R B I L	I Q R X L V A Z O

f. Consider now the letter $\overset{1}{M}_c$; it does not occur either as a prefix of $\overset{2}{W}_c$, or as a suffix of $\overset{5}{Q}_c$. Hence it is most probably a vowel, and on account of its high frequency it may be assumed to be O_p. On the other hand, note that $\overset{1}{Q}_c$ occurs five times as a prefix of $\overset{2}{W}_c$ and three times as a suffix of $\overset{5}{Q}_c$. It is therefore a consonant, most probably R_p, for it would give the digraph $ER_p(=\overset{51}{QQ}_c)$ as occurring three times and $RE_p(=\overset{12}{QW}_c)$ as occurring five times.

g. The letter $\overset{1}{V}_c$ occurs three times as a prefix of $\overset{2}{W}_c$ and twice as a suffix of $\overset{5}{Q}_c$. It is therefore a consonant, and on account of its frequency, let it be assumed to be T_p. The letter $\overset{1}{B}_c$ occurs twice as a prefix of $\overset{2}{W}_c$ but not as a suffix of $\overset{5}{Q}_c$. Its frequency is only medium, and it is probably a consonant. In fact, the twice repeated digraph $\overset{12}{BW}_c$ is once a part of the trigraph $\overset{512}{GBW}$, and $\overset{5}{G}_c$, the letter of second highest frequency in Alphabet 5,

looks excellent for T_p. Might not the trigraph GBW^{512} be $\overline{THE}_p$? It will be well to keep this possibility in mind.

h. The letter G_c^1 occurs only once as a prefix of W_c^2 and does not occur as a suffix of Q_c^5. It may be a vowel, but one cannot be sure. The letter L_c^1 occurs once as a prefix of W_c^2 and once as a suffix of Q_c^5. It may be considered to be a consonant. R_c^1 occurs once as a prefix of W_c^2, and twice as a suffix of Q_c^5, and is certainly a consonant. Neither the letter S_c^1 nor the letter C_c^1 occurs as a prefix of W_c^2 or as a suffix of Q_c^5; both would seem to be vowels, but a study of the prefixes and suffixes of these letters lends more weight to the assumption that C_c^1 is a vowel than that S_c^1 is a vowel. For all the prefixes of C, viz., N^5, T^5, and W^5, are in subsequent analysis of Alphabet 5 classified as consonants, as are likewise its suffixes, viz., T, C, and B in Alphabet 2. On the other hand, only one prefix, L_c^5, and one suffix, B_c^2, of S_c^1 are later classified as consonants. Since vowels are more often associated with consonants than with other vowels, it would seem that C_c^1 is more likely to be a vowel than S_c^1. At any rate C_c^1 is assumed to be a vowel, for the present, leaving S_c^1 unclassified.

i. Going through the same steps with the remaining alphabets, the following results are obtained:

Alphabet	Vowels	Consonants
1	I, M, C.	Q, V, B, L, R, G?
2	W, P, I.	B, C, D, T.
3	G, Z.	J, N, D, Y, F.
4	C, E?, R?, B?.	Y, Z, J, Q.
5	Q, U.	G, N, A, I, W, L, T.

30. Application of principles of direct symmetry of position.--a. The next step is to try to determine a few values in each alphabet. In Alphabet 1, from the foregoing analysis, the following data are on hand:

```
Plain.... 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
Cipher... C?      I      C?          M     Q  V
```

Let the values of E_p already assumed in the remaining alphabets, be set down in a reconstruction matrix, as follows:

		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
Plain																											
	1	C?				I				C?						M			Q		V						
	2					V																					
Cipher	3					G																					
	4					C																					
	5					Q																					

b. It is seen that by good fortune the letter Q is common to Alphabets 1 and 5, and the letter C is common to Alphabets 1 and 4. If it is assumed that one is dealing with a case in which a mixed component is sliding against the normal component, one can apply the principles of direct symmetry of position to these alphabets, as outlined in par. 28. For example, one may insert the following values in Alphabet 5:

		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
Plain																											
	1	C?				I				C?						M			Q		V						
Cipher	5		M			Q		V						C?					I			C?					

c. The process at once gives three definite values: $\overset{5}{M}_c = B_p$, $\overset{5}{V}_c = G_p$, $\overset{5}{I}_c = R_p$. Let these deduced values be substantiated by referring to the frequency distribution. Since B and G are normally low or medium frequency letters in plain text, one should find that M_c and V_c, their hypothetical equivalents in Alphabet 5, should have low frequencies. As a matter of fact, they do not appear in this alphabet, which thus far corroborates the assumption. On the other hand, since $\overset{5}{I}_c = R_p$, if the values derived from symmetry of position are correct, $\overset{5}{I}_c$ should be of high frequency and reference to the distribution shows that I_c is of high frequency. The position of C is doubtful, it belongs either under N_p or V_p. If the former is correct, then the frequency of $\overset{5}{C}_c$ should be high, for it would equal N_p; if the latter is correct, then its frequency should be low, for it would equal V_c. As a matter of fact, $\overset{5}{C}_c$ does not occur, and it must be concluded that it belongs under V_p. This in turn settles the value of $\overset{1}{C}_c$, for it must now be placed definitely under I_p and removed from beneath A_p.

d. The definite placement of C now permits the insertion of new values in Alphabet 4, and one now has the following:

Plain	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
1					I				C						M			Q		V						
2					W																					
3					G																					
4	I				C						M				Q		V									
5		M			Q		V												I				C			

31. Subsequent steps in solution.--a. It is high time that the thus far deduced values, as recorded in the reconstruction matrix, be inserted in the cipher text, for by this time it must seem that the analysis has certainly gone too far upon unproved hypotheses. The following results are obtained:

	5					10					15					20					25					30
A.	Q	W	B	R	I	V	W	Y	C	A	I	S	P	J	L	R	B	Z	E	Y	Q	W	Y	E	U	L	W	M	G	W
	R	E		R		T	E		E		E										R	E				E				
B.	I	C	J	C	I	M	T	Z	E	I	M	I	B	K	N	Q	W	B	R	I	V	W	Y	I	G	B	W	N	B	Q
	E		E	R		O			R		O					R	E		R		T	E		A		E			E	
C.	Q	C	G	Q	H	I	W	J	K	A	G	E	G	X	N	I	D	M	R	U	V	E	Z	Y	G	Q	I	G	V	N
	R		E	N		E	E					E				E					T					R		E	P	
D.	C	T	G	Y	O	B	P	D	B	L	V	C	G	X	G	B	K	Z	Z	G	I	V	X	C	U	N	T	Z	A	O
	I		E								T		E								E			E						
E.	B	W	F	E	Q	Q	L	F	C	O	M	T	Y	Z	T	C	C	B	Y	Q	O	P	D	K	A	G	D	G	I	G
		E		E		R		E			O					I			E							E	A		
F.	V	P	W	M	R	Q	I	I	E	W	I	C	G	X	G	B	L	G	Q	Q	V	B	G	R	S	M	Y	J	J	Y
	T			K		R					E		E				E	N	E		T		E			O				
G.	Q	V	F	W	Y	R	W	N	F	L	G	X	N	F	W	M	C	J	K	X	I	D	D	R	U	O	P	J	Q	Q
	R						E									O					E							N	E	
H.	Z	R	H	C	N	V	W	D	Y	Q	R	D	G	D	G	B	X	D	B	N	P	X	F	P	U	Y	X	N	F	G
			E			T	E			E			E																	
J.	M	P	J	E	L	S	A	N	C	D	S	E	Z	Z	G	I	B	E	Y	U	K	D	H	C	A	M	B	J	J	F
	O							E								E							E			O				
K.	K	I	L	C	J	M	F	D	Z	T	C	T	J	R	D	M	I	Y	Z	Q	A	C	J	R	R	S	B	G	Z	N
			E			O					I					O			E								E			
L.	Q	Y	A	H	Q	V	E	D	C	Q	L	X	N	C	L	L	V	V	C	S	Q	W	B	I	I	I	V	J	R	N
	R			E		T			E	E				E					E		R	E		A	R	E				
M.	W	N	B	R	I	V	P	J	E	L	T	A	G	D	N	I	R	G	Q	P	A	T	Y	E	W	C	B	Y	Z	T
				R		T							E			E		E	N							I				
N.	E	V	G	Q	U	V	P	Y	H	L	L	R	Z	N	Q	X	I	N	B	A	I	K	W	J	Q	R	D	Z	Y	F
			E	N		T									E						E				E					
P.	K	W	F	Z	L	G	W	F	J	Q	Q	W	J	Y	Q	I	B	W	R	X											
		E					E			E	R	E			E	E														

b. The combinations given are excellent throughout and no inconsistencies appear. Note the trigraph $\overset{123}{QWB}$, which is repeated in the following polygraphs (underlined in the foregoing text):

1	2	3	4	5	1				5	1	2	3	4	5	1
Q	W	B	R	I	V	.	.	.	S	Q	W	B	I	I	I
R	E			R	T	.	.	.		R	E		A	R	E

c. The letter B_c is common to both polygraphs, and a little imagination will lead to the assumption of the value $B_c = P_p$, yielding the following:

1	2	3	4	5	1				5	1	2	3	4	5	1
Q	W	B	R	I	V	.	.	.	S	Q	W	B	I	I	I
R	E	P	O	R	T	.	.	.	P	R	E	P	A	R	E

d. Note also (at E29) the polygraph $\overset{4}{I}\ \overset{5}{G}\ \overset{1}{V}\ \overset{2}{P}\ \overset{3}{W}\ \overset{4}{M}$ (A T K), which looks like the word ATTACK. The frequency distributions are consulted to see whether the frequencies given for $\overset{5}{G}_c$ and $\overset{2}{P}_c$ are high enough for T_p and A_p, respectively, and also whether the frequency of $\overset{3}{W}_c$ is good enough for C_p, it is noted that they are excellent. Moreover, the digraph $\overset{51}{GB}_c$, which occurs four times, looks like TH, thus making $\overset{1}{B}_c = H_p$. Does the insertion of these four new values in our diagram of alphabets bring forth any inconsistencies? The insertion of the value $\overset{2}{P}_c = A_p$ and $\overset{1}{B}_c = H_p$ gives no indications either way, since neither letter has yet been located in any of the other alphabets. The insertion of the value $\overset{5}{G}_c = T_p$ gives a value common to Alphabets 3 and 5, for the value $\overset{3}{G}_c = E_p$ was assumed long ago. Unfortunately an inconsistency is found here. The letter I has been placed two letters to the left of G in the mixed component, and has given good results in Alphabets 1 and 5; if the value $\overset{3}{W}_c = C_p$ (obtained above from the assumption of the word ATTACK) is correct, then W, and not I, should be the second letter to the left of G. Which shall be retained? There has been so far nothing to establish the value of $\overset{3}{G}_c = E_p$; this value was assumed from frequency considerations solely. Perhaps it is wrong. It certainly behaves like a vowel, and one may see what happens when one changes its value to O_p. The following placements in the reconstruction matrix result from the analysis, when only two or three new values have been added as a result of the clues afforded by the deductions:

~~CONFIDENTIAL~~

Plain	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
1			S		I		G	B	C					M		P	Q	R	V	W						
2	P	Q	R	V	W							S		I		G	B	C							M	
3	R	V	W							S		I		G	B	C						M		P	Q	
4	I		G	B	C					M		P	Q	R	V	W								S		
5		M		P	Q	R	V	W									S		I		G	B	C				

e. Many new values are produced, and these are inserted throughout the message, yielding the following:

	5					10					15					20					25					30
A.	Q	W	B	R	I		V	W	Y	C	A		I	S	P	J	L		R	B	Z	E	Y		Q	W	Y	E	U		L	W	M	G	W
	R	E	P	O	R		T	E		E			E	M	Y				S	R					R	E						E	W	C	H
B.	I	C	J	C	I		M	T	Z	E	I		M	I	B	K	N		Q	W	B	R	I		V	W	Y	I	G		B	W	N	B	Q
	E	S		E	R		O			R			O	O	P				R	E	P	O	R		T	E		A	T		H	E		D	E
C.	Q	C	G	Q	H		I	W	J	K	A		G	E	G	X	N		I	D	M	R	U		V	E	Z	Y	G		Q	I	G	V	N
	R	S	O	N			E	E					G		O				E		W	O			T			T			R	O	O	P	
D.	C	T	G	Y	O		B	P	D	B	L		V	C	G	X	G		B	K	Z	Z	G		I	V	X	C	U		N	T	Z	A	O
	I		O				H	A		D			T	S	O		T		H				T		E	D		E							
E.	B	W	F	E	Q		Q	L	F	C	O		M	T	Y	Z	T		C	C	B	Y	Q		O	P	D	K	A		G	D	G	I	G
	H	E		E			R			E			O						I	S	P		E		A			G			O	A	T	
F.	V	P	W	M	R		Q	I	I	E	W		I	C	G	X	G		B	L	G	Q	Q		V	B	G	R	S		M	Y	J	J	Y
	T	A	C	K	F		R	O	M		H		E	S	O		T		H		O	N	E		T	R	O	O	P		O				
G.	Q	V	F	W	Y		R	W	N	F	L		G	X	N	F	W		M	C	J	K	X		I	D	D	R	U		O	P	J	Q	Q
	R	D		Q			S	E					G				H		O	S					E		O					A		N	E
H.	Z	R	H	C	N		V	W	D	Y	Q		R	D	G	D	G		B	X	D	B	N		P	X	F	P	U		Y	X	N	F	G
		C		E			T	E			E		S		O		T		H			D			Q		M								T
J.	M	P	J	E	L		S	A	N	C	D		S	E	Z	Z	G		I	B	E	Y	U		K	D	H	C	A		M	B	J	J	F
	O	A					C			E			C				T		E	R							E				O	R			
K.	K	I	L	C	J		M	F	D	Z	T		C	T	J	R	D		M	I	Y	Z	Q		A	C	J	R	R		S	B	G	Z	N
		O		E			O						I		O				O	O			E			S		O	F		C	R	O		
L.	Q	Y	A	H	Q		V	E	D	C	Q		L	X	N	C	L		L	V	V	C	S		Q	W	B	I	I		I	V	J	R	N
	R				E		T			E	E					E				D	B	E	P		R	E	P	A	R		E	D		O	
M.	W	N	B	R	I		V	P	J	E	L		T	A	G	D	N		I	R	G	Q	P		A	T	Y	E	W		C	B	Y	Z	T
	U		P	O	R		T	A							O				E	C	O	N	D					H			I	R			
N.	E	V	G	Q	U		V	P	Y	H	L		L	R	Z	N	Q		X	I	N	B	A		I	K	W	J	Q		R	D	Z	Y	F
		D	O	N			T	A						C			E			O		D			E			E		S				
P.	K	W	F	Z	L		G	W	F	J	Q		Q	W	J	Y	Q		I	B	W	R	X													
		E					G	E			E		R	E			E		E	R		O													

32. Completing the solution.--a. Completion of solution is now a very easy matter. The mixed component is finally found to be the following sequence, based upon the word EXHAUSTING:

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

The completely reconstructed enciphering matrix is shown in Fig. 19.

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

Figure 19.

b. Note that the successive equivalents of A_p spell the word APRIL, which is the key for the message. The plaintext message is as follows:

REPORTED ENEMY HAS RETIRED TO NEWCHESTER. ONE TROOP IS REPORTED AT HENDERSON MEETING HOUSE. TWO OTHER TROOPS IN ORCHARD AT SOUTHWEST EDGE OF NEWCHESTER. SECOND SQ IS PREPARING TO ATTACK FROM THE SOUTH. ONE TROOP OF THIRD SQ IS ENGAGING HOSTILE TROOP AT NEWCHESTER. REST OF THIRD SQ IS MOVING TO ATTACK NEWCHESTER FROM THE NORTH. MOVE YOUR SQ INTO WOODS EAST OF CROSSR/OADS/FIVE THREE NINE AND BE PREPARED TO SUPPORT ATTACK OF SECOND AND THIRD SQ. DO NOT ADVANCE BEYOND NEWCHESTER. MESSAGES HERE.

TREER, COL.

c. The preceding case is a good example of the value of the principles of direct symmetry of position when applied properly to a cryptogram enciphered by the sliding of a mixed component against the normal. The cryptanalyst starts off with only a very limited number of assumptions and builds up many new values as a result of the placement of the few original values in the reconstruction matrix.

33. Solution of subsequent messages enciphered by the same cipher component.--a. Let it be supposed that the correspondents are using the same basic or primary components but with different key words for other messages. Can the knowledge of the sequence of letters in the reconstructed mixed primary component be used to solve the subsequent messages? It has been shown that in the case of a monoalphabetic cipher in which a mixed alphabet was used, the process of completing the plain-component sequence could be applied to solve subsequent messages in which the same components were used, even though the cipher component was set at a different key letter. A modification of the procedure used in that case can be used in this case, where a plurality of cipher alphabets based upon sliding primary components is used.

b. Let it be supposed that the following message passing between the same two correspondents as in the preceding message has been intercepted:

SFDZR YRRKX MIWLL AQRLU RQFRT LJQKF XUWBS MDJZK MICQC UDPTV

TYRNH TRORV BQLTI QBNPR RTUHD PTIVE RMGQN LRATQ PLUKR KGRZF

JCMGP IHSMR GQRFX BCABA OEMTL PCXJM RGQSZ VB

c. The presence in this size sample of a tetragraphic repetition whose interval is 21 letters suggests a key word of three or seven letters; the repeated trigraph at an interval of 28 makes seven as the more probable hypothesis. There are very few other repetitions, and this is to be expected in short messages with a key of such length.

d. Let the message be written in groups of seven letters, in columnar fashion, as shown in Fig. 20a. The letters in each column belong to a single alphabet. Let the first ten letters in each column be converted into their plain-component equivalents by setting the reconstructed cipher component against the normal plain component at any arbitrarily selected point, such as in the following alphabet:

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

The columns of equivalents are now as shown in Fig. 20b.

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

Figure 20a.

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

Figure 20b.

e. It has been shown that in the case of a monoalphabetic cipher involving a mixed cipher component it was merely necessary to complete the normal alphabetic sequence beneath the plain-component equivalents and all the

plain text reappeared on one generatrix. It was also found that in the case of a polyalphabetic cipher involving standard alphabets, the plain-text equivalents of each alphabet reappeared on the same generatrix, and it was necessary only to combine the proper generatrices in order to produce the plain text of the message. In the case at hand both processes are combined: the normal alphabetical sequence is continued beneath the letters of each column and then the generatrices are combined to produce the plain text. The generatrix diagrams for the first five alphabets (i.e., columns) are shown in Fig. 21, below. Only the first ten letters in each generatrix are used in the diagrams, since the application of the generatrix elimination and rough scoring procedures discussed in pars. 21f and g will yield a solution.

Gen.	Alphabet 1	Alphabet 2	Alphabet 3	Alphabet 4	Alphabet 5
	SRLQKMCVRT	FKAFFDQTOI	DXQRXJCYRQ	ZMRTUZURVB	RILIWKDNBN
1	1 FVQUPRLWVG	4 NPDNNMUGSH	1 MBUVBOLIVU	[illegible]	[illegible]
2	[illegible]	7 OQEOONVHTI	1 NCVWCPMZWV	3 ASWHFAFWXL	[illegible]
3	[illegible]	3 PRFPPOWIUJ	[illegible]	[illegible]	[illegible]
4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	[illegible]	[illegible]	[illegible]	[illegible]	3 AMVMCURNPN
7	2 LBWAVXRCBM	6 TVJTTSAMYN	5 SHABHUREBA	[illegible]	5 BNWNDVSOQO
8	2 MCXBWYSDCN	[illegible]	4 TIBCIVSFCB	2 GYCNLGLCDR	5 COXOEWTPRP
9	[illegible]	2 VXLVVUCOAP	[illegible]	3 HZDOMEMDES	[illegible]
10	4 OEZDYAUFEP	0 WYMWWVDPBQ	[illegible]	8 IAEPNINEFT	[illegible]
11	[illegible]	[illegible]	3 WLEFLYVIFE	[illegible]	5 FRARHZWSUS
12	2 QGBFACWHGR	4 YAOYYXFRDS	[illegible]	[illegible]	6 GSBSIAXTVT
13	3 RECGBDXIHS	[illegible]	[illegible]	[illegible]	2 HTCTJBYUWU
14	5 SIDHCEYJIT	[illegible]	4 ZOHIOBYLIH	[illegible]	[illegible]
15	[illegible]	3 BDRBBAIUGV	[illegible]	[illegible]	2 JVEVLDAWYW
16	[illegible]	2 CESCCBJVHW	[illegible]	[illegible]	[illegible]
17	0 VLGKFHBMLW	[illegible]	[illegible]	1 PHLWUPULMA	[illegible]
18	2 WMHLGICNMX	[illegible]	2 DSLMSFCPML	[illegible]	[illegible]
19	[illegible]	[illegible]	5 ETMNTGDQNM	5 RJNYWRWNOC	[illegible]
20	[illegible]	3 GIWGGFNZLA	6 FUNOUHERON	[illegible]	[illegible]
21	[illegible]	[illegible]	4 GVOPVIFSPO	4 TLPAYTYPQE	[illegible]
22	[illegible]	4 IKYIIHPBNC	[illegible]	[illegible]	[illegible]
23	4 BRMQLNHSRC	[illegible]	[illegible]	6 VNRCAVARSG	5 RDMDTLIEGE
24	7 CSNRMOITSD	[illegible]	5 JYRSYLIVSR	4 WOSDBWBSTH	4 SENEUMJFHF
25	6 DTOSNPJUTE	[illegible]	[illegible]	[illegible]	4 TFOFVNKGIG
26	[illegible]	3 MOCMMLTFRG	[illegible]	[illegible]	1 UGPGWOLHJH

Figure 21.

f. A trial of the generatrices with the highest scores in the first three alphabets yields the trigraphs shown in Fig. 22a. The generatrices of the subsequent columns are examined to select those which may be added

to those already selected in order to build up the plain text. The results are shown in Fig. 22b. Note that, in this case, the correct

1 2 3
C O F
S Q U
N E N
R O O
M O U
O N H
I V E
T H R
S T O
D I N

Figure 22a.

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

Figure 22b.

generatrix for Alphabet 5 is not the one with the highest score (6), but one of the four generatrices with a score of 5. The generatrix process is a very valuable aid in the solution of messages after the primary components have been recovered as a result of the longer and more detailed analysis of the frequency distributions of the first message intercepted. Very often a short message can be solved in no other way than the one shown, if the primary components are completely known.

g. It may be of interest to find the key word for the message. Assuming that enciphering method number 1 (see par. 13f, page 20) were known to be employed, all that is necessary is to set the mixed component of the cipher alphabet underneath the plain component so as to produce the cipher letter indicated as the equivalent of any given plaintext letter in each of the alphabets. For example, in the first alphabet it is noted that $C_p = S_c$. Adjust the two components under each other so as to bring S of the cipher component beneath C of the plain component, thus:

Plain: ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher: EXHAUSTINGBCDFJKLMOPQRVWYZEXHAUSTINGBCDFJKLMOPQRVWYZ

It is noted that $A_p = A_c$. Hence, the first letter of the key word to the message is A. The 2d, 3d, 4th,...7th key letters are found in exactly the same manner, and the following is obtained:

When C O F I R S T equals
S F D Z R Y R then A_p successively equals
A Z I M U T H

34. Statistical methods for the determination of correct generatrices. --a. The student has seen the advantages of the simple two-category weighting procedure, as demonstrated in subpars. 21f and g, over the method of ocular inspection. These advantages are that, first of all, the two-category weighting system is very easy to apply mentally, and, secondly,

almost any scoring system will be more accurate and measurable than will a mere "appraisal" by eye which may be rather subjective or intuitive in nature. This second point is especially true when the number of letters in the generatrices is small, that is, around 10 letters or so.

b. Instead of the system of two-category weights, it is possible to use the summation of the relative frequencies of plaintext letters to evaluate generatrices. For convenience in assigning whole numbers as the frequencies, the following scale (summing to 100) has been used:

7	1	3	4	1	3	3	2	3	7	0	0	4	2	8	8	3	0	8	6	9	3	2	2	0	2	0
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	

This system is only a trifle more sensitive than the two-category system, but the convenience of only two categories for mental arithmetic is lost. Besides, the two-category system actually involves ratios of the frequencies of the two classes of letters and in effect gives a multiplication of the weights of a generatrix, whereas the summation of the plaintext frequencies involves an addition of these frequencies which is not a precise mathematical measure of the relative goodness of a generatrix. The summation of the relative frequencies of letters takes into account only the probability of occurrence of each letter in the generatrix, considered separately; that is, the occurrence of an E has a value of 13, regardless of whatever other letters are present in the same generatrix, and this value is added to the frequencies of the other letters.

c. If instead of the summation of the arithmetical frequencies, logarithms of the frequencies are used and these logarithms are added together, then a true picture of the generatrix is obtained. The reason underlying this fact is that the summation of logarithmic weights is equivalent to multiplying the probabilities of occurrence of all the letters in the generatrix taken together, thus giving an accurate evaluation of the generatrix as a whole. This method is especially valuable when generatrices contain as few as 5 or 6 letters. As an aid to the solution of problems wherein the plain component is a standard alphabet, a set of strips has been printed containing the normal sequence and the respective logarithmic weights over each letter. (If the plain component is any other sequence, strips would have to be prepared manually with that particular sequence inscribed.) The logarithmic weights on these strips are as follows:

8	4	7	7	9	6	5	7	8	1	2	7	6	8	8	6	2	8	8	9	6	5	5	3	6	0	
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	

The numbers on these strips are one-digit logarithms (to the base 133) of the relative frequencies of English plaintext letters as found in Table 3, Appendix 2, Military Cryptanalytics, Part I. For the interested student, the derivation of these one-digit logarithms will be discussed in the next two subparagraphs.

d. Let the following table be examined. Column (a) represents the unilliteral frequencies on a basis of 1000 letters; column (b) represents the logarithms (to the base 10) of these frequencies; column (c) contains the figures of the preceding column with the addition of .009 to each logarithm,

column (d) is a conversion of the basic frequencies in column (a) to two-digit logarithms (base 133); and column (e) is a one-digit logarithmic weight based on the logarithms of the preceding column.

	(a)	(b)	(c)	(d)	(e)
A	73.7	1.868	1.877	.87	8
B	9.7	0.987	0.996	.41	4
C	30.7	1.487	1.496	.70	7
D	42.4	1.628	1.637	.76	7
E	130.0	2.113	2.122	.99	9
F	28.3	1.452	1.461	.68	6
G	16.4	1.214	1.223	.57	5
H	33.9	1.530	1.539	.72	7
I	73.5	1.866	1.875	.87	8
J	1.64	0.214	0.223	.12	1
K	2.96	0.471	0.480	.22	2
L	36.4	1.560	1.569	.73	7
M	24.7	1.392	1.401	.65	6
N	79.5	1.900	1.909	.89	8
O	75.3	1.875	1.884	.88	8
P	26.7	1.427	1.436	.67	6
Q	3.50	0.544	0.553	.26	2
R	75.8	1.880	1.889	.88	8
S	61.2	1.787	1.796	.84	8
T	91.9	1.963	1.972	.92	9
U	26.0	1.415	1.424	.66	6
V	15.3	1.184	1.193	.56	5
W	15.6	1.192	1.201	.56	5
X	4.62	0.664	0.673	.31	3
Y	19.3	1.285	1.294	.61	6
Z	.98	0.991-10	0.000	.00	0

e. The addition of .009 to the common logarithms is for the purpose of transforming the letter of the lowest frequency (Z_p) to the value of .000 for convenience; this addition (which is equivalent to an arithmetic multiplication) does not change the ratios between the basic frequencies. Now the highest frequency (E_p) is given the value .99 and all the other logarithms are scaled proportionally down to Z_p which is 0: this is equivalent to expressing the frequencies in logarithms with a base other then 10, which in this case is 133. The new base (C) used to convert each of the uniliteral frequencies to the logarithmic range 0 to 0.99 is derived as follows, when 130 is the highest frequency (E_p):

$$\text{Let } 130 = C^{0.99}$$

$$\text{Log}_{10} 130 = \text{Log}_{10} C^{0.99}$$

$$\text{Log}_{10} 130 = (0.99)(\text{Log}_{10} C)$$

$$C = \text{Antilog} \frac{\text{Log}_{10} 130}{0.99} = \text{Antilog} \frac{2.122}{0.99}$$

$$C = 133$$

The formula for the computation of the logarithm to the new base (C) of any actual frequency (Y) of a series is:

$$\text{Log}_C Y = \frac{\text{Log}_{10} Y}{\text{Log}_{10} C}$$

It is more convenient to use reciprocals in the conversion of a whole series of logarithmic values, as in this instance. The formula is:

$$(\text{Log}_{10} Y)(\text{Log}_{10} C)^{-1} = \text{Log}_C Y$$

After these two-digit logarithms are derived, they are converted into one-digit logarithms by multiplying them by 10, dropping the single decimal.[1]

f. As an example of the application of these logarithmic weights, let us consider the generatrices of Alphabet 5 in Fig. 21. In this example, Generatrix No. 12 (an incorrect generatrix) had a two-category score of 6, and Generatrix Nos. 7, 8, 11, and 23 had scores of 5. If logarithmic weights had been used, these generatrices would have had the following scores:[2]

Gen. 12:	G S B S I A X T V T	
	5 8 4 8 8 8 3 9 5 9	= 67
Gen. 7:	B N W N D V S O Q O	
	4 8 5 8 7 5 8 8 2 8	= 63
Gen. 8:	C O X O E W T P R P	
	7 8 3 8 9 5 9 6 8 6	= 69
Gen. 11:	F R A R H Z W S U S	
	6 8 8 8 7 0 5 8 6 8	= 64
Gen. 23:	R D M D T L I E G E	
	8 7 6 7 9 7 8 9 5 9	= 75

The results clearly point to Generatrix No. 23 as the correct generatrix. Even if these generatrices had contained only six letters instead of 10, the logarithmic weights would have pointed to the correct generatrix.

---

[1] Logarithms multiplied by 10 are called decibans, logarithms multiplied by 100 are called centibans Logarithmic weights are usually expressed in decibans or centibans for convenience in treatment as integral values

[2] It is interesting to determine what is the numerical expectation of the sum of the logarithmic weights for correct generatrices, as well as the expectation for incorrect ones--in other words, a sort of logarithmic ϕ test The expected value for the correct (i e plaintext) generatrices is calculated by multiplying the logarithmic weights by the probabilities of each letter, summing the results, this sum is then multiplied by the number of letters in the generatrix to give the expected value of the sum of the logarithmic weights for the generatrix The random expectation is the sum of the logarithmic weights in the scale (151) divided by 26 multiplied by the number of letters in the generatrix. Thus the plaintext expectation is 7 6N and the random expectation is 5 8N, where N is the number of letters in the generatrix

g. In order to illustrate the degree of refinement between logarithmic weights and the arithmetical frequency weights mentioned in subpar. 34c, let us consider the following case. Below are the best generatrices from the first three alphabets of a seven-alphabet polyalphabetic cipher; the number at the left of the generatrices is the sum of the arithmetical weights, while the number to the right is the sum of the logarithmic weights.

	Alphabet 1			Alphabet 2			Alphabet 3	
40	O I A G N N	45	40	Z O T W E N	39	32	H N H Z T N	39
37	S M E K R R	41	35	O D I L T C	46	31	C I C U O I	44
36	T N F L S S	46	32	D S X A I R	42	31	M S M E Y S	43
34	B V N T A A	42				31	O U O G A U	41
33	U O G M T T	43				29	S Y S K E Y	39

It will be seen that, although the generatrix OIAGNN in Alphabet 1 has the highest arithmetic sum, nevertheless the most probably correct generatrix as shown by the logarithmic weights is TNFLSS. In Alphabet 2, the generatrix ODILTC has the highest probability of being the correct one; and in the third alphabet the logarithmic sum points to CICUOI as the most likely generatrix.[3] These generatrices when juxtaposed yield the following plaintext fragments, attesting to the validity of the selection:

T O C . . . .
N D I . . . .
F I C . . . .
L L U . . . .
S T O . . . .
S C I . . . .

h. In difficult cases wherein generatrices contain very few letters, one more statistical resource is available to the cryptanalyst. Suppose that in a certain short cryptogram the number of letters in each alphabet is only four, and that in this particular case the generatrix RTIS is selected from the generatrices for Alphabet 1. In Alphabet 2, the generatrices EINP, IMRT, and PTYA (logarithmic weights of 31, 31, and 29, respectively) appear to be the most likely candidates for the correct generatrix. The generatrix RTIS is juxtaposed against the three generatrices of the second alphabet, and now we record the logarithmic weights[4]

[3] In Alphabet 1, the difference between the logarithmic scores between OIAGNN (45) and TNFLSS (46) shows that the latter generatrix is $133^{1} = 1$ 6 times better than the first generatrix, in Alphabet 2 the difference (7) between the logarithmic scores of ZOTWEN and ODILTC shows that the latter is $133^{7} = 31$ times better than the first generatrix, and in Alphabet 3, the difference (5) between the logarithmic scores of HNHZTN and CICUOI shows that the latter is $133^{5} = 11$ times better than the first generatrix

[4] These weights are taken from Table 15, Appendix 2, Military Cryptanalytics Part I The table gives two-digit logarithms to the base 224 of the digraphic frequencies of plain text, for convenience, these logarithms are treated here as centibans by dropping the decimal point This logarithmic method is much more precise then a method wherein only the digraphic frequencies are used to obtain a score

of the digraphs thus formed, as follows:

RE	96	RI	75	RP	59
TI	82	TM	45	TT	67
IN	92	IR	73	IY	0
SP	55	ST	88	SA	71
	335		281		197

The logarithmic score of 335 points to the generatrix EINP as the most probable for Alphabet 2.[5] Thus the selection of the correct generatrices has been reduced to a purely statistical basis which is of great assistance in effecting a quick solution. Moreover, an understanding of the principles involved will be of considerable value in subsequent work.

35. Solution by the probable-word method.--a. Occasionally one may encounter a cryptogram which is so short that it contains no recurrences even of digraphs, and thus gives no indications of the number of alphabets involved. If the sliding mixed components are known, one may apply the methods illustrated in par. 22, assuming the presence of a probable word, checking it against the text and the sliding components to establish a key, if the correspondents are using key words.

b. For example, suppose that the presence of the word ENEMY is assumed in the message in subpar. 33b above. One proceeds to check it against an unknown key word, sliding the already-reconstructed mixed component against the normal and starting with the first letter of the cryptogram, in this manner:

When ENEMY equals
SFDZR then A_p successively equals
XENFW

The sequence XENFW spells no intelligible word. Therefore, the location of the assumed word ENEMY is shifted one letter forward in the cipher text, and the test is made again, just as was explained in subpar. 22d. When the group AQRLU is tried, the key letters ZIMUT are obtained, which, taken as a part of a word, suggest the word AZIMUTH. The method must yield solution when the correct assumptions are made.

c. The placement of probable words in polyalphabetic ciphers may be facilitated by considering (1) the frequency patterns of the letters composing cribs, and (2) the partial idiomorphisms which may be produced in the periodic encipherment of certain cribs.

(1) For instance, in the first case, the plaintext frequency pattern of the word CAVALRY has a distinctive relative high- (H), medium- (M), and low-frequency (L) pattern of MHLHMHL; if the individual monoalphabetic

[5] The generatrix EINP is better than IMRT by a factor of $224^{(3\ 35-2\ 81)} = 224^{0.54} = 18$, likewise, the generatrix EINP is better than PTYA by a factor of $224^{(3\ 35-1\ 97)} = 224^{1\ 38} = 1712$

frequencies of a periodic cipher are written over the cipher letters of a cryptogram, possible placements for the word CAVALRY might be seen. It is important to note that, although the frequencies of high- and medium-frequency plaintext letters might be distorted in a polyalphabetic cryptogram, it is not expected that low-frequency letters will be changed appreciably, therefore these low-frequency letters are a more accurate guide to crib-placing than are the other letters. It goes without saying that polygraphic repetitions may be used as a basis on which to assume probable words, depending on the length of the polygraphs; these repetitions, together with the frequency pattern of the cipher letters composing the repetitions, form one of the most valuable means of plaintext entries in a cryptogram.

(2) The aspect of partial idiomorphism is based on the fact that, if there is a pair of repeated letters at a distance of N in a plaintext word, this idiomorphism will show through in the cipher when there has been a polyalphabetic encipherment of a period of N. For example, if the word DIVISION is enciphered by a polyalphabetic substitution of four alphabets, the first and third I_p must of necessity be represented by the same cipher-text equivalent. Thus if in a four-alphabet system an A...A pattern is found in the cipher text, reference may be made to compilations of words containing like letters repeated at various intervals,[6] and under the listing of "A(3)A" will be found DIVISION, among other words, which may be used as possible assumptions.

36. Solution when the plain component is a mixed sequence, the cipher component, the normal.--a. This falls under Case IIb outlined in par. 8. It is not the usual method of employing a single mixed component, but may be encountered occasionally in cipher devices.

b. The preliminary steps, as regards factoring to determine the length of the period, are the same as usual. The message is then transcribed into its periods. Frequency distributions are then made, as usual, and these are attacked by the principles of frequency and recurrence. An attempt is made to apply the principles of direct symmetry of position as demonstrated thus far, but this attempt will be futile, for the reason that the plain component is in this case an unknown mixed sequence. (See par. 28d.) Any attempt to find symmetry in the secondary alphabets based upon the normal sequence can therefore disclose no symmetry because the symmetry which exists is based upon a wholly different sequence.

c. However, if the usual principles of direct symmetry of position are of no avail in this case, there are certain other principles of symmetry which may be employed to great advantage. To explain them an actual example will be used. Let it be assumed that it is known to the cryptanalyst that

[6] In this connection, see Appendix 2

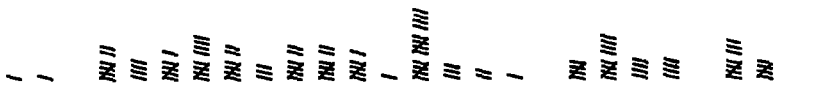
the enemy is using the general system under discussion, viz., a mixed sequence, variable from day to day, is used as plain component; the normal sequence is used as cipher component; and a repeating key, variable from message to message, is used in the ordinary manner.

The following message has been intercepted:

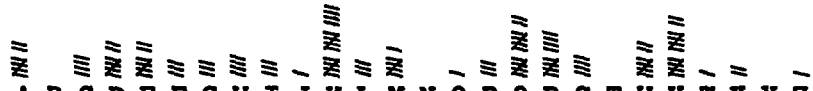
	5					10					15					20					25					30
A.	Q	E	O	V	K	L	R	M	L	Z	J	V	G	T	G	N	D	L	V	K	E	V	N	T	Y	E	R	M	U	E
B.	V	R	Z	M	O	Y	A	A	M	P	D	K	E	I	J	S	F	M	Y	O	Y	H	M	M	E	G	Q	A	M	B
C.	U	Q	A	X	R	H	U	F	B	U	K	Q	Y	M	U	N	E	L	V	T	K	Q	I	L	E	K	Z	B	U	E
D.	U	L	I	B	K	N	D	A	X	B	X	U	D	G	L	L	A	D	V	K	P	O	A	Y	O	D	K	K	Y	K
E.	L	A	D	H	Y	B	V	N	F	V	U	E	E	M	E	F	F	M	T	E	G	V	W	B	Y	T	V	D	Z	L
F.	S	P	B	H	B	X	V	A	Z	C	U	D	Y	U	E	L	K	M	M	A	E	U	D	D	K	N	C	F	S	H
G.	H	S	A	H	Y	T	M	G	U	J	H	Q	X	P	P	D	K	O	U	E	X	U	Q	V	B	F	V	W	B	X
H.	N	X	A	L	B	T	C	D	L	M	I	V	A	A	A	N	S	Z	I	L	O	V	W	V	P	Y	A	G	Z	L
J.	S	H	M	M	E	G	Q	D	H	O	Y	H	I	V	P	N	C	R	R	E	X	K	D	Q	Z	G	K	N	C	G
K.	N	Q	G	U	Y	J	I	W	Y	Y	T	M	A	H	W	X	R	L	B	L	O	A	D	L	G	N	Q	G	U	Y
L.	J	U	U	G	B	J	H	R	V	X	E	R	F	L	E	G	W	G	U	O	X	E	D	T	P	D	K	E	I	Z
M.	V	X	N	W	A	F	A	A	N	E	M	K	G	H	B	S	S	N	L	O	K	J	C	B	Z	T	G	G	L	O
N.	P	K	M	B	X	H	G	E	R	Y	T	M	W	L	Z	N	Q	C	Y	Y	T	M	W	I	P	D	K	A	T	E
P.	F	L	N	U	J	N	D	T	V	X	J	R	Z	T	L	O	P	A	H	C	D	F	Z	Y	Y	D	E	Y	C	L
Q.	G	P	G	T	Y	T	E	C	X	B	H	Q	E	B	R	K	V	W	M	U	N	I	N	G	J	I	Q	D	L	P
R.	J	K	A	T	E	G	U	W	B	R	H	U	Q	W	M	V	R	Q	B	W	Y	R	F	B	F	K	M	W	M	B
S.	T	M	U	L	Z	L	A	A	H	Y	J	G	D	V	K	L	K	R	R	E	X	K	N	A	O	N	D	S	B	X
T.	X	C	G	Z	A	H	D	G	T	L	V	K	M	B	W	I	S	A	U	E	F	D	N	W	P	N	L	Z	I	J
V.	S	R	Q	Z	L	A	V	N	H	L	G	V	W	V	K	F	I	G	H	P	G	E	C	Z	U	K	Q	A	P	

d. A study of the repetitions and of the factors of their intervals discloses that five alphabets are involved. Uniliteral frequency distributions are made and are shown in Fig. 23a:

Alphabet 1



Alphabet 2



Alphabet 3

Alphabet 4

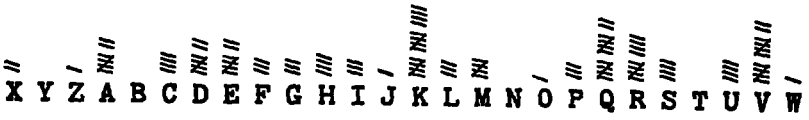
Alphabet 5

Figure 23a.

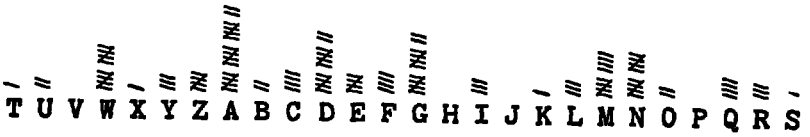
e. Since the cipher component in this case is the normal alphabet, it follows that the five frequency distributions are based upon a sequence which is known, and therefore, the five frequency distributions should manifest a direct symmetry of distribution of crests and troughs. By virtue of this symmetry and by shifting the five distributions relative to one another to proper superimpositions, the several distributions may be combined into a single uniliteral distribution. Note how this shifting has been done in the case of the five illustrative distributions:

Alphabet 1

Alphabet 2



Alphabet 3



Alphabet 4

Alphabet 5

Figure 23b.

f. The superimposition of the respective distributions enables one to convert the cipher letters of the five alphabets into one alphabet. Suppose it is decided to convert Alphabets 2, 3, 4, and 5 in terms of Alphabet 1. It is merely necessary to substitute for the respective letters in the four alphabets those which stand above them in Alphabet 1. For example, in Fig. 23b, X_c in Alphabet 2 is directly under A_c in Alphabet 1; hence, if the superimposition is correct then $\overset{2}{X}_c = \overset{1}{A}_c$. Therefore, in the cryptogram it is merely necessary to replace every X_c in the second position by A_c. Again T_c in Alphabet 3 = A_c in Alphabet 1; therefore in the cryptogram one replaces every T_c in the third position by A_c. The entire process, hereinafter designated as conversion into monoalphabetic terms, gives the following converted message:

	5					10					15					20					25					30
A.	Q	H	V	H	T	L	U	T	X	I	J	Y	N	F	P	N	G	S	H	T	E	Y	U	F	H	E	U	T	G	N
B.	V	U	G	Y	X	Y	D	H	Y	Y	D	N	L	U	S	S	I	T	K	X	Y	K	T	Y	N	G	T	H	Y	K
C.	U	T	H	J	A	H	X	M	N	D	K	T	F	Y	D	N	H	S	H	C	K	T	P	X	N	K	C	I	G	N
D.	U	O	P	N	T	N	G	H	J	K	X	X	K	S	U	L	D	K	H	T	P	R	H	K	X	D	N	R	K	T
E.	L	D	K	T	H	B	Y	U	R	E	U	H	L	Y	N	F	I	T	F	N	G	Y	D	N	H	T	Y	K	L	U
F.	S	S	I	T	K	X	Y	H	L	L	U	G	F	G	N	L	N	T	Y	J	E	X	K	P	T	N	F	M	E	Q
G.	H	V	H	T	H	T	P	N	G	S	H	T	E	B	Y	D	N	V	G	N	X	X	X	H	K	F	Y	D	N	G
H.	N	A	H	X	K	T	F	K	X	V	I	Y	H	M	J	N	V	G	U	U	O	Y	D	H	Y	Y	D	N	L	U
J.	S	K	T	Y	N	G	T	K	T	X	Y	K	P	H	Y	N	F	Y	D	N	X	N	K	C	I	G	N	U	O	P
K.	N	T	N	G	H	J	L	D	K	H	T	P	H	T	F	X	U	S	N	U	O	D	K	X	P	N	T	N	G	H
L.	J	X	B	S	K	J	K	Y	H	G	E	U	M	X	N	G	Z	N	G	X	X	H	K	F	Y	D	N	L	U	I
M.	V	A	U	I	J	F	D	H	Z	N	M	N	N	T	K	S	V	U	X	X	K	M	J	N	I	T	J	N	X	X
N.	P	N	T	N	G	H	J	L	D	H	T	P	D	X	I	N	T	J	K	H	T	P	D	U	Y	D	N	H	F	N
P.	F	O	U	G	S	N	G	A	H	G	J	U	G	F	U	O	S	H	T	L	D	I	G	K	H	D	H	F	O	U
Q.	G	S	N	F	H	T	H	J	J	K	H	T	L	N	A	K	Y	D	Y	D	N	L	U	S	S	I	T	K	X	Y
R.	J	N	H	F	N	G	X	D	N	A	H	X	X	I	V	V	U	X	N	F	Y	U	M	N	O	K	P	D	Y	K
S.	T	P	B	X	I	L	D	H	T	H	J	J	K	H	T	L	N	Y	D	N	X	N	U	M	X	N	G	Z	N	G
T.	X	F	N	L	J	H	G	N	F	U	V	N	T	N	F	I	V	H	G	N	F	G	U	I	Y	N	O	G	U	S
V.	S	U	X	L	U	A	Y	U	T	U	G	Y	D	H	T	F	L	N	T	Y	G	H	J	L	D	K	T	H	B	

The uniliteral frequency distribution for this converted text follows. Note that the frequency of each letter is the sum of the five frequencies in the corresponding columns of Fig. 23b.

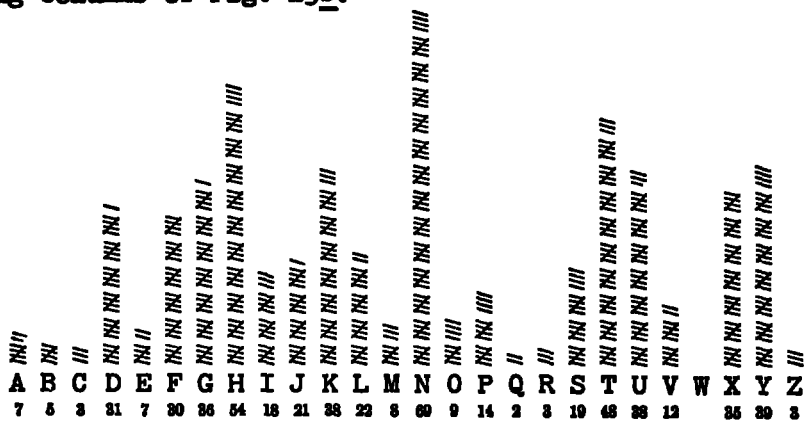


Figure 24.

g. The problem having been reduced to monoalphabetic terms, a triliteral frequency distribution can now be made and solution readily attained by simple principles.[7] It yields the following:

JAPAN CONSULTED GERMANY TODAY ON REPORTS THAT THE COMMUNIST INTERNATIONAL WAS BEHIND THE AMAZING SEIZURE OF GENERALISSIMO CHIANG KAI SHEK IN CHINA. TOKYO ACTED UNDER THE ANTICOMMUNIST ACCORD RECENTLY SIGNED BY JAPAN AND GERMANY. THE PRESS SAID THERE WAS INDISPUTABLE PROOF THAT THE COMINTERN INSTIGATED THE SEIZURE OF GENERAL CHIANG AND SOME OF HIS GENERALS. MILITARY OBSERVERS SAID THE COUP WOULD HAVE BEEN IMPOSSIBLE UNLESS GENERAL CHANG HSUEN LIANG HOTHEADED FORMER WAR LORD OF MANCHURIA HAD FORMED AN ALLIANCE WITH THE COMMUNIST LEADERS HE WAS SUPPOSED TO BE FIGHTING. SUCH AN ALLIANCE THESE OBSERVERS DECLARED OPENED UP A RED ROUTE FROM MOSCOW TO NORTH AND CENTRAL CHINA.

h. The reconstruction of the plain component is now a very simple matter. It is found to be as follows:

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

Note also, in Fig. 23b, the key word for the message, (HEAVY), the letters being in the columns headed by the letter H.

i. The solution of subsequent messages with different keys can now be reached directly, by a simple modification of the principles explained in par. 28. This modification consists in using for the completion sequence the mixed plain component (now known) instead of the normal alphabet, after the cipher letters have been converted into their plain-component equivalents. Let the student confirm this by experiment.

j. The probable-word method of solution discussed under par. 22 is also applicable here, in case of very short cryptograms. This method presupposes, of course, possession of the mixed component and the procedure is essentially the same as that in par. 22. In the example discussed in the present paragraph, the letter A on the plain component was successively set against the key letters HEAVY; but this is not the only possible procedure.

---

[7] An interesting technique is possible at this point to recover the key word for the plain component from the composite uniliteral frequency distribution of the cipher text at one fell swoop if the plain component is a keyword-mixed sequence Note the distribution in Fig 24 If this represents the "profile" of a keyword-mixed sequence then it appears that the key word begins at D_C with the letters Z_C A_C B_C, C_C being the equivalents of four of the five plaintext letters VWXYZ Q_C and R_C are obviously J_P and K_P respectively, and $W_C = Q_P$ The sequence $STUV_C$ represents either $LNOP_P$ or $MNOP_P$, XY_C must be two of the letters RST_P If $N_C = E_P$ then $OP_C = FG_P$, M_C is probably B_P thus delineating the key word of 9 letters beginning at D_C From this analysis it can be conjectured that the key word contains the letters A, C, D, H, I, U, one from the group LM one from the group RST and probably Y (for a more likely percentage of vowels) From this point on anagramming of the key word presents no problem

k. The student should go over carefully the principle of "conversion into monoalphabetic terms" explained in subpar. f above until he thoroughly understands it. Later on he will encounter cases in which this principle is of very great assistance in the cryptanalysis of more complex problems. (Other examples will be found in pars. 38, 61, and 62.)

l. The principle illustrated in subpar. e, above, that is, shifting two or more monoalphabetic frequency distributions relatively so as to bring them into proper alignment for amalgamation into a single monoalphabetic distribution, is called matching. It is a very important cryptanalytic principle. Note that its practical application consists in sliding one monoalphabetic distribution against the other so as to obtain the best coincidence between the entire sequence of crests and troughs of the other distribution. When the best point of coincidence has been found, the two sequences may be amalgamated and theoretically the single resultant distribution will also be monoalphabetic in character. The successful application of the principle of matching depends upon several factors. First, the cryptographic situation must be such that matching is a correct cryptographic step. For example, the distributions in Fig. 23a are properly subject to matching because the cipher component in the basic sequences concerned in this problem is the normal sequence, while the plain component is a mixed sequence. But it would be futile to try to match the distributions in subpar. 29c, for in that case the cipher component is a mixed sequence, the plain component is the normal sequence. Hence, no ammount of shifting or matching can bring the distributions of subpar. 29c into proper superimposition for correct amalgamation. (If the occurrences in the various distributions in subpar. 29c had been distributed according to the sequence of letters in the mixed component, then matching would be possible; but in order to be able to distribute these occurrences according to the mixed component, the latter has to be known--and that is just what is unknown until the problem has been solved.) A second factor involved in successful matching is the number of elements in the two distributions forming the subject of the test. If both of them have very few tallies, there is hardly sufficient information to permit of ocular matching with any degree of assurance that the work is not in vain. If one of them has many tallies, the other only a few, the chances for success are better than before, because the positions of the blanks in the two distributions can be used as a guide for their proper superimposition. Fortunately, there exist certain mathematical and statistical procedures which can be brought to bear upon the matter of cryptanalytic matching. One of these, involving the χ (chi) test, will be discussed in par. 37.

m. The normal conditions existing that permit the employment of direct symmetry of position in polyalphabetic ciphers are those cases already tested wherein the plain component is a known sequence. In such examples the sequence of the plain component is inscribed along the top of the sequence reconstruction matrix, and direct symmetry will manifest itself among the cipher components within the matrix. When the inverse conditions are present, i.e., those cases wherein the cipher component is a known sequence and the plain component unknown, the usual method of solution is,

of course, the matching of distributions and the conversion of the cryptogram to monoalphabetic terms. However, when the number of elements in the distribution is very small, this method is inapplicable, and the only feasible solution involves the assumption of probable words. If the usual type of matrix is made (with an A to Z sequence outside for the plain component), only indirect symmetry of position will be manifested. But if an inverse matrix is made having the known cipher component on the outside, then direct symmetry will be evident among the various plain-component alphabets. We will now consider an example to illustrate this technique.

(1) Let it be assumed that the enemy has been using for his low-echelon cryptosystems a small cipher disc in which the cipher component is a standard alphabet, the plain component a mixed sequence which is changed daily. The following is the beginning of an intercepted message, the remaining portion having been lost through operational difficulties:

K O L N T E Q Z D F I X K T K X K Y M B J J G B R H R T A F

R W W V C F M K B Y Q B D T .....

This message having originated from a headquarters that has frequently been guilty of stereotypic phraseology, it is suspected that the plain text begins with the opening phrase "REFERENCE YOUR MESSAGE NUMBER ..." Superimposing the assumed plain text against the cipher text,

```
          5        10        15        20        25
K O L N T E Q Z D F I X K T K X K Y M B J J G B R H
R E F E R E N C E Y O U R M E S S A G E N U M B E R
```

it is observed that R_p is enciphered as K_c at the first and thirteenth positions, thereby tentatively establishing the period-length as 12:

```
K O L N T E Q Z D F I X
R E F E R E N C E Y O U

K T K X K Y M B J J G B
R M E S S A G E N U M B

R H R T A F R W W V C F
E R

M K B Y Q B D T . . . .
```

The student will observe that all other periods from 2 to 14 are ruled out because of coincidences in the plain text which are not substantiated by like coincidences in the cipher for the assumed period-length. For example, the blocks for the periods 9, 10, and 11 yield the following:

```
(9)              (10)             (11)
R E F E R E N C E   R E F E R E N C E Y   R E F E R E N C E Y O
Y O U R M E S S A   O U R M E S S A G E   U R M E S S A G E N U
G E N U M B E R     N U M B E R           M B E R
```

(2) An inverse matrix is made, and the plaintext values from the plain-cipher relationships above are inscribed within the matrix, as follows:

Plain:																									
	1	.	.	.	.	.	.	.	.	.	R	.	.	.	.	.	E	.	.	.	.	.	.	.	.
	2	.	.	.	.	.	.	.	R	.	.	.	.	.	E	.	.	.	.	M	.	.	.	.	.
	3	.	.	.	.	.	.	.	.	.	E	F	.	.	.	.	.	.	.	.	.	.	.	.	.
	4	.	.	.	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	S	.	.		
	5	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	R	.	.	.	.	.	
	6	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	
	7	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	N	.	.	.	.	.	.	.	.	
	8	.	E	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	
	9	.	.	.	E	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
	10	.	.	.	.	.	Y	.	.	.	U	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
	11	.	.	.	.	.	.	M	.	O	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
	12	.	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	U	.	.	

Cipher: 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

After application of the principles of direct symmetry of position, the matrix will now look like this:

Plain:																									
	1	.	S	.	.	.	.	.	.	.	R	A	.	.	.	C	.	E	F	G	.	.	M	N	O	.
	2	.	.	.	.	.	.	.	R	A	.	.	.	C	.	E	F	G	.	.	M	N	O	.	.	S	.
	3	.	.	.	R	A	.	.	.	C	.	E	F	G	.	.	M	N	O	.	.	S	.	.	.	.	.
	4	.	.	.	.	.	.	.	R	A	.	.	.	C	.	E	F	G	.	.	M	N	O	.	.	S	.	.
	5	E	F	G	.	.	M	N	O	.	.	S	.	.	.	.	.	.	.	.	R	A	.	.	.	C	.
	6	.	.	C	.	E	F	G	.	.	M	N	O	.	.	S	.	.	.	.	.	.	.	R	A	.	
	7	.	.	.	R	A	.	.	.	C	.	E	F	G	.	.	M	N	O	.	.	S	.	.	.	.	.
	8	.	E	F	G	.	.	M	N	O	.	.	S	.	.	.	.	.	.	.	R	A	.	.	.	C	
	9	.	C	.	E	F	G	.	.	M	N	O	.	.	S	.	.	.	.	.	.	.	R	A	.	.	
	10	.	.	.	.	.	Y	.	.	.	U	.	.	.	B	.	.	.	.	.	.	.	.	.	.	.	
	11	.	E	F	G	.	.	M	N	O	.	.	S	.	.	.	.	.	.	.	R	A	.	.	.	C	
	12	.	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	.	U	.	.	

Cipher: 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

When these new derived values are substituted in the cryptogram,

K O L N T E Q Z D F I X
R E F E R E N C E Y O U

K T K X K Y M B J J G B
R M E S S A G E N U M B

R H R T A F R W W V C F
E R O N E F O R F

M K B Y Q B D T
R

the U_p of the segment ONE FOUR in the third line permits the amalgamation of the two partial sequences into one, viz.:

R A U . . C B E F G . . M N O . . S . . . . . Y .

(3) With but little further experimentation, the entire plain text is synthesized and the plain component is found to be based on HYDRAULIC. The complete message fragment is now as follows:

K O L N T E Q Z D F I X
R E F E R E N C E Y O U

K T K X K Y M B J J G B
R M E S S A G E N U M B

R H R T A F R W W V C F
E R O N E F O U R O F J

M K B Y Q B D T . . . .
U L Y T H I R D

There is no evidence of a key word for the repeating key in the inverse matrix; but if the matrix is rewritten in the usual enciphering form, the key word HEADQUARTERS will be apparent under H_p. Thus:

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

37. The χ (chi) test for evaluating the relative matching of distributions.--a. The student by now is well familiar with the ϕ test which is used to determine the monoalphabeticity of a single distribution. If two messages were enciphered monoalphabetically by the same cipher alphabet, it follows not only that their corresponding distributions would be monoalphabetic, but also that these distributions would be strikingly similar in respect to their corresponding peaks and troughs. Likewise, if in a polyalphabetic cipher there are repeated letters in the key, then the distributions appertaining to the repeated letters will show identical spatial relationships of the positions of the peaks and troughs. Furthermore, in situations wherein the cipher component is a standard alphabet (or any other

known sequence), these spatial relationships are relative and require only the correct juxtaposition of the distributions to make the relationships absolute. When the number of tallies in the distributions is large, ocular matching is a simple matter; however if the number of tallies is small, ocular matching becomes difficult and recourse must be had to statistical means for evaluating the relative matching of distributions.

b. One of the simplest means for determining the relative matching or non-matching attributes of distributions is the χ (chi) test, sometimes called the "cross-products sum." With this test, which is related to the ϕ test[8], the "observed value of χ" is compared with the "expected value of χ for matching distributions" (symbolized by χ_m) and the "expected value of χ for non-matching distributions" (symbolized by χ_r). The formulas used for the expected values of χ for matching and non-matching distributions, respectively, are:

$$\chi_m = .0667(N_1N_2) \quad \text{and} \quad \chi_r = .0385(N_1N_2)$$

where (N_1N_2) represents the products of the total number of tallies in each distribution. The observed value of χ is calculated by multiplying the frequency of each element in the first distribution by its homologous counterpart in the second distribution, and totalling the result; i.e., the frequency of A_c in the first distribution is multiplied by the frequency of A_c in the second distribution, etc., and then the sum of these cross products is obtained.

c. The use of the χ test is best illustrated by an example. Suppose the following two distributions are to be matched:

No. 1:	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
No. 2:	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

Now let the frequencies be juxtaposed, for convenience in finding the cross products, thus:

f_1.....	1	4	0	3	0	1	0	0	1	0	0	1	0	0	1	0	0	3	2	2	1	0	1	3	0	2
	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
f_2.....	0	2	0	0	0	3	0	0	1	0	1	0	0	1	1	0	0	3	1	1	0	0	0	0	1	2
f_1f_2...	0	8	0	0	0	3	0	0	1	0	0	0	0	0	1	0	0	9	2	2	0	0	0	0	0	4

$N_1 = 26$; $N_2 = 17$; $\Sigma f_1f_2 = 30$

$$\chi_m = .0667(26 \times 17) = 29.5; \quad \chi_r = .0385(26 \times 17) = 17$$

The fact that the observed value of χ (30) agrees very closely with the expected value for matching distributions (29.5) means that the two distributions very probably belong together or are properly matched. Note

[8] The derivation of the ϕ and χ tests will be treated in Military Cryptanalytics, Part III

the qualifying phrase "very probably." It implies that there is no certainty about this business of matching distributions by mathematical methods; the mathematics serve only as measuring devices, so to speak, which can be employed to measure the degree of similarity that exists. There are other mathematical or statistical tests for matching, in addition to the χ test.[9] Moreover, it is possible to go further with the χ test and find a measure of reliance that may be placed upon the value obtained; but these points will be left for discussion in the next text.

d. One more point will, however, here be added in connection with the χ test. Suppose the very same two distributions in the preceding subparagraph are again juxtaposed, with f_2 shifted one interval to the left of the position shown above, and let us take the cross-products sum. Thus:

f_1.....	1 4 0 3 0 1 0 0 1 0 0 1 0 0 1 0 0 3 2 2 1 0 1 3 0 2 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	N = 26
f_2.....	2 0 0 0 3 0 0 1 0 1 0 0 1 1 0 0 3 1 1 0 0 0 0 1 2 0 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 A	N = 17
f_1f_2...	2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 2 0 0 0 0 3 0 0	Σf_1f_2 = 10

Since the observed value of χ (10) more closely approximates the expected value of χ for non-matching distributions (17), it may be concluded that if the two distributions pertain to the same primary components they are not properly superimposed. In other words, the χ test may also be applied in cases where two or more frequency distributions must be shifted relatively in order to find their correct superimposition. The theory underlying this application of the χ test is, of course, the same as before: two monoalphabetic distributions when properly combined will yield a single distribution which should still be monoalphabetic in character. In applying the χ test in such cases it may be necessary to shift two 26-element distributions to various superimpositions, make the χ test for each superimposition, and take as correct that one which yields the best value for the test. The nature of the problem will, of course, determine whether the frequency distributions which are to be matched should be compared (1) by direct superimposition, that is, setting the A to Z tallies of one distribution directly opposite the corresponding tallies of the other distribution, as in subpar. c; or (2) by shifted superimposition, that is, keeping the A to Z tallies of the first distribution fixed and sliding the whole sequence of tallies of the second distribution to various superimpositions against the first.

e. A very common method of expressing the relative matching quality of a pair of distributions involves the ratio of the observed χ to the expected value for χ_r. This ratio of $\frac{\chi_o}{\chi_r}$ is called the "cross I.C." (abbr. $\frac{\chi}{\chi}$ I.C.). The cross I.C. is usually the preferred expression, rather than the χ value, since the ratio gives a quick measure (when compared with

[9] The most important of these is the chi-square test based on the χ^2 distribution

the expected ξ I.C.'s of 1.73 and 1.00 for matching and nonmatching distributions, respectively) of the relative goodness of a particular matching. For instance, in the example given in subpar. c, the ξ I.C. of the two distributions is $\frac{30}{17} = 1.76$; in the example in subpar. d, the ξ I.C. is $\frac{10}{17} = 0.59$. The ordinary monographic I.C., i.e., $\frac{\phi o}{\phi r}$, is often referred to as the δ I.C. (read "delta I.C."), in order to distinguish between this and the cross I.C.

38. Modified Porta systems.--a. Variations of the Porta system are possible, wherein either the A-M sequence is left undisturbed and the N-Z portion is mixed, or the A-M portion is mixed and the N-Z sequence is the normal; these situations are exemplified in Figs. 25a and b, below:

	A	B	C	D	E	F	G	H	I	J	K	L	M
AB	S	P	U	R	T	N	O	Q	V	W	X	Y	Z
CD	P	U	R	T	N	O	Q	V	W	X	Y	Z	S
EF	U	R	T	N	O	Q	V	W	X	Y	Z	S	P
GH	R	T	N	O	Q	V	W	X	Y	Z	S	P	U
IJ	T	N	O	Q	V	W	X	Y	Z	S	P	U	R
KL	N	O	Q	V	W	X	Y	Z	S	P	U	R	T
MN	O	Q	V	W	X	Y	Z	S	P	U	R	T	N
OP	Q	V	W	X	Y	Z	S	P	U	R	T	N	O
QR	V	W	X	Y	Z	S	P	U	R	T	N	O	Q
ST	W	X	Y	Z	S	P	U	R	T	N	O	Q	V
UV	X	Y	Z	S	P	U	R	T	N	O	Q	V	W
WX	Y	Z	S	P	U	R	T	N	O	Q	V	W	X
YZ	Z	S	P	U	R	T	N	O	Q	V	W	X	Y

Figure 25a.

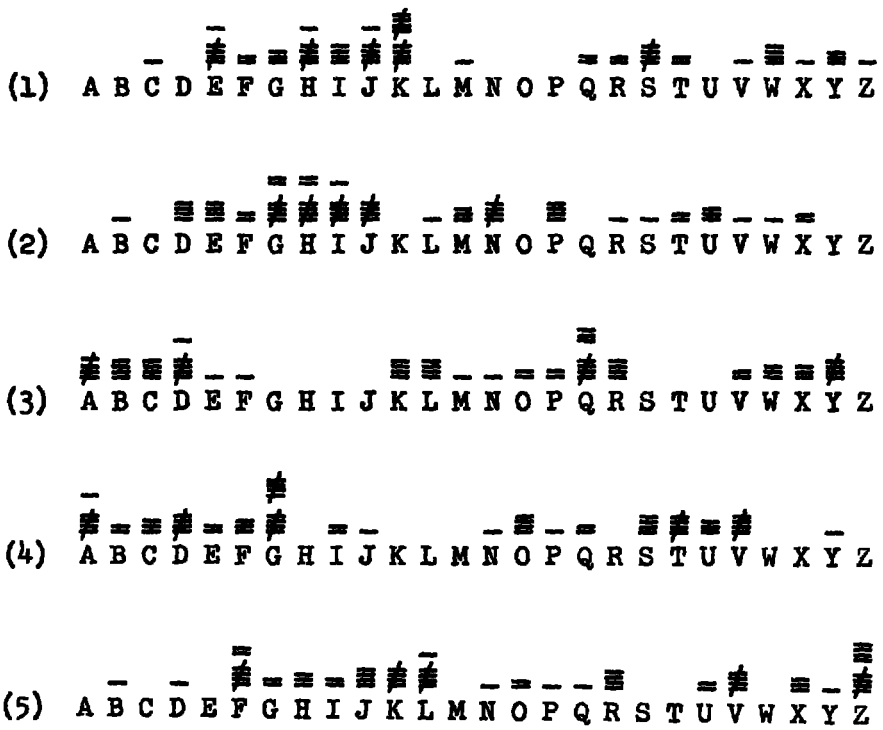
	F	L	A	M	E	B	C	D	G	H	I	J	K
AB	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
CD	O	P	Q	R	S	T	U	V	W	X	Y	Z	N
EF	P	Q	R	S	T	U	V	W	X	Y	Z	N	O
GH	Q	R	S	T	U	V	W	X	Y	Z	N	O	P
IJ	R	S	T	U	V	W	X	Y	Z	N	O	P	Q
KL	S	T	U	V	W	X	Y	Z	N	O	P	Q	R
MN	T	U	V	W	X	Y	Z	N	O	P	Q	R	S
OP	U	V	W	X	Y	Z	N	O	P	Q	R	S	T
QR	V	W	X	Y	Z	N	O	P	Q	R	S	T	U
ST	W	X	Y	Z	N	O	P	Q	R	S	T	U	V
UV	X	Y	Z	N	O	P	Q	R	S	T	U	V	W
WX	Y	Z	N	O	P	Q	R	S	T	U	V	W	X
YZ	Z	N	O	P	Q	R	S	T	U	V	W	X	Y

Figure 25b.

In such situations, one half of the cipher text may be converted into monoalphabetic terms, while in the other half there will be manifested direct symmetry of position.

b. Let the following cryptogram and its accompanying distributions be studied:

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



It is noted that the A-M halves of the cipher distributions may be matched by sliding them at appropriate intervals; this is proof that a Porta matrix of the type shown in Fig. 25a has been used in the encipherment. The correct matching is obvious:

c. The cipher letters A-M of the cryptogram may now be converted into monoalphabetic terms, using Alphabet 1 as a base, as shown in Fig. 26 below:

```
                5              10              15              20              25              30
A.  W P A V V   Q E X G J   K R A S G   R D B S N   S I D I Z   H P C T Q
      H           F  K I   K  H  F       E I         J K M     H  J

B.  C D L G F   T H Y G K  E I D D J   R M Q A J   K U O T V   Z W R F F
    C E F K E     I  K J  E J K H I     A  E I   K               J E

C.  H N O O Z   E T D J K   S M N G K   E H A C X   I N D F R   J E A G P
    H           E  K A J     A  K J   E I H G     I  K J       J F H K

D.  H P K A F   I J L G H   H G Q U L   J I R V F   E X C U Z   F V R F Z
    H  E E E    I K F K G   H H   K     J J   E     E  J         F   J

E.  E M Y Y O   Y J C E R   J F Q B U   K H D O I   W E X V X   V N K A R
    E A         K J I       J G   F     K I K  H     F           E E

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

G.  J G Q A L   Q L Q A G   G G L G F   S J V U L   J D B D H   Y H Y G K
    J H  E K      M  E F   G H F K E     K     K   J E I H G     I  K J

H.  E H D B X   K E Y T F   K U L Q L   I J W D V   F T K P Z   T I W A Z
    E I K F     K F   E     K  F  K     I K  H     F  E         J  E

J.  J H B G D   K F A Q Z   W S E C V   H H Q D F   X B R V I   Y J M D R
    J I I K C   K G H         L G     H I  H E     C    H     K G H

K.  S G P T V   M N C O U   S G Q S J   K I Q N L   G G V G H   H U Y I Z
      H         M  J         H   I     K J     K   G H   K G   H    M
```

Figure 26.

The conversion process makes patent several new polygraphic repetitions which were previously latent in the cipher text. The threefold occurrence of the sequence F.KIK.H.F. at A6, E14, and H13 can be established as probably comprising a 10-letter repetition. The X_c at A8 and E23 shows that θ_c^2 and θ_c^{10} of the repetition are identical plaintext letters, thus establishing the partial idiomorphic pattern as ABCDC...AB which may be identified in a pattern list as belonging to the plaintext word PHOTOGRAPH. At D3, the $\overline{EEE}_c$ is most probably $\overline{SSS}_p$, with the preceding P_c most likely an E_p. With these entries, reconstruction of the matrix and recovery of the rest of the plain text is an easy matter.

39. Additional remarks.--a. It might be well to bring in at this point several observations in connection with the solution of the systems discussed thus far in this text. These observations are treated as brief notes below.

b. When factoring does not indicate a period uniquely, write out the cipher text on the fewest number of widths that will accomodate the possible periods. For instance if factoring indicates a maximum of 12 alphabets, then writing the cipher text on the widths of 12, 10, and 8 will facilitate examination and taking of distributions on the basis of 2, 3, 4, 5, 6, 8, 10, and 12 alphabets. If only two periods are possible and one is a multiple of the other, write out the text on the longer period; in other cases where there are only two possible periods, select the least common multiple for the first trial.

c. When confronted by a polyalphabetic cipher of a fairly lengthy period, and the problem involves the matching of comparatively small distributions, the χ test should be used to match the distributions, beginning with a pair of distributions having the best "profiles", and after all the distributions have been matched, the cipher text is converted to monoalphabetic terms. Even if some of the distributions are mismatched, it will usually be possible to solve the resulting monoalphabet; systematic garbles every nth position will point to the distributions incorrectly matched. However, if the repeating key is a plaintext key, a search for plaintext fragments in one of the columns of the matched distributions might make possible a quick recovery of the repeating key and thereby bypass difficulties in matching some of the "less good" distributions.

d. If the repeating key of a polyalphabetic cipher is not found under the first letter of the plain component, and it is known or assumed that the normal equation $\theta_{k/2} = \theta_{1/1}$, $\theta_{p/1} = \theta_{c/2}$ is used, then completing the plain-component sequence on the "key" under any θ_p will disclose the key word if one was used.

e. If in a polyalphabetic cryptogram there are two or more sets of long polygraphic repetitions of equal length, consider the possibility that these two sets might be different encipherments of the same plain text, and look for corroborating evidence. For example, if Set "A" was partially recovered as .EA...A.TE.S$_p$ and Set "B" was recovered as .EA....R.ER.$_p$, these values may be amalgamated into .EA...ARTERS$_p$ and further expanded into HEADQUARTERS.

f. The student should keep his mind open to possible variations of a basic idea, even if a particular variation might seem at first blush to contradict a general principle. For instance, although in Porta encipherment it is not expected that a letter may be enciphered by itself, nevertheless in the matrix illustrated below such a contingency is possible.

	A	B	C	D	E	F	G	H	I	K	L	M	*
AB	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
CD	O	P	Q	R	S	T	U	V	W	X	Y	Z	N
EF	P	Q	R	S	T	U	V	W	X	Y	Z	N	O
							•						
							•						
WX	Y	Z	N	O	P	Q	R	S	T	U	V	W	X
YZ	Z	N	O	P	Q	R	S	T	U	V	W	X	Y

The A-M component has been shortened by combining I and J, so by convention a letter appearing in the asterisked column may be represented by itself.

g. Polyalphabetic numerical systems with "standard" alphabets have been discussed in par. 25. When however the cipher component consists of a mixed numerical sequence, then direct symmetry of position will of course be manifested, and this fact can be exploited in the solution of a cryptogram. If however the cipher component is a "normal" numerical sequence (say, a sequence of the dinomes 01-26 or the dinomes 10-45 in numerical order) and the plain component is an unknown mixed sequence, then the methods discussed in par. 36 are applicable.

(BLANK)

CHAPTER VI

REPEATING-KEY SYSTEMS WITH MIXED CIPHER ALPHABETS, II; INDIRECT SYMMETRY OF POSITION

40. Further cases to be considered.--a. Thus far Cases II a and b of the mixed-alphabet cases mentioned in par. 8 have been treated. There remains Case II c which has been further subdivided as follows:

Case II c. Both components are mixed sequences.

1. Components are identical mixed sequences.

(a) Sequences proceed in the same direction. (The secondary alphabets are mixed alphabets.)

(b) Sequences proceed in opposite directions. (The secondary alphabets are reciprocal mixed alphabets.)

2. Components are different mixed sequences. (The secondary alphabets are mixed alphabets.)

b. The first of the foregoing subcases, i.e., Case II c 1 (a), will now be examined. Case II c 1 (b) will be taken up in subpar. 44n, and Case II c 2 will be treated in subpar. 44o.

41. Identical primary mixed components proceeding in the same direction.--a. It is often the case that the mixed components are derived from an easily remembered word or phrase, so that they can be reproduced at any time from memory. Thus, for example, given the key word QUESTIONABLY, the following mixed sequence is derived:

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

b. By using this sequence as both plain and cipher component, that is, by sliding this sequence against itself, a series of 26 secondary mixed alphabets may be produced. In enciphering a message, sliding strips may be

employed with a key word to designate the particular and successive positions in which the strips are to be set, the same as was the case in previous examples of the use of sliding components. The method of designating the positions, however, requires a word or two of comment at this point. In the examples thus far shown, the key letter, as located on the cipher component, was always set opposite A, as located on the plain component; possibly an erroneous impression has been created, viz., that this is invariably the rule. This is decidedly not true, as has already been explained in par. 13c. If it has seemed to be the case that θ_1 always equals A_p, it is only because the text has dealt thus far principally with cases in which the plain component is the normal sequence and its initial letter, which usually constitutes the index for juxtaposing cipher components, is A. It must be emphasized, however, that various conventions may be adopted in this respect; but the most common of them is to employ the initial letter of the plain component as the index letter. That is, the index letter, θ_1, will be the initial letter of the mixed sequence, in this case, Q. Furthermore, to prevent the possibility of ambiguity it will be stated again that the pair of enciphering equations employed in the ensuing discussion will be the first of the 12 set forth under par. 13f, viz., $\theta_k/2 = \theta_1/1$; $\theta_p/1 = \theta_c/2$. In this case the subscript "1" means the plain component, the subscript "2", the cipher component, so that the enciphering equation is the following: $\theta_k/c = \theta_1/p$; $\theta_p/p = \theta_c/c$.

c. By setting the two sliding components against each other in the two positions shown below, the cipher alphabets labeled (1) and (2) given by two key letters, A and B, are seen to be different.

Key letter = A θ_1

Plain component.... QUESTIONABLYCDFGHJKMPRVWXZ
Cipher component... QUESTIONABLYCDFGHJKMPRVWXZQUESTIONABLYCDFGHJKMPRVWXZ
θ_k

Secondary alphabet (1):

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

Key letter = B θ_1

Plain component.... QUESTIONABLYCDFGHJKMPRVWXZ
Cipher component... QUESTIONABLYCDFGHJKMPRVWXZQUESTIONABLYCDFGHJKMPRVWXZ
θ_k

Secondary alphabet (2):

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

d. Very frequently a square table is employed by the correspondents, instead of sliding strips, but the results are the same. The cipher square based upon the word QUESTIONABLY is shown in Fig. 27. It will be noted that it does nothing more than set forth the successive positions of the two primary sliding components; the top line of the square is the plain component, the successive horizontal lines below it, the cipher component in its various juxtapositions. The usual method of employing such a square (i.e., corresponding to the enciphering equations $\theta_{k/c} = \theta_{i/p}$; $\theta_{p/p} = \theta_{c/c}$) is to take as the cipher equivalent of a plaintext letter that letter which lies at the intersection of the vertical column headed by the plaintext letter and the horizontal row begun by the key letter. For example, the cipher equivalent of E_p with key letter T is the letter O_c; or $E_p\ (T_k) = O_c$. The method given in subpar. b, for determining the cipher equivalents by means of the two sliding strips yields the same results as does the cipher square.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z
U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q
E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U
S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E
T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S
I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T
O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I
N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O
A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N
B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A
L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B
Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L
C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y
D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C
F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D
G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F
H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G
J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H
K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J
M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K
P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M
R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P
V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R
W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V
X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W
Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X

Figure 27.

42. Enciphering and deciphering by identical primary mixed components. There is nothing of special interest to be noted in connection with the use either of identical mixed components or of an equivalent square table such as that shown in Fig. 27, in enciphering or deciphering a message. The basic principles are the same as in the case of the sliding of one mixed component against the normal, the displacements of the two components being controlled by changeable key words of varying lengths. The components may be changed at will and so on. All this has been discussed adequately enough in Chapter II.

43. Principles of solution.--a. Basically the principles of solution in the case of a cryptogram enciphered by two identical mixed sliding components are the same as in the preceding case. Primary recourse is had to the principles of frequency and repetition of single letters, digraphs, trigraphs, and longer polygraphs. Once an entering wedge has been forced into the problem, the subsequent steps may consist merely in continuing along the same lines as before, building up the solution bit by bit.

b. Doubtless the question has already arisen in the student's mind as to whether any principles of symmetry of position can be used to assist in the solution and in the reconstruction of the cipher alphabets in cases of the kind under consideration. This phase of the subject will be taken up in the succeeding paragraphs and will be treated in a detailed manner, because the theory and principles involved are of very wide application in cryptanalytics.

44. Theory of indirect symmetry of position in secondary alphabets.--a. Note the two secondary alphabets (1) and (2) given in subpar. 41c. Externally they show no resemblance or symmetry despite the fact that they were produced from the same primary components. Nevertheless, when the matter is studied with care, a symmetry of position is discoverable. Because it is a hidden or latent phenomenon, it may be termed latent symmetry of position. However, the phenomenon has a long-standing designation in cryptologic literature as an indirect symmetry of position and this terminology has grown into usage, so that a change now is perhaps inadvisable. Indirect symmetry of position is a very interesting and exceedingly useful phenomenon in cryptanalytics.

b. Consider the following secondary alphabet (the one labeled (2) in subpar. 41c):

(2)
Plain:	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
Cipher:	J	K	R	V	Y	W	X	Z	F	Q	U	M	E	H	G	S	B	T	C	D	L	I	O	N	P	A

c. Assuming it to be known that this is a secondary alphabet produced by two identical mixed primary components, it is desired to reconstruct the latter. Construct a chain of alternating plaintext and ciphertext equivalents, beginning at any point and continuing until the chain has been completed. Thus, for example, beginning with $A_p = J_c$, $J_p = Q_c$, $Q_p = B_c$, . . .,

and dropping out the letters common to successive pairs, there results the sequence A J Q B . . .. By completing the chain the following sequence of letters is established:

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

d. This sequence consists of 26 letters. When slid against itself it will produce exactly the same secondary alphabets as do the primary components based upon the word QUESTIONABLY. To demonstrate that this is the case, compare the secondary alphabets given by the two settings of the externally different components shown below:

Plain component.... QUESTIONABLYCDFGHJKMPRVWXZ
Cipher component... QUESTIONABLYCDFGHJKMPRVWXZQUESTIONABLYCDFGHJKMPRVWXZ

Secondary alphabet (1):

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

Plain component.... AJQBKULMEYPSCRTDVIFWOGXNHZ
Cipher component... AJQBKULMEYPSCRTDVIFWOGXNHZAJQBKULMEYPSCRTDVIFWOGXNHZ

Secondary alphabet (2):

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

e. Since the sequence A J Q B K . . . gives exactly the same equivalents in the secondary alphabets as does the sequence Q U E S T . . . X Z, the former sequence is cryptographically equivalent to the latter sequence. For this reason the A J Q B K . . . sequence is termed an equivalent primary component.[1] If the real or original primary component is a keyword-mixed sequence, it is hidden or latent within the equivalent primary sequence, but it can be made patent by decimation of the equivalent primary component. The procedure is as follows: Find three letters in the equivalent primary component such as are likely to have formed an unbroken sequence in the original primary component, and see if the interval between the first and second is the same as that between the second and third. Such a case is presented by the letters W, X, and Z in the equivalent primary component above. Note the sequence . . . W O G X N H Z . . . , the distance or interval between the letters W, X, and Z is three letters. Continuing the chain by adding letters three intervals removed, the latent original primary component is made patent. Thus:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
W X Z Q U E S T I O N A B L Y C D F G H J K M P R V

[1] Such an equivalent component is merely a sequence which has been or can be derived from the original sequence or basic primary component by applying a decimation process to the latter, conversely, the original or basic component can be derived from an equivalent component by applying the same sort of process to the equivalent component By decimation is meant the selection of elements from a sequence according to some fixed interval For example, the sequence A E I M is derived, by decimation from the normal alphabet by selecting every fourth letter

f. It is possible to perform the steps given in c and e in a combined single operation when the original primary component is a keyword-mixed sequence. Starting with any pair of letters (in the cipher component of the secondary alphabet) likely to be sequent in the keyword-mixed sequence, such as JK_c in the secondary alphabet labeled (2), the following chain of digraphs may be set up. Thus, J and K in the plain component stand over Q and U, respectively, in the cipher component; Q and U in the plain component stand over B and L, respectively, in the cipher component, and so on. Connecting the pairs in a series, the following results are obtained:

JK→QU→BL→KM→UE→LY→MP→ES→YC→PR→ST→CD→RV→

TI→DF→VW→IO→FG→WX→ON→GH→XZ→NA→HJ→ZQ→AB→JK . . .

These may now be united by means of their common letters:

JK→KM→MP→PR→RV→etc. = J K M P R V W X Z Q U E S T I O N A B L Y C D F G H

The original primary component is thus completely reconstructed.

g. Not all of the 26 secondary alphabets of the series yielded by two sliding primary components may be used to develop a complete equivalent primary component. If examination be made, it will be found that only 12 of these secondary alphabets will yield complete equivalent primary components when the method of reconstruction shown in subpar. c above is followed. For example, the following secondary alphabet, which is also derived from the primary components based upon the word QUESTIONABLY, will not yield a complete chain of 26 plaintext-ciphertext equivalents:

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

Equivalent primary components:

1 2 3 4 5 6 7 8 9 10 11 12 13 | 1 2 3
A C H P X E O L F K V Q T | A C H . . . (The A C H sequence begins again.)

h. It is seen that only 13 letters of the chain have been established before the sequence begins to repeat itself. It is evident that exactly one-half of the chain has been established. The other half may be established by beginning with a letter not in the first half. Thus:

1 2 3 4 5 6 7 8 9 10 11 12 13 | 1 2 3
B D J R Z S N Y G M W U I | B D J . . . (The B D J sequence begins again.)

i. There are several methods for combining two 13-letter chains. The simplest method, applicable when the primary component is a keyword-mixed sequence, will now be described. If we assume two letters to be sequent in

the original keyword-mixed sequence, such as for example J and K, and if these letters are not in the same 13-letter chain, we would then write the two 13-letter chains over one another, with the J of the one chain superimposed over the K in the other chain. Thus, using for an example the chains A C H P X E O L F K V Q T and B D J R Z S N Y G M W U I established in subpars. g and h above, we would have the following:

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

The vertical digraphs thus formed are not incompatible with the manifestations to be expected if the primary component were a keyword-mixed sequence. Now noting the vertical digraph $\begin{matrix} W \\ X \end{matrix}$, we may assume that Y will follow it in the mixed component; the Y is not in the same 13-letter chain that contains the X, so this looks promising. The chain containing the Y is now written beneath the diagram, properly juxtaposed so that Y is under the X, thus:

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

The resulting vertical trigraphs are not satisfactory as portions of a keyword-mixed sequence, so it appears that Y must be in the key word and not in the remaining semi-alphabetical portion of the sequence. If we now assume that WX is followed by Z (which is not in the same chain as X) in the sequence, we have the following:

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

The "good" trigraphs produced (DFG, JKM, etc.) attest to the correctness of the trial. We now have in effect a series of three-letter chains, which may be interconnected by the common letters in the first and third rows, thus: WXZ, ZQU, UES,...; this quickly yields the QUESTIONABLY...XZ sequence. Note that these three-letter chains will always be at a constant interval apart; in this case, the interval was +7.

j. The reason why a complete chain of 26 letters cannot be constructed from the secondary alphabet given under subpar. g is that it represents a case in which two primary components of 26 letters were slid an even number of intervals apart. (This will be explained in further detail in subpar. p below.) There are 12 such cases in all, none of which will admit of the construction of a complete chain of 26 letters. In addition, there is one case wherein, despite the fact that the primary components are an odd number of intervals apart, the secondary alphabet cannot be made to yield a complete chain of 26 letters for an equivalent primary component. This is the case in which the displacement is 13 intervals. Note the secondary

alphabet based upon the primary components below (which are the same as those shown in subpar. d):

Primary Components

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

Secondary alphabet

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

k. If an attempt is made to construct a chain of letters from this secondary alphabet alone, no progress can be made because the alphabet is completely reciprocal. However, the cryptanalyst need not at all be baffled by this case. The attack will follow along the lines shown below in subpars. l and m.

l. If the original primary component is a keyword-mixed sequence, the cryptanalyst may reconstruct it by attempting to "dovetail" the 13 reciprocal pairs (AR, BV, CZ, DQ, EG, FU, HS, IK, JT, LW, MO, NP, and XY) into one sequence. The members of these pairs are all 13 intervals apart. Thus:

```
 0  1  2  3  4  5  6  7  8  9 10 11 12 13
A .  .  .  .  .  .  .  .  .  .  .  .  . R
B .  .  .  .  .  .  .  .  .  .  .  .  . V
C .  .  .  .     .  .  .  .  .  .  .  . Z
D .  .  .  .  .  .  .  .  .  .  .  .  . Q
E .  .  .  .  .  .  .  .  .  .  .  .  . G
F .  .  .  .     .  .  .  .  .     .  . U
H .  .  .  .  .  .  .  .  .  .  .  .  . S
I .  .  .  .  .  .  .  .  .  .  .  .  . K
J .  .  .  .  .  .  .  .  .  .  .  .  . T
L .  .  .  .  .  .  .  .  .  .  .  .  . W
M .  .  .  .  .  .     .  .  .     .  . O
N .  .  .  .  .  .  .     .  .  .  .  . P
X .  .     .  .     .  .  .  .     .  . Y
```

Write out the series of numbers from 1 to 26 and insert as many pairs into position as possible, being guided by considerations of probable partial sequences in the keyword-mixed sequence. Thus:

```
 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
A  B  C  D  .  .  .  .  .  .  .  .  R  V  Z  Q
```

It begins to look as though the key word commences with the letter Q, in which case it should be followed by U. This means that the next pair to be inserted is FU. Thus:

```
 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17
A  B  C  D  F  .  .  .     .  .  .  .  R  V  Z  Q  U
```

The sequence A B C D F means that E is in the key. Perhaps the sequence is A B C D F G H. Upon trial, using the pairs EG and HS, the following placements are obtained:

```
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
A B C D F G H . . . . . R V Z Q U E S
```

This suggests the word QUEST or QUESTION. The pair JT is added:

```
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
A B C D F G H J . . . . R V Z Q U E S T
```

The sequence G H J suggests G H J K, which places an I after T. Enough of the process has been shown to make the steps clear.

m. Another method of circumventing the difficulties introduced by the 14th secondary alphabet (displacement interval, 13) is to use it in conjunction with another secondary alphabet which is produced by an even-interval displacement. For example, suppose the following two secondary alphabets are available.[2]

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

The first of these secondaries is the 13-interval secondary; the second is one of the even-interval secondaries, from which only half-chain sequences can be constructed. But if the construction be based upon the two sequences, 1 and 2 in the foregoing diagram, the following is obtained:

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

This is a complete equivalent primary component. The original keyword-mixed component can be recovered from it by decimation at an interval of +9.

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

n. (1) When the primary components are identical mixed sequences proceeding in opposite directions, all the secondary alphabets will be reciprocal alphabets. Reconstruction of the primary component can be accomplished by the procedure indicated under subpar. m above. Note the

[2]The method of writing down the secondary alphabets shown in the diagram below will hereafter be followed in all cases when alphabet reconstruction matrices are necessary The top line will be understood to be the plain component, it is common to all the secondary alphabets, and is set off from the cipher components by the heavy black line This top line of letters will be designated by the digit ø, and will be referred to as "the zero line ' in the diagram The successive lines of letters, which occupy the space below the zero line and which contain the various cipher components of the several secondary alphabets, will be numbered serially These numbers may then be used as reference numbers for designating the horizontal lines in the diagram The numbers standing above the letters may be used as reference numbers for the vertical columns in the diagram Hence, any letter in the reconstruction matrix may be designated by coordinates, giving the row coordinate first Thus D (2-11) means the letter D standing in row 2, column 11

following three reciprocal secondary alphabets:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Ø	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
1	P	M	H	G	Q	F	D	C	W	Y	L	K	B	R	V	A	E	N	Z	X	U	O	I	T	J	S
2	W	V	M	K	S	J	H	G	Q	F	D	R	C	X	Z	Y	I	L	E	U	T	B	A	N	P	O
3	T	S	S	Z	L	X	W	V	N	R	P	E	M	I	O	K	C	J	B	A	Y	H	G	F	U	D

(2) Using lines 1 and 2, the following chain can be constructed (equivalent primary component):

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

Or, using lines 2 and 3:

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

The original keyword-mixed primary component (based on the word QUESTIONABLY) can be recovered from either of the two foregoing equivalent primary components. But if lines 1 and 3 are used, only half-chains can be constructed:

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

This is because 1 and 3 are both odd-interval secondary alphabets, whereas 2 is an even-interval secondary. It may be added that odd-interval secondaries are characterized by having two cases in which a plaintext letter is enciphered by itself; that is, θ_p is identical with θ_c. This phrase "identical with" will be represented by the symbol $\equiv$; the phrase "not identical with" will be represented by the symbol $\not\equiv$. (Note that in secondary alphabet number 1 above, $F_p \equiv F_c$ and $U_p \equiv U_c$; in secondary alphabet number 3 above, $M_p \equiv M_c$ and $O_p \equiv O_c$). This characteristic will enable the cryptanalyst to select at once the proper two secondaries to work with in case several are available; one should show two cases where $\theta_p \equiv \theta_c$; the other should show none.

o. (1) When the primary components are different mixed sequences, their reconstruction from secondary cipher alphabets follows along the same lines as set forth above, under b to j, inclusive, with the exception that the selection of letters for building up the chain of equivalents for the primary cipher component is restricted to those below the zero line in the reconstruction matrix. Having reconstructed the primary cipher component, the plain component can readily be reconstructed. This will become clear if the student will study the following example:

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

(2) Using only lines 1 and 2, the following chain is constructed:

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

This is an equivalent primary cipher component. By finding the value of the successive letters of this chain in terms of the plain component of secondary alphabet number 1 (the zero line), the following is obtained:

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

The sequence A S P T . . . is an equivalent primary plain component. The original keyword-mixed components may be recovered from each of the equivalent primary components. That for the primary plain component is based upon the key PUBLISHERS MAGAZINE; that for the primary cipher component is based upon the key QUESTIONABLY.

(3) Another method of accomplishing the process indicated above can be illustrated graphically by the following two chains, based upon the two secondary alphabets set forth in subpar. o (1):

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Ø	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
1	T	V	A	B	U	L	I	Q	X	Y	C	W	S	N	D	P	F	E	Z	G	R	H	J	K	M	O
2	Z	J	S	T	V	I	Q	R	M	O	N	K	X	E	A	G	B	W	P	L	H	Y	C	D	F	U

Col 1		Col 2.				
A (Ø–1)	→	T (1–1),	→	T (2–4)	→	D (Ø–4), →
D (Ø–4)	→	B (1–4),	→	B (2–17)	→	Q (Ø–17), →
Q (Ø–17)	→	F (1–17),	→	F (2–25)	→	Y (Ø–25), →
Y (Ø–25)	→	M (1–25),	→	M (2–9)	→	I (Ø–9), →
I (Ø–9)	→	X (1–9),	→	X (2–13)	→	M (Ø–13), →
M (Ø–13)	→	S (1–13),	→	S (2–3)	→	C (Ø–3), →
etc.		etc				

(4) By joining the letters in Column 1, the following chain is obtained: A D Q Y I M, etc. If this be examined, it will be found to be an equivalent primary of the sequence based upon PUBLISHERS MAGAZINE. By joining the letters in Column 2, the following chain is obtained: T B F M X S. This is an equivalent primary of the sequence based upon QUESTIONABLY.

p. A final word concerning the reconstruction of primary components in general may be added. It has been seen that in the case of a 26-element component sliding against itself (both components proceeding in the same direction), it is only the secondary alphabets resulting from odd-interval displacements of the primary components which permit of reconstructing a single 26-letter chain of equivalents. This is true except for the 13th interval displacement, which, in spite of its being an odd number, still acts like an even-number displacement in that no complete chain of equivalents can be established from the secondary alphabet. This exception gives the clue to the basic reason for this phenomenon: it is that the number 26 has two factors, 2 and 13, which enter into the picture. With the exception of

displacement-interval 1, any displacement interval which is a sub-multiple of, or has a factor in common with, the number of letters in the primary sequence will yield a secondary alphabet from which no complete chain of 26 equivalents can be derived for the construction of a complete equivalent primary component. This general rule is applicable only to components which progress in the same direction; if they progress in opposite directions, all the secondary alphabets are reciprocal alphabets and they behave exactly like the reciprocal secondaries resulting from the 13-interval displacement of two 26-letter identical components progressing in the same direction.

q. The foregoing remarks give rise to the following observations based upon the general rule pointed out above. Whether or not a complete equivalent primary component is derivable by decimation from an original primary component (and if not, the lengths and numbers of chains of letters, or incomplete components, that can be constructed in attempts to derive such equivalent components) will depend upon the number of letters in the original primary component and the specific decimation interval selected. For example, in a 26-letter original primary component, decimation interval 5 will yield a complete equivalent primary component of 26 letters, whereas decimation intervals 4 or 8 will yield 2 chains of 13 letters each. In a 24-letter component, decimation interval 5 will also yield a complete equivalent primary component (of 24 letters), but decimation interval 4 will yield 6 chains of 4 letters each, and decimation interval 8 will yield 3 chains of 8 letters each. It also follows that in the case of an original primary component in which the total number of characters is a prime number, all decimation intervals will yield complete equivalent primary components. The following table has been drawn up in the light of these observations, for original primary sequences from 16 to 32 elements. (All prime-number sequences have been omitted.) In this table, the column at the extreme left gives the various decimation intervals, omitting in each case the first interval. which merely gives the original primary sequence, and the last interval, which merely gives the original sequence reversed. The top line of the table gives the various lengths of original primary sequences from 32 down to 16. (The student should bear in mind that sequences containing characters in addition to the letters of the alphabet may be encountered; he can add to this table when he is interested in sequences of more than 32 characters.) The numbers within the table then show, for each combination of decimation interval and length of, original sequence, the lengths of the chains of characters that can be constructed. (The student may note the symmetry in each column.) The bottom line shows the total number of complete equivalent primary components which can be derived for each different length of original component.

Decimation interval	Number of characters in original primary component											
	32	30	28	27	26	25	24	22	21	20	18	16
2	16	15	14	27	13	25	12	11	21	10	9	8
3	32	10	28	9	26	25	8	22	7	20	6	16
4	8	15	7	27	13	25	6	11	21	5	9	4
5	32	6	28	27	26	5	24	22	21	4	18	16
6	16	5	14	9	13	25	4	11	7	10	3	8
7	32	30	4	27	26	25	24	22	3	20	18	16
8	4	15	7	27	13	25	6	11	21	5	9	2
9	32	10	28	9	26	25	8	22	7	20	2	16
10	16	3	14	27	13	5	12	11	21	2	9	8
11	32	30	28	27	26	25	24	2	21	20	18	16
12	8	5	7	9	13	25	2	11	7	5	3	4
13	32	30	28	27	2	25	24	22	21	20	18	16
14	16	15	2	27	13	25	12	11	3	10	9	8
15	32	2	28	9	26	5	8	22	7	4	6	
16	2	15	7	27	13	25	6	11	21	5	9	
17	32	30	28	27	26	25	24	22	21	20		
18	16	5	14	9	13	25	4	11	7	10		
19	32	30	28	27	26	25	24	22	21			
20	8	3	7	27	13	5	6	11				
21	32	10	4	9	26	25	8					
22	16	15	14	27	13	25	12					
23	32	30	28	27	26	25						
24	4	5	7	9	13							
25	32	6	28	27								
26	16	15	14									
27	32	10										
28	8	15										
29	32											
30	16											
Total number of complete sequences	14	6	10	16	10	18	16	8	10	6	4	6

45. Reconstruction of primary components by employing principles of indirect symmetry of position.--a. Let us now consider the application of indirect symmetry in a typical example. In a certain periodic polyalphabetic cryptogram under study which factored to five alphabets, the following assumptions based on repetitions in the cipher text have been made:

12345123	34512345	123451234	512345
ZFOOWATF	XSMAQUHX	YINOBRKXF	IPZZPO
DIVISION	REGIMENT	ARTILLERY	ATTACK

These values are inserted in a sequence reconstruction matrix, as illustrated below:

∅	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
1	Y			Z					A			R								P						
2					K				F				Q		T				I	Z						
3	Z				U									F					X	N		O				
4			P		S				O					H											F	
5	I						M				O	B									W	X				

Figure 28.

It is clear from the intervals (X...O), (S...O), and (M...O) in rows 3, 4, and 5 that the plain component is not the normal sequence. If an inverse matrix were constructed, it would disclose that neither is the cipher component the normal sequence, if this possibility had not already been ruled out by the absence of matching qualities of the distributions for the five alphabets. It is also evident that this is not a case of mixed sequences proceeding in opposite directions, since reciprocity between the plain and cipher is contradicted (e.g., in Alphabet 1, $A_p = Y_c$, but $A_c = I_p$). At the moment, we do not know whether the plain component is identical with the cipher component, or whether it is a different mixed sequence, in the absence of evidence to the contrary[3], we will assume the former hypothesis.

b. In order to derive additional values for possible insertion in the cryptogram, around which values further assumptions may be made and thus speed up the process of solution, certain relationships among the letters in the matrix may be studied. For convenience, we may refer to these relationships as "proportions", arising from the process of "proportioning." Proportioning should be done in a systematic manner, if it is to be efficient; this is especially true in the initial stages of solution, when it is important not to overlook a possible derived value. The following procedure is suggested to insure thoroughness of method.

(1) Each vertical pairing in the matrix is transferred to the horizontal, and a four-element proportion is utilized to complete another proportion which has three elements in common with the first proportion. Referring to Fig. 28, we shall start with the vertical pairing AY (∅-1, 1-1) and transfer whatever data is available to the horizontal pairing AY (∅-1, ∅-25). The vertical pairing AY includes proportionally, between rows ∅ and 1, the pairs DZ, IA, LR, and TP; or, expressed differently, A:Y::D:Z::I:A::L:R::T:P. The horizontal pairing AY (∅-1, ∅-25) includes no other proportional pairs--however there are present the latent proportions A:Y::Y:Θ, A:Y::Z:Θ, A:Y::Θ:F, and A:Y::I:Θ. Now as we have already established the four-element

[3] An hypothesis of identical components proceeding in the same direction would be ruled out by a situation in which, for example, in one of the alphabets $A_p = A_c$, and in the same alphabet $B_p \neq B_c$, or this hypothesis would be ruled out if in one alphabet there were evidences of only partial reciprocity (such as $A_p = L_c$, $A_c = L_p$, $N_p = R_c$, but $N_c \neq R_p$).

proportion A:Y::I:A from the vertical pairing, then we know that the θ in the last horizontal proportion A:Y::I:θ must be the letter A; therefore we place A in the matrix at position 5-25.

(2) This process of transferring values is then attempted for the vertical pairing AZ (ϕ-1, 3-1), and then for AK, but with negative results. When the vertical pairing AI is considered, the proportion A:I::Y:A turns up once again, but this yields no additional information since this proportion has already been established when we placed the A at 5-25. The subsequent vertical pairings CP, DZ, EK, EU, ES, and GM yield nothing further until IA is reached, wherein IA (ϕ-9, 1-9) = AY (ϕ-1, 1-1). But it is noted that the homologous proportion IA (ϕ-9, ϕ-1) = AY (1-9, 1-1) has its elements in exactly the same locations as IA (ϕ-9, ϕ-1) = AY (ϕ-1, 1-1), so no transference of data from one alphabet to another is possible. In other words, the data within a proportion of four elements in one particular "rectangular" reading may be transferred to a <u>different</u> rectangular reading having three of the elements in common. The need, of course, for seeking a <u>four</u>-element proportion first is that proportions in cryptanalytics must be defined by four elements (for a given set of enciphering components), because unlike the field of mathematics wherein the missing member of 2:4::3:x <u>must</u> be a 6, the cryptanalytic proportion A:B::C:θ could be satisfied by <u>any</u> letter, depending upon the components involved. What we were actually saying by the proportion A:Y::I:A is that in a <u>certain</u> pair of components (the case under study) we have established an empirical cryptanalytic proportion, and that this relationship will be true in <u>all</u> cases involving these same four elements.

(3) The vertical pairing IA nevertheless yields a good proportion, namely IA (ϕ-9, 1-9) = DZ (ϕ-4, 1-4), which permits the insertion of the letter D in position 3-9 after the homologous rectangular reading IA (ϕ-9, ϕ-1) = OZ (3-9, 3-1). Then the vertical pairing IF will yield the value T at the position 4-6 from the proportion I:F::O:T. This process is continued until the last vertical pairings at the extreme right of the matrix have been treated.

c. The matrix will now appear as in Fig. 29, below, after we have systematically proportioned once straight across the matrix from left to right:

ϕ	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
1	Y	L		Z	I				A		F	R	G		K				O	P			S	T		
2					K				F		P	E	Q		T			I	X	Z		H		D		
3	Z				U			T	D					F				X		N		O				
4	K		P		S	T			O					H											F	
5	I			U		K	M		E		O	B			S	T		L	W	X					A	D

Figure 29.

It should be emphasized that this has been a rather theoretical treatment of the problem in order to demonstrate the principles of proportioning. In actual practice, of course, the derivation of additional values would be a basis upon which to make further plaintext assumptions, probably rendering unnecessary the rigorous systematization of the process of proportioning across the width of the matrix, as just described.

d. From the matrix in Fig. 29, the following partial chains, fragments of equivalent primary components, are derived:

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

It is observed that the chains Ø-3 and 1-2 bear a 1:3 relationship, i.e., one is an expansion at an interval of 3 of the other (cf. AZ and ID in line Ø-3, and AFPZ and IKTD in line 1-2). Let then the chains in 1-2 be considered as a relative +1 decimation of the primary sequence, and inscribe AFPZ in the first four positions of a line of 26 cells on cross-section paper, thus:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	F	P	Z																						

Now since the chains in line Ø-3 are a 1:3 contraction of the chains in line 1-2, then in order to equate line Ø-3 with line 1-2 we must expand or decimate the former at an interval of 3. That is, the interval between the letters AZ must be 3 instead of 1. This expansion, of course also applies to the remaining chains in line Ø-3. Thus HTNF must be expanded into H..T..N..F; and since we already have an F in our basic AFPZ sequence, these letters are then interpolated as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	F	P	Z															H			T			N	

e. With these new values, the sequence IKTD from the assumed +1 decimation in line 1-2 may be added, because of the presence of T in both sequences,

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	F	P	Z															H	I	K	T	D		N	

and the remaining fragments in the various chains may be amalgamated to permit the completion of the sequence as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	F	P	Z	L	J	S	Y	C	M	V	R	E	O	X	U	G	Q	H	I	K	T	D	B	N	W

This complete equivalent primary component, reconstructed from a series of smaller chains derived from a number of other equivalent primary components, may now be decimated to produce the original primary component which is a keyword-mixed sequence.

f. It must be noted that proportioning does not yield any new basic data, but merely gives a re-statement of data already latent or inherent in the matrix; what the process does is to make all the relationships patent and this re-statement facilitates the derivation of chains. Furthermore, in the sets of partial chains given in subpar. d, the first five chains (∅-1, to ∅-5) give all the latent relationships present in the matrix. However, the addition of one more set of chains (line 1-2) brings out the existing relationships in a much clearer light than would have otherwise been possible, and this materially speeds up solution.

g. Another example may be presented to demonstrate further the principles of indirect symmetry. In a polyalphabetic cryptogram which factored to six alphabets, the following message beginning is assumed, based on collateral information:

123456	123456	123456	123456	123456	1
EKIIBK	KGOZVZ	DIWBBH	LFCRKZ	WKODBB	Q...
ONEFIV	EFIVEM	MHOWIT	ZERAMM	UNITIO	N

The values from this crib are put into a sequence reconstruction matrix, as follows:

∅	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
1					K								D	Q	E						W					L
2					F	G		I						K												
3					I				O						W			C								
4	R					I														D		Z	B			
5					V				B				K													
6													Z		B					H		K				

Noting an apparent +1 decimation of a keyword-mixed primary component between lines ∅-2, EF:FG::HI, we may begin to reconstruct the original primary component by inspection directly without deriving additional values by proportions:[4]

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
(∅-2)	EF:FG::HI	E	F	G	H	I																					
(∅-3)	EI:IO:OW	E	F	G	H	I	.	.	.	O	.	.	.	W													
(∅-4)	FI::WB	E	F	G	H	I	.	.	.	O	.	.	.	W	.	.	B										
(3-5)	IV::OB	E	F	G	H	I	.	.	.	O	.	.	V	W	.	.	B										
(∅-1)	OE::UW	E	F	G	H	I	.	.	.	O	.	.	V	W	.	.	B	.	.	.	.	U					

Continuing in this vein, the original primary component may be quickly recovered; let the student finish the solution as an exercise.

[4] It must be pointed out that there is no proof at this stage that the HI ties in with EFG, the HI might be in the key word, but there is a greater probability that it is a part of the remaining alphabetical sequence after the occurrence of the key word.

46. Theory of a graphical method of indirect symmetry.[5]--a. It has been shown that the interval between letters of a sequence obtained from a secondary alphabet is a constant function of the interval separating the letters in the original primary component. Consider the following sequence:

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

Assume that this component is slid against itself and that the following groups of partial sequences are obtained from three secondary alphabets:

Group 1--S T I; U E; N A
Group 2--I N; E T; O A
Group 3--T N; Q S O

Figure 30.

Referring to the primary component, it will be seen that the letters of the partial sequences obtained from group 1 coincide in their interval (i.e., a +1 decimation) with that in the primary component; the letters of the partial sequences obtained from group 2 represent a decimation interval of two in the primary component; and those obtained from group 3, a decimation interval of three.

b. In the foregoing case, decimation was accomplished by taking intervals to the right along a horizontal component. Referring to the square based on QUESTIONABLY given in Fig. 27 on p. 107, let a portion of that square or matrix be considered, as shown in Fig. 31 below:

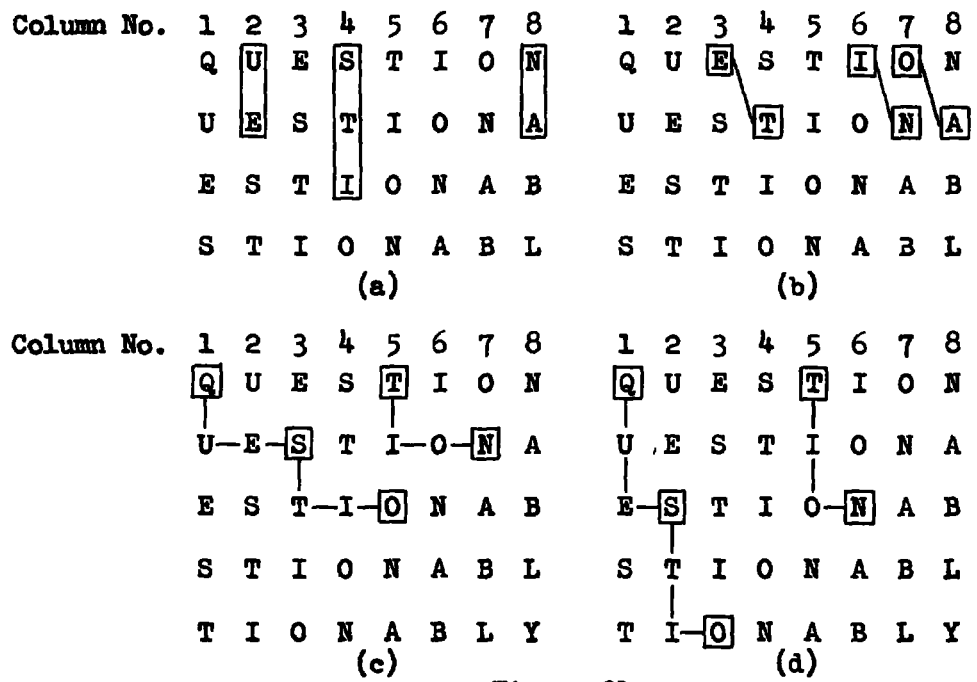


Figure 31.

[5] The basic theory underlying this modified method of applying the principles was first set forth in a brief paper in November 1941 by 1st Lt. Paul E Neff, Sig. C. His original notes, slightly modified, comprise pars. 46 and 52, and subpars. 53a to d, inclusive.

c. Again referring to Fig. 30, the partial sequences STI, UE, and NA can be obtained from Fig. 31a by reading down columns 4, 2, and 8, respectively. This can be represented graphically by the symbol ↓1, which means that all partial sequences obtained from Fig. 31a by proceeding downward in any column would be in the same group (i.e., secondary alphabet) and have the same decimation interval.

d. The partial sequences IN, ET, and OA can be represented graphically by $1\underset{1}{\rightarrow}$, or simply ↘1, which indicates that all partial sequences obtained by taking letters one space down and one space to the right, or one space down a diagonal to the right would represent the same decimation interval.

e. The partial sequences TN and QSO can be represented by the symbol $1\underset{2}{\rightarrow}$; but they can also be represented by $2\underset{1}{\rightarrow}$ and, if the entire matrix of Fig. 27 is considered, by other possible routes.

f. The decimation interval of a secondary sequence derived from a primary is the sum of the horizontal and vertical components of the route selected. Since the partial sequence TN can be represented by $1\underset{2}{\rightarrow}$, the decimation interval of this sequence is equal to the vertical decimation interval of the basic square plus twice the horizontal decimation interval in that square. Any other route selected for the same sequence would give an equivalent of this.

g. It is seen, therefore, that the decimation interval of a component can be represented graphically in various ways other than along the horizontal, by use of diagrams such as in Fig. 31, in which the successive juxtaposed components have the same relative displacement. In this case the successive horizontal lines had a one-letter displacement to the left.

h. Not being limited to one dimension, reconstruction of the primary component or an equivalent should be possible in one combined matrix by reversing the foregoing process and graphically integrating partial sequences from different secondary alphabets into a single diagram. Suppose the partial sequences in Fig. 30 are given and it is desired to reconstruct the primary component.

Group ----------	1	2	3
Partial sequences----	S T I U E N A	I N E T O A	T N Q S O
	→ 1	↓1	↘1

Figure 32.

i. (1) Using cross-section paper one can arbitrarily select the STI sequence in group 1 and write this sequence horizontally, making the graphical notation $\underset{1}{\rightarrow}$ below group 1.

(2) Proceeding to group 2, the partial sequence IN contains one letter in common with the sequence STI already entered, but since NA forms a sequence in group 1 and OA forms a sequence in group 2, it is clear that two different decimations are involved and therefore it would be incorrect to integrate the STI and the IN into STIN However, the letter N can arbitrarily be placed in any position other than along the horizontal line on which STI has been placed. It will be placed directly below the letter I and the group will be denoted graphically by $\downarrow 1$, giving:

S T I
. . N

Figure 33a.

(3) The skeleton of the matrix or diagram is now fixed in two dimensions, and no further letters can be arbitrarily placed within it. However, additional sequences from groups 1 and 2 can be added, provided a common letter is available in the diagram; sequences from other groups can be added, provided one pair is already entered in the diagram which would fix the proper graphical decimation.

(4) Moving to group 3, there is the partial sequence TN and it is noted that this pair of letters is present in the diagram. The symbol $\searrow 1$ can therefore be placed under group 3.

(5) In group 3 the partial sequence QSO appears and the letter S is in the diagram. It therefore follows that the letters Q and O can be placed thus:

(1) Q . . .
(2) . S T I
(3) . . O N

Figure 33b.

(6) Similarly the letter E of the partial sequence ET in group 2 goes directly above the T:

(1) Q . E .
(2) . S T I
(3) . . O N

Figure 33c.

(7) The letter U of the sequence UE in group 1 goes before the E:

(1) Q U E .
(2) . S T I
(3) . . O N

Figure 33d.

(8) Likewise the letter A of NA in group 1 follows N:

(1) Q U E . .
(2) . S T I .
(3) . . O N A

Figure 33e.

(9) The sequence OA in group 2 remains to be entered. Since both these letters are already in the diagram, the letter A can be placed under the existing O or the letter O can be placed above the existing A. Either alternative would be correct. Selecting the latter alternative yields the following:

(1) Q U E . .
(2) . S T I O
(3) . . O N A

Figure 33f.

j. All the original information has now been entered in the diagram seen in Fig. 33f and the letter O appears twice therein. This letter O may be termed the "tie-in" letter since it indicates the horizontal interval between the juxtaposed reconstructed sequences of the basic matrix. The absence of a tie-in letter in the diagram would indicate that insufficient data are present for the reconstruction of a complete sequence.

k. (1) By sliding the last row of Fig. 33f two intervals to the right the two O's can be superimposed, giving:

(1) Q U E . . . .
(2) . S T I O . .
(3) . . . . O N A

Figure 33g.

(2) Since each horizontal sequence must be shifted two intervals to the right of its initial position in relation to the line above, row (1) must be moved two intervals to the left of its original position. Thus:

(1) Q U E . . . . . .
(2) . . . S T I O . .
(3) . . . . . O N A

Figure 33h.

(3) Since the three rows involve the same decimation, and since the O of ONA coincides with the O of STIO, the ONA sequence may be raised up one row and united with the STIO sequence. If this is legitimate then the new row (2) may likewise be raised up one row. This yields the united

sequence QUESTIONA... . This last step may be more clearly understood by studying the following partially reconstructed matrix:

(1) Q U E S T I O N . .
(2) E S T I O N A B . .
(3) T I O N A B L Y . .
(4) O N A B L Y C D . .

Figure 33i.

1. The application of this graphical method of indirect symmetry to a specific example will be illustrated in par. 52, in the next chapter.

47. Further remarks.--a. A study of the principles and techniques discussed in this chapter should impress the student with the importance and value of indirect symmetry of position as a tool in cryptanalytics. Admittedly, indirect symmetry is a difficult subject to treat in writing, as it lends itself much better to blackboard demonstration in a classroom to insure thorough understanding of the principles. In any case, it is only by practice on a multitude of different examples and cases that these principles can be firmly implanted in the mind of the student--and even then the practice must be a continuous process, as it is only too easy to lose adroitness and facility in the application of these principles.

b. It is recommended that the student prepare as training aids five strips bearing the following sequences, double length:

(1) A normal A-Z sequence.
(2) A keyword-mixed sequence based on QUESTIONABLY.
(3) A keyword-mixed sequence based on QUESTIONABLY.
(4) A QUESTIONABLY keyword-mixed sequence, running in reverse.
(5) A keyword-mixed sequence based on HYDRAULIC.

With these strips the phenomena arising in all cases of direct and indirect symmetry may be duplicated, and the strips will be found useful in further experimentation and study. For example, strips (1) and (2) may be used to produce the phenomena of direct symmetry; (2) and (3) may oe employed to produce the manifestations inherent in indirect symmetry extending to the Ø (plaintext) alphabet; (3) and (4) will bring out the peculiarities inherent in cases of indirect symmetry within the matrix only, but with the added feature of reciprocity between the plain and cipher components; and (4) and (5) will duplicate the idiosyncracies of indirect symmetry within the matrix only.

c. The student has seen the two principal methods of the application of indirect symmetry, and perhaps the question will be asked: "Which method is preferable?" The answer is--both are useful. In most cases it

is usually easier to employ the graphical method, but occasionally there is encountered a problem which is easier to solve by the linear method.[6] But both methods should be practiced constantly, in order to maintain facility in the application of these principles.

d. Thus far there has been treated the recovery of primary components based on keyword-mixed sequences only. What would happen if the primary component were, say, a transposition-mixed sequence? In the previous text[7] it has been shown how to recover the key word in various types of sequences, including transposition-mixed sequences. If, however, a transposition-mixed sequence were decimated, as it would be in the case of an equivalent primary component, a slight modification of procedure is necessary.

(1) Let us consider the following sequence:

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

This can easily be reduced to its original transposition rectangle:

4	9	3	7	1	8	6	5	2
H	Y	D	R	A	U	L	I	C
B	E	F	G	J	K	M	N	O
P	Q	S	T	V	W	X	Z	

However, if the basic sequence were decimated at an interval of three,

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

the usual procedures of uncovering the transposition do not apply. The sequence must be decimated at the intervals of +3, +5, +7, +9, and +11; these resultant sequences are then examined in turn in an attempt to remove the transposition, reading these sequences both forwards and backwards. When the +9 decimation is considered, it will yield the original transposition.

(2) In unusual circumstances wherein all the secondary alphabets consist of even decimations, thus giving rise exclusively to 13-letter chains, there is an approach which may be used to cope with this distressing situation. For instance, let us suppose that we have the two chains (AEPTOQDCHYNMK) and (BZWXLIJRVUSGF). We will assume that the letters VWXYZ are at the end of the transposition matrix, and we will complete

[6] As an example, it should be noted that the chains in subpar. 45d permit of easier treatment by the linear method than by the graphical method, let the student confirm this by experiment.

[7] Military Cryptanalytics, Part I, par. 51.

the sequences on these letters from the foregoing chains, thus:

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

The sequence GIJKL on the fourth line certainly looks like part of the transposition matrix two rows above the VWXYZ row. The diagram may now be expanded to the left and right, as is shown below:

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

The key word PREACH (or PREACHER) is manifested, and, with a little experimentation, the original transposition matrix is recovered as follows:

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

This example admittedly is a simple case because of the brevity and particular composition of the key word; longer key words quickly complicate

the problem of recovery, but the general lines just indicated will apply. Each case must be treated as a special case; the cryptanalyst has to be on the alert to capitalize on any peculiarities or phenomena manifested.

e. In treating a periodic cipher, the first step is, of course, to determine the number of alphabets involved by factoring the intervals between the various polygraphic repetitions. The cryptogram is then written out in proper period-lengths, and distributions for each alphabet are made; if necessary (that is, where the number of tallies per distribution is small), I.C.'s of the distributions are calculated, for futher proof of the correctness of the factoring. These distributions are now carefully examined for (1) a possible fitting of all the distributions to the normal frequency distribution or its reverse, which would show that standard alphabets are involved; (2) a possible matching of all the distributions in respect to each other, which would show that the cipher component is the normal sequence; (3) a possible direct matching, "head on", of two or more distributions, which would show that the repeating key has repeated letters in the homologous positions; and (4) a possible fitting of only one or so of the distributions to the normal, "head on." If case (4) is present, then it is proof that the cryptogram is either a normal Porta with that particular distribution being in the key (A,B), or else the cryptogram involves indirect symmetry extending to the ϕ (plaintext) row of the reconstruction matrix, since this manifestation is brought about by the interaction of a mixed sequence against itself, running in the same direction, with the juxtaposition $A_p = A_c$, so that every letter in that particular alphabet would be enciphered by itself.

f. After the distributions have been examined, assumptions of high frequency letters or of probable words are inserted in the sequence reconstruction matrix, which in turn is examined for evidences or contradictions of direct symmetry or of indirect symmetry either within or without the matrix. (If reciprocity in more than one alphabet is observed, then it may be assumed that the cryptogram involves a mixed sequence running against itself in reverse.[8]) Whenever an assumption of a plaintext value for a cipher letter is made, the student should be sure to finish four things before making any further assumptions: (1) the plaintext value should be entered below all occurrences of the cipher letters; (2) the value should be entered in the reconstruction matrix; (3) examination should be made if any inconsistencies are produced either in the plain text or in the matrix; and (4) an attempt should be made to derive new values by direct symmetry or by proportioning within the matrix. Adherence to the foregoing systematization of method will save much time, contribute to the proper cryptanalytic education of the student, and will prove of considerable importance in the solution of difficult problems encountered in actual operations.

[8] Reciprocity in only one alphabet could be caused by a sequence shifted 13 positions against itself, in the case of 26-letter components.

g. As a final remark on indirect symmetry, it must be noted that Porta matrices, as well as Vigenère-type matrices, might be encountered in which indirect symmetry will be manifested; this situation will obtain when both "families" of the Porta matrix are mixed sequences. It also follows that indirect symmetry will be present in schemes wherein the cipher component is a mixed numerical sequence and the plain component is also a mixed sequence.

CHAPTER VII

APPLICATION OF PRINCIPLES OF INDIRECT SYMMETRY OF POSITION

48. Applying the principles to a specific example.--a. The preceeding chapter, with the many details covered, now forms a sufficient base for proceeding with an exposition of how the principles of indirect symmetry of position can be applied very early in the solution of a polyalphabetic substitution cipher in which sliding primary components were employed to produce the secondary cipher alphabets for the enciphering of the cryptogram.

b. The case described below will serve not only to explain the method of applying these principles but will at the same time show how their application greatly facilitates the solution of a single, rather difficult, polyalphabetic substitution cipher. It is realized, of course, that the cryptogram could be solved by the usual methods of frequency and long, patient experimentation. However, the method to be described was actually applied and very materially reduced the amount of time and labor that would otherwise have been required for solution.

49. The cryptogram employed in the exposition.--a. The problem that will be used in this exposition involves an actual cryptogram submitted for solution in connection with a cipher device having two concentric disks upon which the same random-mixed alphabet appears, both alphabets progressing in the same direction. This was obtained from a study of the descriptive circular accompanying the cryptogram. By the usual process of factoring, it was determined that the cryptogram involved 10 alphabets. The message as arranged according to its period is shown in Fig. 34, in which all repetitions of two or more letters are indicated.

b. The triliteral frequency distributions are given in Fig. 35. It will be seen that on account of the brevity of the message, considering the number of alphabets involved, the frequency distributions do not yield many clues. By a very careful study of the repetitions, tentative individual determinations of values of cipher letters, as illustrated in Figs. 36, 37, 38, and 39, were made. These are given in sequence and in detail in order

to show that there is nothing artificial or arbitrary in the preliminary stages of analysis here set forth.

THE CRYPTOGRAM

(Repetitions underlined)

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
B	G	B	Z	D	P	F	B	O	U	O
C	G	R	F	T	Z	M	Q	M	A	V
D	K	Z	U	G	D	Y	F	T	R	W
E	G	J	X	N	L	W	Y	O	U	X
F	I	K	W	E	P	Q	Z	O	K	Z
G	P	R	X	D	W	L	Z	I	C	W
H	G	K	Q	H	O	L	O	D	V	M
I	G	O	X	S	N	Z	H	A	S	E
J	B	B	J	I	P	Q	F	J	H	D
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	Q	R	Q	F	V	M	L	H
M	S	R	Q	E	W	M	L	N	A	E
N	G	S	X	E	R	O	Z	J	S	E
O	G	V	Q	W	E	J	M	K	G	H

	1	2	3	4	5	6	7	8	9	10
P	R	C	V	O	P	N	B	L	C	W
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
S	C	F	B	S	C	V	X	C	H	Q
T	Z	T	Z	S	D	M	X	W	C	M
U	R	K	U	H	E	Q	E	D	G	X
V	F	K	V	H	P	J	J	K	J	Y
W	Y	Q	D	P	C	J	X	L	L	L
X	G	H	X	E	R	O	Q	P	S	E
Y	G	K	B	W	T	L	F	D	U	Z
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
BB	H	S	P	O	P	N	M	D	L	M
CC	G	C	K	W	D	V	B	L	S	E
DD	G	S	U	G	D	P	O	T	H	X

	1	2	3	4	5	6	7	8	9	10
EE	B	K	D	Z	F	M	T	G	Q	J
FF	L	F	U	Y	D	T	Z	V	H	Q
GG	Z	G	W	N	K	X	J	T	R	N
HH	Y	T	X	C	D	P	M	V	L	W
II	B	G	B	W	W	O	Q	R	G	N
JJ	H	H	V	L	A	Q	Q	V	A	V
KK	J	Q	W	O	O	T	T	N	V	Q
LL	B	K	X	D	S	O	Z	R	S	N
MM	Y	U	X	O	P	P	Y	O	X	Z
NN	H	O	Z	O	W	M	X	C	G	Q
OO	J	J	U	G	D	W	Q	R	V	M
PP	U	K	W	P	E	F	X	E	N	F
QQ	C	C	U	G	D	W	P	E	U	H
RR	Y	B	W	E	W	V	M	D	W	J
SS	R	Z	X							

Figure 34.

Triliteral Frequency Distributions

I

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
	EB	FF			XK	YB	ES	XK	ZC	VZ	WQ			ZC	ZR	DC	HC	HR		MK		-F		YQ	QT
	HZ	FC				OR	NH		VQ	ZL	JF						MK							NT	QG
	XK					WJ	ZO		QJ								JZ							NU	
	WG					WK																		HB	
	QK					MO																			
						ES																			
						EV																			
						LH																			
						EK																			
						MC																			
						ES																			

II

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
	GZ	QB			WU	ZW	GX		GX	IW	KB			GX		LZ	GF	GX	ZZ	YX	GQ				KU
	BJ	JQ			CB	BB	HV		JU	GQ				HZ		YD	PX	HP	YX						BZ
	YW	RV			LU					RU						JW	SQ	GU							RX
		OD								FV															
		GK								GB															
		CU								BD															
										BX															
										UW															

III

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
	CZ		QP		RT				BI	CW					SO	KH				FP	CO	KE	JN		BD
	FS		CH													CR				ZG	KH	GN	RD		QA
	KW		KZ													RE				KH	HL	QO	OS		ZC
	LP															VW				SG		KP	SE		TS
	GW																			FY		BE	HE		OO
																				JG			TC		
																				CG			KD		
																							UO		
																							Z-		

IV

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
ZA		ZK	ZP	WP		UD	QO	JP			VA		XL	VP	UC		QQ	XN	FZ			QE		UD	BE
		XD	XW	QW		UD	UE						WK	PP	DC			BC				BT			DF
			XS	XR		UD	VP							WO	BC			ZD				KD			
				XR		UD	DW							XP	WE							BW			
				WW										ZW											

V

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
AA		PF	GY	ZX	ZM					CQ	NW		SZ	HL	DF	RF	EO	DO	WL			DL			TM
LQ		SV	SM	WJ						NX				OT	EQ		EO					EM			
		PJ	WV	HQ											IQ							HM			
		PJ	GP	PF											ON							WO			
			YT												HJ							OM			
			GP												ON							EV			
			GW												OP										
			GW																						

VI

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
AM					CO				EM		WZ	ZQ	PB	RZ	DO	PZ			DZ		CX	LY	EQ	DF	NH
					PB				PJ		OO	WL	PM	RQ	DM	PF			OT		DB	DQ	KJ		
					QV				CX		TF	DX		WQ	PY	KO					WM	DP			
					EX				CO			WZ		SZ		EE									
												FT				AQ									
												WX													

VII

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
	FO			QD	YT		ZA		JK		MN	JK		FC	WE	MM			MG		FM		VC	WO	QO
	NL				QJ				XT			AD		LD		XT			TN				MW	PO	LI
	VL				LD							ND		QI		OP							JL		OJ
												PV		JT		OR							MC		MT
												VD		PT		QV							FE		TV
																WR									OR

VIII

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
HS		OJ	OV	XN		TQ		ZC	FH	MG	BC	QA	LA	BU	QS		QG		FR		ZH	XC			
		XH	MC	PU				OK	ZS	JJ	XL	VL	TV	YU			ZS		QX		ML				
		XG	EG								BS			ZK			QV		ZU		QA				
			FU											YX					OX						
			ML																OH						
			MY																JR						

IX

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
MV		IW				KH	JD		CY	OZ	MH		EF			GJ	TW	AE		OO	DM		TZ	DJ	
NE		LW				DX	CQ		KY	IF	LL						TN	JE		OX	NQ		TE		
VV		DH				RN	TX				DM							PE		DZ	RM		OZ		
		WM				CQ	VQ				VW							LE		TZ					
																		RN		EH					

X

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	Z	Y	Z
			HQ	SB	KC		LS		QL		LG	VG	RY	UG		HZ					AK	RG	UI	JG	KP
				AG	NC		GR		YR			CR	GH			HZ					AJ	CG	GF	JY	XJ
				SG			CB					LG	SY			VB						CL	HB		UO
				SG			UY					VU				GJ						LB			UK
				XH																					XH
				SG																					

Figure 35.

~~CONFIDENTIAL~~

INITIAL VALUES FROM ASSUMPTIONS

$G_c=E_p$, $K_c=E_p$, $X_c=E_p$, and $D_c=E_p$, from frequency considerations

UGD=THE, PCJ=THE, and SEG=THE, from study of repetitions

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
			T	T	H					
B	G	B	Z	D	P	F	B	O	U	O
	E									
C	G	R	F	T	Z	M	Q	M	A	V
	E									
D	K	Z	U	G	D	Y	F	T	R	W
			T	H	E					
E	G	J	X	N	L	W	Y	O	U	X
	E		E							
F	I	K	W	E	P	Q	Z	O	K	Z
		E								
G	P	R	X	D	W	L	Z	I	C	W
			E							
H	G	K	Q	H	O	L	O	D	V	M
	E	E								
I	G	O	X	S	N	Z	H	A	S	E
	E		E						T	H
J	B	B	J	I	P	Q	F	J	H	D
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	Q	R	Q	F	V	M	L	H
M	S	R	Q	E	W	M	L	N	A	E
										H
N	G	S	X	E	R	O	Z	J	S	E
	E		E						T	H
O	G	V	Q	W	E	J	M	K	G	H
	E				E					

	1	2	3	4	5	6	7	8	9	10
P	R	C	V	O	P	N	B	L	C	W
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
S	C	F	B	S	C	V	X	C	H	Q
					H					
T	Z	T	Z	S	D	M	X	W	C	M
					E					
U	R	K	U	H	E	Q	E	D	G	X
		E	T							
V	F	K	V	H	P	J	J	K	J	Y
		E				E				
W	Y	Q	D	P	C	J	X	L	L	L
				T	H	E				
X	G	H	X	E	R	O	Q	P	S	E
	E		E						T	H
Y	G	K	B	W	T	L	F	D	U	Z
	E	E								
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
				T	H	E				H
BB	H	S	P	O	P	N	M	D	L	M
CC	G	C	K	W	D	V	B	L	S	E
	E				E				T	H
DD	G	S	U	G	D	P	O	T	H	X
	E		T	H	E					

	1	2	3	4	5	6	7	8	9	10
EE	B	K	D	Z	F	M	T	G	Q	J
		E								
FF	L	F	U	Y	D	T	Z	V	H	Q
			T		E					
GG	Z	G	W	N	K	X	J	T	R	N
HH	Y	T	X	C	D	P	M	V	L	W
			E		E					
II	B	G	B	W	W	O	Q	R	G	N
JJ	H	H	V	L	A	Q	Q	V	A	V
KK	J	Q	W	O	O	T	T	N	V	Q
LL	B	K	X	D	S	O	Z	R	S	N
		E	E						T	
MM	Y	U	X	O	P	P	Y	O	X	Z
NN	H	O	Z	O	W	M	X	C	G	Q
OO	J	J	U	G	D	W	Q	R	V	M
			T	H	E					
PP	U	K	W	P	E	F	X	E	N	F
		E		T						
QQ	C	C	U	G	D	W	P	E	U	H
			T	H	E					
RR	Y	B	W	E	W	V	M	D	Y	J
SS	R	Z	X							
			E							

Figure 36.

Additional Values from Assumptions (I)

Refer to line DD in Figure 29, S_2 assumed to be N_p.

Refer to line M in figure 29, A_9 assumed to be W_p

Then in lines C–D, A V K Z U G D (9 10 1 2 3 4 5) is assumed to be WITH THE

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
			T	T	H					
B	G	B	Z	D	P	F	B	O	U	O
	E									
C	G	R	F	T	Z	M	Q	M	A	V
	E								W	I
D	K	Z	U	G	D	Y	F	T	R	W
	T	H	T	H	E					
E	G	J	X	N	L	W	Y	O	U	X
	E		E							
F	I	K	W	E	P	Q	Z	O	K	Z
		E								
G	P	R	X	D	W	L	Z	I	C	W
			E							
H	G	K	Q	H	O	L	O	D	V	M
	E	E								
I	G	O	X	S	N	Z	H	A	S	E
	E		E						T	H
J	B	B	J	I	P	Q	F	J	H	D
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	Q	R	Q	F	V	M	L	H
M	S	R	Q	E	W	M	L	N	A	E
									W	H
N	G	S	X	E	R	O	Z	J	S	E
	E	N	E						T	H
O	G	V	Q	W	E	J	M	K	G	H
	E					E				

	1	2	3	4	5	6	7	8	9	10
P	R	C	V	O	P	N	B	L	C	W
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
		H								
S	C	F	B	S	C	V	X	C	H	Q
					H					
T	Z	T	Z	S	D	M	X	W	C	M
					E					
U	R	K	U	H	E	Q	E	D	G	X
		E	T							
V	F	K	V	H	P	J	J	K	J	Y
		E				E				
W	Y	Q	D	P	C	J	X	L	L	L
				T	H	E				
X	G	H	X	E	R	O	Q	P	S	E
	E		E						T	H
Y	G	K	B	W	T	L	F	D	U	Z
	E	E								
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
	T			T	H	E				
BB	H	S	P	O	P	N	M	D	L	M
		N								
CC	G	C	K	W	D	V	B	L	S	E
	E				E				T	H
DD	G	S	U	G	D	P	O	T	H	X
	E	N	T	H	E					

	1	2	3	4	5	6	7	8	9	10
EE	B	K	D	Z	F	M	T	G	Q	J
		E								
FF	L	F	U	Y	D	T	Z	V	H	Q
			T		E					
GG	Z	G	W	N	K	X	J	T	R	N
HH	Y	T	X	C	D	P	M	V	L	W
			E		E					
II	B	G	B	W	W	O	Q	R	G	N
JJ	H	H	V	L	A	Q	Q	V	A	V
									W	I
KK	J	Q	W	O	O	T	T	N	V	Q
LL	B	K	X	D	S	O	Z	R	S	N
		E	E						T	
MM	Y	U	X	O	P	P	Y	O	X	Z
NN	H	O	Z	O	W	M	X	C	G	Q
OO	J	J	U	G	D	W	Q	R	V	M
			T	H	E					
PP	U	K	W	P	E	F	X	E	N	F
		E		T						
QQ	C	C	U	G	D	W	P	E	U	H
			T	H	E					
RR	Y	B	W	E	W	V	M	D	Y	J
SS	R	Z	X							
		H	E							

Figure 37.

~~CONFIDENTIAL~~

Additional Values from Assumptions (II)

Refer to Figure 30, line A, W F U P C F O C J Y (1 2 3 4 5 6 7 8 9 10), assume to be BUT THOUGH

W	F	U	P	C	F	O	C	J	Y
		T	T	H					

Refer to Figure 30, lines N and X, where repetition X E R O (3 4 5 6) occurs, assume EACH

X	E	R	O	
E				

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
	B	U	T	T	H	O	U	G	H	
B	G	B	Z	D	P	F	B	O	U	O
	E					O				
C	G	R	F	T	Z	M	Q	M	A	V
	E								W	I
D	K	Z	U	G	D	Y	F	T	R	W
	T	H	T	H	E					
E	G	J	X	N	L	W	Y	O	U	X
	E		E							
F	I	K	W	E	P	Q	Z	O	K	Z
		E		A						
G	P	R	X	D	W	L	Z	I	C	W
			E							
H	G	K	Q	H	O	L	O	D	V	M
	E	E				U				
I	G	O	X	S	N	Z	H	A	S	E
	E		E						T	H
J	B	B	J	I	P	Q	F	J	H	D
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	Q	R	Q	F	V	M	L	H
						O				
M	S	R	Q	E	W	M	L	N	A	E
				A					W	H
N	G	S	X	E	R	O	Z	J	S	E
	E	N	E	A	C	H			T	H
O	G	V	Q	W	E	J	M	K	G	H
	E					E				

	1	2	3	4	5	6	7	8	9	10
P	R	C	V	O	P	N	B	L	C	W
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
		H					U			
S	C	F	B	S	C	V	X	C	H	Q
		U			H			G		
T	Z	T	Z	S	D	M	X	W	C	M
					E					
U	R	K	U	H	E	Q	E	D	G	X
		E	T							
V	F	K	V	H	P	J	J	K	J	Y
		E					E		H	
W	Y	Q	D	P	C	J	X	L	L	L
				T	H	E				
X	G	H	X	E	R	O	Q	P	S	E
	E		E	A	C	H			T	H
Y	G	K	B	W	T	L	F	D	U	Z
	E	E								
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
	T			T	H	E	U			H
BB	H	S	P	O	P	N	M	D	L	M
		N								
CC	G	C	K	W	D	V	B	L	S	E
	E				E				T	H
DD	G	S	U	G	D	P	O	T	H	X
	E	N	T	H	E		U			

	1	2	3	4	5	6	7	8	9	10
EE	B	K	D	Z	F	M	T	G	Q	J
		E								
FF	L	F	U	Y	D	T	Z	V	H	Q
		U	T		E					
GG	Z	G	W	N	K	X	J	T	R	N
HH	Y	T	X	C	D	P	M	V	L	W
			E		E					
II	B	G	B	W	W	O	Q	R	G	N
						H				
JJ	H	H	V	L	A	Q	Q	V	A	V
									W	I
KK	J	Q	W	O	O	T	T	N	V	Q
LL	B	K	X	D	S	O	Z	R	S	N
		E	E			H			T	
MM	Y	U	X	O	P	P	Y	O	X	Z
NN	H	O	Z	O	W	M	X	C	G	Q
								G		
OO	J	J	U	G	D	W	Q	R	V	M
			T	H	E					
PP	U	K	W	P	E	F	X	E	N	F
		E		T		O				
QQ	C	C	U	G	D	W	P	E	U	H
			T	H	E					
RR	Y	B	W	E	W	V	M	D	Y	J
				A						
SS	R	Z	X							
		H	E							

Figure 38.

Additional Values From Assumptions (III)

4 5 6
OPN—assume ING from repetition and frequency

9 0 1
HQZ—assume ING from repetition and frequency

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
	B	U	T	T	H	O	U	G	H	
B	G	B	Z	D	P	F	B	O	U	O
	E				N	O				
C	G	R	F	T	Z	M	Q	M	A	V
	E								W	I
D	K	Z	U	G	D	Y	F	T	R	W
	T	H	T	H	E					
E	G	J	X	N	L	W	Y	O	U	X
	E		E							
F	I	K	W	E	P	Q	Z	O	K	Z
		E		A	N					
G	P	R	X	D	W	L	Z	I	C	W
			E							
H	G	K	Q	H	O	L	O	D	V	M
	E	E					U			
I	G	O	X	S	N	Z	H	A	S	E
	E		E						T	H
J	B	B	J	I	P	Q	F	J	H	D
					N				I	
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	O	R	Q	F	V	M	L	H
						O				
M	S	R	Q	E	W	M	L	N	A	E
				A					W	H
N	G	S	X	E	R	O	Z	J	S	E
	E	N	E	A	C	H			T	H
O	G	V	Q	W	E	J	M	K	G	H
	E					E				

	1	2	3	4	5	6	7	8	9	10
P	R	C	V	O	P	N	B	L	C	W
				I	N	G				
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
		H					U			
S	C	F	B	S	C	V	X	C	H	Q
		U			H			G	I	N
T	Z	T	Z	S	D	M	X	W	C	M
	G				E					
U	R	K	U	H	E	Q	E	D	G	X
		E	T							
V	F	K	V	H	P	J	J	K	J	Y
		E			N	E			H	
W	Y	Q	D	P	C	J	X	L	L	L
				T	H	E				
X	G	H	X	E	R	O	Q	P	S	E
	E		E	A	C	H			T	H
Y	G	K	B	W	T	L	F	D	U	Z
	E	E								
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
	T			T	H	E	U			H
BB	H	S	P	O	P	N	M	D	L	M
		N		I	N	G				
CC	G	C	K	W	D	V	B	L	S	E
	E				E				T	H
DD	G	S	U	G	D	P	O	T	H	X
	E	H	T	H	E		U		I	

	1	2	3	4	5	6	7	8	9	10
EE	B	K	D	Z	F	M	T	G	Q	J
		E								
FF	L	F	U	Y	D	T	Z	V	H	Q
		U	T		E				I	N
GG	Z	G	W	N	K	X	J	T	R	N
	G									
HH	Y	T	X	C	D	P	M	V	L	W
			E		E					
II	B	G	B	W	W	O	Q	R	G	N
						H				
JJ	H	H	V	L	A	Q	Q	V	A	V
									W	I
KK	J	Q	W	O	O	T	T	N	V	Q
				I						N
LL	B	K	X	D	S	O	Z	R	S	N
		E	E			H			T	
MM	Y	U	X	O	P	P	Y	O	X	Z
				I	N					
NN	H	O	Z	O	W	M	X	C	G	Q
				I				G		N
OO	J	J	U	G	D	W	Q	R	V	M
				T	H	E				
PP	U	K	W	P	E	F	X	E	N	F
		E		T		O				
QQ	C	C	U	G	D	W	P	E	U	H
				T	H	E				
RR	Y	B	W	E	W	V	M	D	Y	J
				A						
SS	R	Z	X							
		H	E							

Figure 39.

c. From the initial and subsequent tentative identifications shown in Figs. 36, 37, 38, and 39, the values obtained were arranged in the form of the secondary alphabets in a reconstruction matrix, shown in Fig. 40.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
ø	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
1		W			G		Z													K						
2					K			Z						S							F					
3					X															U						
4	E							G	O											P						
5			R		D			C						P												
6					J		N	O							F											
7																					O					
8							C																			
9								J	H											S				A		
10								E	V					Q												

Figure 40.

50. Application of principles.--a. Throughout this paragraph reference will be made to Fig. 40, above. Hereafter, in order to avoid all ambiguity and for ease in reference, the position of a letter in Fig. 40 will be indicated as stated in footnote 2, p. 113. Thus, N(6-7) refers to the letter N in row 6 and in column 7 of Fig. 40.

b. (1) Let us consider the following pairs of letters:

E(ø-5)	J(6-5)
G(ø-7)	N(6-7)
H(ø-8)	O(6-8)
O(ø-15)	F(6-15)

} HO, OF = HOF

(We are able to use the row marked "ø" in Fig. 40 since this is a case of a mixed sequence sliding against itself.)

(2) The immediate results of this set of values will now be given. Having HOF as a sequence, with EJ as belonging to the same displacement interval, suppose HOF and EJ are placed into juxtaposition as portions of sliding components. Thus:

Plain.... . . . H O F . . .
Cipher... . . . E J . . . .

When $H_p = E_c$, then $O_p = J_c$.

(3) Refer now to alphabet 10, Fig. 40, where it is seen that $H_p = E_c$. The derived value, $O_p = J_c$, can be inserted immediately in the same alphabet and substituted in the cryptogram.

(4) The student may possibly get a clearer idea of the principles involved if he will regard the matter as though he were dealing with arithmetical proportion. For instance, given any three terms in the proportion 2:8=4:16, the 4th term can easily be found. Furthermore, given the pair of values on the left-hand side of the equation, one may find numerous pairs of values which may be inserted in the right-hand side, or vice versa. For instance, 2:8=4:16 is the same as 2:8=5:20, or 9:36=4:16, and so on. An illustration of each of these principles will now be given, reference being made to Fig. 40. As an example of the first principle, note that E(Ø-5):H(Ø-8) = J(6-5):O(6-8). Now find E(10-8):H(Ø-8) = ?(10-15):O(Ø-15). It is clear that J may be inserted as the 3d term in this proportion, thus giving the important new value $O_p^{10} = J_c$, which is exactly what was obtained directly above, by means of the partial sliding components. As an example of the second principle, note the following pairs:

E(Ø-5)	H(Ø-8)
K(2-5)	Z(2-8)
D(5-5)	C(5-8)
J(6-5)	O(6-8)

These additional pairs are also noted:

K(1-20)	Z(1-7)
T(Ø-20)	G(Ø-7)

Therefore, E:H=K:Z=D:C=J:O=T:G, and T may be inserted in position (4-5).

c. (1) Again, GN belongs to the same set of displacement-interval values as do EJ and HOF. Hence, by superimposition:

Plain....	. . . H O F . . .
Cipher...	. . . G N . . . .

(2) Referring to alphabet 4, when $H_p = G_c$, then $O_p = N_c$. Therefore, the letter N can be inserted in position (4-15) in Fig. 40, and the value $N_c^4 = O_p$ can be substituted in the cryptogram.

(3) Furthermore, note the corroboration found from this particular superimposition:

H(Ø-8)	G(Ø-7)
O(6-8)	N(6-7)

This checks up the value in alphabet 6, $G_p = N_c$.

d. (1) Again superimpose HOF and GN:

. . . H O F . . .
. . . . G N . . .

(2) Note this corroboration:

O(6-8) G(4-8)
F(6-15) N(4-15)

which has just been inserted in Fig. 40, as stated above.

e. (1) Again using HOF and EJ, but in different superimposition:

. . . H O F . .
. . E J . . . .

(2) Refer now to H(9-9), J(9-8). Directly under these letters is found V(10-9), E(10-8). Therefore, the V can be added immediately before H O F, making the sequence V H O F.

f. (1) Now take V H O F and juxtapose it with E J, thus:

. . . V H O F . . .
. . . E J . . .

(2) Refer now to Fig. 40, and find the following:

V(10-9) E(10-8)
H(9-9) J(9-8)
O(4-9) G(4-8)
I(ϕ-9) H(ϕ-8)

(3) From the value O G it follows that G can be set next to J in E J. Thus:

. . . V H O F . . .
. . . E J G . . .

(4) But G N already is known to belong to the same set of displacement-interval values as E J. Therefore, it is now possible to combine E J, J G, and G N into one sequence, E J G N, yielding:

. . . V H O F . . .
. . . E J G N . . .

g. (1) Refer now to Fig. 40.

V(ø-22)	E(ø-5)
?(1-22)	G(1-5)
?(2-22)	K(2-5)
?(3-22)	X(3-5)
?(5-22)	D(5-5)
?(6-22)	J(6-5)

(2) The only values which can be inserted are:

O(1-22)	G(1-5)
H(6-22)	J(6-5)

(3) This means that $V_p = O_c$ in alphabet 1 and that $V_p = H_c$ in alphabet 6. There is one O_c in the frequency distribution for Alphabet 1, and no H_c in that for Alphabet 6. The frequency distribution is, therefore, corroborative insofar as these values are concerned.

h. (1) Further, taking E J G N and V H O F, superimpose them thus:

. . . E J G N . . .
. . . V H O F . . .

(2) Refer now to Fig. 40.

E(ø-5)	H(ø-8)
G(1-5)	?(1-8)

(3) From the diagram of superimposition the value G(1-5) F(1-8) can be inserted, which gives $H_p = F_c$ in alphabet 1.

i. (1) Again, V H O F and E J G N are juxtaposed:

. . . V H O F . . .
. . . E J G N . . .

(2) Refer to Fig. 40 and find the following:

H(ø-8)	G(4-8)
A(ø-1)	E(4-1)

This means that it is possible to add A, thus:

. . . A V H O F . . .
. . . E J G N . . .

(3) In the set there are also:

E(ø-5) G(1-5)
G(ø-7) Z(1-7)

Then in the superimposition

. . . E J G N . . .
. . . E J G N . . .

it is possible to add Z under G, making the sequence E J G N Z.

(4) Then taking

. . . A V H O F . . .
. . . E J G N Z . . .

and referring to Fig. 40:

H(ø-8) N(ø-14)
O(6-8) ?(6-14)

It will be seen that O = Z from superimposition, and hence in alphabet 6 $N_p = Z_c$, an important new value, but occurring only once in the cryptogram. Has an error been made? The work so far seems too corroborative in interlocking details to think so.

j. (1) The possibilities of the superimposition and sliding of the AVHOF and the EJGNZ sequences have by no means been exhausted as yet, but a little different trail this time may be advisable.

E(ø-5) T(ø-20)
G(1-5) K(1-20)
X(3-5) U(3-20)

(2) Then:

. . . E J G N Z . . .
. . . T . K . . .

(3) Now refer to the following:

E(ø-5) K(2-5)
N(ø-14) S(2-14)

whereupon the value S can be inserted:

. . . E J G N Z . . .
. . . T . K . . S . . .

k. (1) Consider all the values based upon the displacement interval corresponding to JG:

J(6-5)	G(1-5) →	J(9-8)	G(4-8)				
N(6-7)	Z(1-7)	H(9-9)	O(4-9)				
		S(9-20)	P(4-20) →	S(2-14)	P(5-14)		
				Z(2-8)	C(5-8)		
				K(2-5)	D(5-5)		

(2) Since J and G are sequent in the E J G N Z sequence, it can be said that all the letters of the foregoing pairs are also sequent. Hence Z C, S P, and K D are available as new data. These give E J G N Z C and T . K D . S P.

(3) Now consider:

T(ø-20)	P(4-20)
A(ø-1)	E(4-1)
H(ø-8)	G(4-8)
I(ø-9)	O(4-9)

Now in the T . K D . S P sequence the interval between T and P is

1 2 3 4 5 6
T . . . . . P. Hence, the interval between A and E is 6 also. It follows therefore that the sequences A V H O F and E J G N Z C should be united, thus:

1 2 3 4 5 6
. . . A V H O F . E J G N Z C . . .

(4) Corroboration is found in the interval between H and G, which is also 6. The letter I can be placed into position, from the relation I(ø-9) O(4-9), thus:

. . . I . . A V H O F . E J G N Z C . . .

l. (1) From Fig. 40:

H(ø-8)	Z(2-8)
E(ø-5)	K(2-5)
N(ø-14)	S(2-14)
U(ø-21)	F(2-21)

(2) Since in the I . . A V H O F . E J G N Z C sequence the letters H and Z are separated by 8 intervals one can write:

1 2 3 4 5 6 7 8
. . . H . . . . Z . . .
. . . E . . . . K . . .
. . . N . . . . S . . .
. . . U . . . . F . . .

(3) Hence one can make the sequence

```
                                    1 2 3 4 5 6 7 8
. . . I . . A V H O F . E J G N Z C . . K . . .
Then . . . I . . A V H O F . E J G N Z C T . K D . S P . . .
and . . U I . . A V H O F . E J G N Z C T . K D . S P . . .
      1 2 3 4 5 6 7 8           1 2 3 4 5 6 7 8
```

m. (1) Subsequent derivations can be indicated very briefly as follows:

E(ø-5) C(ø-3)
D(5-5) R(5-3)

```
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
From U I . . A V H O F . E J G N Z C T . K D . S P . . .
one can write      . . . E . . . . C . . .
                         1 2 3 4 5
and                                  . . . D . . . . R .
                                           1 2 3 4 5
```

making the sequence

```
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
U I . . A V H O F . E J G N Z C T . K D . S P . R .
```

(2) Another derivation:

U(3-20) T(ø-20)
X(3-5) E(ø-5)

```
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
From U I . . A V H O F . E J G N Z C T . K D . S P . R .
one can write
      U I . . . . . . . . . . . . . . T . . .
and   . . . . . . . . . E . . . . . . . . . X . . . . .
```

making the sequence

```
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
U I . . A V H O F . E J G N Z C T . K D X S P . R .
```

(3) Another derivation:

E(ø-5) G(1-5)
B(ø-2) W(1-2)

```
From          . . . E J G . . .
one can write . . . E . G . . .
and then      . . . B . W . . .
```

There is only one place where B . W can fit, viz., at the end:

```
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
U I . . A V H O F . E J G N Z C T . K D X S P B R W
```

n. Only four letters remain to be placed into the sequence, viz., L, M, Q, and Y. Their positions are easily found by application of the primary component to the message. The complete sequence is as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
U	I	M	Y	A	V	H	O	F	L	E	J	G	N	Z	C	T	Q	K	D	X	S	P	B	R	W

Having the primary component fully constructed, decipherment of the cryptogram can be completed with speed and precision. The text is as follows:

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

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

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

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

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

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

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

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

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

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

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

J C Q R Q F V M L M K L B P C J O T X E U K W P E F X E N F
L O O K F O R W A R T I N T H E U N C H R E C T I O N O F A

S R Q E W M L N A E H S P O P N M D L M C C U G D W P E U H
D T O A T I M E W H A N G I N G S T A R N O T H E R C O S M

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

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

Figure 41.

o. The primary component appears to be a random-mixed sequence, no key word for the repeating key is to be found, at least none reappears on experimentation with various hypotheses as to enciphering equations. Nevertheless, the random construction of the primary component did not complicate or retard the solution.

p. Some analysts may prefer to work exclusively with the reconstruction matrix, rather than with sliding strips. One method is as good as the other and personal preferences will dictate which will be used by the individual student. If the reconstruction matrix is used, the original letters should be inserted in red pencil, so as to differentiate them from derived letters.

51. General remarks on the foregoing solution.--a. It is to be stated that the sequence of steps described in the preceding paragraphs corresponds quite closely with that actually followed in solving the problem. It is also to be pointed out that this method can be used as a control in the early stages of analysis because it will allow the cryptanalyst to check assumptions for values. For example, the very first value derived in applying the principles of indirect symmetry to the problem herein described was $H_c = A_p$ in alphabet 1. As a matter of fact the writer had been inclined toward this value, from a study of the frequency and combinations which H_c showed; when the indirect-symmetry method actually substantiated his tentative hypothesis he immediately proceeded to substitute the value given. If he had assigned a different value to H_c, or if he had assumed a letter other than H_c for A_p in that alphabet, the conclusion would immediately follow that either the assumed value for H_c was erroneous, or that one of the values which led to the derivation of $H_c = A_p$ by indirect symmetry was wrong. Thus, these principles aid not only in the systematic and nearly automatic derivation of new values (with only occasional, or incidental references to the actual frequencies of letters), but they also assist very materially in serving as corroborative checks upon the validity of the assumptions already made.

b. Furthermore, while the writer has set forth, in the reconstruction matrix in Fig. 40, a set of 30 values apparently obtained before he began to reconstruct the primary component, this was done for purposes of clarity and brevity in exposition of the principles herein described. As a matter of fact, what he did was to watch very carefully, when inserting values in the reconstruction matrix to find the very first chance to employ the principles of indirect symmetry; and just as soon as a value could be derived, he substituted the value in the cryptographic text. This is good procedure for two reasons. Not only will it disclose impossible combinations but also it gives opportunity for making further assumptions for values by the addition of the derived values to those previously assumed. Thus, the processes of reconstructing the primary component and finding additional data for the reconstruction proceed simultaneously in an ever-widening circle.

c. It is worth noting that the careful analysis of only 30 cipher equivalents in the reconstruction matrix shown in Fig. 40 results in the derivation of the entire table of secondary alphabets, 676 values in all. And while the elucidation of the method seems long and tedious, in its actual

application the results are speedy, accurate, and gratifying in their corroborative effect upon the mental activity of the cryptanalyst.

d. (1) The problem here used as an illustrative case is by no means one that most favorably presents the application and the value of the method, for it has been applied in other cases with much speedier success. For example, suppose that in a cryptogram of 6 alphabets the equivalents of only $\overline{THE}_p$ in all 6 alphabets are fairly certain. As in the previous case, it is supposed that the secondary alphabets are obtained by sliding a mixed alphabet against itself. Suppose the secondary alphabets to be as follows:

∅	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
1	B Q E
2	C L X
3	I V C
4	N P B
5	X O P
6	T Z V

Figure 42.

(2) Consider the following chain of derivatives arranged diagrammatically:

H(∅-8) O(5-8)
T(∅-20) P(5-20)
E(∅-5) X(5-5)→ E(1-20) X(2-20)
Q(1-8) L(2-8)
B(1-5) C(2-5)→ B(4-20) C(3-20)
N(4-5) I(3-5)
P(4-8) V(3-8)→

→ P(5-20) V(6-20)
O(5-8) Z(6-8)
X(5-5) T(6-5)→ X(2-20) T(∅-20)
L(2-8) H(∅-8)
C(2-5) E(∅-5)→ C(3-20) E(1-20)
V(3-8) Q(1-8)
I(3-5) B(1-5)

Figure 43.

(3) These pairs manifestly all belong to the same displacement interval, and therefore unions can be made immediately. The complete list is as follows:

EX, QL, NI, LH, HO, BC, OZ, CE, TP, PV, XT, VQ, IB.

(4) Joining pairs by their common letters, the following sequence is obtained:

. . . N I B C E X T P V Q L H O Z . . .

e. With this as a nucleus the cryptogram can be solved speedily and accurately. When it is realized that the cryptanalyst can assume THE's rather readily in some cases, the value of this principle becomes apparent. When it is further realized that if a cryptogram has sufficient text to enable the THE's to be found easily, it is usually also not at all difficult to make correct assumptions of values for two or three other high-frequency letters; it is then clear that the principles of indirect symmetry of position may often be used with gratifyingly quick success to reconstruct the complete primary component.

f. When the probable-word method is combined with the principles of indirect symmetry the solution of a difficult case is often accomplished with astonishing ease and rapidity.

52. Use of the graphical method in the foregoing example.--a. As an illustration of the application of the graphical method of indirect symmetry, we shall use as an example the cryptogram given in par. 49. It is desired to reconstruct the original primary component, or an equivalent, from the values entered in the reconstruction matrix shown in Fig. 40 on p. 139. Since a mixed sequence is sliding against itself, all the partial sequences (pairs or greater) which can be established by studying the reconstruction matrix are listed as shown in Fig. 44a, below. The single pairs in ∅-7 and ∅-8 are crossed out since they offer no data for reconstruction. This yields the following groups of partial sequences:

∅-	1	2	3	4	5	6	7	8	9	10
	BW	EK	EX	AE	ED	EJ	~~UO~~	~~GC~~	IHJ	HE
	EGZ	HZ	TU	HG	HCR	GN			TS	IV
	TK	NS		IO	NP	HOF			WA	NQ
		UF		TP						

Figure 44a.

b. (1) The sequences HOF and EJ in group 6 and HE in group 10 are noted. The HOF will be placed horizontally and the notation $\underset{1}{\rightarrow}$ is made under group 6. The letter E of the pair HE of group 10 will be placed under the H, and the notation $\downarrow 1$ added under group 10. Thus:

∅-	1	2	3	4	5	6	7	8	9	10
	BW	EK	EX	AE	ED	EJ	~~UO~~	~~GC~~	IHJ	HE
	EGZ	HZ	TU	HG	HCR	GN			TS	IV
	TK	NS		IO	NP	HOF			WA	NQ
		UF		TP		$\underset{1}{\rightarrow}$				$\downarrow 1$

Figure 44b.

Since the sequence EJ belongs to the same displacement interval as HOF, the letter J can be inserted after the letter E, giving:

H O F
E J .

Figure 45a.

No more pairs can be added immediately from groups 6 or 10. Those pairs already entered are crossed out in their respective groups and an inspection is made for additional data in another group.

(2) The sequence IHJ is noted in group 9. The letters H and J are already entered in the diagram. One can therefore place the letter I, and the notation $\searrow$1 is placed under group 9. The addition of the letter I now permits the insertion of the letter V of the sequence IV in group 10, giving:

```
I . . .
V H O F
. E J .
```

Figure 45b.

(3) In group 4 there is the sequence IO which is obtainable in the diagram by the route 1$\underset{2}{\llcorner\!\!\rightarrow}$. This notation is made beneath group 4; the letter A of the sequence AE and the letter G of the sequence HG can now also be entered. The addition of the letter A permits the placement of the letter W of the pair WA of group 9; likewise the addition of the letter G permits the insertion of the letter N of the sequence GN of group 6; finally, the placement of the letter N permits the placement of the Q of group 9. One now has:

```
W . I . . . .
. A V H O F .
. . . E J G N
. . . . . . Q
```

Figure 45c.

(4) Referring to group 1, the sequence EGZ is noted, of which EG appears in the diagram at $\underset{2}{\rightarrow}$. The letter Z can therefore be placed and the letter B of the sequence BW can be inserted two intervals to the left of the letter W, giving:

```
B . W . I . . . . .
. . . A V H O F . .
. . . . . E J G N Z
. . . . . . . . Q .
```

Figure 45d.

(5) Noting the sequence HZ of group 2 as being graphically represented in the diagram by 1$\underset{4}{\llcorner\!\!\rightarrow}$, the letters K, S, and U of the sequence EK, NS and UF may be placed. Thus:

```
B . W U I . . . . . . . .
. . . A V H O F . . . . .
. . . . . E J G N Z . . .
. . . . . . . . Q K . . S
```

Figure 45e.

(6) The letter T of the sequence TK of group 1 can now be placed, which permits the addition of the letter P of the sequence TP of group 4. A study of the diagram shows the pair TU of group 3 at interval $3\underset{4}{\rightarrow}$, which allows the placing of the letter X of the pair EX of the same group. One then has:

```
. X . . . . . . . . . . .
B . W U I . . . . . . . .
. . . A V H O F . . . . .
. . . . . E J G N Z . . .
. . . . . . . T Q K . . S
. . . . . . . . . P . . .
```

Figure 45f.

(7) The diagram now shows the pair NP of group 5 at $2\underset{1}{\rightarrow}$, the letter D of the sequence ED and the letters C and R of HCR can therefore be inserted. Thus:

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

Figure 45g.

(8) Pair TS of group 9 remains. It has already been noted that the notation $\searrow 1$ has been applied to group 9. Hence the letter S can also be placed one interval to the right and below the T, as shown in Fig. 45h, in which all the available data are now entered.

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

Figure 45h.

c. (1) The letter S appears in rows (5) and (6) at a displacement interval of four. This letter then serves as the "tie-in" letter. Marking off 26 squares on cross-section paper the D.SP of row (6) is written, and

row (5) is moved four intervals to the left, at which position the letter S is properly superimposed as follows:

```
          1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Row (5)   C  T  Q  K  .  .  S  .
Row (6)   .  .  .  .  D  .  S  P
```

(2) Likewise row (4) is moved four intervals to the left of its original relative position to row (5) and dropped into position. Row (3) is moved the same distance in relation to row (4), etc. These steps may be illustrated as follows:

```
          1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Row (4)   .  .  .  .  E  J  G  N  Z  C  T  Q  K  D  .  S  P  .  .  .  .  .  .  .  .
Row (3)   H  O  F  .  E  J  G  N  Z  C  T  Q  K  D  .  S  P  .  .  .  .  .  .  A  V
Row (2)   H  O  F  .  E  J  G  N  Z  C  T  Q  K  D  .  S  P  B  .  W  U  T  .  .  A  V
```

(3) The placing of the letter X of row (1) and the letter R of row (7) gives the final sequence:

```
          1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
          H  O  F  .  E  J  G  N  Z  C  T  Q  K  D  X  S  P  B  R  W  U  I  .  .  A  V
```

(4) It will be noted that the foregoing component is identical with that obtained in subpar. 50m(3).

53. Additional remarks on the graphical method.--a. In the example given above only one tie-in letter was available and it was located in adjacent rows. Although only one is necessary, in most cases several tie-in letters are present after all pairs of letters have been entered in the diagram; then the superimposed sequences can be easily connected by their common letters. If the tie-in letter had appeared in adjacent columns instead of adjacent rows as in the foregoing example, the columns would have been shifted vertically and the sequence taken from the diagram in that manner.

b. When only a few pairs of letters forming partial sequences are available, frequently only one tie-in letter may be encountered. If it does not occur in adjacent rows or columns the component can still be written with additional considerations. For example, adjacent diagonals might be used. However, the student will experience no difficulty after the application of this method to a few problems.

c. Since all the data are entered in one diagram, the graphical method of reconstruction quickly discloses erroneous assumptions and enables one to ascertain in a short time whether sufficient data are present for the reconstruction of the component. Even if this is not the case, the diagram automatically offers new values which may be substituted in the cryptogram. One may then assume additional values which can be entered in the diagram or which will serve to corroborate sequences already entered.

d. The placing of the first two sequences of different displacement

intervals in the diagram determines the type of sequences that will be established. If the original sequence entered horizontally in the diagram is an odd decimation of the primary component, a 26-letter sequence can be obtained horizontally. If this original sequence is initially tied in vertically with another sequence of an odd decimation interval, a 26-letter sequence can also be obtained vertically from the diagram.

e. (1) In certain instances, however, it will happen that the available partial sequences have all resulted from even decimations of the basic sequence and that no tie-in letters are present to permit the integration of all the data into a single diagram. In such cases the reconstruction of the basic sequence may take place by taking data from two or more different diagrams, and then, using the relative positions of the letters with respect to each other in these diagrams, the basic sequence may be established. This method can best be demonstrated by means of an example, and the following one is based upon the QUEST... sequence of subpar. 46a. Suppose the reconstruction diagram from the derivation of a few plaintext-ciphertext relationships yields the following partial sequences:

Group.....	1	2	3
Sequences...	Q H O	Q X V	Q T A
	F T	O T	X E
	C E	P K	F K
	J N	F C	U I B
	W D S	U Z W	Z S
		N I	Y G M
		G D	

Figure 46.

(2) The partial sequences in the three groups can be combined to form two diagrams. This may be accomplished by considering the sequences of group 1 as parts of a horizontal component and those of group 2 as parts of a vertical component of a cipher square based upon the original or an equivalent primary sequence. When all the letters of these two groups have been entered into the two resultant diagrams in Figs. 47a and b, it will be observed that the positions occupied in these two diagrams by the letters of group 3 represent the interval 1 $\underset{2}{\llcorner\!\!\rightarrow}$. Thus:

Q H O . .	Y U J N . .
X F T P .	. Z G I . .
V C E K A	. W D S M B
(a)	(b)

Figure 47.

(3) It will be noted that there are 12 letters in each of the two diagrams and that all the letters appearing in the original partial sequences have been included in these two groups. It appears, first, that two 13-letter sequences are involved and second, that the partial sequences in all

three groups represent even decimations of the basic component. The problem now remains to reconstruct the original or an equivalent primary cipher square to which these diagrams belong, or to find the original or an equivalent primary component of which the partial sequences in groups 1, 2, and 3 are derivatives.

(4) Since the two diagrams are linked by the partial sequences of group 3 (because the interval $1\underset{2}{\rightarrow}$ is common to both of them), it follows that any two letters in one of the diagrams will be separated from each other in the basic sequence by the same interval as any two letters occupying the same relative positions in the other diagram. Another way of saying the same thing is, that while the intervals between V and C, C and E, E and K, and K and A, in the basic component (or an equivalent thereof) are unknown, whatever they may be they are identical and the same as that between W and D, D and S, S and M, M and B (from WDSMB), or between Y and U, U and J, J and N (from YUJN), and so on. Likewise, Q and K (interval $2\underset{3}{\rightarrow}$) are separated by the same interval as Y and S, or U and M, and so on.

(5) Making the easiest assumption first, suppose the basic sequence is a keyword-mixed sequence, and that the letter Z is the final letter thereof. If it is preceded by Y, then, because of the relative positions occupied by Y and Z in Fig. 47b, the following would also be sequent in the basic sequence: QF, HT, OP, XC, FE, TK, PA; UG, JI, ZD, GS, and IM. Since the majority of these are hardly likely to occur in a keyword-mixed sequence, the assumption that Y precedes Z is discarded. Suppose X precedes Z (implying that Y is in the keyword). But X and Z are not in the same diagram, so no test can be made. Suppose the sequence is W.Z. Then the following sequences would be valid:

W . Z . U	V . X . Q
D . G . J	C . F . H
S . I . N	E . T . O
	K . P

These look very likely. In fact, noting the D.G.J and the C.F.H sequences it seems logical to integrate or "dovetail" them thus: CDFGHJ. This then suggests that W.Z.U and V.X.Q may be integrated into VWXZQU; S.I.N and E.T.O may be integrated into ESTION. From this point on the matter of extending the partial sequences into the basic one is simple and rather obvious.

f. (1) Suppose, however, that the basic sequence is not a keyword-mixed sequence, so that clues of the nature of those employed in the preceding subparagraph are no longer available. Then what?

(2) Referring back to subpar. 53e(2), it has already been noted that the two diagrams, each containing 12 letters, represent half-sequences (of

13 letters) derived from an even decimation of the original component. (The decimation must be the same in both cases because the interval 1 is common to them.) Suppose an attempt is made to integrate the QHO, XFTP, and VCEKA sequences of Fig. 47a into a 13-letter cycle in a number of ways but the correct integration will be that which will satisfy all the conditions set up by the partial sequences in groups 1, 2, and 3. After a bit of experimentation it is found that the only one which will satisfy all conditions is this:

1	2	3	4	5	6	7	8	9	10	11	12	13
Q	H	O	V	C	E	K	A	X	F	T	P	.

Note, for example, that the conditions represented by QXV in group 2 are satisfied in that the intervals between these letters are the same in the 13-letter cycle; the same is true as regards the intervals between O and T, P and K, and so on. Likewise, the conjugate sequence from Fig. 47b is established as

1	2	3	4	5	6	7	8	9	10	11	12	13
Y	U	J	N	W	D	S	M	B	Z	G	I	.

Thus there have been established the two half sequences involved. The problem now remains to integrate them into a single sequence which is either the primary one or an equivalent primary component.

(3) Each of these sequences may, of course, be expanded to form a 26-element sequence, the elements of which will satisfy the interval relationships among the letters in each 13-letter sequence. Thus:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
(1)	Q	.	O	.	C	.	K	.	X	.	T	.	.	.	H	.	V	.	E	.	A	.	F	.	P	.
(2)	Y	.	J	.	W	.	S	.	B	.	G	.	.	.	U	.	N	.	D	.	M	.	Z	.	I	.

Figure 48.

There remains the problem of integrating these two sequences into a single sequence.

(4) Suppose a start is made thus·

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Q	Y	O	J	C	W	K	S	X	B	T	G	.	.	H	U	V	N	E	D	A	M	F	Z	P	I

Figure 49.

All the interval relationships of groups 1, 2, and 3 of Fig. 46 are satisfied by this sequence. If the sequence is written on a pair of sliding strips, any even-interval displacement of one of the strips will produce plaintext-ciphertext relationships fully satisfied by the requirements of

the sequences in Fig. 46 or Fig. 47. Thus:

(1) Q Y O J C W K S X B T G . . H U V N E D A M F Z P I
H U V N E D A M F Z P I Q Y O J C W K S X B T G . .

(2) Q Y O J C W K S X B T G . . H U V N E D A M F Z P I
X B T G . . H U V N E D A M F Z P I Q Y O J C W K S

(3) Q Y O J C W K S X B T G . . H U V N E D A M F Z P I
T G . . H U V N E D A M F Z P I Q Y O J C W K S X B

Figure 50.

The foregoing three juxtapositions will satisfy all the requirements of the sequences indicated in groups 1, 2, and 3 of Fig. 46, as well as those indicated in Figs. 47a and b. Without further restrictions or additional data, therefore, it is impossible to tell whether the reconstructed single sequence is correct or not. In fact, there are 13 possible integrations of the two expanded 13-letter sequences which will yield equivalent results, since there are 13 positions in which the "dovetailing" of the second sequence may be commenced with respect to the first sequence. Only one of these, however, will be correct in that it will yield a single sequence which, when slid against itself at all juxtapositions (both odd and even displacements) will invariably yield the full quota of plaintext-ciphertext relationships that the original basic or an equivalent primary component yields when slid against itself. (An incorrect integration will often yield a series of equivalents of which only a few are wrong.)

(5) The correct integration will, however, be disclosed quickly enough when the cryptanalyst refers to the cipher text and one or two additional values are derived. Thus, suppose an additional word is deciphered and it yields a pair of values in a new secondary alphabet, for example, $A_p = D_c$ and $U_p = O_c$. The single sequence reconstructed as shown in Fig. 49 will not yield this pair of values, as seen in the following juxtaposition of the sliding strips:

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

Figure 51.

Here $A_p = D_c$ but $U_p = H_c$, not O_c. However, if the "dovetailing" is commenced with the letter S of Fig. 48 and the resultant 26-letter sequence is juxtaposed against itself as shown in Fig. 52, it will be found that the sequence will now satisfy all the requirements.

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

Figure 52.

The sequence is, of course, a decimation of the QUESTIONABLY... sequence, at the third interval.

54. Solution of subsequent messages enciphered by the same primary components.--a. In the discussion of the methods of solving repeating-key ciphers using secondary alphabets derived from the sliding of a mixed component against the normal component (Chapter V), it was shown how subsequent messages enciphered by the same pair of primary components but with different keys could be solved by application of principles involving the completion of the plain-component sequence (pars. 33, 34). The present paragraph deals with the application of these same principles to the case where the primary components are identical mixed sequences.

b. Suppose that the following primary component has been reconstructed from the analysis of a lengthy cryptogram:

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

A new message exchanged between the same correspondents is intercepted and is suspected of having been enciphered by the same primary components but with a different key. The message is as follows:

N F W W P N O M K I W P I D S C A A E T Q V Z S E Y O J S C

A A A F G R V N H D W D S C A E G N F P F A N B N K R V S A

C W D S L O U F A Z N C V X B I U W A G S J C F G

c. Factoring discloses that the period is 7 letters. The text is transcribed accordingly, and is as follows:

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

Figure 53.

d. The letters belonging to the same alphabet are then employed as the initial letters of completion sequences, in the manner shown in par. 33e, using the already reconstructed primary component. The completion

diagrams for the first 10 letters of the first three alphabets, together with the 2-category scores, are as follows:

Gen.	Alphabet 1	Alphabet 2	Alphabet 3
1	7 NMSVSRSPRS	~~FKOZGVGFVL~~	9 WIASANAASO
2	6 QPTWTVTRVT	ø GMDQDWDGWY	5 XOBTBABBTN
3	5 BRIXIWIVWI	~~HPTUFXFHKO~~	4 ZNLILBLLIA
4	~~LVOZOXOWXO~~	~~JRGEGZGJZD~~	3 QAYOYLYYOB
5	~~YWNQNZNXZN~~	~~KVHBSHQHKQF~~	2 HBCNCYCCNL
6	~~GXAUAQAZQA~~	~~MWJTUUJMUG~~	3 ELDADCDDAY
7	~~DZBEBUBQUB~~	~~PXKIKEKPEH~~	1 SYFBFDFFBC
8	3 FQLSLELUEL	~~RZMOMSMRSJ~~	1 TCGLGFGGLD
9	4 GUYTYSYESY	~~VQPNPTPVTK~~	1 IDHYHGHHYF
10	5 HECICTCSTC	6 WURARIRWIM	~~OFJCJHIJCC~~
11	5 JSDODIDTID	~~XEVBVOVXOP~~	~~NGKDKJKKDH~~
12	5 KTFNFOFIOF	~~ZSWLWNWZNR~~	~~AHMFMKMMFJ~~
13	5 MIGAGNGONG	~~QTXYXAXQAV~~	~~BJPGPMPTGK~~
14	4 POHBHAHNAH	~~UIZCZBZUBW~~	4 LKRHRPRRHM
15	~~RNJLJBJABJ~~	~~EOQDQLQELX~~	~~YMVJVRVVJP~~
16	~~VAKYKLKBLK~~	3 SNUFUYUSYZ	~~GPWKWVWWKR~~
17	ø WBMCMYMLYM	6 TAEGECETCQ	~~DRXMXWXXMV~~
18	ø XLPDPCPYCP	5 IBSHSDSIDU	~~FVZPZXZZPW~~
19	4 ZYRFRDRCDR	6 OLTJTFTOFE	~~GWQRQZQQRX~~
20	ø QCVGVFVDFV	6 NYIKIGINGS	~~HXUVUQUUVZ~~
21	ø UDWHWGWFGW	6 ACOMOHOAHT	~~JZEWEUEEWQ~~
22	~~EFKJXHXGHX~~	~~BDNPNJNBJI~~	~~KQSXSESSXU~~
23	~~GGZKZJZHJZ~~	~~LFARAKALKO~~	~~MUTZTSTTZE~~
24	~~THQMQKQJKQ~~	1 YGBVBMBYMN	~~PEIQITIIQS~~
25	~~IJUPUMUKMU~~	1 CHLWLPLCPA	8 RSOUOIOOUT
26	6 OKEREPEMPE	~~DJYXYRYDRB~~	9 VTNENONNEI

Figure 54.

e. The determination of the correct generatrices in Fig. 54 is now an easy matter, however, since in this case the 2-category scores do not at once point to the correct generatrices, a slight experimentation is necessary to arrive at the solution. Logarithmic weights may here be used to by-pass further experimentation; the few generatrices having the best 2-category scores in Fig. 54 are set forth below, with their logarithmic scores:

Gen.		Alphabet 1		Gen.		Alphabet 2		Gen.		Alphabet 3	
1	7	NMSVSRSPRS		10	6	WURARIRWIM		1	9	WIASANAASO	
		8685888688	= 73			5688888586	= 70			5888888888	= 77
2	6	APIWTVTRVT		17	6	TAEGECETCQ		26	9	VTNENONNEI	
		8695959859	= 73			9895979972	= 74			5989888898	= 80
26	6	OKEREPEMPE		19	6	OLTJTFTOFE		25	8	RSOUOIOOUT	
		8298969669	= 72			8791969869	= 72			8886888869	= 77
3	5	BRIXIWIVWI		20	6	NYIKIGINGS					
		4883858558	= 62			8682858858	= 66				
10	5	HECICTCSTC		21	6	ACOMOHOAHT					
		7978797897	= 78			8786878879	= 76				
11	5	JSDODIDTID		18	5	IBSHSDSIDU					
		1878787987	= 70			8487878876	= 71				
12	5	KTFNFOFIOF									
		2968686886	= 67								
13	5	MIGAGNGONG									
		6858585885	= 66								

The correct generatrices, as shown by the highest logarithmic scores, are now assembled in columnar fashion and yield the following plain text:

1	2	3	4	5	6	7
H	A	V				
E	C	T				
C	O	N				
I	M	E				
C	O	N				
T	H	O				
C	O	N				
S	A	N				
T	H	E				
C	T	I				

Figure 55.

f. The corresponding key letters are sought, using enciphering equations $\theta_k/c = \theta_i/p$; $\theta_p/p = \theta_c/c$, and are found to be JOU, which suggests the key word JOURNEY, among others. Testing the key-letters RNEY for alphabets 4, 5, 6, and 7, the following results are obtained:

1	2	3	4	5	6	7
J	O	U	R	N	E	Y
N	F	W	W	P	N	O
H	A	V	E	D	I	R
M	K	I	W	P	I	D
E	C	T	E	D	S	E

Figure 56.

The message may now be completed with ease. It is as follows:

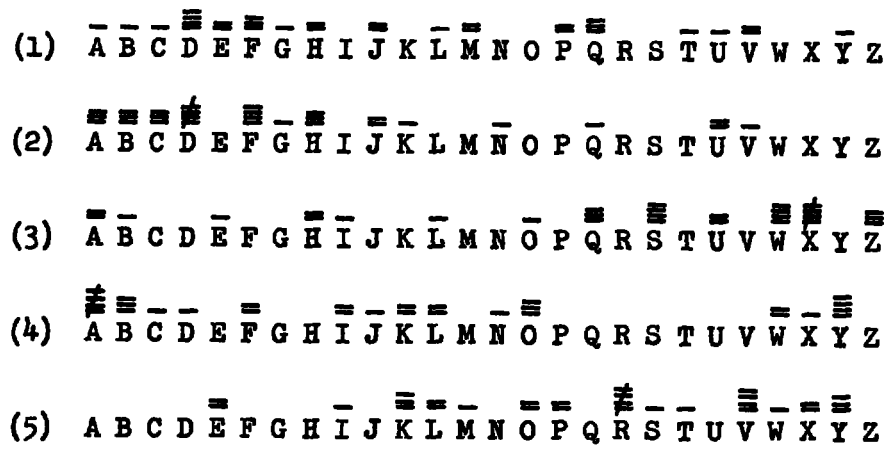
J O U R N E Y	J O U R N E Y
N F W W P N O	P F A N B N K
H A V E D I R	S A N C E I N
M K I W P I D	R V S A C W D
E C T E D S E	T H E D I R E
S C A A E T Q	S L O U F A Z
C O N D R E G	C T I O N O F
V Z S E Y O J	N C V X B I U
I M E N T T O	H O R S E S H
S C A A A F G	W A G S J C F
C O N D U C T	O E F A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

Figure 57.

g. Another method for the solution of cryptograms when the primary components have been recovered might be mentioned at this point. This method, based on the analysis of the uniliteral frequency distributions of the individual monoalphabets, is applicable when there are a sufficient number of tallies (say, at least 25 or so) in each distribution; in such situations this method is often easier and quicker than the generatrix method treated in the preceding subparagraphs.

(1) Let us assume that the enemy has been using keyword-mixed sequences based on QUESTIONABLY for the primary components, and that the following message (factoring to five alphabets) with its accompanying frequency distributions are at hand:

P F O F R	V V Z D V	Q G Q Y I	E F Q J M	H J I C Y	V A B L Y
Q F Z B V	F B U K V	A U S B Y	M U S K P	M C E A R	F N W I L
D D W Y K	Q J L I R	P A A W R	L B Q F K	C D X A X	J H S A R
D D X B W	F C H A O	3 F X A K	E D X O E	Y C H N P	D Q X O E
D D A X O	G B W L T	U H S Y X	H H W Y V	T K U W L	J A Z O S



In spite of the flatness of the distribution for Alphabet 1 (I.C. = 1.08), there is no doubt but that the period is 5.

(2) Consider the frequency distribution for the second alphabet, which has several pronounced peaks. The cipher letters D, F, A, B, C, and H are high, with D_c being the highest. In general, these letters should represent most of the high-frequency plaintext letters such as E, T, N, R, O, A, I, S, etc. Now prepare two strips bearing keyword-mixed sequences based on QUESTIONABLY, the plain-component strip should be 26 letters long, the cipher-component strip doubled length of 52 letters. Place the D_c on the cipher (long) strip under the E_p on the plain (short) strip, and note what plaintext values of F_c, A_c, B_c, C_c, and H_c are concomitant with $D_c = E_p$. Then place D_c on the cipher strip under T_p, N_p, R_p, etc. in turn, noting what plaintext values of the other cipher letters correspond to each setting. When the correct juxtaposition is made, the values of all the cipher letters in Alphabet 2 become known, and the frequencies of the plaintext letters will approximate fairly closely their normal frequencies.

(3) After the correct placement of Alphabet 2 is found, the values for the cipher letters are entered in their proper places in the message. Then the same procedure is applied to each alphabet in turn, and the plaintext values are entered in the message when the correct juxtaposition for the strip is determined. It will be found that the easiest process is to treat the distributions with the most striking peaks (such as those of Alphabets 2 and 3) first, leaving the flattest distributions (such as that of Alphabet 1) until last. Furthermore, after several alphabets have been correctly determined, the clusters of plaintext fragments in the message might suggest complete words, or recovery of part of the key might suggest the entire repeating key, thus rendering unnecessary the placement of the remaining alphabets by the analytic process just described. The solution of this problem is left to the student as an exercise in the foregoing method.

55. Solution of repeating-key ciphers in which the identical mixed components proceed in opposite directions.--The secondary alphabets in this case (Case II c 1 (b) of par. 8) are reciprocal. The steps in solution are essentially the same as in the preceding case (par. 41), the principles of indirect symmetry of position can also be applied with the necessary modifications introduced by virtue of the reciprocity existing within the respective secondary alphabets (subpar. 44p).

56. Solution of repeating-key ciphers in which the primary components are different mixed sequences.--This is Case II c 2 of par. 8. The steps in solution are essentially the same as in pars. 41 and 44, except that in applying the principles of indirect symmetry of position it is necessary to take cognizance of the fact that the primary components are different mixed sequences (subpar. 44q).

57. Solution of subsequent messages after the primary components have been recovered.--a. In the case in which the primary components are identical mixed sequences proceeding in opposite directions, as well as in the case in which the primary components are different mixed sequences, the solution of subsequent messages[1] is a relatively easy matter. In both cases, however, the student must remember that before the method illustrated in par. 54 can be applied it is necessary to convert the cipher letters into their plain-component equivalents before completing the plain-component sequence. From there on, the process of selecting and assembling the proper generatrices is the same as usual.

b. Perhaps an example may be advisable. Suppose the enemy has been found to be using primary components based upon the keyword QUESTIONABLY, the plain component running from left to right, the cipher component in the reverse direction. The following new message has arrived from the intercept station:

M V X O X B Z I Y Z N L W Z H O X I E O O O E P Z F X S R X

E J B S H B O N A U R A P Z I N R A M V X O X A I J Y X W F

K N D O W J E R C U R A L V B Z A Q U W J W X Y I D G R K D

Q B D R M Q E C Y V Q W

c. Factoring discloses that the period is 6 and the message is accordingly transcribed into 6 columns, Fig. 58. The first 10 letters

[1] That is messages intercepted after the primary components have been reconstructed and enciphered by keys different from those used in the messages upon which the reconstruction of the primary components was accomplished

of these columns are then converted into their plain-component equivalents by juxtaposing the two primary components at any point of coincidence, for example, $Q_p = Z_c$. The converted letters are shown in Fig. 59. The letters

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

Figure 58.

1	2	3	4	5	6
O	S	U	M	U	H
Q	P	F	Q	K	G
E	Q	B	M	U	P
W	M	M	M	W	I
Q	Y	U	V	T	U
W	A	H	V	B	H
M	K	J	X	T	J
I	Q	P	K	T	J
O	S	U	M	U	J
P	A	F	U	E	Y

Figure 59.

of the individual columns are then used as the initial letters of completion sequences, using the QUESTIONABLY primary sequence. The final step is the selection and assembling of the selected generatrices. The results for the first ten letters of the first three columns are shown below:

Gen.	Alphabet 1	Alphabet 2	Alphabet 3
1	OQEWQWMIOP	SPQMYAKQSA	UFBMUHJPUF
2	NUSKUKPONR	3 TRUPCBMUTB	EGLPBJKREG
3	ABTZEZRNAV	5 IVERDLPEIL	4 SHYRSKMVSH
4	BGIQSQVABW	5 OWSVFYRSOY	3 TJCVTMPWTJ
5	3 LTOUTUWBLX	4 NXTWGCVTNC	IKDWIPRXIK
6	YINEIEXLYZ	AZIXHDWIAD	OMFXORVZOM
7	COASOGZYCQ	BQOZJFXOBF	NPGZNVWQNP
8	4 DNBTNTQCDU	LUNQKGZNLG	ARHQAWXUAR
9	5 FALLIAIUDFE	3 YEAUMHQAYH	BVJUBXZEBV
10	4 GBYOBOEFGS	GSBEPJUBGJ	LWKELZQSLW
11	4 HLCNLNSGHT	DTLGRKGLDK	YXMSYQUTYX
12	JYDAYATHJI	3 FIYTVMSYFM	GZPTGUEIGZ
13	KCFBCBIJKO	3 GOCIWPTCGP	DQRIDESODQ
14	2 MDGLDLOKMN	5 HNDOXRIDHR	4 FUVOFSTNFU
15	2 PFHYFYNMPA	JAFNZVOFJV	6 GEWNGTIAGE
16	3 RGJCGCAPRB	KBGAQWNGKW	5 HSXAHIOBHS
17	1 VHKDHDBRVL	MLHBUXAHMX	JTZBJONLJT
18	WJMFJFLVWY	PYJLEZBJPZ	KIQLKNAYKI
19	XKPGKGYWXC	RGKYSQLKRQ	3 MOUYMABCMO
20	ZMRHMHGXZD	1 VDMCTUYMVU	3 PNECPBLDPN
21	QPVJPJDZQP	3 WFPDIECPWE	6 RASDRLYFRA
22	URWKRKFQUC	XGRFOSDRXS	1 VBTFVYCGVB
23	2 EVXMVMGUEH	ZHVGNTFVZP	1 WLIGWCDHWL
24	SWZPWPHESJ	QJWHAIGWOI	XYOHXDFJXY
25	TXQRSRJSTK	UKXJBOHXUO	ZGNJZFGKZC
26	IZUVZVKTIM	EMZKLNJZEN	QDAKQGHNQD

Figure 60.

Columnar assembling of selected generatrices gives what is shown in Fig. 61.

1	2	3	4	5	6
F	I	R	.	.	.
A	V	A	.	.	.
L	E	S	.	.	.
I	R	D	.	.	.
A	D	R	.	.	.
I	L	L	.	.	.
U	P	Y	.	.	.
D	E	F	.	.	.
F	I	R	.	.	.
E	L	A	.	.	.

Figure 61.

d. The key letters are sought, and found to be NUM, which suggests NUMBER. The entire message may now be read with ease. It is as follows:

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

Figure 62.

e. If the primary components are different mixed sequences, the procedure is identical with that just indicated. The important point to note is that one must not fail to convert the cipher letters into their plain-component equivalents before the completion-sequence method is applied.

~~CONFIDENTIAL~~

(BLANK)

CHAPTER VIII

SPECIAL SOLUTIONS FOR PERIODIC CIPHERS

58. General remarks.--The preceding two chapters have been devoted to an elucidation of the general principles and procedure in the solution of typical cases of repeating-key ciphers. This chapter will be devoted to a consideration of the variations in cryptanalytic procedure arising from special circumstances. It may be well to add that by the designation "special circumstances" it is not meant to imply that the latter are necessarily unusual circumstances. The student should always be on the alert to seize upon any opportunities that may appear in which he may apply the methods to be described. In practical work such opportunities are by no means rare and are seldom overlooked by competent and experienced cryptanalysts.

59. Deriving the secondary alphabets, the primary components, and the repeating key, given a cryptogram with its plain text.--a. It may happen that a cryptogram and its equivalent plain text are at hand, as the result of capture, pilferage, compromise, etc. This, as a general rule, affords a very easy attack upon the whole system.

b. Taking first the case where the plain component is the normal sequence, the cipher component a mixed sequence, the first thing to do is to write out the cipher text with its letter-for-letter decipherment. From this, by a slight modification of the principles of "factoring", one discovers the length of the key. It is obvious that when a word of three or four letters is enciphered by the same cipher text, the interval between the two occurrences is almost certainly a multiple of the length of the key.[1]

[1] Again, as a note of caution see the remark made in footnote 6 on p 32

By noting a few recurrences of plain text and cipher letters, one can quickly determine the length of the key (assuming of course that the message is long enough to afford sufficient data). Having determined the length of the key, the message is rewritten according to its periods, with the plain text likewise in periods under the cipher letters. From this arrangement one can now reconstruct complete or partial secondary alphabets. If the secondary alphabets are complete, they will show direct symmetry of position; if they are but fragmentary in several alphabets, then the primary component can be reconstructed by the application of the principles of direct symmetry of position.

c. If the plain component is a mixed sequence, and the cipher component the normal (direct or reversed sequence), the secondary alphabets will show no direct symmetry unless they are arranged in the form of deciphering alphabets (that is, $A_c...Z_c$ above the zero line, with their equivalents below). The student should be on the lookout for such cases.

d. (1) If the plain and cipher primary components are identical mixed sequences proceeding in the same direction, the secondary alphabets will show indirect symmetry of position, and they can be used for the speedy reconstruction of the primary components (subpars. 44a to m).

(2) If the plain and the cipher primary components are identical mixed sequences proceeding in opposite directions, the secondary alphabets will be completely reciprocal secondary alphabets and the primary component may be reconstructed by applying the principles outlined in subpar. 44n.

(3) If the plain and cipher primary components are different mixed sequences, the secondary alphabets will show indirect symmetry of position and the primary components may be reconstructed by applying the principles outlined in subpar. 44o.

e. In all the foregoing cases, after the primary components have been reconstructed, the keys can be readily recovered.

60. Solution of isologs involving the same pair of unknown primary components but with different key words of identical length.--a. The simplest case of this kind is that involving two monoalphabetic substitution ciphers with mixed alphabets derived from the same pair of sliding components. An understanding of this case is necessary to that of the case involving repeating-key ciphers.

(1) A message is transmitted from Station "A" to Station "B". "B" then sends "A" some operating signals which indicate that "B" cannot decipher the message, and soon thereafter "A" sends a second message, identical in length with the first. This leads to the suspicion that the plain text of both messages is the same. The intercepted messages are superimposed. Thus:

1. NXGRV MPUOF ZQVCP VWERX QDZVX WXZQE TBDSP VVXJK RFZWH ZUWLU IYVZQ FXOAR
2. EMLHJ FGVUB PRJNG JKWHM RAPJM KMPRW ZTAXG JJMCD HBPKY PVKIV QOJPR BMUSH

(2) Initiating a chain of ciphertext equivalents from Message 1 to Message 2, the following complete sequence is obtained:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
N	E	W	K	D	A	S	X	M	F	B	T	Z	P	G	L	I	Q	R	H	Y	O	U	V	J	C

(3) Experimentation along already-indicated lines soon discloses the fact that the foregoing component is an equivalent primary component of the original primary cipher component based upon the keyword QUESTIONABLY, decimated on the 21st interval. Let the student decipher the cryptogram.

(4) The foregoing example is somewhat artificial in that the plain text was consciously selected with a view to making it contain every letter of the alphabet. The purpose in doing this was to permit the construction of a complete chain of equivalents from only two short messages, in order to give a simple illustration of the principles involved. If the plain text of the message does not contain every letter of the alphabet, then only partial chains of equivalents can be constructed. These may be united, if circumstances will permit, by recourse to the various principles elucidated in par. 44.

(5) The student should carefully study the foregoing example in order to obtain a thorough comprehension of the reason why it was possible to reconstruct the primary component from the two cipher messages without having any plain text to begin with at all. Since the plain text of both messages is the same, the relative displacement of the same primary components in the case of Message 1 differs from the relative displacement of the same primary components in the case of Message 2 by a fixed interval. Therefore, the distance between N and E (the first letters of the two messages), on the primary component, regardless of what plaintext letter these two cipher letters represent, is the same as the distance between E and W (the 18th letters), W and K (the 17th letters), and so on. Thus, this fixed interval permits of establishing a complete chain of letters separated by constant intervals and this chain becomes an equivalent primary component.

b. With the foregoing basic principles in mind the student is ready to note the procedure in the case of two repeating-key ciphers having identical plain texts. First, the case in which both messages have key words of identical length but different compositions will be studied.

c. (1) Given the following two cryptograms suspected to contain the same plain text:

Message 1

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

Message 2

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

(2) The first step is to try to determine the length of the period. The usual method of factoring cannot be employed because there are no long repetitions and not enough repetitions even of digraphs to give any convincing indications. However, a subterfuge will be employed based upon the theory of factoring.

d. (1) Let the two messages be superimposed:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1.	Y	H	Y	E	X	U	B	U	K	A	P	V	L	L	T	A	B	U	V	V	D	Y	S	A	B	P	C	Q	T	U
2.	C	G	S	L	Z	Q	U	B	M	N	C	T	Y	B	V	H	L	Q	F	T	F	L	R	H	L	M	T	A	I	Q

	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
1.	N	G	K	F	A	Z	E	F	I	Z	B	D	J	E	Z	A	L	V	I	D	T	R	O	Q	S	U	H	A	F	K
2.	Z	W	M	D	Q	N	S	D	W	N	L	C	B	L	Q	N	E	T	O	C	V	S	N	Z	R	B	J	N	O	Q

(2) Now let a search be made of cases of identical superimposition. For example, $\begin{matrix}4\\E\\L\end{matrix}$ and $\begin{matrix}44\\E\\L\end{matrix}$ are separated by 40 letters, $\begin{matrix}6\\U\\Q\end{matrix}$, $\begin{matrix}18\\U\\Q\end{matrix}$, and $\begin{matrix}30\\U\\Q\end{matrix}$ are separated by 12 letters. Let these intervals between identical superimpositions be factored, just as though they were ordinary repetitions. That factor which is the most frequent should correspond with the length of the period for the following reason. If the period is the same and the plain text is the same in both messages, then the condition of identity of superimposition can only be the result of identity of encipherments by identical cipher alphabets. This is only another way of saying that the same relative position in the keying cycle has been reached in both cases of identity. Therefore, the distance between identical superimpositions must be either equal to or else a multiple of the length of the period. Hence, factoring the intervals must yield the length of the period. The complete list of intervals and factors applicable to cases of identical superimposed pairs is as follows:

Repetition	Interval	Factors
1st EL to 2d EL...	40	2, 4, 5, 8, 10, 20.
1st UQ to 2d UQ...	12	2, 3, 4, 6.
2d UQ to 3d UQ....	12	2, 3, 4, 6.
1st UB to 2d UB...	48	2, 3, 4, 6, 8, 12, 24.
1st KM to 2d KM...	24	2, 3, 4, 6, 8, 12.
1st AN to 2d AN...	36	2, 3, 4, 6, 9, 12, 18.
2d AN to 3d AN....	12	2, 3, 4, 6.
1st VT to 2d VT...	8	2, 4.
2d VT to 3d VT....	28	2, 4, 7, 14.
1st TV to 2d TV...	36	2, 3, 4, 6, 9, 12, 18.
1st AH to 2d AH...	8	2, 4.
1st BL to 2d BL...	8	2, 4.
2d BL to 3d BL....	16	2, 4, 8.
1st SR to 2d SR...	32	2, 4, 8, 16.
1st FD to 2d FD...	4	2.
1st ZN to 2d ZN...	4	2.
1st DC to 2d DC...	8	2, 4.

(3) The factors 4 and 2 are the only ones common to every one of these intervals, and, since a period of 2 is not very probable, it may be taken as beyond question that the length of the period is 4.

e. Let the messages now be superimposed according to their periods:

1. Y H Y E X U B U K A P V L L T A B U V V D Y S A B P C Q
2. C G S L Z Q U B M N C T Y B V H L Q F T F L R H L M T A

1. T U N G K F A Z E F I Z B D J E Z A L V I D T R O Q S U
2. I Q Z W M D Q N S D W N L C B L Q N E T O C V S N Z R B

1. H A F K
2. J N O Q

f. (1) Now distribute the superimposed letters into a reconstruction matrix, thus:

∅	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
1		L		F	S			J	O		M	Y			N					I					Z	C	Q
2	N		C		D		G					B				M	Z				Q					L	
3	Q	U	T			O			W	B		E		Z		C			R	V		F				S	
4	H				L		W				Q						A	S				B	T				N

(2) By the usual methods, construct the primary or an equivalent primary component. Taking lines ∅ and 1, the following sequences are noted:

BL, DF, ES, HJ, IO, KM, LY, ON, TI, XZ, YC, ZQ.

which, when united by means of common letters and study of other sequences, yield the complete original primary component based upon the key word QUESTIONABLY:

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

(3) The fact that the pair of lines with which the process was commenced yield the original primary sequence is purely accidental, it might have just as well yielded an equivalent primary sequence.

g. (1) Having the primary cipher component, the solution of the messages is now a relatively simple matter. An application of the method elucidated in par. 54 is made, involving the completion of the plain-component sequence and the selection of those generatrices which contain the best assortment of high frequency letters.[2] Thus, using Message 1:

[2] We are assuming in this case that the plain component is identical with the cipher component If this is not the case subpars 61f and g outline the procedure to be followed in such situations

~~CONFIDENTIAL~~

Gen.	Alphabet 1	Alphabet 2	Alphabet 3	Alphabet 4
1	~~YXKLBDBTKE~~	1 HUALUYPUFF	5 YBPTVSCNAI	~~EUVAVAQGZZ~~
2	2 CZMYLFLIMS	4 JEBYECREGG	5 CLRIWTDABO	~~SEWBWBUHQQ~~
3	2 DQPCYGYOPT	3 KSLCSDVSHH	3 DYVOXIFBLN	~~TSXLXLEJUU~~
4	4 FURDCHCNRI	~~MTYDTFWTJJ~~	3 FCWNZOGLYA	~~ITZYZYSKEE~~
5	3 GEVFDJDAVO	~~PICFIGXIKK~~	~~GDXAQNHYCB~~	~~OIQLQGTMSS~~
6	2 HSWGFKFBWN	4 RODGOHZOMM	~~HFZBUAJGDL~~	5 NOUDUDIPTT
7	~~JTXHGMGLXA~~	~~VNFHNJQNPP~~	~~JGQLEBKDFY~~	8 ANEFEFORII
8	~~KIZJHPHYZB~~	~~WAGJAKUARR~~	1 KHUYSLMFGC	6 BASGSGNVOO
9	~~MOQKJRJCQL~~	~~XBHKBMEBVV~~	2 MJECTYPGHD	5 LBTHTHAWNN
10	~~PNUMGVKDUY~~	~~ZLJMLPSLWW~~	~~PKSDICRHJF~~	~~YLIJIJBXAA~~
11	4 RAEPMWMFEC	~~QYKPYRTFXX~~	~~RMTFODVJKO~~	~~CYOKOKLZBB~~
12	3 VBSRPXPGSD	~~UGMRGVIGZZ~~	2 VPIGNFWKMH	2 DCNMNMYQLL
13	4 WLTVRZRHTF	~~EDPVDWODQQ~~	~~WROHAGXMPJ~~	2 FDAPAPCUYY
14	~~XYIWVQVJIG~~	3 SFRWFXNFUU	~~XVNJBHZPRK~~	3 GFBRBRDECC
15	~~ZGOXWUWKOH~~	~~TGVXGZAGEE~~	~~ZWAKLJQRVM~~	1 HGLVLVFSDD
16	~~QDNZXEXMNJ~~	~~IHWZHQBHSS~~	~~QXBMYKUVWP~~	1 JHYWYWGTFF
17	~~UFAQZSZPAK~~	~~OJXQJULJTT~~	~~UZLPCMEWXR~~	~~KJCXCXHIGG~~
18	~~EGBUQTQRBM~~	~~NKZUKEYKII~~	~~EQYRDPSXZV~~	~~MKDZDZJOHH~~
19	3 SHLEUIUVLP	5 AMQEMSCMOO	~~GUCVFRTZQV~~	~~PMFQFQKNJJ~~
20	6 TJYSEOEWYR	4 BPUSPTDPNN	~~TEDWGVIQUX~~	~~RPGUGUMAKK~~
21	~~IKCTSNSXCV~~	8 LRETRIFRAA	~~ISFXHWOUEZ~~	3 VRHEHEPBMM
22	5 OMDITATZDW	3 YVSIVOGVBB	~~OTGZJYNESQ~~	~~WVJSJSRLPP~~
23	~~NPFOIBIQFX~~	3 CWTOWNHWLL	~~NIHQKZASTU~~	~~XWKTKTVYRR~~
24	5 ARGNOLOUGZ	~~DXINXAJXYY~~	~~AOJUMQBTIE~~	~~ZXMIMIWCVV~~
25	4 BVHANYNEHQ	~~FZOAZBKZCC~~	5 BNKEPULIOS	~~QZPAPOXDWW~~
26	~~LWJBACASJU~~	~~GQNBQLMQDD~~	7 LAMSREYONT	~~UQRNRNZFXX~~

(2) In this particular case, it is easy to pick out the correct generatrices for Alphabets 2, 3, and 4, since the correct ones have the highest two-category scores. These generatrices are assembled in columnar fashion in Fig. 63a, below; from this step it is easy to see that the correct generatrix for Alphabet 1 is Generatrix No. 24, as is shown in Fig. 63b:

1	2	3	4
.	L	L	A
.	R	A	N
.	E	M	E
.	T	S	F
.	R	R	E
.	I	E	F
.	F	Y	O
.	R	O	R
.	A	N	I
.	A	T	I

Figure 63a.

1	2	3	4
A	L	L	A
R	R	A	N
G	E	M	E
N	T	S	F
O	R	R	E
L	I	E	F
O	F	Y	O
U	R	O	R
G	A	N	I
Z	A	T	I

Figure 63b.

(3) The key letters are sought and give the key word SOUP. The plain text for the second message is now known, and by reference to the cipher text and the primary components, the key word for this message is found to be TIME. The complete texts are as follows:

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

Figure 64.

61. Solution of isologs involving the same pair of unknown primary components but with key words of different lengths.--a. In the foregoing case the key words for the two messages, although different, were identical in length. When this is not true and the key words are of different lengths, the procedure need be only slightly modified.

b. Given the following two cryptograms suspected of containing the same plain text enciphered by the same primary components but with different key words of different lengths, solve the messages.

Message No. 1

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

Message No. 2

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

c. The messages are long enough to show a few short repetitions which permit factoring. The latter discloses that Message 1 has a period of 4 and Message 2, a period of 6 letters. The messages are superimposed, with numbers marking the position of each letter in the corresponding period, as shown below:

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No 1	V	M	Y	Z	G	E	A	U	N	T	P	K	F	A	Y	J	I	Z	M	B	U	M	Y	K
No 2	Z	G	A	N	W	I	O	M	O	A	C	O	D	H	A	C	L	R	L	P	M	O	Q	O
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No 1	B	V	F	I	V	V	S	E	O	A	F	S	K	X	K	R	Y	W	C	A	C	Z	O	R
No 2	J	E	M	O	Q	U	D	H	X	B	Y	U	Q	M	G	A	U	V	G	L	Q	D	B	S
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No 1	D	O	Z	R	D	E	F	B	L	K	F	E	S	M	K	S	F	A	F	E	K	V	Q	U
No 2	P	U	O	A	B	I	R	P	W	X	Y	M	O	G	G	F	T	M	R	H	V	F	G	W
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No 1	R	C	M	Y	Z	V	O	X	V	A	B	T	A	Y	Y	U	O	A	Y	T	D	K	F	E
No 2	K	N	I	V	A	U	P	F	A	B	R	V	I	L	A	Q	E	M	Z	D	J	X	Y	M
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No 1	N	W	N	T	D	B	Q	K	U	L	A	J	L	Z	I	O	U	M	A	B	O	A	F	S
No 2	E	D	D	Y	B	O	S	V	M	P	N	L	G	X	X	D	Y	D	O	P	X	B	Y	U
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No 1	K	X	Q	P	U	Y	M	J	P	W	Q	T	D	B	T	O	S	I	Y	S	M	I	Y	K
No 2	Q	M	N	K	Y	F	L	U	Y	Y	G	V	P	V	R	D	N	C	Z	E	K	J	Q	O
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No 1	U	R	O	G	M	W	C	T	M	Z	Z	V	M	V	A	J
No 2	R	W	J	X	R	V	G	D	K	D	S	X	C	E	E	C
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4

d. A reconstruction matrix of "secondary alphabets" is now made (cf. Fig. 65) by distributing the letters in respective lines corresponding to the 12 different superimposed pairs of numbers. For example, all pairs corresponding to the superimposition of position 1 of Message 1 with position 1 of Message 2 are distributed in lines ∅ and 1 of the matrix. Thus, the very first superimposed pair is $\begin{cases} 1 \\ V \\ Z \\ 1 \end{cases}$; the letter Z is inserted in line 1 under the letter V. The next $\begin{cases} 1 \\ 1 \end{cases}$ pair is the 13th superimposition, with $\begin{cases} F \\ D \end{cases}$; the letter D is inserted in line 1 under the letter F, and so on. The matrix is then as follows:

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

Figure 65.

e. There are more than sufficient data here to permit of the reconstruction of a complete equivalent primary component, for example, the following:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
I	T	K	N	P	Z	H	M	W	B	Q	E	U	L	F	C	S	J	A	X	R	G	D	V	O	Y

f. The subsequent steps in the actual decipherment of the text of either of the two messages are of considerable interest. Thus far the cryptanalyst has only the cipher component of the primary sliding components. The plain component may be identical with the cipher component and may progress in the same direction, or in the reverse direction; or, the two components may be different. If different, the plain component may be the normal sequence, direct or reversed; or it may be a different mixed sequence. Tests must be made to ascertain which of these various possibilities is true.

g. (1) It will first be assumed that the primary plain component is the normal direct sequence. Applying the procedure outlined in par. 33 to the message with the shorter key (Message No. 1, to give the

most data per secondary alphabet), an attempt is made to solve the message. It is unnecessary here to go further into detail in this procedure, suffice it to indicate that the attempt is unsuccessful and it follows that the plain component is not the normal direct sequence. A normal reversed sequence is then assumed for the plain component and the proper procedure applied. Again the attempt is found useless. Next, it is assumed that the plain component is identical with the cipher component, and the procedure outlined in par. 54 is tried. This also is unsuccessful. Another attempt, assuming the plain component runs in the reverse direction, is likewise unsuccessful. There remains one last hypothesis, viz., that the two primary components are different mixed sequences.

(2) Below is given Message No. 1 transcribed in periods of four letters. Uniliteral frequency distributions for the four secondary alphabets are shown below in Fig. 66a, labeled 1a, 2a, 3a, and 4a. These distributions are based upon the normal sequence A to Z. But since the reconstructed cipher component is at hand, these distributions can be rearranged according to the sequence of the cipher component, as shown in distributions labeled 1b, 2b, 3b, and 4b in Fig. 66b. The latter distributions may be combined by shifting distributions 2b, 3b, and 4b to proper superimpositions with respect to 1b so as to yield a single monoalphabetic distribution for the entire message. In other words, the polyalphabetic message can be converted into monoalphabetic terms, thus very considerably simplifying the solution.

Message No. 1

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

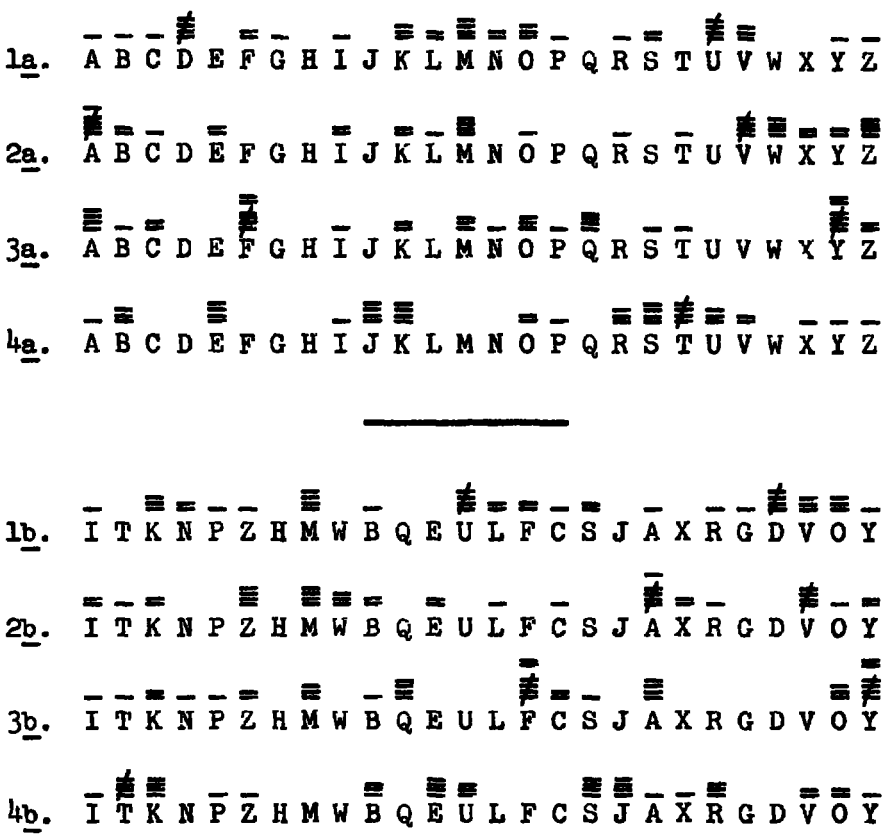


Figure 66a.

(3) Note in Fig. 66b how the four distributions are shifted for superimposition and how the combined distribution presents the characteristics of a typical monoalphabetic distribution.

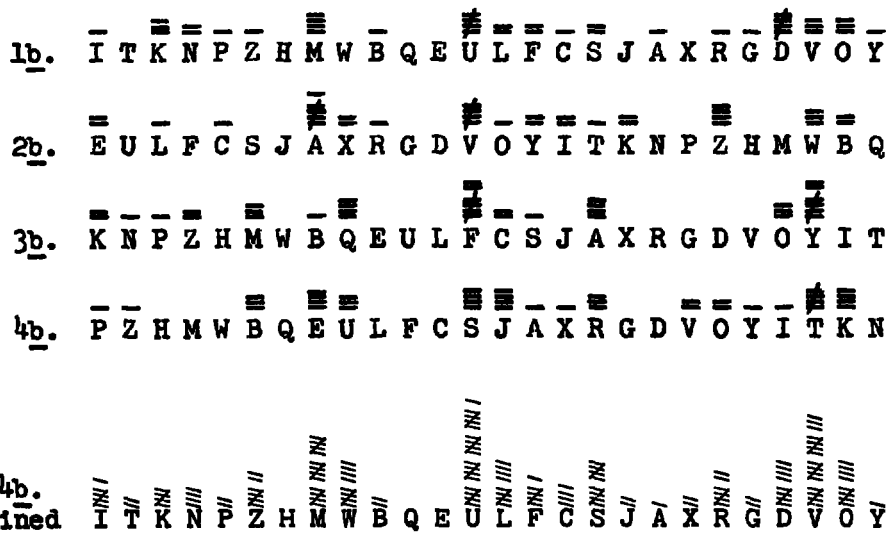


Figure 66b.

(4) The letters belonging to Alphabets 2, 3, and 4 of the message may now be transcribed in terms of Alphabet 1. That is, the two E's of Alphabet 2 become I's; the L of Alphabet 2 becomes a K; the C becomes a P, and so on. Likewise, the two K's of Alphabet 3 become I's, the N becomes a T, and so on. The entire message is then a monoalphabet and can readily be solved. It is as follows:

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

(5) Having the plain text, the derivation of the plain component (an equivalent) is an easy matter. It is merely necessary to base the reconstruction upon any of the secondary alphabets, since the plaintext-ciphertext relationship is now known directly, and the primary cipher component is at hand. The primary plain component is found to be as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
H	M	P	C	B	L	.	R	S	W	.	.	O	D	U	G	A	F	Q	K	I	Y	N	E	T	V

(6) The key words for both messages can now be found, if desirable, by finding the equivalent of A_p in each of the secondary alphabets of the original polyalphabetic messages. The key word for No. 1 is STAR; that for No. 2 is OCEANS.

(7) The student may, if he wishes, try to find out whether the primary components reconstructed above are the original components or are equivalent components, by examining all the possible decimations of the two components for evidence of derivation from key words.

h. As already treated in par. 37, the χ test may be brought to bear in the process of matching distributions to ascertain proper superimpositions for monoalphabeticity. In the case just considered there were sufficient data in the distributions to permit the process to be applied successfully by eye, without necessitating statistical tests. Where, however, the distributions contain relatively few tallies, the use of statistical methods is imperative.

1. This case is an excellent illustration of the application of the process of converting a polyalphabetic cipher into monoalphabetic terms. Because it is a very valuable and important cryptanalytic "trick," the student should study it most carefully in order to gain a good understanding of the principle upon which it is based and its significance in cryptanalysis. The conversion in the case under discussion was possible because the sequence of letters forming the cipher component had been reconstructed and was known, and therefore the uniliteral distributions for the respective secondary cipher alphabets could theoretically be shifted to correct superimpositions for monoalphabeticity. It also happened that there were sufficient data in the distributions to give proper indications for their relative displacements. Therefore, the theoretical possibility in this case became an actuality. Without these two necessary conditions the superimposition and conversion cannot be accomplished. The student should always be on the lookout for situations in which this is possible.

62. Solution of isologs involving different pairs of unknown primary components.--a. If each message of a pair of isologs has been enciphered with a different set of primary components, the repeating keys being of different lengths, there are two procedures available for attacking such a situation. The first procedure involves a modification of the principles demonstrated in par. 61; the second procedure involves an entirely different technique, one which effects a direct conversion of the text to monoalphabetic terms. These two procedures will now be treated in the subparagraphs below.

b. Given the following two cryptograms suspected of containing the same plain text enciphered by different sets of primary components and with key words of different lengths, solve the messages.

Message No. 1

B W X P S O B Y I I U Y H L F K F S O P V G E Y W P B V X O

U G J P B W D X U G H S W D H K H K H C U A Y K P N F S P D

O B B Y B I N K F L W A B O X P J X U V W Q F X R W X Y W S

S D Y Z Q Z H E T A J X X Z W X J R O S P D E E W O J O N K

G I R X R W U Y D K N T J W R E V B U R D L I S J B L C K K

F O D E V D Y Z Q Z S H C T W D I E X Z

Message No. 2

J H L E J M W U A H J H U I V Y N C H C H L P K D E W Z J J

J N A H B H Z B I M T U B Q E F J A K M J V B E F X N C T L

F A A K G K I A B G C V F N Y F W B I Q G E R S A T Z U S D

S X B U D S H A W A Y X L J D C Q L E D H X G Z L Z W H N B

V T J S A T S U U C M I A K K J E M I Y D S K G B V T J Y C

X Y L Z E C X L S U M V M N D O N F J Y

c. Factoring discloses that Message No. 1 has a period of 4, and Message No. 2 a period of 5. The messages are superimposed on a width of the least common multiple (20), and a reconstruction matrix is made, following the method outlined in subpar. 61d. This matrix is shown below:

Values Message No 1

Values Message No 2	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
1-1		J		D			V	T		Y					F				M				H	G		
2-2		A					L	V	T			S					E		U				H	X	
3-3		A	M		P	R			K									J					B	L		
4-4				Q												E			G	N				S	K	J
1-5		G						E		B								A	J				D			
2-1		V		O					K		F				M	E							T	C		
3-2		W						J	N	Q		T		I							S			Z		
4-3			J		F						A							L				Z			U	
1-4				U		B		K	A		Y				E								S	J		
2-5			M						H		C	G			J				D							Y
3-1					X									M		H			S		J		C			
4-2	V			X			N								Y					I				H		
1-3		F		L	G			U		A														B		
2-4					Z						E	I			N	H							K			U
3-5		B				V										F	D	K				E	L	Y		
4-1				C	J						Y			X	Z	F								H			S
1-2				Z		N		H		W												E			X	
2-3		M			A										H				C					B		L
3-4														N	H	T	S			W	I					
4-5	A			L			M				B					C		Y					Q			U

Figure 67.

d. Since the pairs of components for the two messages are different, indirect symmetry will in this case not extend to the ∅ line, so all chaining must be done within the matrix. Apparent conflicts in the matrix are noted, such as the A's in lines 2-2 and 3-3, the rest of the letters in these lines not being identical as might at first be expected. However, if we restrict our treatment only to the homogeneous lines 1-1, 1-2, 1-3, 1-4, and 1-5 (see the matrix in Fig. 68, below), we will have data which may be interrelated

and which will produce an equivalent primary component, either from these data alone or facilitated by another family of related rows of

	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
1-1	J			D			V	T		Y					F				M				H	G		
1-2				Z		N		H		W													E			X
1-3	F			L	G			U		A																B
1-4				U		B		K	A		Y				E									S	J	
1-5	G							E		B								A	J				D			

Figure 68.

the matrix, such as 2-1, 2-2, ... 2-5. The equivalent primary component recovered will be that for Message No. 2, since it is the values for this message which are entered within the matrix and which are manipulated. By inverting the matrix so that the values for Message No. 1 are written within the matrix, a similar procedure will yield an equivalent primary component for the first message.

e. An entirely different technique for treating these isologs will now be described. We have factored the two messages as periods of 4 and 5; and now we write out the messages on the width of the least common multiple, retaining indications of the alphabets to which the cipher letters belong, thus:

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	B	W	X	P	S	O	B	Y	I	I	U	Y	H	L	F	K	F	S	O	P
No. 2	J	H	L	E	J	M	W	U	A	H	J	H	U	I	V	Y	N	C	H	C
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	V	G	E	Y	W	P	B	V	X	O	U	G	J	P	B	W	D	X	U	G
No. 2	H	L	P	K	D	E	W	Z	J	J	J	N	A	H	B	H	Z	B	I	M
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	H	S	W	D	H	K	H	K	H	C	U	A	Y	K	P	N	F	S	P	D
No. 2	T	U	B	Q	E	F	J	A	K	M	J	V	B	E	F	X	N	C	T	L
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	O	B	B	Y	B	I	N	K	F	L	W	A	B	O	X	P	J	X	U	V
No. 2	F	A	A	K	G	K	I	A	B	G	C	V	F	N	Y	F	W	B	I	Q
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	W	Q	F	X	R	W	X	Y	W	S	S	D	Y	Z	Q	Z	H	E	T	A
No. 2	G	E	R	S	A	T	Z	U	S	D	S	X	B	U	D	S	H	A	W	A
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	J	X	X	Z	W	X	J	R	O	S	P	D	E	E	W	O	J	O	N	K
No. 2	Y	X	L	J	D	C	Q	L	E	D	H	X	G	Z	L	Z	W	H	N	B
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	G	I	R	X	R	W	U	Y	D	K	N	T	J	W	R	E	V	B	U	R
No. 2	V	T	J	S	A	T	S	U	U	C	M	I	A	K	K	J	E	M	I	Y
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	D	L	I	S	J	B	L	C	K	K	F	O	D	E	V	D	Y	Z	Q	Z
No. 2	D	S	K	G	B	V	T	J	Y	C	X	Y	L	Z	E	C	X	L	S	U
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

	1	2	3	4	1	2	3	4	1	2
No. 1	S	H	C	T	W	D	I	E	X	Z
No. 2	M	V	M	N	D	O	N	F	J	Y
	1	2	3	4	5	1	2	3	4	5

f. Let us arbitrarily assign the value of A_p to the first letter of the plain text. Since then, in Message No. 1, $B_c = A_p$ of Alphabet 1, every B_c in Alphabet 1 must equal A_p; these values are entered on the

work sheet. Now since the 65th and 73d cipher letters of Message No. 1 are A_p, this establishes that the 65th and 73d letters of Message No. 2 ($^{5}G_c$ and ^{3}Fc) are also A_p; therefore, these latter values are entered throughout the work sheet where they occur. Similarly, since every J_c of Alphabet 1 in Message No. 2 equals A_p, this value is entered on the work sheet under every occurrence of J_c. By continuing this process, all the A_p's of the pseudo-plain text shall have been recovered, and the work sheet will now look as shown in Fig. 69, below:

```
        1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1   B W X P S O B Y I I U Y H L F K F S O P
No. 2   J H L E J M W U A H J H U I V Y N C H C
        1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
        A                   A     A
        1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1   V G E Y W P B V X O U G J P B W D X U G
No. 2   H L P K D E W Z J J J N A H B H Z B I M
        1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
                            A                 A
        1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1   H S W D H K H K H C U A Y K P N F S P D
No. 2   T U B Q E F J A K M J V B E F X N C T L
        1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
                            A
        1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1   O B B Y B I N K F L W A B O X P J X U V
No. 2   F A A K G K I A B G C V F N Y F W B I Q
        1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
              A           A     A           A
        1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1   W Q F X R W X Y W S S D Y Z Q Z H E T A
No. 2   G E R S A T Z U S D S X B U D S H A W A
        1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5

        1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1   J X X Z W X J R O S P D E E W O J O N K
No. 2   Y X L J D C Q L E D H X G Z L Z W H N B
        1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5

        1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1   G I R X R W U Y D K N T J W R E V B U R
No. 2   V T J S A T S U U C M I A K K J E M I Y
        1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
                    A                 A     A
        1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1   D L I S J B L C K K F O D E V D Y Z Q Z
No. 2   D S K G B V T J Y C X Y L Z E C X L S U
        1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
          A
        1 2 3 4 1 2 3 4 1 2
No. 1   S H C T W D I E X Z
No. 2   M V M N D O N F J Y
        1 2 3 4 5 1 2 3 4 5
                    A
```

Figure 69.

g. We will now arbitrarily assign the value B_p to the V_c at the 21st position of Message No. 1; the other V_c of Message No. 1 establishes the $\overset{2}{E}_c$ of Message No. 2 also as B_p. This procedure is continued, until all the B_p's in the pseudo-plain text are recovered. Continuing in this vein, assigning arbitrary plaintext values to all the cipher letters of Alphabet 1 of Message No. 1, we are able to reduce almost the entire text[3] to monoalphabetic terms. The work sheet will now look as follows:

	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	B	W	X	P	S	O	B	Y	I	I	U	Y	H	L	F	K	F	S	O	P
No. 2	J	H	L	E	J	M	W	U	A	H	J	H	U	I	V	Y	N	C	H	C
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	A	C	H	D	I	I	F	C	K		A	C	C	A		F	M	E		D
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	V	G	E	Y	W	P	B	V	X	O	U	G	J	P	B	W	D	X	U	G
No. 2	H	L	P	K	D	E	W	Z	J	J	J	N	A	H	B	H	Z	B	I	M
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	B			C	E		F		L	I	A	M	F		F	B	H	O	A	M
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	H	S	W	D	H	K	H	K	H	C	U	A	Y	K	P	N	F	S	P	D
No. 2	T	U	B	Q	E	F	J	A	K	M	J	V	B	E	F	X	N	C	T	L
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	C	E	O	O	C	D		F	C	M	A	J	O	D	B		M	E	B	O
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	O	B	B	Y	B	I	N	K	F	L	W	A	B	O	X	P	J	X	U	V
No. 2	F	A	A	K	G	K	I	A	B	G	C	V	F	N	Y	F	W	B	I	Q
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	D	G	F	C	A		I	F	M	A	O	J	A	I	H	D	F	O	A	
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	W	Q	F	X	R	W	X	Y	W	S	S	D	Y	Z	Q	Z	H	E	T	A
No. 2	G	E	R	S	A	T	Z	U	S	D	S	X	B	U	D	S	H	A	W	A
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	E	B		E	J	C	H	C	E	E	L	O	O	H	E	L	C	F		J
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	J	X	X	Z	W	X	J	R	O	S	P	D	E	E	W	O	J	O	N	K
No. 2	Y	X	L	J	D	C	Q	L	E	D	H	X	G	Z	L	Z	W	H	N	B
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	F	O	H	L	E	O		H	D	E	B	O	P	F	O		F	I	I	F
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	G	I	R	X	R	W	U	Y	D	K	N	T	J	W	R	E	V	B	U	R
No. 2	V	T	J	S	A	T	S	U	U	C	M	I	A	K	K	J	E	M	I	Y
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	G			E	J	C	A	C	H	D	I	I	F	C		A	B	G	A	H
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
No. 1	D	L	I	S	J	B	L	C	K	K	F	O	D	E	V	D	Y	Z	Q	Z
No. 2	D	S	K	G	B	V	T	J	Y	C	X	Y	L	Z	E	C	X	L	S	U
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	H	A	M		F	G			N	D			H	F	C	O	O	H	E	L
	1	2	3	4	1	2	3	4	1	2										
No. 1	S	H	C	T	W	D	I	E	X	Z										
No. 2	M	V	M	N	D	O	N	F	J	Y										
	1	2	3	4	5	1	2	3	4	5										
	I	J	G	I	E		M	A	L	H										

[3] Actually in this particular case the reduction is 85% complete

Note the idiomorphic repetition (representing the word ARTILLERY), previously latent, which now becomes patent in the reduction process.

h. At this point, sequence reconstruction matrices may be made of the two messages, the ∅ line representing the pseudo-plain text and the values inside the matrix being the cipher text. These matrices are illustrated in Figs. 70a and b, below:

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

Figure 70a.

∅	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
1	J	H	T	F	G	Y	V	D	M			S			C											
2	S	E	H		U	W	A	Z	I	V			N		X											
3	F		U		C	A	M	L	H				K		B	G										
4	I	T	K	E	S	Z		U	N		A	J	B	Y	Q											
5	G	F	E	C	D	B		Y	J	A		U	M		L											

Figure 70b.

From these matrices, it is a simple matter to chain out the equivalent primary cipher components used for each message. Having reconstructed the cipher component for a message, the alphabets may be aligned and now the entire text converted to monoalphabetic terms. After solution of the messages, it is found that Message No. 1 is a case of direct symmetry with the cipher component being based on the key word HYDRAULIC, and Message No. 2 is a case of indirect symmetry with both components being keyword-mixed sequences based on QUESTIONABLY.

i. The method described in subpars. 62e to h, above, involves techniques which have a broad application in cryptanalytics, in other fields besides the solution of periodic polyalphabetic cryptograms. But even in this latter field, it is the only approach to solution where the cryptosystem involves non-related, random-mixed secondary alphabets among which no symmetry of any sort exists.

j. The two messages used in the example had periods prime to each other. If this had not been the case, only a slight modification of either of the two methods would have been necessary. For the student who cares to investigate this matter further, there is given in Fig. 71, below, a new Message No. 2 to be paired with Message No. 1 in subpar. 62b; he can then solve this pair of isologs by either of the two methods demonstrated in this paragraph.

Message No. 2

T Q L G C F X Q J M L Z J D R C F L G L D Z H Z E X A S V F

T J A S T Y Q A F V W L Z A U G X O J J F Y I B S P K H B N

U K A I L D O N K K I I T G L G T N T X C J J L W Q L I A L

H A B F A X J N O Y T N Q Z C D N A U Y D D Y O Z R A P C O

B F J H S Z T Q L G C F X Q J K B J T B L K F V X K G R H B

Y R L C W N Z B C T C I B G C N F C H B

Figure 71.

63. Solution of a pair of periodic cryptograms involving a "stagger". --a. It happens occasionally that the cryptanalyst has two messages with identical beginnings, but after a few letters the cipher texts diverge; and the group counts of the two messages are either identical or nearly identical. This situation could arise in a pair of isologs, when the first message has a letter omitted (or added) at the point of divergence, and the second message (with the identical beginning) has this error corrected. In other words, we have a pair of true isologs except for the deletion (or addition) of a single letter. Such a situation is called a "stagger".[4] By treating the isologous portions of the two messages, we may recover the primary cipher component by the process of indirect symmetry. This is best illustrated by an example.

b. Let us suppose the following two messages are at hand:

Message "A"

K O I P Q I H G I S P Q O K D M P V K S Y K E Q V S K P U S

E P P S F K P E E E P Y V X B P I S W Y E T D Q S P I M X K

H G F T J G O G J T X I E Q E H P G C G O E O B E Y E T E W

J E E E P U U M D K A Q V O L M B

Message "B"

K O I P Q I H G I S P Q O K P S T I H N E N H Q P Q E Q I M

C U D T B X R V M F Q I E Q L R I Y C C F F O W P G D O T G

R P U V O C O S U G O G G U T F L I O O U U Q N K U F F O Q

U T M F Q I B Y A S H S P Q B C T V

[4] Such errors are more prone to happen in machine cipher systems when the addition or deletion of a word separator in a pair of otherwise identical cryptograms (enciphered with identical initial settings or keys) causes a "stagger" Sometimes the stagger may be progressive, i e , the interval of the displacement becoming greater and greater as additional word separators are omitted

We note the identical 14-letter beginnings; we also note that Message "B" is one letter longer than Message "A", and that the trigraphic and tetragraphic repetitions in Message "B" are spaced identically as in Message "A", except for the fact that their position in the cipher text is one letter more than their counterparts in Message "A". Both messages factor to 6 alphabets.

c. These phenomena clearly point to a case of a stagger, with Message "B" containing one more letter than Message "A". If this is the case, and the plain texts are otherwise identical, then the P_C in the 15th position of Message "B" seems to be the extra letter, coming as it does after the identical beginning. The two texts are now superimposed, and the equivalencies are inserted into a sequence reconstruction matrix, as is shown below:

```
          3456123456123456123456123456123456123456123456123456
"A"       DMPVKSYKEQVSKPUSEPPSFKPEEEPYVXBPISWYETDQSPIMXKHG
"B"       STIHNENHQPQEQIMCUDTBXRVMFQIEQLRIYCCFFOWPGDOTGRPU
          4561234561234561234561234561234561234561234561234

          3456123456123456123456123456123456123456123456123456
"A"       FTJGOGJTXIEQEHPGCGQEOBEYETEWJEEEPUUMDKAQVOLMB
"B"       VOCOSUGOGGUTFLIOOUUQNKUFFOQUTMFQIBYASHSPQBCTV
          4561234561234561234561234561234561234561234561234
```

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

The solution, which proceeds in the usual manner, is left as an exercise for the interested student.[5]

[5] Note, in matrices of this kind, how easy it is to align properly the cipher components after the primary cipher component (or an equivalent) has been recovered, thereby expediting the reduction of the cipher text to monoalphabetic terms. From the data in the sequence reconstruction matrix, it is observed that the U_C of Alphabet 2 is under the E_C of Alphabet 1, the M_C of Alphabet 3 is under the E_C of Alphabet 2, the F_C of Alphabet 4 is under the E_C of Alphabet 3, etc., thus all the cipher components may be quickly put together at their proper relative displacements, revealing the repeating key in the process in one of the columns.

64. Solution of a periodic cryptogram containing a long latent repetition.--a. It sometimes happens that a periodic cryptogram contains a long passage repeated in its plain text, the second occurrence of which is enciphered at a cyclical offset from the first occurrence. The recognition and delineation of the latent repetition may be made possible by the spatial relationships of ciphertext repetitions present within the message.[6] If such a latent repetition is found, and it is long enough, the equivalencies from the two corresponding sequences may be chained together to yield an equivalent primary cipher component, and thus by-pass a more laborious process of solution by the usual method of frequency analysis or making assumptions in the plain text of a polyalphabetic cipher.

b. As an example, let us suppose the following message has been intercepted:

A S E X Z L R C W C H X R U G L X A X W I N Y E H P J R D X

R D X W M L A I E K K T A G Z F Z N C E L L T F W C R B J T

Z L C V H M A A J R D X R T F M T V J H P G V P I M W Y R L

R R B J T Z L C V H D K J T A I B L F P A W N Y E H P E M G

F I R V A R X V J F J A H D F I V V T N A E Z A Z J X V L Y

P J T N Q A B K X D J X H A X Y W P I M I C G A N I W E F G

W M I Z J H V I X V L Y P G A Y X Z M E K L I S B O T F O M

V W E F G W M I Z J D Z A C G J V M P V N D K G K V M A I B

L F P A W N Y E H P

An examination of the cipher text, which factors to a period of 7, reveals the following striking sets of repetitions with identical spatial relationships of the repetitions (beginning at positions 27 and 147 in the cipher text) in the two sets:

Set "A": JRDXR..(25)..RBJTZLCVH..(3)..JRDXR..(18)..RBJTZLCVH
Set "B": XVLYP..(25)..WEFGWMIZJ..(3)..XVLYP..(18)..WEFGWMIZJ

This phenomenon could arise from a repetition of a long section of plain text within the message. The presence of the repetitions at the beginning

[6] It is also possible that, because of highly stereotyped and redundant plain text of a particular message, there might be two sets of long polygraphic repetitions of the same length in the cipher text. If these sets actually represent the same plain text, they may be chained together as described below to derive a partial or even complete equivalent primary cipher component.

and end of these sets of sequences delineates the limits, insofar discernible, of the repeated plaintext passage. Since these two sets are equivalent to each other, it is possible to superimpose these sections and distribute the equivalencies into a sequence reconstruction matrix, from which an equivalent primary component may be chained, following a procedure similar to that outlined in subpar. 63c. The completion of the solution of this problem, too, is left as an exercise for the student.

65. Solution by superimposition.--a. In solving an ordinary repeating-key cipher the first step, that of ascertaining the length of the period, is of no significance in itself. It merely paves the way for and makes possible the second step, which consists in allocating the letters of the cryptogram into individual monoalphabetic distributions. The third step then consists in solving these distributions. Usually, the text of the message is transcribed into its periods and is written out in successive lines corresponding in length with that of the period. The diagram then consists of a series of columns of letters, and the letters in each column belong to the same monoalphabet. Another way of looking at the matter is to conceive of the text as having thus been transcribed into superimposed periods; in such a case the letters in each column have undergone the same kind of treatment by the same elements (plain and cipher components of the cipher alphabet).[7]

b. Suppose, however, that the repetitive key is very long and that the message is short, so that there are only a very few, if any, complete cycles in the text. Then the solution of the message becomes difficult, if not impossible (unless the alphabets are known), because there is not a sufficient number of superimposable periods to yield monoalphabetic distributions which can be solved by frequency principles. But suppose also that there are many short cryptograms all enciphered by the same key, each message beginning at identical starting points in the key. Then it is clear that if these messages are superimposed "head on" or "in flush depth", (1) the letters in the respective columns will all belong to individual alphabets, and (2) if there is a sufficient number of such superimposable messages (say 25-30, for English), then the frequency distributions applicable to the successive columns of text can be solved--without knowing the length of the key.[8] In other words, any difficulties that may have arisen on account of failure or inability to ascertain the length of the period have been circumvented. The second step in normal solution is thus by-passed.

c. Furthermore, and this is a very important point, even if an extremely long key is employed and a series of messages beginning at different initial points are enciphered by such a key, this method of solution by superimposition can be employed, provided the messages can be superimposed correctly, that is, so that the letters which fall in one column really belong to one cipher alphabet. Just how this can be done will be treated in Chapters IX and XIV.

[7] In operational parlance, the superimposed periods are said to be "in depth".

[8] The assumption of probable initial words of messages and stereotyped beginnings is a powerful method of attack in such situations.

66. Additional remarks.--a. We have seen in pars. 60-64 that the chaining process between cipher texts applies to the latent characteristics of the cipher components, regardless of the identity of the plain components and regardless whether direct or indirect symmetry is involved in the cryptosystems.

b. The observant student will have noted that a large part of the text thus far is devoted to the elucidation and application of a very few basic principles. These principles are, however, extremely important and their proper usage in the hands of a skilled cryptanalyst makes them practically indispensable tools of his art. The student should therefore drill himself in the application of these tools by practicing upon problem after problem, until he acquires facility in their use and feels competent to apply them in practice whenever the least opportunity presents itself. This will save him much time and effort in the solution of bona fide messages.

CHAPTER IX

PROGRESSIVE ALPHABET SYSTEMS

67. Preliminary remarks.--a. In progressive alphabet systems the basic principle is quite simple. Two primary components are arranged or provided for according to a key which may be varied from time to time; the interaction of the primary components results in making available for cryptographic purposes a set of cipher alphabets; all the latter are employed in a fixed sequence or progression; hence the designation progressive alphabet system. Since the number of alphabets available for such use is rather small (usually 26), if the text to be enciphered is much longer than the sequence of alphabets, then the system reduces to a periodic method. But if the number of alphabets is large as compared with the text to be enciphered,[1] so that the sequence of alphabets is not repeated, then, of course, the cryptographic text will exhibit no periodic phenomena.

b. The series of cipher alphabets in such a system constitutes a keying sequence. Once set up, often the only remaining element in the key for a specific message is the starting point in the sequence, that is, the initial cipher alphabet employed in enciphering a given message. If this keying sequence must be employed by a large group of correspondents, and if all messages employ the same starting point in the keying sequence, obviously the cryptograms may simply be superimposed without any preliminary testing to ascertain proper points for superimposition. It has already been indicated (cf. par. 65) how cases of this sort may be solved. However, if messages are enciphered with varying starting points, the matter of superimposing them properly takes on a different aspect. This matter will be treated in par. 72.

68. Solution of a progressive alphabet cipher when the cipher alphabets are known.--a. The simplest case of a progressive alphabet system involves two interacting primary components which slide against each other to produce a set of 26 secondary alphabets, which are employed one after the other consecutively in the simplest type of progression. Beginning at

[1] For instance if the cipher component of a disc cipher device were composed of 100 dinomes and no message were longer than 100 letters, no periodicity would be manifested

an initial juxtaposition, producing say, Alphabet 1, the subsequent secondary alphabets are in the sequence 2, 3, ... 26, 1, 2, 3, ..., and so on. If a different initial juxtaposition is used, say Alphabet 10 is the first one, the sequence is exactly the same as before, only beginning at a different point.

b. Suppose that the two primary components are based upon the key word HYDRAULIC. A message is to be enciphered, beginning with Alphabet 1. Thus:

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

Letter No:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Alphabet No:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Plain text:	E	N	E	M	Y	H	A	S	P	L	A	C	E	D	H	E	A	V	Y	I
Cipher text:	E	O	G	P	U	U	E	Y	H	M	K	Q	V	M	K	Z	S	J	Q	H

Letter No:	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Alphabet No:	21	22	23	24	25	26	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Plain text:	N	T	E	R	D	I	C	T	I	O	N	F	I	R	E	U	P	O	N	Z
Cipher text:	E	N	L	H	H	L	C	V	B	S	S	N	J	E	P	K	D	D	D	G

Letter No:	41	42	43	44	45	46	47	48	49	50	51	52	53
Alphabet No:	15	16	17	18	19	20	21	22	23	24	25	26	1
Plain text:	A	N	E	S	V	I	L	L	E	R	O	A	D
Cipher text:	P	U	H	F	K	H	H	Y	L	H	M	R	D

c. This method reduces to a periodic system involving 26 secondary cipher alphabets and the latter are used in simple progression. It is obvious therefore that the 1st, 27th, 53d, ... letters are in the 1st alphabet; the 2d, 28th, 54th, ... letters are in the 2d alphabet, and so on.

d. To solve such a cryptogram, knowing the two primary components, is hardly a problem at all. The only element lacking is a knowledge of the starting point. But this is not necessary, for merely by completing the plain-component sequence and examining the diagonals of the diagram, the plain text becomes evident. For example, let us consider that the first two groups of an intercepted message are HIDCT EHUXI..., and let us assume that the components are keyword-mixed sequences based upon HYDRAULIC. Completing the plain-component sequences inititated by the successive cipher letters, the plain text ENEMY MACHI ... is seen to come out in successive steps upward in Fig. 72. Had the cipher component been shifted in the opposite direction in encipherment, the steps would have been downward instead of upward. If the sliding strips had been set up according to the sequence of cipher letters but on a diagonal, then, of course, the plaintext letters would have reappeared on one generatrix.

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

Figure 72.

e. If the components were two different known mixed sequences, it would of course first be necessary to convert the cipher letters into their plain-component equivalents before completing the plain-component sequences. In any case, faced with an unknown type of progressive alphabet cipher, completing the plain-component sequences on the hypothesis of direct or reversed standard alphabets is the logical thing to do, and will quickly prove or disprove these hypotheses. If the primary components are not known sequences, the methods given in this paragraph obviously cannot apply; fortunately, however, there exist several methods which can be used in such situations, as will be treated in the succeeding paragraphs.

69. Solution by a method involving the χ test.--a. An interesting general solution of a statistical nature of a progressive alphabet system will now be discussed. The problem involves secondary alphabets derived from the interaction of two identical mixed primary components. It will be assumed that the enemy has been using a system of this kind and that the primary components are changed daily.

b. Before attacking an actual problem of this type, suppose a few minutes be devoted to a general analysis of its elements. It is here assumed that the primary components are based upon the HYDRAULIC....Z sequence and that the cipher component is shifted toward the right one step at a time. Consider a cipher square such as that shown in Fig. 73, which is applicable to the type of problem under study. It has been arranged in the form of a deciphering square. In this square, the horizontal sequences are all identical but merely shifted relatively; the letters inside the square are plaintext letters.

Alphabet No.

Cipher Letters	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R
B	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C
C	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I
D	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y
E	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B
F	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E
G	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F
H	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z
I	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L
J	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G
K	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J
L	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U
M	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K
N	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M
O	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N
P	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O
Q	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P
R	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D
S	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q
T	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S
U	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A
V	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T
W	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V
X	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W
Y	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H
Z	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X

(Plaintext letters are within the square proper)

Figure 73.

c. If, for mere purposes of demonstration, instead of letters within the cells of the square there are placed tallies corresponding in number with the normal frequencies of the letters occupying the respective cells, the cipher square becomes as follows (showing only the first three rows of the square):

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	8	2	3	5	2	1	13	2	1			2	8	9	2		6	10	2	2			2	1	3	13
B	1	9	2	1			1	5	7	2		4	9	1	1			2	1	3	5	4	5	3	2	1
C	2	1	7	2	1			1	5	4	2		4	5	1	1			2	1	2	5	4	3	2	4

Figure 74a.

d. It is obvious that here is a case wherein if two distributions pertaining to the square are isolated from the square, the χ test can be applied to ascertain how the distributions should be shifted relative to each other so that they can be superimposed and made to yield a monoalphabetic composite. There is clearly only one correct superimposition out of 25 possibilities. In this case, the "B" row of tallies must be displaced 5 intervals to the right in order to match it and amalgamate it with the "A" row of tallies. Thus:

A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
B	22	23	24	25	26	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21

Figure 74b.

e. Note that the amount of displacement, that is, the number of intervals the "B" sequence must be shifted to make it match the "A" sequence in Fig. 74b, corresponds exactly to the distance between the letters A and B in the primary cipher component, which is 5 intervals. Thus: ... $\overset{0}{A}\ \overset{1}{U}\ \overset{2}{L}\ \overset{3}{I}\ \overset{4}{C}\ \overset{5}{B}$ ... . The fact that the primary plain component is in this case identical with the primary cipher component has nothing to do with the matter. The displacement interval is being measured on the cipher component. It is important that the student see this point very clearly. He can, if he like, prove the point by experimenting with two different primary components.

f. Assuming that a message in such a system is to be solved, the text is transcribed in rows of 26 letters. A uniliteral frequency distribution is made for each column of the transcribed text, the 26 separate distributions being compiled within a single square such as that shown in Fig. 75. Such a square may be termed a frequency distribution square.

g. Now the vertical columns of tallies within such a distribution square constitute frequency distributions of the usual type: They show the distribution of the various cipher letters in each cipher alphabet. If there were many lines of text, all arranged in periods of 26 letters, then each column of the frequency square could be solved in the usual manner, by the application of the simple principles of monoalphabetic frequency. But what do the horizontal rows of tallies within the square represent? Is it not clear that the first such row, the "A" row, merely shows the distribution of A_c throughout the successive cipher alphabets? And does not this graphic picture of the distribution of A_c correspond to the sequence of letters composing the primary plain component? Furthermore, is it not clear that what has been said of the "A" row of tallies applies equally to the B, C, D, ... Z rows? Finally, is it not clear that the graphic pictures of all the

distributions correspond to the same sequence of letters, except that the sequence begins with a different letter in each row? In other words, all the horizontal rows of tallies within the distribution square apply to the same sequence of plaintext letters, the sequences in one row merely beginning with a different letter from that with which another row begins. The sequences of letters to which the tallies apply in the various rows are merely displaced relative to one another. Now if there are sufficient data for statistical purposes in the various horizontal sequences of tallies within the distribution square, these sequences, being approximately similar, can be studied by means of the χ test to find their relative displacements. And in finding the latter, a method is provided whereby the primary cipher component may be reconstructed, since the correct assembling of the displacement data will yield the sequence of letters constituting the primary cipher component. If the plain component is identical with the cipher component, the solution is immediately at hand; if the components are different, the solution is but one step removed. Thus, there has been elaborated a method of solving this type of cipher system without making any assumptions of values for cipher letters.

h. We will now take up a typical problem, employing the procedures just discussed. The following cryptogram has been enciphered according to the method indicated, by progressive, simple, uninterrupted shifting of a primary cipher component against an identical primary plain component.

~~CONFIDENTIAL~~

Cryptogram

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

1. The message is transcribed in lines of 26 letters, since that is the total number of secondary alphabets in the system. The transcribed text is shown below:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	W	G	J	J	M	M	M	J	X	E	D	G	C	O	C	F	T	R	P	B	M	I	I	I	K	Z
2	R	Y	N	N	B	U	F	R	W	W	W	W	Y	O	I	H	F	J	K	O	K	H	T	T	A	Z
3	C	L	J	E	P	P	F	R	W	C	K	O	O	F	F	F	G	E	P	Q	R	Y	Y	I	W	X
4	M	X	U	D	I	P	F	E	X	M	L	L	W	F	K	G	Y	P	B	B	X	C	H	B	F	Y
5	I	E	T	X	H	F	B	I	V	D	I	P	N	X	I	V	R	P	W	T	M	G	I	M	P	T
6	E	C	J	B	O	K	V	B	U	Q	G	V	G	F	F	F	K	L	Y	Y	C	K	B	I	W	X
7	M	X	U	D	I	P	F	F	U	Y	N	V	S	S	I	H	R	M	H	Y	Z	H	A	U	Q	W
8	G	K	T	I	U	X	Y	J	J	A	O	W	Z	O	C	F	T	R	P	P	O	Q	U	S	G	Y
9	C	X	V	C	X	U	C	J	L	M	L	L	Y	E	K	F	F	Z	V	Q	J	Q	S	I	Y	S
10	P	D	S	B	B	J	U	A	H	Y	N	W	L	O	C	X	S	D	Q	V	C	Y	V	S	I	L
11	I	W	N	J	O	O	M	A	Q	S	L	W	Y	J	G	T	V	P	Q	K	P	K	T	L	H	S
12	R	O	O	N	I	C	F	E	V	M	N	V	W	N	B	N	E	H	A	M	R	C	R	O	V	S
13	T	X	E	N	H	P	V	B	T	W	K	U	Q	I	O	C	A	V	W	B	R	Q	N	F	J	V
14	N	R	V	D	O	P	U	Q	R	L	K	Q	N	F	F	F	Z	P	H	U	R	V	W	L	X	G
15	S	H	Q	W	H	P	J	B	C	N	N	J	Q	S	O	Q	O	R	C	B	M	R	R	A	O	N
16	R	K	W	U	H	Y	Y	C	I	W	D	G	S	J	C	T	G	P	G	R	M	I	Q	M	P	S
17	G	C	T	N	M	F	G	J	X	E	D	G	C	O	P	T	G	P	W	Q	Q	V	Q	I	W	X
18	T	T	T	C	O	J	V	A	A	A	B	W	M	X	I	H	O	W	H	D	E	Q	U	A	I	N
19	F	K	F	W	H	P	J	A	H	Z	I	T	W	Z	K	F	E	X	S	R	U	Y	Q	I	O	V
20	R	E	R	D	J	V	D	K	H	I	R	Q	W	E	D	G	E	B	Y	B	M	L	A	B	J	V
21	T	G	F	F	G	X	Y	I	V	G	R	J	Y	E	K	F	B	E	P	B	J	O	U	A	H	C
22	U	G	Z	L	X	I	A	J	K	W	D	V	T	Y	B	F	R	U	C	C	C	U	Z	Z	I	N
23	N	D	F	R	J	F	M	B	H	Q	L	X	H	M	H	Q	Y	Y	Y	M	W	Q	V	C	L	I
24	P	T	W	T	J	Y	Q	B	Y	R	L	I	T	U	O	U	S	R	C	D	C	V	W	D	G	I
25	G	G	U	B	H	J	V	V	P	W	A	B	U	J	K	N	F	P	P	F	Y	W	V	Q	Z	Q	F
26	L	H	T	W	J	P	D	R	X	Z	O	W	U	S	S	G	A	M	H	N	C	W	H	S	W	W
27	L	R	Y	Q	Q	U	S	Z	V	D	N	X	A	N	V	N	K	H	F	U	C	V	V	S	S	S
28	P	L	Q	U	P	C	V	V	V	W	D	G	S	J	O	G	T	C	H	D	E	V	Q	S	I	J
29	P	H	Q	J	A	W	F	R	I	Z	D	W	X	X	H	C	X	Y	C	T	M	G	U	S	E	S
30	N	D	S	B	B	K	R	L	V	W	R	V	Z	E	E	P	P	P	A	T	O	I	A	N	E	E
31	E	E	J	N	R	C	Z	B	T	B	L	X	P	J	J	K	A	P	P	M	J	E	G	I	K	R
32	T	G	F	F	H	P	V	V	V	Y	K	J	E	F	H	Q	S	X	J	Q	D	Y	V	Z	G	R
33	R	H	Z	Q	L	Y	X	K	X	A	Z	O	W	R	R	X	Y	K	Y	G	M	G	Z	B	Y	N
34	V	H	Q	B	R	V	F	E	F	Q	L	L	W	Z	E	Y	L	J	E	R	O	Q	S	O	Q	K
35	O	M	W	I	O	G	M	B	K	F	F	L	X	D	X	T	L	W	I	L	P	Q	S	E	D	Y
36	I	O	E	M	O	I	B	J	M	L	N	N	S	Y	K	X	J	Z	J	M	L	C	Z	B	M	S
37	D	J	W	Q	X	T	J	V	L	F	I	R	N	R	X	H	Y	B	D	B	J	U	F	I	R	J
38	I	C	T	U	U	U	S	K	K	W	D	V	M	F	W	T	T	J	K	C	K	C	G	C	V	S
39	A	G	Q	B	C	J	M	E	B	Y	N	V	S	S	J	K	S	D	C	B	D	Y	F	P	P	V
40	F	D	W	Z	M	T	B	P	V	T	T	C	G	B	V	T	Z	K	H	Q	D	D	R	M	E	Z
41	O	O																								

1. A frequency distribution square is then compiled, each column of the text forming a separate distribution in columnar form in the square. The latter is shown in Fig. 75.

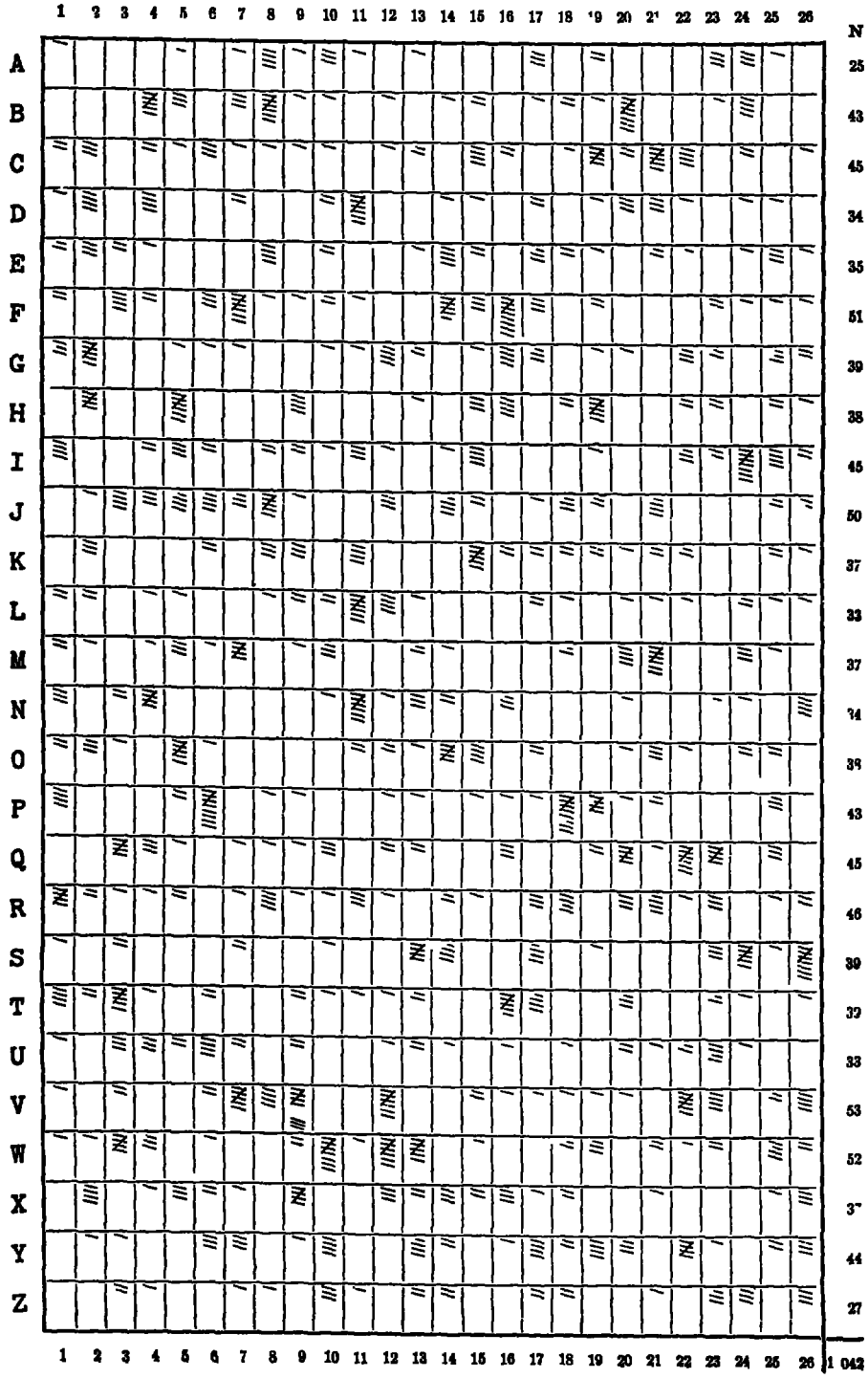


Figure 75.

k. The χ test will now be applied to the horizontal rows of tallies in the distribution square, in accordance with the theory set forth in subpar. 69g. Since this test is purely statistical in character and becomes increasingly reliable as the size of the distributions increases, it is best to start by working with the two distributions having the greatest total numbers of tallies. These are the V and W distributions, with 53 and 52 occurrences, respectively. The results of three of the 25 possible relative displacements of these two distributions are shown below, labeled "First test," "Second test," and "Third test." For convenience in estimating the matching propensities, the χ value is expressed in terms of the ξI.C.[2]

First test

f_V	1	0	2	0	0	2	6	4	8	0	0	7	0	0	2	1	1	1	1	1	0	6	4	0	2	4
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
f_W	24	25	26	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
	0	4	2	1	1	5	3	0	1	0	0	2	8	1	7	6	0	1	0	0	2	3	0	2	1	2
$f_V f_W$	0	0	4	0	0	10	18	0	8	0	0	14	0	0	14	6	0	1	0	0	0	18	0	0	2	8

$N_V = 53$; $N_W = 52$; $\Sigma f_V f_W = 103$

$$\chi_o = 103 \qquad \chi_r = \frac{53\ 52}{26} = 106 \qquad \xi \text{I C} = \frac{103}{106} = 0\ 97$$

Second test

f_V	1	0	2	0	0	2	6	4	8	0	0	7	0	0	2	1	1	1	1	1	0	6	4	0	2	4
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
f_W	18	19	20	21	22	23	24	25	26	1	2	3	4	5	6	6	8	9	10	11	12	13	14	15	16	17
	2	3	0	2	1	2	0	4	2	1	1	5	3	0	1	0	0	2	8	1	7	6	0	1	0	0
$f_V f_W$	2	0	0	0	0	4	0	16	16	0	0	35	0	0	2	0	0	2	8	1	0	36	0	0	0	0

$N_V = 53$; $N_W = 52$; $\Sigma f_V f_W = 122$

$$\chi_o = 122 \qquad \chi_r = \frac{53\ 52}{26} = 106 \qquad \xi \text{I C} = \frac{122}{106} = 1\ 15$$

Third test

f_V	1	0	2	0	0	2	6	4	8	0	0	7	0	0	2	1	1	1	1	1	0	6	4	0	2	4
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
f_W	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	1	2	3
	3	0	1	0	0	2	8	1	7	6	0	1	0	0	2	3	0	2	1	2	0	4	2	1	1	5
$f_V f_W$	3	0	2	0	0	4	48	4	56	0	0	7	0	0	4	3	0	2	1	2	0	24	8	0	2	20

$N_V = 53$; $N_W = 52$; $\Sigma f_V f_W = 190$

$$\chi_o = 190 \qquad \chi_r = \frac{53\ 52}{26} = 106 \qquad \xi \text{I C} = \frac{190}{106} = 1\ 79$$

[2] See subpar 37e on p 98

l. Since the last of the three foregoing tests gives a value somewhat better than the expected ξI.C. of 1.73, it looks as though the correct position of the W distribution with reference to the V distribution has been found. In practice, several more tests would be made to insure that other close approximations to 1.73 are not found, but these will here be omitted. The test indicates that the primary cipher component has the letters V and W in these positions: $\overset{1\ 2\ 3\ 4}{\text{V . . W}}$, since the correct superimposition requires that the 4th cell of the W distribution must be placed under the 1st cell of the V distribution (see the last superimposition above).

m. The next best distribution with which to proceed is the F distribution, with 51 occurrences. Therefore, the F sequence is matched against the W and V sequences separately, and then against both W and V sequences at their correct superimposition; this procedure serves as a check on the correct matching of the W and V sequences. The following shows the correct relative positions of the three distributions:

f_V	1	0	2	0	0	2	6	4	8	0	0	7	0	0	2	1	1	1	1	1	0	6	4	0	2	4	$N_V=53$
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
f_F	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	1	2	3	4	5	6	7	
	1	1	2	1	0	0	6	3	9	3	0	2	0	0	0	2	1	1	1	2	0	4	2	0	3	7	$N_F=51$
$f_V f_F$	1	0	4	0	0	0	36	12	72	0	0	14	0	0	0	2	1	1	1	2	0	24	8	0	6	28	$\Sigma f_V f_F=212$

$$\chi_0 = 212 \quad \chi_r = \frac{53\ 51}{26} = 104 \qquad \text{ξI C} = \frac{212}{104} = 2\ 04$$

f_W	1	1	5	3	0	1	0	0	2	8	1	7	6	0	1	0	0	2	3	0	2	1	2	0	4	2	$N_W=52$
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
f_F	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	1	2	3	4	
	0	3	7	1	1	2	1	0	0	6	3	9	3	0	2	0	0	0	2	1	1	1	2	0	4	2	$N_F=51$
$f_W f_F$	0	3	35	3	0	2	0	0	0	48	3	63	18	0	2	0	0	0	6	0	2	1	4	0	16	4	$\Sigma f_W f_F=210$

$$\chi_0 = 210 \quad \chi_r = \frac{52\ 51}{26} = 102 \qquad \text{ξI C} = \frac{210}{102} = 2\ 06$$

$f_{(V+W)}$	4	0	3	0	0	4	14	5	15	6	0	8	0	0	4	4	1	3	2	3	0	10	6	1	3	9	$N_{V+W}=105$
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
f_F	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	1	2	3	4	5	6	7	
	1	1	2	1	0	0	6	3	9	3	0	2	0	0	0	2	1	1	1	2	0	4	2	0	3	7	$N_F=51$
$f_{(V+W)} f_F$	4	0	6	0	0	0	84	15	35	18	0	16	0	0	0	8	1	3	2	6	0	40	12	0	9	63	$\Sigma f_{(V+W)} f_F=422$

$$\chi_0 = 422 \quad \chi_r = \frac{105\ 51}{26} = 206 \qquad \text{ξI C} = \frac{422}{206} = 2\ 05$$

The test yields the sequence $\overset{1\ 2\ 3\ 4\ 5\ 6\ 7\ 8}{\text{V . . W . . . F .}}$

n. The process is continued in the foregoing manner until the entire primary cipher component has been reconstructed. It is obvious that as the work progresses the cryptanalyst is forced to employ smaller and smaller distributions, so that statistically the results are apt to become less and less certain. But to counterbalance this, there is the fact that the number of possible superimpositions becomes progressively smaller as the work progresses. For example, at the commencement of operations the number of possible points for superimposing a second sequence against the first is 25; after the relative positions of 5 distributions have been ascertained and a 6th distribution is to be placed in the primary sequence being reconstructed, there are 20 possible positions; after the relative positions of 20 distributions have been ascertained, there are only 5 possible positions for the 21st distribution, and so on.

o. In the foregoing case the completely reconstructed primary cipher component is as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K

Since it was stated that the problem involves identical primary components, both components are now at hand.[3]

p. Of course, it is probable that in practical work the process of matching distributions would be interrupted soon after the positions of only a few letters in the primary component had been ascertained. For by trying partially reconstructed sequences on the cipher text, the skeletons of some words would begin to show. By filling in these skeletons with the words suggested by them, the process of reconstructing the components is much facilitated and hastened.

q. The components having been reconstructed, only a moment or two is necessary to ascertain their initial position in encipherıng the message. It is only necessary to juxtapose the two components so as to give "good" values for any one of the vertical distributions of Fig. 75. This then gives the juxtaposition of the components for that column, and the rest follows very easily, for the plain text may now be obtained by direct use of the components. The decipherment of the beginning of the cipher text is as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	W	G	J	J	M	M	M	J	X	E	D	G	C	O	C	F	T	R	P	B	M	I	I	I	K	Z
	W	I	T	H	T	H	E	I	M	P	R	O	V	E	M	E	N	T	S	I	N	T	H	E	A	I
2	R	Y	N	N	B	U	F	R	W	W	W	W	Y	O	I	H	F	J	K	O	K	H	T	T	A	Z
	R	P	L	A	N	E	A	N	D	T	H	E	M	E	A	N	S	O	F	C	O	M	M	U	N	I
3	C	L	J	E	P	P	.....																			
	C	A	T	I	O	N	.....																			

[3] If we did not know in advance that identical primary components were involved this fact could have been deduced from a study of the frequency distributions in Fig 75 Note that the distribution for col 1 may be fitted to the normal, this shows that we have identical components running in the same direction and that the setting for col 1 is $A_p = A_c$

r. The student should clearly understand the real nature of the matching process employed to such good advantage in this problem. In practically all the previous cases frequency distributions were made of cipher letters occurring in a cryptogram, and the tallies in those distributions represented the actual occurrences of cipher letters. Furthermore, when these distributions were compared or matched, what were being compared were actually cipher alphabets. That is, the text was arranged in a certain way, so that letters belonging to the same column and the frequency distribution for a specific cipher alphabet was made by tabulating the letters in that column. Then if any distributions were to be compared, usually the entire distribution applicable to one cipher alphabet was compared with the entire distribution applying to another cipher alphabet. But in the problem just completed, what were compared in reality were not frequency distributions applying to the columns of the cipher text as transcribed in subpar. 69e, but graphic representations of the variations in the frequencies of plaintext letters falling in identical sequences, the identities of these plaintext letters being unknown for the moment. Only after the reconstruction has been completed do their identities become known, when the plain text of the cryptogram is established.

70. Solution by the probable word method.--a. The foregoing method of solution is, of course, almost entirely statistical in nature. There is, however, another method of attack which should be brought to notice because in some cases the statistical method, involving the study of relatively large distributions, may not be feasible for lack of sufficient text. Yet in these cases there may be sufficient data in the respective alphabets to permit of some assumptions of values of cipher letters, or there may be good grounds for applying the probable word method. The present paragraph will therefore deal with a method of solving progressive alphabet cipher systems which is based upon the application of the principles of indirect symmetry to certain phenomena arising from the mechanics of the progressive alphabet encipherment method itself.

b. Take the two sequences below and encipher the phrase FIRST BATTALION by the progressive alphabet method, sliding the cipher component to the left one interval after each encipherment.

Components

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

Message

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Plain.....	F	I	R	S	T	B	A	T	T	A	L	I	O	N
Cipher....	E	I	C	N	X	D	S	P	Y	T	U	K	Y	Y

c. Certain letters are repeated in both plain text and cipher text. Consider the former. There are two I's, three T's, and two A's. Their encipherments are isolated below, for convenience in study.

	F	I	R	S	T	B	A	T	T	A	L	I	O	N	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Plain.....	.	I	.	.	.	.	.	.	.	.	.	I	.	.	(1)
Cipher....	.	I	.	.	.	.	.	.	.	.	.	K	.	.	(2)
Plain.....	.	.	.	.	T	.	.	T	T	.	.	.	.	.	(3)
Cipher....	.	.	.	.	X	.	.	P	Y	.	.	.	.	.	(4)
Plain.....	.	.	.	.	.	.	A	.	.	A	.	.	.	.	(5)
Cipher....	.	.	.	.	.	.	S	.	.	T	.	.	.	.	(6)

The two I's in line (1) are 10 letters apart; reference to the cipher component will show that the interval between the cipher equivalent of the first I_p (which happens to be I_c) and the second I_p (which is K_c) is 10. Consideration of the mechanics of the enciphering system soon shows why this is so: since the cipher component is displaced one step with each encipherment, two identical letters n intervals apart in the plain text must yield cipher equivalents which are n intervals apart in the cipher component. Examination of the data in lines (3) and (4), (5) and (6) will confirm this finding. Consequently, it would appear that in such a system the successful application of the probable word method of attack, coupled with indirect symmetry, can quickly lead to the reconstruction of the cipher component.

d. Now consider the repeated cipher letters in the example under subpar. b. There happens to be only two cases of repetition, both involving Y's. Thus:

1	2	3	4	5	6	7	8	9	10	11	12	13	14
.	.	.	.	.	.	.	.	T	.	.	.	O	N
.	.	.	.	.	.	.	.	Y	.	.	.	Y	Y

Reference to the plain component will show that the plaintext letters represented by the three Y's appear in the order N O . . . T, that is, reversed with respect to their order in the plain text. But the intervals between these letters is correct. Again a consideration of the mechanics of the enciphering system shows why this is so: since the cipher component is displaced one step with each encipherment, two identical letters n intervals apart in the cipher text must represent plaintext letters which are n intervals apart in the plain component. In the present case the direction in which these letters run in the plain component is opposite to that in which the cipher component is displaced. That is, if the cipher component is displaced toward the left, the values obtained from a study of repeated plaintext letters give letters which coincide in sequence (interval and direction) with the same letters in the cipher component; the values obtained from a study of repeated ciphertext letters give letters the order of which must be reversed in order to make these letters coincide in sequence (interval and direction) with the same letters in the plain component. If the cipher component is displaced toward the right, this relationship is merely reversed: the values obtained from a study of the repeated plaintext letters must be reversed in their order when placing them in the cipher component;

those yielded by a study of the repeated ciphertext letters are inserted in the plain component in their original order.

e. Of course, if the primary components are identical sequences the data from the two sources referred to in subpars. c and d need not be kept separate but can be combined and made to yield the primary component very quickly.

f. With the foregoing principles as background, and given the first few groups of an intercepted message, which is assumed to begin with COMMANDING GENERAL FIRST ARMY (probable word method of attack), the data yielded by this assumed plain text are shown in Fig. 76.

```
I K M K I   L I D O L   W L P N M   V W P X W   D U F F T   F N I I G
X G A M X   C A D U V   A Z V I S   Y N U N L.....

                        1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Assumed plain text      C O M M A N D I N G G E N E R A L F I R S T A R M Y
Cipher.............     I K M K I L I D O L W L P N M V W P X W D U F F T F
```

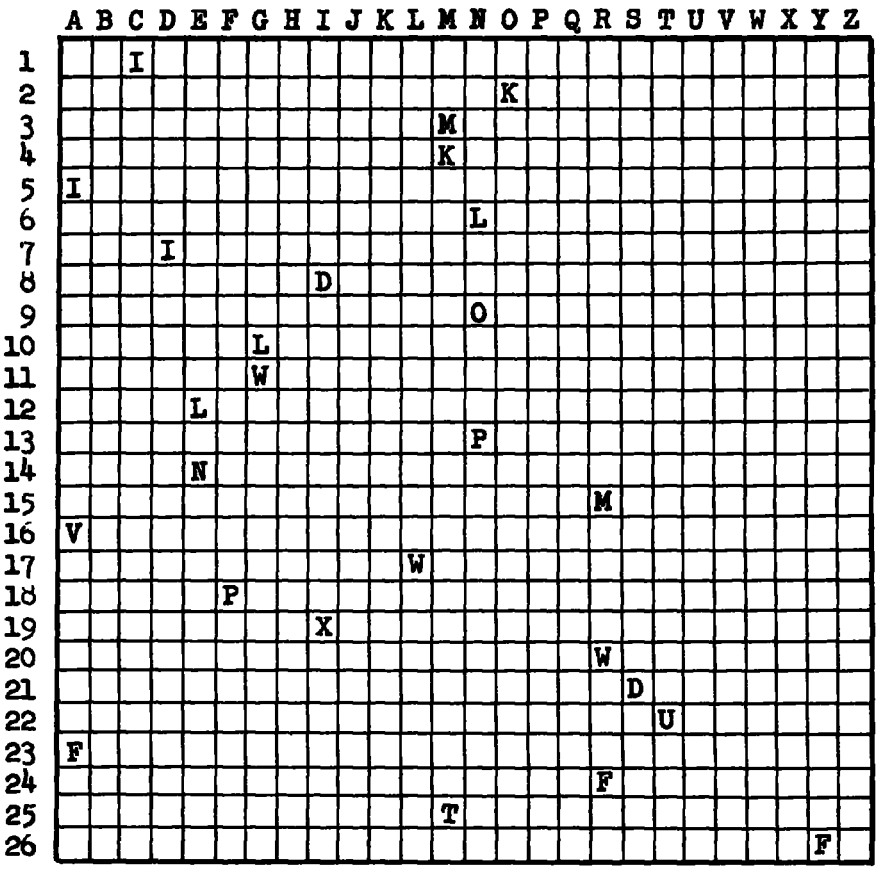


Figure 76.

Analysis of the data afforded by Fig. 76, in conjunction with the principles of indirect symmetry, yields the following partial components:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Plain.....	A	.	L	I	C	.	E	F	G	.	.	M	N	O	.	.	S	.	.	.	.	.	.	Y	D	R
Cipher.....	.	.	M	K	V	.	L	W	N	O	.	F	.	P	.	.	.	.	.	I	.	.	.	T	.	
	D	.	.	.	.	.	.	.	.	.	.	X														

Setting the two partial components into juxtaposition so that $C_p = I_c$ (first encipherment) the 8th value, $I_p = D_c$, gives the position of D in the cipher component and permits the addition of X to it, these being two letters which until now could not be placed into position in the cipher component. With these two partial sequences it becomes possible now to decipher many other letters in the message, gaps being filled in from the context. For example, the first few letters after ARMY decipher as follows:

	1	2	3	4	5	6	7	8	9	10	11	12
Cipher....	N	I	I	G	X	G	A	M	X	C	A	D
Plain.....	.	I	L	.	.	.	.	E	O	.	.	R

The word after ARMY is probably WILL. This leads to the insertion of the letter W in the plain component and G in the cipher component. In a short

time both components can be completely established.

g. In passing, it may be well to note that in the illustrative message in subpar. 69h the very frequent occurrence of tripled letters (MMM, WWW, FFF, etc.) indicates the presence of a frequently used short word, a frequently used ending, or the like, the letters of which are sequent in the plain component. An astute cryptanalyst who has noted the frequency of occurrence of such triplets could assume the value THE for them, go through the entire text replacing all triplets by THE, and then, by applying the principles of indirect symmetry, build up the plain component in a short time. With that much as a start, solution of the entire message would be considerably simplified.

71. Solution by means of isomorphs.--a. One of the most powerful attacks in cryptanalytics involves the exploitation of isomorphs; i.e., ciphertext sequences which exhibit an idiomorphism identical with that of another ciphertext sequence; this method finds applicability in many varieties of manual cipher systems, and it also takes on a very important aspect in the solution of many machine cipher systems. In progressive alphabet ciphers, the presence of isomorphs, if they are of fair length and proper composition, might enable the cryptanalyst to derive the complete primary cipher component directly and thus reach a quick and easy solution of a problem; in any case, isomorphs will enable the partial reconstruction of the cipher component, facilitating further analysis and solution.

b. Isomorphic sequences in the cipher text of progressive alphabet systems may be brought about by identical plaintext beginnings[4] of a pair of messages, by identical endings[5], by a stagger situation, or by a latent repetition occurring within a message or between a pair of messages. Isomorphism may be discovered by examining all pronounced idiomorphic patterns in the cipher text and comparing patterns so disclosed for exact correspondences of repeated letters, i.e., isomorphs. For instance, the two isomorphic ciphertext sequences below may be isolated by searching for all the

(1) ..... C V C N A U H Y H H I T N L C .....
(2) ..... E X E P L I D R D D B W P C E .....

AA patterns[6] in the cipher texts under examination, and inspecting the letters which precede and follow these AA patterns for further evidences of

[4] These beginnings need not necessarily be stereotyped beginnings The composition of the plain text is unimportant, the only requisite factor is that the beginnings consist of identical plain text

[5] This situation is, of course often brought about by identical signatures at the ends of messages

[6] In searching for isomorphs, it might be also necessary to examine all A-A patterns and then A--A patterns if the AA patterns do not bear fruit

apparent isomorphism. Obviously, not all sequences surrounding AA patterns will be causal isomorphs; this is especially true in the case of short isomorphic sequences, just as short repetitions in cipher text may occur by chance and not be due to causal factors. But if an isomorph can be extended sufficiently (i.e., if further isomorphic patterns about the basic AA patterns are noticed), then the isomorphs may be considered valid, and the delineation of the isomorphs to the left and right, insofar as discernible, may be established.

c. As an example of a solution by means of isomorphs, let us consider the following beginnings of three messages:

Message "A"

V N N P H S M X W I P U C W R S T G U C R M L J J T U Q R E
H S F V O J R R T D.....

Message "B"

R W W Z I Y V U A K Z G M A E Y D Q G M E V J S S D G W G Z
S T T D B G T O C N.....

Message "C"

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

It is noted that Messages "A" and "B" are isomorphic from their beginnings to their 27th letters, and that Message "C" is isomorphic with the other two from its beginning to the 20th letter.[7] The isomorphic portions are now superimposed, as is shown in the diagram below:

"A": VNNPHSMXWIPUCWRSTGUCRMLJJTU
"B": RWWZIYVUAKZGMAEYDQGMEVJSSDG
"C": UZZYBRXILNYKOLGRATKO

Chains from the foregoing diagram are derived, as follows:

"A"-"B"	"A"-"C"	"B"-"C"
CMVRE	VUK	WZYRUIB
NWA	MXINZ	VX
PZ	PY	DAL
HIK	SRGTA	EGKN
LJSY	WL	MO
XUGQ	CO	QT
TD		

Using the principles of indirect symmetry, an equivalent primary cipher component is recovered as follows: CMVREOXUGQHIKTDBNWAFPZLJSY. Decimation

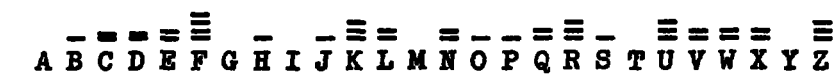
[7] The isomorphism manifested actually begins with the second letters of the messages, it cannot be proved at this point that it includes the first letters, but the absence of contradictory evidence in isomorphs of this length plus the fact that the isomorphs are at the beginning of the messages, makes this a safe assumption

of this sequence at an interval of -11 brings out the original keyword-mixed sequence, based on HYDRAULIC.

d. With the cipher component now at hand, the text of any one of the messages may now be reduced to monoalphabetic terms, if we can assume the correct motion of the cipher component;[8] in other words, a known motion makes conversion to monoalphabetic terms possible. Taking the fragment of Message "A" as an example, and assuming that the cipher component is slid to the right after each encipherment, we have the following conversion (in terms of an arbitrary A-Z sequence for the plain component) and its accompanying unıliteral frequency distribution:

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

C: U Q R E H S F V O J R R T D.....
P: F U F N E Z R D Z W N O H P



I.C. = 0.90

This is certainly not satisfactory. Assuming that the cipher component is slid to the left, we have the following:

C: V N N P H S M X W I P U C W R S T G U C R M L J J T
P: W P O P W P J R P Y I U W K P F F V N P J U K Q P W

C: U Q R E H S F V O J R R T D.....
P: F S B H W P F P J E T S J P

I.C. = 2.33

Obviously this is the correct case. After solution of the monoalphabet, which is facilitated in this case by the idiomorphic patterns now revealed, it is found that the recovered plain component is the same as the cipher component, except that it runs in the reverse direction.

e. It should be clear why isologous sequences in progressive alphabet systems, unlike isologous sequences in other types of periodic ciphers, produce isomorphs which may be chained without regard to the particular alphabets involved, and also why the conversion process is not affected by the identity of the particular alphabets. In the usual type of repeating-key cipher, the selection of the alphabets used is determined by a key word

[8] This very important consideration forms the basis of solution of cryptograms produced by many types of cipher devices and cipher machines This principle will be elaborated upon in much greater detail in Military Cryptanalytics, Part III

which is used for this purpose; the letters of the key bear no constant displacement-relationship to each other. However, in the case of a progressive alphabet cipher the successive elements of the key, 26 in number, are a constant interval apart from each other as measured on the cipher component; this accounts for the fact that, if one encipherment of a plaintext passage produces an idiomorphic ciphertext sequence, the remaining 25 other possible encipherments will also produce idiomorphic cipher texts which will all be isomorphic to one another.

72. Solution by superimposition.--a. The discussion in this chapter thus far has, except for special solutions, been limited to cases wherein there is available a long message in a progressive alphabet system. Suppose that in the traffic there are no long messages, what then? If a number of short messages are available, then there should be a way to superimpose the messages properly, that is, put them in depth, even if no two messages begin with the same initial key letter, i.e., start at the same point in the key sequence.

b. There are three principal means for superimposing messages in progressive alphabet systems.[9] These are: (1) superimposition by means of known indicators[10]; (2) superimposition by ciphertext repetitions; and (3) superimposition by a comparison of columnar frequency distributions. The first of these methods is rather obvious: it goes without saying that if the enemy were still using a compromised or recovered indicator system, then of course all messages could be put in depth without any analysis whatsoever. The second method, that of superimposition by repetitions, is also quite obvious: since long repetitions (i.e., long for a given sample size) have a high probability of being causal, then the alignment of messages to make the polygraphic repetitions fall into identical columns of the width of the period will result in the correct superimposition of the messages. The third method, that of comparison of the columnar frequency distributions, will be discussed in detail below.

c. Let us consider a long message in a progressive alphabet cipher, such as that given in subpar. 69i, and let us also consider its columnar frequency distributions given in Fig. 75. If we had at hand another long message, which however began at a point in the keying sequence 5 places to the right of the first message, it is clear that col. 1 of the second message would not resemble col. 1 of the first message. Nevertheless, col. 1 of the second message would bear a close resemblance to col. 6 of the first, and col. 2 of the second message would be very similar to col. 7 of the first, and so on. If our two messages had about 40 tallies per distribution (as in the example), there would be little trouble in finding the correct

---

[9] These means to be described are also applicable for the superimposition of messages in other types of repeating-key systems

[10] Indicators play an important role in cryptography An indicator is a symbol (consisting of a letter, a group of letters, a figure, or a group of figures) which indicates the specific key used under the general cryptosystem or it may indicate which one of a number of general systems has been used or it may indicate both.

matching of the columns, since this could be done easily by ocular inspection. If however we had at hand a pair of short messages (say between 100 and 200 letters each), then mere ocular inspection would be of no avail, and recourse must be had to statistical methods to find the proper superimposition.

d. Let us assume that we have for study a set of short messages intercepted on a particular day on a naval circuit known to be passing traffic in progressive alphabet systems. It is further known that the primary components are changed daily; therefore the traffic of one day is expected to be homogeneous with respect to the primary components involved. Two of the longest messages are given below, of lengths 190 and 170 letters, respectively:

Message "A"

HFDCS	WTQOO	YCPXF	NWLGP	ULRIU	RHFDQ	HCXPS	SNIPG	NXUVL	CUDAV
WAYNK	ZHKXS	BIPDM	BNKKI	FBLWT	RDAAH	YQSSJ	VSODY	EFFBI	UGXLB
IAYRH	RNMBM	VSUAS	CMFKM	LAFBL	OICZK	KEZVH	JSAGT	ZNEBX	VERGF
ZIAUJ	ZSJFT	WSOQF	GQOKZ	WBREC	EIYCD	VUYXD	MKZKT		

Message "B"

KVCRV	FUBOX	SYFGV	ZWTQO	OEAQP	ZKBJW	SPLEN	WDKJW	WNHLT	PEOYD
PLGRC	UAYVR	RSLAH	OPWYL	WWTRS	QIFFA	DBQSA	IURYA	DZEZS	BXXAE
OPNFF	UKIEL	IVSUA	YGEDI	HSVMP	SQMLI	GEGID	BZEMA	YPNZR	CZTGG
NDAKP	NMKGB	SBLPH	AYAHX						

e. Since the period of the messages is 26, the messages are written out on this width. What we will now do is align the messages in flush depth and perform a χ test of the corresponding columns, arriving at a value of χ (or a $\S$I.C.) for that particular alignment.[11] After this test is completed, we will slide Message "B" over one position with respect to Message "A", and perform the test again; and so on for the 26 possible alignments. In order to facilitate the comparison of the texts, we will write out Message "B" in doubled length, as is shown in Fig. 77 below (for the first comparison); the χ_o values are derived for each column and are indicated under the respective columns.

[11] This procedure is somewhat akin to that demonstrated in subpar 18e in connection with the use of the ϕ test to determine the number of alphabets of a relatively short cryptogram of a lengthy period. See also footnote 11 on p. 39.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
H	F	D	C	S	W	T	Q	O	O	Y	C	P	X	F	N	W	L	G	P	U	L	R	I	U	R
H	F	D	Q	H	C	X	P	S	S	N	I	P	G	N	X	U	V	L	C	U	D	A	V	W	A
Y	N	K	Z	H	K	X	S	B	I	P	D	M	B	N	K	K	I	F	B	L	W	T	R	D	A
A	H	Y	Q	S	S	J	V	S	O	D	Y	E	F	F	B	I	U	G	X	L	B	I	A	Y	R
H	R	N	M	H	M	V	S	U	A	S	C	M	F	K	M	L	A	F	B	L	O	I	C	Z	K
K	E	Z	V	H	J	S	A	G	T	Z	N	E	B	X	V	E	R	G	F	Z	I	A	U	J	Z
S	J	F	T	W	S	O	Q	F	G	Q	O	K	Z	W	B	R	E	C	E	I	Y	C	D	V	U
Y	X	D	M	K	Z	K	T																		

1	2	3	24	25	26	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
						K	V	C	R	V	F	U	B	O	X	S	Y	F	G	V	Z	W	T	Q	O	O	E	A	Q	P	Z
K	V	C	Q	P	Z	K	B	J	W	Z	P	L	E	N	W	D	K	J	W	W	N	H	L	T	P	E	O	Y	D	P	L
K	B	J	D	P	L	G	R	C	U	A	Y	V	R	R	S	L	A	H	O	P	W	Y	L	W	W	T	R	S	Q	I	F
G	R	C	Q	I	F	F	A	D	B	Q	S	A	I	U	R	Y	A	D	Z	E	Z	S	B	X	X	A	E	O	P	N	F
F	A	D	P	N	F	F	U	K	I	E	L	I	V	S	U	A	Y	G	E	D	I	H	S	V	M	P	S	Q	M	L	I
F	U	K	M	L	I	G	E	G	I	D	B	Z	E	M	A	Y	P	N	Z	R	C	Z	T	G	G	N	D	A	K	P	N
G	E	G	K	P	N	M	K	G	B	S	B	L	P	H	A	Y	A	H	X												
M	K	G																													
						2	2	4	0	2	2	1	2	4	3	5	2	0	4	1	0	1	2	3	2	0	2	4	1	0	1

$\xi\chi = 50$

Figure 77.

f. In the foregoing comparison, we note that the observed value of χ for all the columns is 50. We will now compute the expected values of χ_m and χ_r with which to compare the χ_o. Since in Fig. 77 we have 8 columns of 8x7 letters, 6 columns of 7x7 letters, and 12 columns of 7x6 letters, the total number of comparisons is $8(8\cdot7)+6(7\cdot7)+12(7\cdot6) = 1246$. Thus $\chi_m = .0667(1246) = 83.11$, and $\chi_r = \frac{1246}{26} = 47.92$. The ξI.C. $= \frac{50}{47.92} = 1.04$, so instead of arching his eyebrows the cryptanalyst will merely shrug his shoulders and go on to the next test.

g. For the second test, we will move Message "B" one column to the right so that the first letter of Message "B" is under the second letter of Message "A", as shown below, and the χ values are determined as before.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
H	F	D	C	S	W	T	Q	O	O	Y	C	P	X	F	N	W	L	G	P	U	L	R	I	U	R
H	F	D	Q	H	C	X	P	S	S	N	I	P	G	N	X	U	V	L	C	U	D	A	V	W	A
Y	N	K	Z	H	K	X	S	B	I	P	D	M	B	N	K	K	I	F	B	L	W	T	R	D	A
A	H	Y	Q	S	S	J	V	S	O	D	Y	E	F	F	B	I	U	G	X	L	B	I	A	Y	R
H	R	N	M	H	M	V	S	U	A	S	C	M	F	K	M	L	A	F	B	L	O	I	C	Z	K
K	E	Z	V	H	J	S	A	G	T	Z	N	E	B	X	V	E	R	G	F	Z	I	A	U	J	Z
S	J	F	T	W	S	O	Q	F	G	Q	O	K	Z	W	B	R	E	C	E	I	Y	C	D	V	U
Y	X	D	M	K	Z	K	T																		

1	2	3	24	25	26	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
						K	V	C	R	V	F	U	B	O	X	S	Y	F	G	V	Z	W	T	Q	O	O	E	A	Q	P	Z
K	V	C	Q	P	Z	K	B	J	W	Z	P	L	E	N	W	D	K	J	W	W	N	H	L	T	P	E	O	Y	D	P	L
K	B	J	D	P	L	G	R	C	U	A	Y	V	R	R	S	L	A	H	O	P	W	Y	L	W	W	T	R	S	Q	I	F
G	R	C	Q	I	F	F	A	D	B	Q	S	A	I	U	R	Y	A	D	Z	E	Z	S	B	X	X	A	E	O	P	N	F
F	A	D	P	N	F	F	U	K	I	E	L	I	V	S	U	A	Y	G	E	D	I	H	S	V	M	P	S	Q	M	L	I
F	U	K	M	L	I	G	E	G	I	D	B	Z	E	M	A	Y	P	N	Z	R	C	Z	T	G	G	N	D	A	K	P	N
G	E	G	K	P	N	M	K	G	B	S	B	L	P	H	A	Y	A	H	X												
M	K	G																													

0 4 1 2 1 3 1 2 1 3 1 4 3 3 2 1 2 0 2 1 0 1 1 2 1 0 $\{\chi = 42$

Figure 78.

For this case, the number of comparisons is $7(8\cdot7)+1(8\cdot6)+7(7\cdot7)+11(7\cdot6) = 1245$, thus $\chi_m = .0667(1245) = 83.04$, and $\chi_r = \frac{1245}{26} - 47.88$. Since $\chi_o = 42$, the $\{$I.C. in this case is $\frac{42}{47.88} = 0.88$, which means another shrug.

h. Subsequent tests still give §I.C.'s in the close vicinity of 1.00, until the 15th test is made, as follows:

```
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
 H  F  D  C  S  W  T  Q  O  O  Y  C  P  X  F  N  W  L  G  P  U  L  R  I  U  R
 H  F  D  Q  H  C  X  P  S  S  N  I  P  G  N  X  U  V  L  C  U  D  A  V  W  A
 Y  N  K  Z  H  K  X  S  B  I  P  D  M  B  N  K  K  I  F  B  L  W  T  R  D  A
 A  H  Y  Q  S  S  J  V  S  O  D  Y  E  F  F  B  I  U  G  X  L  B  I  A  Y  R
 H  R  N  M  H  M  V  S  U  A  S  C  M  F  K  M  L  A  F  B  L  O  I  C  Z  K
 K  E  Z  V  H  J  S  A  G  T  Z  N  E  B  X  V  E  R  G  F  Z  I  A  U  J  Z
 S  J  F  T  W  S  O  Q  F  G  Q  O  K  Z  W  B  R  E  C  E  I  Y  C  D  V  U
 Y  X  D  M  K  Z  K  T
```

```
 1  2  3  10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26  1  2  3  4  5  6  7  8  9 10 11 12 13  24 25 26
                                                             K  V  C  R  V  F  U  B  O  X  S  Y  F   Q  P  Z
 K  V  C  X  S  Y  F  G  V  Z  W  T  Q  O  O  E  A  Q  P  Z  K  B  J  W  Z  P  L  E  N  W  D  K  J   D  P  L
 K  B  J  W  D  K  J  W  W  N  H  L  T  P  E  O  Y  D  P  L  G  R  C  U  A  Y  V  R  R  S  L  A  H   Q  I  F
 G  R  C  S  L  A  H  O  P  W  Y  L  W  W  T  R  S  Q  I  F  F  A  D  B  Q  S  A  I  U  R  Y  A  D   P  N  F
 F  A  D  R  Y  A  D  Z  E  Z  S  B  X  X  A  E  O  P  N  F  F  U  K  I  E  L  I  V  S  U  A  Y  G   M  L  I
 F  U  K  U  A  Y  G  E  D  I  H  S  V  M  P  S  Q  M  L  I  G  E  G  I  D  B  Z  E  M  A  Y  P  N   K  P  N
 G  E  G  A  Y  P  N  Z  R  C  Z  T  G  G  N  D  A  K  P  N  M  K  G  B  S  B  L  P  H  A  Y  A  H
 M  K  G  A  Y  A  H  X

 8  6  2  1  1  4  3  3  2  3  2  2  4  0  6  2  4  0  0  1  0  2  3  2  2  2   §χ = 65
```

Figure 79.

At this alignment, the number of comparisons is $2(8\cdot7)+6(8\cdot6)+12(7\cdot7)+6(7\cdot6) = 1240$, thus $\chi_m = .0667(1240) = 82.71$, and $\chi_r = \frac{1240}{26} = 47.69$. The observed χ is 97, so the §I.C. here is $\frac{97}{47.69} = 2.03$, which should gladden the heart of the cryptanalyst. This is it.

i. Just when the cryptanalyst is about to relax and enjoy life, being on the verge of cryptanalytic fructification, he observes with consternation that at the very next alignment of the messages, a pentagraphic repetition would come in phase, a repetition which somehow he had previously overlooked: at the 16th test, the pentagraph WTQOO would be lined up between the two messages.[12] So with a dejected feeling he superimposes the messages anew, notes that $\chi_m = .0667(1241) = 82.77$, $\chi_r = \frac{1241}{26} = 47.73$, and $\chi_o = 65$. This makes the §I.C. $= \frac{65}{47.73} = 1.36$, which in turn makes the cryptanalyst quite unhappy.

[12] The mathematical expectation of this phenomenon is calculated as follows Message "A" is 190 letters long, and thus contains 186 pentagraphs, Message "B", of 170 letters, contains 166 pentagraphs The number of pentagraphic comparisons is therefore 186x166, or 30,876 Since there are 26^5 different pentagraphs, the probability of a pentagraphic coincidence between these two messages is $\frac{30,876}{26^5}$ or 0026, in other words, less than 3 chances in 1,000

j. Grasping at straws, the cryptanalyst tries to save the situation with the ϕ test.[13] For the 15th superimposition, he obtains the following:

$$\phi_o = 346$$

$$\phi_r = \frac{2(15\cdot14)+18(14\cdot13)+6(13\cdot12)}{26} = \frac{4632}{26} = 178$$

$$\phi_p = .0667(4632) = 309$$

$$\delta \text{ I.C.} = \frac{346}{178} = 1.94$$

Fine. Now for the 16th superimposition:

$$\phi_o = 298$$

$$\phi_r = \frac{3(15\cdot14)+16(14\cdot13)+7(13\cdot12)}{26} = \frac{4634}{26} = 178$$

$$\phi_p = .0667(4634) = 309$$

$$\delta \text{ I.C.} = \frac{298}{178} = 1.67$$

What did not arch the eyebrows, shrug the shoulders, or gladden the heart, now gives a sinking feeling in the pit of the stomach.

k. Impaled on the horns of a dilemma,[14] the cryptanalyst is forced to try both hypotheses; at least there are only two--it could have been worse. The 15th superimposition is actually the correct alignment of the messages, so if he trusts the high δI.C., the cryptanalyst will be right the first time. The authors hasten to assure the reader that, in spite of dark suspicions to the contrary, the accidental pentagraphic repetition was not manipulated or forced, but really did happen accidentally. What this experience teaches us is that the δI.C. is more to be trusted than the δI.C. in matching distributions[15], and that, evidently, in samples of this size the probability of a very high δI.C. being reached in an incorrect case is

---

[13] The procedure here is that demonstrated in subpar 18e

[14] Of, however, a small order of magnitude

[15] In Statistical Methods in Cryptanalysis, par 22, Dr Kullback demonstrates the fact that the χ test is preferable to the ϕ test insofar as matching distributions is concerned He also shows (p 49) that if two monoalphabetic distributions have been merged, the expected value of ϕ is given by the formula $\phi = .0667(N-1) - .0564N_1N_2$, where $N = N_1+N_2$ If the two distributions are of equal size (i e , if $N_1 = N_2$), the expected δI.C. of two merged non-related monoalphabetic distributions will usually be in the vicinity of $\frac{1.73-1.00}{2} = 1.37$ Thus, when Message 'A' and Message "B' are incorrectly superimposed in flush depth as in subpar e, the average δI.C of the merged distributions if found to be 1.42, which is just about what is expected

less than the probability of an accidental pentagraphic repetition. In passing, it might be well to keep in mind the following excerpt from Kullback:[16]

"It is not very often that statistical analysis alone will enable the cryptanalyst to arrive at the solution of a cryptogram Statistical analysis will, however, enable the cryptanalyst to evaluate the desirability of pursuing certain procedures and will indicate the most likely order in which to try various possible steps in solution "

While we're at it, it might not hurt to keep in mind the following quotation from an unidentified author:

"There are three kinds of lies lies, damned lies, and statistics "

1. It was stated in subpar. d that the two messages under discussion were from a set of short messages; the foregoing procedures would be continued, adding more messages to the already established depth, until a sufficient number of messages were put in depth to permit of solution by means of the χ test as treated in par. 69. These first two messages might by themselves probably be unsolvable; for the student who is interested in solving them, it will be added that the signature TOMLINSON is present in Message "B".

73. Additional remarks.--a. As has already been indicated in subpar. 67a, the number of different alphabets in progressive alphabet systems is not necessarily confined to 26. It is possible to have N = 25, 27, 30, 32, 36, ... 100; obviously, where N is greater than 26, the cipher characters cannot be restricted merely to the 26 letters of the alphabet, but must either include additional symbols, or else the cipher text must be represented by digit groups such as dinomes. If a Baudot system incorporated a progressive alphabet principle, then of course the components would involve the 32 characters of the Baudot alphabet.

b. The principles elucidated in this chapter may, of course, also be applied to cases of progressive alphabet systems in which the progression is by regular intervals greater than 1, and, with necessary modifications, to cases in which the progression is not regular but follows a specific pattern, such as the successive displacements 1-2-3-4-5, 1-2-3-4-5, ..., or 2-5-1-3-1-4-2-3, 2-5-1-3-1-4-2-3, and so on.[17] The latter types of progression are encountered in certain mechanical cipher devices, the study of which will be reserved for the next text.

c. There has been a liberal sprinkling of elementary cryptomathematics as applied to specific situations in the text thus far. This topic will be treated in full detail in Military Cryptanalytics, Part III. In the mean-

[16] Ibid , p 1

[17] Cases may be encountered in which the selection of alphabets is controlled by a 26-element key comprised by a keyword-mixed sequence as a mnemonic device.

while, the student who is interested in pursuing the subject further will profit if he consults the works listed below and gleans from them what he can, depending upon his mathematical background:

S. Kullback, Statistical Methods in Cryptanalysis, (Revised Edition), Washington, 1938. (Unclassified)

H. Campaigne, The Index of Coincidence, Washington, 1955. (Confidential/Modified)

H. Campaigne, Statistics for Cryptology, Washington, 1951. (Confidential)

For a first book on statistics, the following is highly recommended:

S. S. Wilks, Elementary Statistical Analysis, Princeton University Press, 1948.

Three valuable tables exceedingly useful for cryptanalytic work are those cited below. These documents are prefaced with an introduction which shows how the tables are used in typical cryptanalytic applications.

Cryptanalyst's Manual, Section 5-1, Table of the Poisson Distribution (Individual and Cumulative Terms), Washington, 1955. (Unclassified)

Cryptanalyst's Manual, Section 5-3, Expected Number of x-fold Repetitions (Binomial Distribution), Washington, 1950. (Unclassified)

Cryptanalyst's Manual, Section 5-4, Abridged Binomial Tables Applicable to Single-Character Cryptanalytic Distributions (P = 1/32, 1/30, 1/26, 1/25, 1/10), Washington, 1954. (Unclassified)

(BLANK)