By Order of the Secretary of the Army:

CARL E. VUONO
General, United States Army
Chief of Staff

Official:

THOMAS F. SIKORA
Brigadier General, United States Army
The Adjutant General

DISTRIBUTION:

Active Army, USAR, and ARNG: To be distributed in accordance with DA Form 12-11E, requirements for FM 34-40-2, Basic Cryptanalysts, (Qty rqr block no. 4607) and FM 34-3, Intelligence Analysis (Qty rqr block no, 1119).

☆ U.S. GOVERNMENT PRINTING OFFICE; 1990-729/952

PREFACE

This field manual is intended as a training text in basic cryptanalytics and as a reference for cryptanalysts in military occupational specialty (MOS) 98C and related MOSs.

The proponent of this publication is Headquarters, United States Army Training and Doctrine Command (TRADOC). Send comments and recommendations on DA Form 2028 (Recommended Changes to Publications and Blank Forms) directly to Commander, United States Army Intelligence School, Fort Devens (USAISD), ATTN: ATSI-ETD-PD, Fort Devens, MA 01433-6301.

INTRODUCTION

This manual presents the basic principles and techniques of cryptanalysts and their relation to cryptography. Cryptography concerns the various ways of protecting messages from being understood by anyone except those for whom the messages are intended. Cryptographers are the people who create and use codes and ciphers. Cryptanalytics is the art and science of solving unknown codes and ciphers. Cryptanalysts try to break the codes and ciphers created and used by cryptographers.

This publication is organized into six parts. Part One explains basic principles which apply to all the parts that follow. The following five parts each cover a major type of system and the cryptanalytic techniques that apply to it. Parts Two, Three, and Four each build on the techniques explained in the parts that precede them. A new student should study these in order. Parts Five and Six are largely independent of Parts Two through Four and can be used separately after Part One.

For practice in the techniques explained in this manual, the Army Correspondence Course Program offers a course in basic cryptanalysts. See the References Section at the back of this manual for further information.

FIELD MANUAL
NO 34-40-2

HEADQUARTERS
DEPARTMENT OF THE ARMY
Washington, DC, 13 September 1990

BASIC CRYPTANALYSIS

TABLE OF CONTENTS

*This publication supersedes TM 32-220, 20 August 1970.

PART TWO ● MONOGRAPHIC SUBSTITUTION SYSTEMS

PART THREE ● POLYGRAPHIC SUBSTITUTION SYSTEMS

PART FOUR ● POLYALPHABETIC SUBSTITUTION SYSTEMS

PART FIVE ● TRANSPOSITION SYSTEMS

PART SIX ● ANALYSIS OF CODE SYSTEMS

PART ONE

Introduction to Cryptanalyst

CHAPTER 1

TERMINOLOGY AND SYSTEM TYPES

Section I
Basic Concepts

1-1. Cryptology

Cryptology is the branch of knowledge which concerns secret communications in all its aspects. Two major areas of cryptology are cryptography and *cryptanalytics.*

1-2. Cryptography

Cryptography is the branch of cryptology concerned with protecting communications from being read by the wrong people. Codes and ciphers that are used to protect communications are called cryptographic systems. The application of codes and ciphers to messages to make them unreadable is called encryption. The resulting messages are called cryptograms. The people who create and use cryptographic systems are called cryptographers.

1-3. Cryptanalytics

Cryptanalytics is the branch of cryptology concerned with solving the cryptographic systems used by others. The objects of cryptanalysts are to read the text of encrypted messages and to recover the cryptographic systems used. The text is recovered for its potential intelligence value. The systems are recovered for application to future messages in the same or similar systems.

1-4. Signal Communications

In military applications most encrypted messages are sent by electronic means rather than physically carried or mailed. The electronic means include those sent by wire and those transmitted by radio. Whether wire or radio is used, they can be sent by telephone, telegraph (Morse code), teletypewriter, facsimile, or computer. The electronic means provide greater speed than physical means, but make the communications more vulnerable to intercept by others.

Section II
Cryptographic Systems

1-5. Ciphers and Codes

There are two major categories of cryptographic systems, called ciphers and codes. Nearly all military systems fall into one or the other of these categories or a combination of the two. Cipher systems are those in which the encryption is carried out on single characters or groups of characters without regard to their meaning. Codes, on the other hand, are more concerned with meanings than characters. The basic unit of encryption in a code system is a word or phrase. When a message is encrypted by a code system, code groups primarily replace words and phrases. Code groups may also replace single characters where necessary, but the substitution for complete words is the key distinction that separates a code from a cipher. Because of this, the cryptanalytic approaches to codes and ciphers are quite different from each other.

a. Messages encrypted by a cipher system are said to be enciphered. Similarly, messages encrypted by a code system are encoded. The resulting text is called ciphertext or codetext. When a cryptogram is translated back into readable form or *plaintext,* it is said to be decrypted, or more specifically, decoded or deciphered.

b. The term code in this manual is given the formal meaning as explained above and in more detail in Part Six. You will often see and hear the term code used with other meanings that do not apply here. Code, in its more general sense, can mean any cryptographic system or any system of replacing one set of values with another. The terms Morse code, binary code, Baudot code, and computer code are examples of the more general usage of the term.

1-6. Enciphered Codes

Some code systems are further encrypted by a cipher system to produce a hybrid type called enciphered codes. This second encryption process is called superencryption or superencipherment. Such systems are normally much more secure than singly encrypted systems, but because of the added complexity take longer to encrypt and are more prone to errors.

1-7. Other Means of Security Communications

Although most military requirements to secure communications are met through the use of codes and ciphers, there are other approaches that can be used in special situations. One such approach is the use of concealment systems. In a concealment system, the plaintext is hidden within another longer text by a predetermined rule or pattern. Other approaches to concealing messages are to use invisible inks or to reduce a message photographically to a dot-sized piece of film. Another approach is to transmit a message from a tape played so fast that it sounds to the ear like a burst of static on the radio. Security for all these methods depends on concealing the fact that a secret

message is being sent at all. Once the existence of the communications is suspected or anticipated, the security is significantly lessened.

1-8. Types of Ciphers

There are hundreds of types of cipher systems ranging from very simple paper-and-pencil systems to very complex cipher machine or computer enciphered systems. These can be categorized as either transposition or substitution or a combination of the two.

a. Transposition. In a transposition system, the plaintext characters of a message are systematical1 y rearranged, After transposing a message, the same characters are still present, but the order of the letters is changed.

b. Substitution. In a substitution system, the plaintext characters of a message are systematically replaced by other characters. After the substitution takes place, the order of the underlying plaintext is unchanged, but the same characters are no longer present. In the simplest substitution systems, the replacement is consistent; a given plaintext character always receives the same replacement character or characters. More secure systems change the replacements so that the equivalents change each time the same character is encrypted.

1-9. Substitution Cipher Alphabets

In everyday usage, an alphabet is a list of the letters used by a lan uage. They vary by language. Many European and Latin American languages share tiie same alphabet as ours or have minor variations. Russian, Greek, Arabic, and Oriental languages have recognizably different alphabets. The term cipher *alphabets* has a slightly different meaning. Instead of a list of characters, a cipher alphabet has two parts; a list of plaintext characters and their cipher equivalents. In the simplest ciphers, an English cipher alphabet will have 26 plaintext letters and 26 ciphertext equivalents, as in the example below.

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

p: send help
c : BLMI ULGS

In the example, *p:* designates plaintext and *c:* designates ciphertext. For clarity, the plaintext is shown in lower case and the ciphertext in capitals. A more secure alphabet may have more ciphertext equivalents than plaintext characters to provide for some variation in encipherment. Whether or not there is variation, a single alphabet system is called a *monoalphabetic* system. A system which gains more security by systematically using more than one alphabet is called a polyalphabetic system.

CHAPTER 2

SECURITY OF CRYPTOGRAPHIC SYSTEMS

Section I
Requirements of Military Systems

2-1. Practical Requirements

Military cryptographic systems must meet a number of practical considerations.

a. An ideal cryptographic system for military purposes is a single all-purpose system which is practical for use from the highest headquarters to the individual soldier on the battlefield. It is secure no matter how much message traffic is sent using the system. It is easy to use without special training. It presents no logistics problems in keeping the users supplied with the system's keys. It operates under all weather conditions, on all means of communication, and in the dark. Little of value is compromised if the enemy captures the system. No system exists that meets all these requirements.

b. Cryptographic system selection for military use depends on much more than its degree of security. While protecting information from unfriendly eyes, a system must still allow communications to take place rapidly, to be reliable, and to be usable by all who need to conduct communications. It must be usable under all conditions that the communications must take place. For example, a system requiring an hour of pains-taking encryption would go unused by a combat military force on the move. A system that has no tolerance for errors in its use would be inappropriate for soldiers under fire in severe weather conditions. A system that only supports a low volume of messages would be inappropriate for a major message center handling thousands of messages daily. A system that requires expensive, sophisticated equipment would be inappropriate for a military force that can barely afford to buy ammunition. No single system meets all the requirements of security, speed, reliability, flexibility, and cost. The need for security must be balanced against the practical requirements when systems are selected for use. Breakable systems are found today, despite technological advances, because of these practical requirements.

2-2. Security Requirements of Military Systems

When security must be balanced against practical considerations, how much security is enough security?

a. Almost any cryptographic system, given enough time and resources can eventually be solved. The only exception to this is a system which uses absolutely random changing keys with every character encrypted and never repeated. Such a system can be achieved under very limited conditions, but is in practice impossible on any large scale.

b. Even the most sophisticated machine or computer based cryptographic system cannot produce random, nonrepeating keys. The requirement for each communicating machine to generate the same keys prevents truly random keys. At best, a machine system can produce keys by so sophisticated a process that it appears to be random and resists efforts to recover the key generation process.

c. Given the practical considerations, a military system is expected to delay successful analysis, not prevent it. When the system is finally solved, the information obtained has lost most of its value.

2-3. Factors Affecting Cryptographic Security

As discussed above, given enough time and resources, almost any system can be solved. No nation has unlimited resources to devote to the effort. If the potential intelligence payoff is timely enough and valuable enough and the resource costs reasonable, the necessary resources will usually be devoted to the effort. A number of factors affect the vulnerability of cryptographic systems to successful cryptanalytic attack.

a. The most obvious factor is the cryptographic soundness of the system or systems in use. Systems with minimal key repetition and limited orderly usage patterns provide the most resistance.

b. The volume of traffic encoded or enciphered with a given set of keys affects system security. The longer the keys are used without change, the more chance an analyst has of finding exploitable repetition and patterns to build the attack upon,

c. The discipline of system users can play a major role in system security. A system that is very sound when used correctly can often be quickly compromised when rules are broken. An obvious example is when a user retransmits a message in the clear that has also been transmitted in encrypted form. When it is recognized, the comparison of the plaintext message with its encrypted form makes key recovery much easier. Other typical examples of undisciplined usage are--

- To mix plaintext and encrypted text in the same transmission.
- To use the same keys longer than prescribed.

- To make unauthorized changes or simplifications to the system.
- To openly discuss the contents of an encrypted message.
- To openly discuss the system or its keys.

d. The amount of collateral information available about the message sender and the situation under which the message was sent affect the security of a system. The more that is known about the sender, the more likely the contents of a message can be determined.

Section II
Cryptanalytic Attack

2-4. Role of Cryptanalysts in Communications Intelligence Operations

Communications intelligence (COMINT) operations study enemy communications for the purpose of obtaining information of Intelligence value. COMINT includes the collection, processing, evaluation, and reporting of intelligence information gathered from enemy communications. When cryptanalysts is successful on a timely basis, it provides the most direct indication of the enemy's intentions. Cryptanalysis is most likely to be successful when other COMINT techniques are also productive. Collection of communications signals, transmitter location and identification, traffic analysis, and translation and analysis of cleartext transmissions all play a part in the production of COMINT.

2-5. Comparison Between Cryptanalysts and Traffic Analysis

Cryptanalysis is the study of encrypted messages. These messages, when passed as part of radio communications, or traffic, are considered the internals of the communications. Traffic analysis is the study of the externals of the communications.

a. The externals of a communications include the following:

- Call signs and call words.
- Call up procedures between operators.
- Radio frequencies.
- Times of transmissions and total volume of traffic.
- Routing information indicating where a message is to be sent.

- Chatter between radio operators.
- Serial numbers or other filing information.
- Indications of precedence or importance of the messages.
- Indicators designating what cryptographic systems or what key settings are in use.

These externals can be a rich source of information about an enemy, regardless of encrypted message recovery. The systems that communicators use to provide this external information can give substantial clues to unit type, organization, and the purpose of communications.

b. The last category of externals mentioned above, indicators of the cryptographic systems or keys in use, is of particular interest to both the traffic analyst and the cryptanalyst. For the traffic analyst, the indicators help establish patterns of usage which give clues to the enemy's organization and structure. For the cryptanalyst, the indicators help group messages into those encrypted by the same system or keys. In some cases, they may even aid directly in the solution of the system.

2-6. Steps in Cryptanalysis

The solution of nearly every cryptogram involves four basic steps-

@ Determination of the language used.

- Determination of the general system used.
- Reconstruction of the specific keys to the system.
- Reconstruction of the plaintext.

a. Determination of the language used normally accompanies identification of the sender through traffic analysis or radio direction finding. If these forms of support are unavailable, or if an enemy uses several languages, the determination of the language may have to be made at a later stage of analysis.

b. Determination of the general system can come from several sources, such as–

- A detailed study of the system characteristics, aided where necessary by character fre uency counts, searches for repeated patterns, and various statistical tests. Ta e study can extend beyond single messages to searching for patterns and repetitions between different messages with similar characteristics. This single step of system determination can be the most time consuming part of the analysis.
- Past history of system usage by the sender. In most cases, the user does not change systems regularly but uses the same system or set of systems from one day to the next. The specific keys may change regularly, but the general systems remain unchanged except at longer interva ls.

- System indicators included with the traffic. Whenever the user has a choice of systems or a choice of keys within the system, the choice must be made known to the receiving cryptographer. The choice is usually communicated by some form of indicators, which can appear within the text of a message or as part of the externals. When the indicators reveal the choice of system, they are called system indicators or discriminants. When they denote specific frequently changing keys to the system, they are called message indicators. Once you learn just how indicators are used from day to day, they can provide a substantial assist to cryptanalysts.

c. Reconstruction of the specific keys to the system is an important step. Although the following step of plaintext recovery produces the most intelligence information, the full key reconstruction can speed recovery of future messages. The approach used to recover keys will vary greatly from system to system.

d. Reconstruction of the plaintext, although listed as the final step, will usually proceed simultaneously with the key reconstruction. Either step can come first, depending on the system and situation. Partial recovery of one aids in the recovery of the other. The two steps often proceed alternately, with each recovery of one helping in recovery of the other until a full solution is reached.

Section III
Analytic Aids

2-7. Analytic Aids to Identification and Solution

There are a number of aids to identification and solution available to help you as a cryptanalyst. By preparing character frequency counts, performing statistical tests, and recording observed repetitions and patterns in messages, you can compare the data to established norms for various systems and languages. The appendixes to this manual include charts, lists, and tables of normal data for the English language. Similar data are available for other languages. The counting of character frequencies, performance of statistical tests, and search for repetition and patterns can be done manually or with computer assistance, where available. This section outlines the aids that apply to many types of systems. Procedures that apply to specific systems are explained in individual sections.

2-8. Language Characteristics

Each language has characteristics that aid successful cryptanalysts.

a. The individual letters of any language occur with greatly varying frequencies. Some letters are used a great deal. Others are used only a small percentage of the time. In English, the letter E is the most common letter used. It occurs about 13 percent of the time, or about once in every eight letters. In small samples, other letters may be more common, but in almost any sample of 1,000 letters of text or more, E will be the most frequent letter. In other languages, other letters sometimes dominate. In Russian, for example, 0 is the most common letter. The eight highest frequency letters in English, shown in descending order, are E, *T, N,* A, 0, *A, I* and S. The eight highest frequency letters make up about 67 percent of our language. The remainmg 18 letters only make up 33 percent of English text. The lowest frequency letters are J, K, Q, X, and Z. These five letters makeup only a little over 1 percent of English text. The vowels, A, E, I, 0, U and Y, make up about 40 percent of English text. In many cryptographic systems, these frequency relationships show through despite the encryption. The analysis techniques explained in the following chapters make repeated use of these frequency relationships. In particular, you should remember the high frequency letters, ETNROAIS, and the low frequency letters, JKQXZ, for their repeated application. The word *SENORITA,* which includes the high frequency letters is one way to remember them. Some people prefer to remember the pronounceable ETNORIAS as a close approximation of the descending frequency order. Choose the method you prefer. The high frequency letters are referred to frequently.

b. Just as single letters have typical frequency expectations, multiple letter combinations occur with varying, but predictable frequencies, too. The most common pair of letters, or digraph, is EN. After EN, RE and ER are the most common digraphs. There are 676 different possible digraphs in English, but the most common 18 make up 25 percent of the language. Appendix A lists the expected frequencies of English language digraphs. Some cryptographic systems do not let individual letter frequencies show through the encryption, but let digraphic frequencies come through. The systems explained in Part Three of this manual show this characteristic.

c. Appendixes B and C list frequency expectations for sets of three letters (trigraphs) and four letters (tetragraphs). Each of these can be useful when studying cryptograms in which three and four letter repeated segments of text occur.

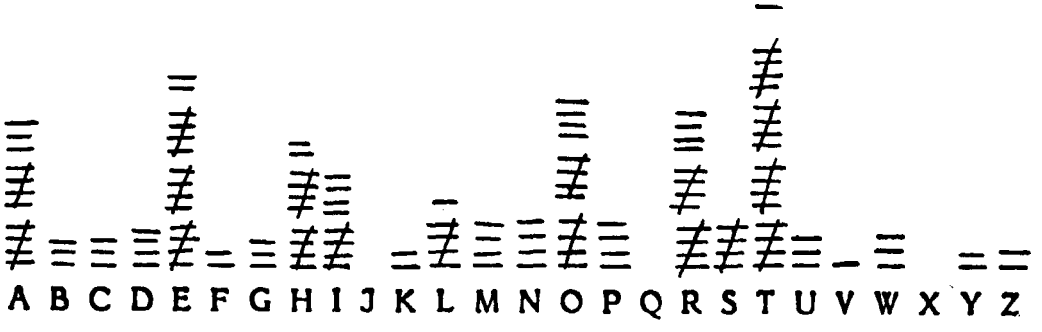
d. Repeated segments of two to four letters will often occur because they are common letter combinations, whether or not they are complete words by themselves. Longer repeated segments readily occur when words and phrases are reused in plaintext. When words are reused in plaintext, they may or may not show up as repeated segments in ciphertext. For a word to show through as a repeat in ciphertext, the same keys must be applied to the same plaintext more than once. Even complex systems which keep changing keys will sometimes apply the same keys to the same plaintext and a repeated ciphertext segment will result. Finding such repeats gives many

clues to the type of system and to the plaintext itself. The search can extend beyond single messages to all messages that you believe may have been encrypted with the same set of keys. If computer support is available to search for repeats for you, a great deal of time can be saved. If not, time spent scanning text to search for repeats will reward you for your time when you find them.

2-9. Unilateral Frequency Distribution

The most basic aid to identification and solution of cipher systems is the unilateral frequency distribution. The term unilateral means one letter at a time. A unilateral frequency distribution is a count of all the letters in selected text, taken one letter at a time.

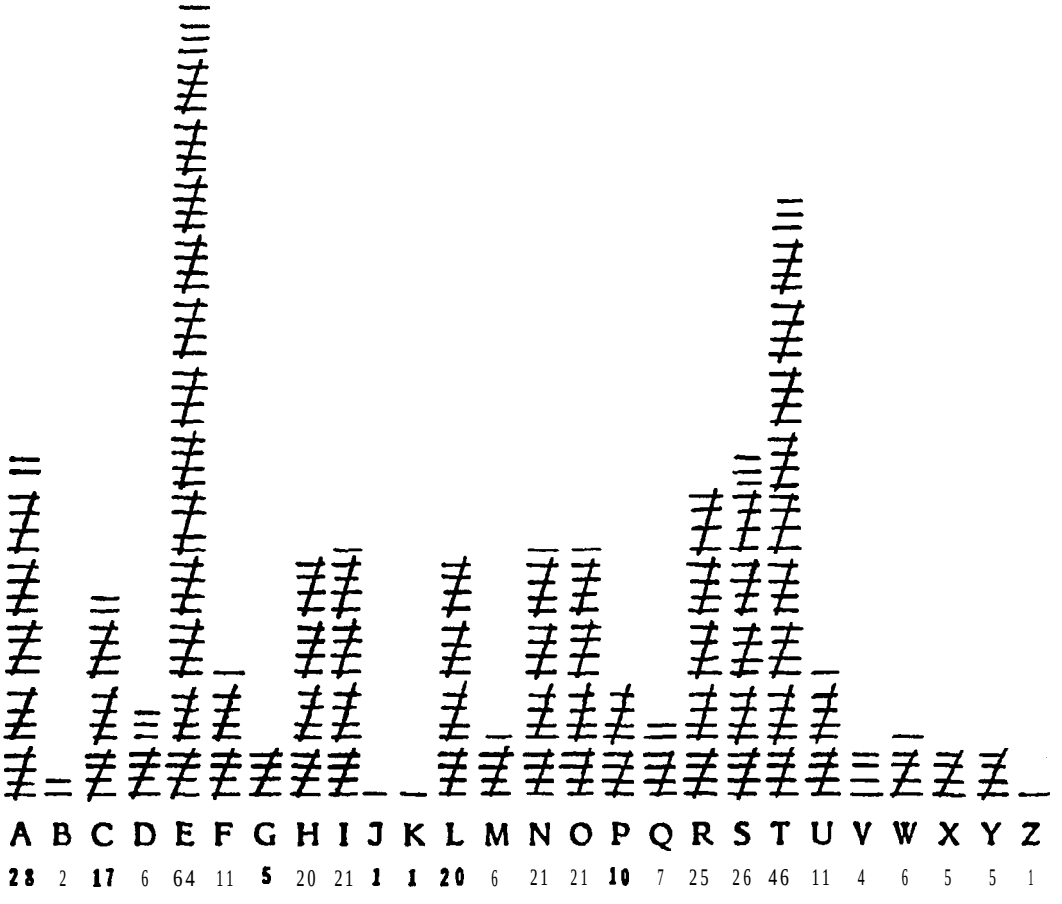
a. The customary method of taking the distribution is to write the letters A through Z horizontally and mark each letter of the cryptogram with a dash above or below the appropriate letter. Proceed through the message from the first letter to the last, marking each letter in the distribution. Avoid the alternate method of counting all the As, Bs, Cs, and so forth, which is very subject to errors. For convenience, each group of five is crossed off by a diagonal slash. The unilateral frequency distribution for the first sentence in this paragraph is shown below.



For comparison, the next example shows the frequency count for the fourth and fifth sentences in paragraph 2-9a.

b. Although individual letter frequencies differ, the pattern of high and low frequency letters is quite similar. The letters that stand above the others in each tally are,

with few exceptions, the expected high frequency letters-ETNROAIS. The expected low frequency letters, JKQXZ, occur once or twice at most. Even in as small a sample as one or two sentences, expected patterns of usage start to establish themselves, Compare this to a frequency count of all letters in this paragraph.



c. When a larger sample is taken, such as the above paragraph, the letters occur much closer to the expected frequency order of ETNROAIS. As expected, E and T are the two highest frequency letters. but the next series of high frequency letters in descending order of occurrence, ASRINO, differs slightly from the expected order of NROAIS. It would take a sample thousands of letters long to produce frequencies exactly in the expected order. Even then, differences in writing style between a field manual and military message texts could produce frequency differences. For example, the word the is often omitted from military message traffic for the sake of brevity. More frequent use of the raises the expected frequency of the letter H.

2-10. Letter Frequencies in Cryptograms

As different cipher systems are explained in this manual, the ways in which letter frequencies can be used to aid identification and solution will be shown. Some basic considerations should be understood now.

a. In transposition systems, the letter frequencies of a cryptogram will be identical to that of the plaintext. A cryptogram in which the ciphertext letters occur with the expected frequency of plaintext will usually be enciphered by a transposition system.

b. In the simplest substitution systems, each plaintext letter has one ciphertext equivalent. The ciphertext letter frequencies will not be identical to the plaintext frequencies, but the same numbers will be present in the frequency count as a whole. For example, if there are 33 Es in the plaintext of a message, and if E is enciphered by the letter K, then 33 Ks will appear in the ciphertext frequency count.

c. More complex substitution cipher systems, such as the polyalphabetic systems in Part Four of this manual, will keep changing the equivalents. E might be enciphered by a K the first time it occurs and by different cipher letters each time it recurs. This will produce a very different looking frequency count.

d. To illustrate the differences in appearance of frequency counts for different types of systems, examine the four frequency counts in Figure 2-1. Each one is a frequency count of the message listed above it. The four messages are different, but each has the same plaintext. The first shows the plaintext and its frequency count. The second shows the frequencies of the same message enciphered by a transposition system. The third shows a simple substitution system encipherment. The fourth shows a polyalphabetic substitution encipherment.

2-11. Roughness

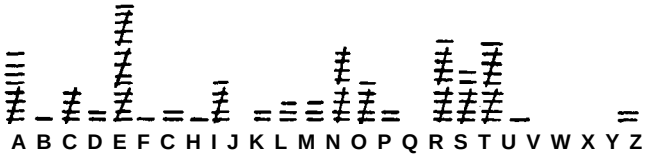
The four examples in Figure 2-l show another characteristic of frequency counts which is useful in system identification. The first three distributions all contain the same letter frequencies. In the first two, the plaintext and the transposition examples, there are 16 Es. In the third, where E has been replaced by W, there are 16 Ws. Where there were 9 As, there are now 9 Ls. Where there was 1 K, there is now 1 C. The first three distributions show the same wide differences between the highest frequency letters and the lowest. The fourth distribution is very different. The distribution lacks the wide differences between the highest and lowest frequency letters. Where the first three showed distinct highs and lows, or peaks and troughs, in the distributions, the fourth is relatively flat.

a. Frequency counts which show the same degree of difference between peaks and troughs as plaintext are considered to be rough distributions. Systems which suppress the peaks and troughs of plaintext letters by changing their equivalents

produce flatter distributions. If letters were selected randomly from the 26 letters of the English alphabet, the resulting distribution would look very much like the fourth example. Random selection will not produce a perfectly level distribution, but it will appear quite flat in comparison to plaintext.

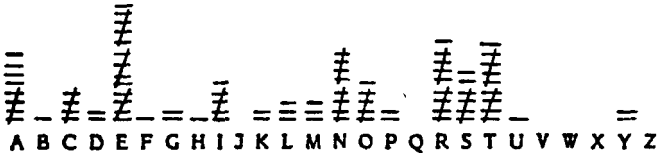
Plaintext:

AERIAL RECONNAISSANCE REPORTS ENEMY REINFORCEMENTS ESTIMATED
AT BATTALION STRENGTH ENTERING YOUR SECTOR PD CLARKE



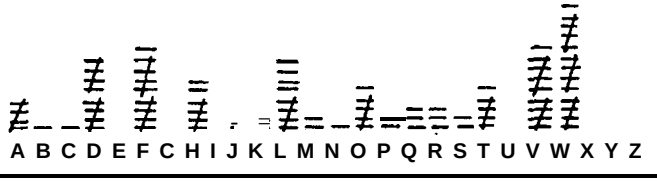
Transposition:

ANRME MTNNO ENEYM AAGGR RAPRE TLTYP IIOEN EIHOD ASRIT DOEUC
LSTNS ANNRL RASFE TSTSA ENEOS BTEER CCNRT ARRCK OEECI TEITE



Simple substitution:

LWVOL QVWAT DDLOH HLMW VWPTV FHWDW RSVWO DNTVA WRWDF HWHFO
RLFWK LFJLF FLQOT DHFVW DMFBW DFWVO DMSTX VHWAF TVPKA· QLVCW



Polyalphabetic substitution:

TARAB CZPNW TNNLL ZEFNM KLNHF OWWQM PEPVM NKRXK QNPRB FXZXE
MBXEO LFJML RWPZS GZXSS EUZYS IXWRV QZFSG FEITT HYHRW EGIKF

Figure 2-l. Frequency count comparison.

b. The simplest substitution systems tend to produce rough distributions. The most secure tend to produce flat distributions. Many other systems tend to fall in between. You can use the degree of roughness as one of the aids to system identification

2-12. Coincidence Tests

Judging whether a given frequency distribution has the same degree of roughness as plaintext or random text is not easy to do by eye alone. To help you make this determination, a number of statistical tests have been developed for your use. The tests are based in probability theory, but you can use the tests whether or not you understand the underlying theories. The most common tests are called coincidence tests.

a. If you pick any two letters from a message, compare them together, and they happen to be the same letter, they are said to coincide. A comparison of the same letters, for example, two As is a coincidence. This comparison can be made of single letters or pairs of letters or longer strings of letters.

b. If you compare two single letters selected at random from the English alphabet, the probability of their being the same is 1 in 26. One divided by 26 is .0385. Expressed as a percentage, 1/26 is slightly less than 4 percent. You would expect to find a coincidence 3.85 times on the average in every 100 comparisons.

c. If you select two letters from English plaintext, however, the probability of their being the same is higher than 1 in 26. Frequency studies have shown that the probability of a coincidence in English plaintext is .0667. In other words, in every 100 comparisons, you would expect to find 6.67 coincidences in plaintext. Each language has its own probabilities, but similar traits occur in each alphabetic language.

d. Different coincidence tests use different methods of comparing letters with each other, but each rests on the probabilities of random and plaintext comparisons. The actual number of coincidences in a cryptogram can be compared with the random and plaintext probabilities to help make judgments about the cryptogram.

2-13. Index of Coincidence

A common way of expressing the results of a coincidence test is the index of coincidence (XC). The index of coincidence is the ratio of observed coincidences to the number expected in a random distribution. For plaintext, the expected index of coincidence for single letters in English is the ratio of .0667 to .0385, which is 1.73.

2-14. Monographic Phi Test

The most common coincidence test is the monographic phi test, which provides a mathematical way of measuring the roughness of a frequency count. *Monographic* is a fancy synonym for one letter. The term monographic distinguishes the test from the digraphic phi test, performed on two letter pairs, and other forms of the phi test. Phi is the English spelling of the Greek letter ϕ. The monographic phi test is based on the coincidence probabilities that occur when every letter in a cryptogram is compared with every other letter in the cryptogram.

a. Fortunately, the phi test can be calculated without actually comparing every letter with every other letter. Both the total number of comparisons and the total number of coincidences can be calculated from the frequency count.

b. The total number of comparisons when every letter is compared with every other letter is the total number of letters multiplied by the total number minus one. Expressed as a formula, it looks like this–

$$\text{Comparisons} = N\,(N - 1).$$

c. Since one out of every 26 comparisons in a random distribution is expected to be a coincidence, the formula for the expected random value of phi is as follows:

$$\phi r = \frac{N\,(N - 1)}{26}$$

or

$$\phi r = .0385\ N\,(N - 1).$$

d. The expected value for plaintext coincidences is–

$$\phi p = .0667\ N\,(N - 1).$$

e. Just as the total number of comparisons is N (N – 1), the total number of coincidences for each letter is f (f – 1), where f is the frequency of the individual letter. The total number of coincidences is the sum of the coincidences for all the letters. The total number of coincidences is labeled phi observed or øo, and can be expressed as either–

$$\phi o = \phi A + \phi B + \phi C + \ldots + \phi Z$$

or

$$\phi o = \Sigma f\,(f - 1).$$

(The Greek letter sigma (Σ) is used to mean sum *of.)*

f. To calculate ϕo, take each letter frequency greater than 1 and multiply it times the frequency minus 1, as the formula suggests. (You can ignore letters with a frequency of 1, because they will be multiplied by 0.) Then add the results of all the multiplications.

g. The index of coincidence for the phi test is called the delta IC. The delta IC is the ratio of phi observed to phi random. It can be expressed using the Greek letter delta (Δ).

$$\Delta IC = \frac{28\ \Sigma f\ (f - 1)}{N\ (N - 1)}$$

h. The results of a phi test can be expressed in terms of ϕ o, ϕ p, and or as the AIC. Where computer support is available to perform the calculations, the AIC is the form usually shown. Where paper and pencil methods are used, either form may be used. Both methods are shown in the next example.

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
f:	3	3	0	7	2	1	1	4	0	0	1	0	0	0	4	1	6	3	0	4	1	0	5	1	0	3
f - 1 :	2	2		6	1			3							3		5	2		3			4			2
f(f-1):	6	6		42	2			12							12		30	6		12			20			6

$$\phi o = \Sigma f\ (f - 1)$$
$$= 6 + 6 + 42 + 2 + 12 + 12 + 30 + 6 + 12 + 20 + 8$$
$$= 154$$

$$\phi p = .0667\ N\ (N - 1)$$
$$= .0667 \times 50 \times 49$$
$$= 183$$

$$\phi r = .0385\ N\ (N - 1)$$
$$= .0385 \times 50 \times 49$$
$$= 94$$

$$AIC = \phi o / \phi r$$
$$= 154/94$$
$$= 1.84$$

2-15. Interpreting the Phi Test

The previous example showed results close to the expected value for plaintext. This indicates the frequency count it was based on had the same approximate degree of

roughness as expected for plaintext. It does not show that it was plaintext or that it was enciphered in a simple substitution system, although the latter is possible. It must be considered as just one piece of evidence in deciding what system was used.

a. In plaintext of 50 to 200 letters, the delta IC will usually fall between 1.50 and 2.00. Shorter text can vary more, and longer text will be consistently closer to 1.73. Since simple monoalphabetic systems have the same frequency distribution as plaintext, these simple systems follow the same guidelines as plaintext.

b. Random text centers around a IC of 1.00 but is subject to the same variability as plaintext. Small samples of under 50 letters vary widely. Samples in the 50 to 200 letter range will usually fall between 0.75 and 1.25. Larger samples approach 1.00 more consistently.

c. Polyalphabetic systems tend to resemble random text, and the more different alphabets that are used, the more likely the AIC is to approach 1.00.

d. The four frequency counts in Figure 2-l follow these guidelines closely. Each one is 100 letters long. The first three, the plaintext, the transposed text, and the simple monoalphabetic substitution each have a AIC of 2.00. The fourth example, the polyalphabetic substitution example, has a AIC of 1.05. The system used in the example has 26 different alphabets, and the underlying plaintext frequencies have been thoroughly suppressed.

PART TWO

Monographic Substitution Systems

CHAPTER 3

MONOALPHABETIC UNILATERAL SUBSTITUTION SYSTEMS USING STANDARD CIPHER ALPHABETS

Section I
Basis of Substitution Systems

3-1. Substitution Systems

The study of analysis of substitution systems begins with the simplest of systems. The systems explained in Part Two are monographic substitution systems. The systems in Chapters 3 and 4 are further categorized as monoalphabetic unilateral substitution systems.

a. Both *monographic* and unilateral mean one letter by their construction. The prefixes *mono-* and *uni-* mean one, and *graphic* and literal refer to letters or other characters. Monographic systems are those in which one plaintext letter at a time is encrypted. Unilateral systems are those in which the ciphertext value is always one character long. Note that the term monographic refers to single plaintext letters and the term unilateral refers to single ciphertext letters.

b. Monoalphabetic systems are those in which a given ciphertext value always equals the same plaintext value. One alphabet is used.

c. Chapter 5 deals with monoalphabetic multilateral systems, which substitute more than one ciphertext character for each plaintext character. Later parts of this manual present the analysis of polygraphic and polyalphabetic systems. Polygraphic systems substitute values for more than one plaintext letter at a time. In polyalphabetic systems, a given ciphertext character will have different plaintext equivalents at different times through the use of multiple alphabets.

d. The techniques used with these simplest of systems carry over to the more complicated systems. Whether or not you will ever see the very simple systems in use, the same skills are used in combination with other techniques to solve more secure systems as well.

3-2. Nature of Alphabets

A cipher alphabet lists all the plaintext values to be enciphered paired with their ciphertext equivalents. Cipher alphabets can take many different forms from a simple listing of 26 letters with 26 equivalent letters to much more complex charts. Chapters 3 and 4 deal with the simple 26 letter for 26 letter types and Chapter 5 introduces some of the more complex chart type multilateral systems.

a. The simple 26 letter for 26 letter cipher alphabets are composed of two sequences of letters: the plain component sequence and the cipher component sequence. The letter sequences can be in standard A through Z order, systematically mixed order, or randomly sequenced. Alphabets are classed as standard, mixed, or random according to the types of sequences they contain. The techniques used to solve the system depend to some extent on the type of alphabet. Alphabets in which both components are standard A through Z sequences are called standard alphabets.

b. A standard sequence does not have to be written beginning with A and ending with Z. A sequence is considered to have no beginning or ending, but continues as if it were written in a circle. The letter that follows Z in a standard sequence is A. Each of the following examples is a standard sequence.

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

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

c. If the alphabetic progression is in the normal left to right order, it is called a direct standard sequence. If the alphabetic progression proceeds from right to left, it is called a reverse standard sequence. Each of the following examples is a reverse standard sequence.

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

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

d. Standard alphabets are also classed as direct or reverse. If the two standard sequences (plaintext and ciphertext) run in the same direction, the alphabet is called a direct standard alphabet. Each of the following alphabets is a direct standard alphabet. Notice that the second one has the identical equivalents to the first and can be rewritten in left to right order without changing its substitution at all.

p: 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: 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

p: z y x w v u t s r q p o n m l k j i h g f e d c b a

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

p: j i h g f e d c b a z y x w v u t s r q p o n m l k

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

e. If the two standard sequences (plaintext and ciphertext) run in opposite directions, the alphabet is called a reverse standard alphabet. Notice that the two following examples of reverse standard alphabets are also equivalent.

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

p: g f e d c b a z y x w v u t s r q p o n m l k j i h
c: 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. An alphabet, in which the plain component is shown in A through Z order, is called an enciphering alphabet. The first alphabet after paragraph 3-2e is an enciphering alphabet. If the cipher component is in A through Z order, it is called a deciphering alphabet. The second alphabet is a deciphering alphabet.

g. Standard alphabet cryptograms are the easiest to solve. The rest of Chapter 3 explains the techniques of cryptography and cryptanalysts of standard monoalphabetic ciphers.

Section II
Monoalphabetic Unilateral Substitution

3-3. Cryptography

The users of a monoalphabetic unilateral substitution system must know three things about the keys to the system. They must know what sequence of letters is used for the plain component, what sequence is used for the cipher component, and how the two components line up with each other. The alignment is termed the *specific key*. Whatever keys are put into use by the originating cryptographer must be known by the receiving cryptographer, too. The key selection must either be prearranged or sent along with the cryptogram itself.

a. Prearranged keys are normally included in published operating instructions, known variously as the Signal Operation Instructions (SOI) or Communications-Electronics Operation Instructions (CEOI). For example, an SOI might specify the use of direct standard sequences for an extended period and a new alignment of the two sequences at regular shorter intervals. A portion of an SOI might look like this example.

31 May 1989, 0001-0600Z

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

31 May 1989, 0601-1200Z

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

Another way to provide exactly the same information in a more abbreviated form is shown below,

31 May 1989

Plain component: Direct standard sequence.
Cipher component: Reverse standard sequence.

0001-0600Z: A p = Q c
0601-1200Z: A p = T c

In this example, the alphabet construction is left to the cryptographer, who writes out the sequences and aligns them with each other according to the specific keys for each key period.

b. Transmitted keys are used whenever the cryptographer is given some choice of the specific key selections. For example, if the alignment of the sequences were left to the cryptographer, the alignment would need to be transmitted. One way to do this is to agree that the first group of the message is always the cipher equivalent of plaintext A repeated five times. This group then tells the receiving cryptographer how to align the alphabet. The example is simple, but more complex systems can be used for greater security.

3-4. Message Preparation

The cryptographer normally prepares a message for encryption by writing the plaintext in regular length groups. Four or five letter groups are common for this type of system.

a. Word lengths are not preserved normally, because they provide strong clues to the plaintext when they appear. It is easier for a cryptanalyst to figure out the plaintext for example 1 in Figure 3-1 than example 2.

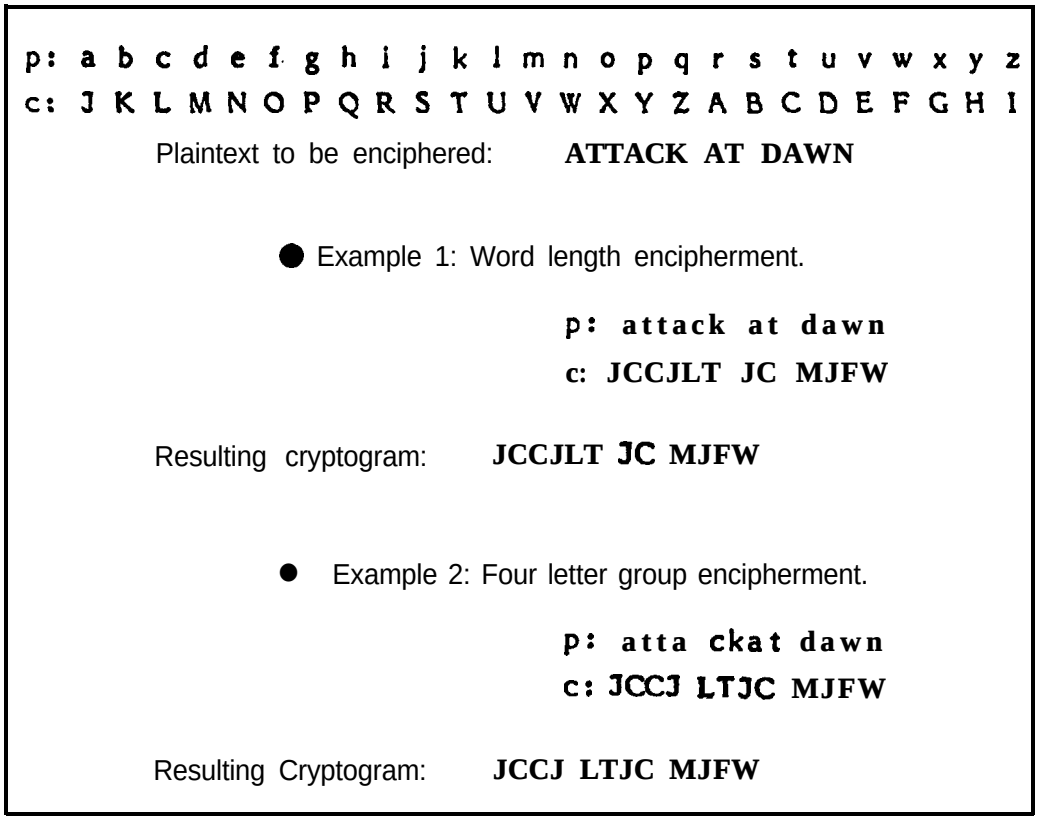


Figure 3-l. Word and group length encipherment.

b. In writing out the message for encipherment with a simple system, any numbers in the text must be spelled out or left in the clear. Punctuation must be spelled out or omitted. At the end of sentences, PD or STOP is often used in English. Commas are replaced by COMMA or CMA.

c. Whenever the text does not break evenly into groups, the text will generally be padded to fill out the groups. The filler letters are usually added at the end of the last group. For clarity, they are often just a repeated low frequency letter such as X or Z. The above cryptogram, broken mto five letter groups, appears below.

JCCJL TJCMJ FWXXX

Section III

Solution of Monoalphabetic Unilateral Ciphers Using Standard Cipher Alphabets

3-5. Methods of Solution

Because of the extreme simplicity of standard alphabets, cryptograms enciphered with them can always be solved. There are two general approaches to solving these simple ciphers. One makes use of the frequency characteristics discussed in Chapter 2. The other uses the orderly progression of the alphabet to generate all possible decipherments from which you can pick the correct plaintext. Each method is explained in the following paragraphs.

3-6. Frequency Matching

The first approach consists of matching expected plaintext letter frequencies with the observed ciphertext letter frequencies.

a. As explained in Chapter 2, monoalphabetic unilateral ciphers preserve exactly the same letter frequencies as found in plaintext. The frequencies occur with the cipher equivalents, not the plaintext letters, but the numbers are unchanged. If E was the most common plaintext letter in a cryptogram, then E's replacement will be the highest frequency ciphertext letter.

b. With standard alphabets, another characteristic is preserved in addition to the individual letter frequencies. The order of highs and lows is also preserved. With a direct standard alphabet, the pattern of peaks and troughs remains, although shifted to the right or left. With a reverse standard alphabet, the pattern also remains, but it runs in the opposite direction. Figure 3-2 illustrates the expected frequency distribution of 100 letters of plaintext. It then shows what happens to the distribution when it is enciphered by a direct and a reverse standard alphabet.

c. As shown in Figure 3-2, there are several recognizable patterns in plaintext. First is the three peak pattern formed by the letters A through I. The pattern is a peak (A), a three letter trough (BCD), a peak (E), a three letter trough (FGH), and a peak (I). The second easy to recognize pattern is formed by the letters N through T. The pattern is a double peak (NO), a trough (PQ), and a triple peak (RST). When you compare the plaintext distribution with the two ciphertext distributions, the patterns are still evident.

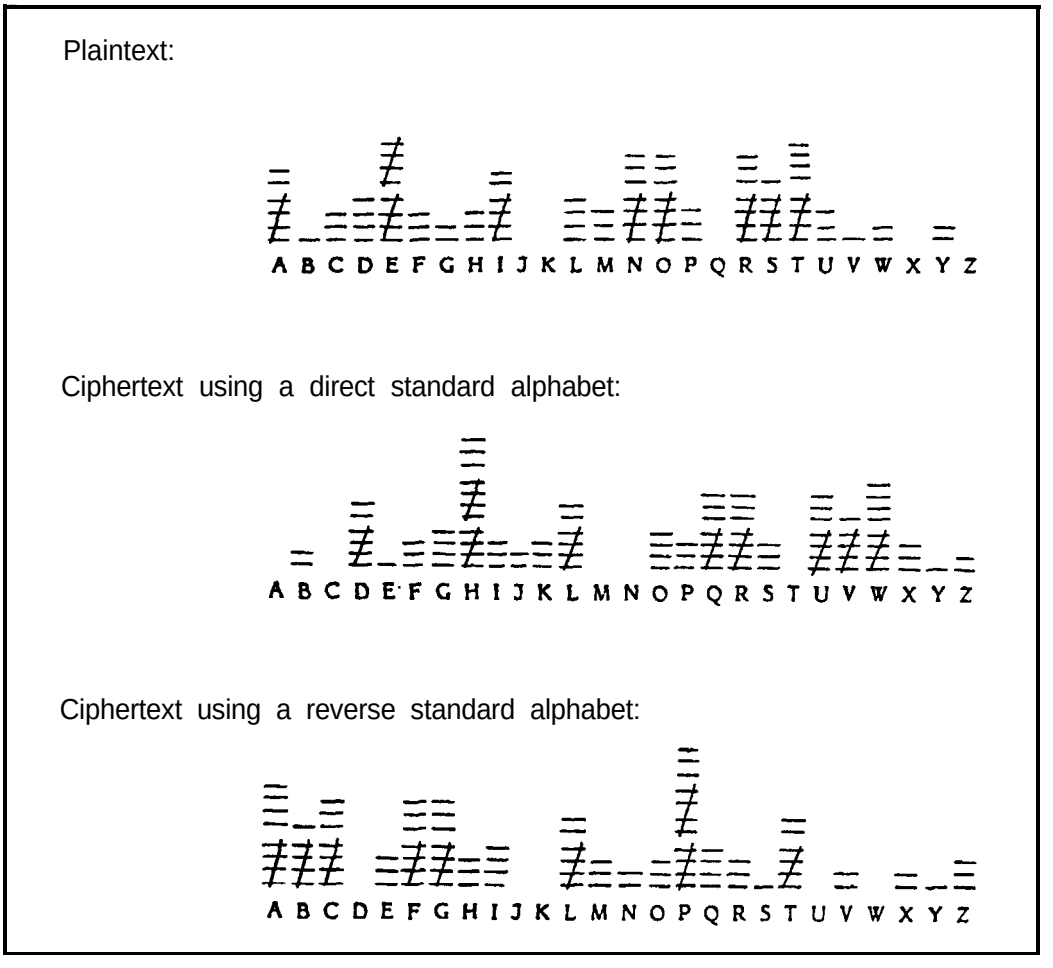


Figure 3-2. Frequency distributions.

d. Not all plaintext frequency distributions show the patterns clearly. The examples in Figure 3-2 show a perfect 100 character frequency distribution with every letter appearing exactly as many times as expected. Actual frequency counts will vary considerably, particularly with small samples. It is easier to recognize the overall patterns by their frequency than it is to recognize individual letters, however. If you can recognize even a partial pattern, it is easy to write the whole alphabet and see if the frequencies are close to expectations. Consider the cryptogram shown below.

CDRDC IPRIS JGXCV EPHII LDUDJ GWDJG HXXXX

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

The four Xs at the end are almost certainly fillers, so they are not counted. The cryptogram is too short for the complete pattern to appear. The cluster of higher frequency letters from C through I could represent the N through T pattern, though. We will write the full sequence of letters on that assumption.

```
p: 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
c: 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 frequency match fits the plaintext letters reasonably well. E does not appear at all, but other vowels make up for it, keeping the vowels near the expected 40 percent. No low frequency letters appear with unexpectedly high frequency. The confirmation of the match occurs when the alphabet is tried with the cryptogram.

```
nocon tactd uring pastt wofou rhour s
CDRDC IPRIS JCXCV EPHII LDUDJ GWDJG HXXXX
```

or

NO CONTACT DURING PAST TWO FOUR HOURS

e. This method depends on knowing or suspecting that standard alphabets are used. With a long message, the frequency count will usually make it obvious. The A-E-I and the NO-RST peaks will stand out. With a short message like the above example, it is not obvious, but it is an easy step to try if you think you spot a partial match.

3-7. Generating All Possible Solutions

The frequency matching technique only works if the text is long enough to produce a recognizable frequency count. A second technique always leads to the solution. With a known standard alphabet, there are only 26 different ways the alphabet can be aligned. It does not take very long to try all 26 settings to find the correct solution.

a. As an example, consider the solution of the following cryptogram.

SIZUX VJFLK

With no repeated letters, frequency matching is not likely to help. Suppose the alphabet was a direct standard with p:a=c: Z.

```
p: a b c d e f g h l j k l m n o p q r s t u v w x y z
c: 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
```

Using the above alphabet, SIZUX VJFLK deciphers as TJAVY WKGML. Obviously, this is not the correct plaintext. The text the trial decipherment produces is called *pseudoplaintext* or *pseudotext.* Suppose the alphabet used p:a=c:Y.

p: a b c d e f g h i j k i m n o p q r s t u v w x y z
c: 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

This alphabet produces UKBWZ XLHNM.
The next alphabet with p:a=c:X gives the text VLCXA YMION.
The next alphabet with p:a=c:W gives the text WMDYB ZNJPO.
The next alphabet with p:a=c:v gives the text XNEZC AOKQP.

Clearly, not one of these is the correct setting, but notice the effect of trying each alphabet in turn. The columns of letters from each successive trial alphabet are in alphabetical order. You can achieve the same effect as trying each alphabet in turn by listing the letters vertically in alphabetical order. Figure 3-3 lists the results of trying all possible alphabets.

SIZUX VJFLK

TJAVY WKGML
UKBWZ XLHNM
VLCXA YMION
WMDYB ZNJPO
XNEZC AOKQP
YOFAD BPLRQ
ZPGBE CQMSR
AQHCF DRNTS
BRIDG ESOUT
CS JEH FTPVU
DTKFI GUQWV
EULGJ HVRXW
FVMHK IWSYX
GWNIL JXTZY
HXOJM KYUAZ
IYPKN LZVBA
JZQLO MAWCB
KARMP NBXDC
LBSNQ OCYED
MCTOR PDZFE
NDUPS QEACF
OEVQT RFBHG
PFWRU SGCIH
QGXSV THDJI
RHYTW UIEKJ

Figure 3-3. All possible **decipherments.**

The plaintext, BRIDGES *OUT,* appears about halfway down the columns. In practice, you would only write enough to recognize the plaintext. Generally, write a column at a time, and only write as many columns as you need. Once you have spotted plaintext, set up the alphabet and complete the decipherment.

b. With a reverse standard alphabet, another step must be added. You cannot generate the columns until you try deciphering first at any alphabet setting of your choice. Then generate the columns starting with your trial decipherment. As you will see in the following chapters, this technique can be used with any known alphabets, not just standard ones. The procedures, which will be illustrated in Chapter 4, are—

- Set up the known alphabet at any alignment.
- Perform a trial decipherment to produce pseudotext.
- Using the trial decipherment as the letters at the head of the columns, generate all possible decipherment by listing the plain component sequence vertically for each column.

CHAPTER 4

MONOALPHABETIC UNILATERAL SUBSTITUTION SYSTEMS USING MIXED CIPHER ALPHABETS

Section I
Generation and Use of Mixed Cipher Alphabets

4-1. Mixed Cipher Alphabets

Mixed cipher alphabets differ from standard alphabets in that one or both sequences are mixed sequences. A mixed sequence is any sequence not in normal alphabetical order. The two main types of mixed sequences are systematically mixed and random mixed sequences.

a. Systematically mixed sequences are produced by an orderly process based on easily remembered keywords, phrases, or simple rules. There are a number of mixed sequence types, which will be explained in this section. Their advantage is that the keys can be easily memorized and reconstructed for use when needed. Their disadvantage is that the orderliness in construction can be used by the opposing cryptanalyst to aid in their recovery.

b. Random mixed sequences are not based on any orderly generation process. They can be produced by various means ranging from pulling the 26 letters out of a hat to complex machine generation. Their advantage is that their structure offers no help to the opposing cryptanalyst. Their disadvantage is that the keys cannot be memorized easily or produced from simple directions as systematically mixed sequences can. They must be printed out in full and supplied to every user.

4-2. Keyword Mixed Sequences

One of the simplest types of systematic sequences is the keyword mixed sequence. The sequence begins with the keyword, which may be a word or a phrase. Any letters repeated in the keyword are used only once, dropping the repeating letters. After the keyword, the rest of the letters are listed in alphabetic order, omitting those already used.

Keyword— **CRYPTOGRAPHIC**

Repeated letters dropped: **CRYPTOGAHI**

Remaining letters added in normal order:

CRYPTOGAHIBDEFJKLMNQSUVWXZ

Keyword- **MILITARY INTELLIGENCE**

Repeated letters dropped: **MILTARYNEGC**

Remaining letters added in normal order:

MILTARYNEGCBDFHJKOPQSUVWXZ

4-3. Transposition Mixed Sequences

Transposition mixed sequences are produced by writing a letter sequence into a matrix and extracting it from the matrix by a different route. The most common types are called simple columnar, numerically keyed columnar, and route transposition sequences.

a. Simple columnar transposition is usually based on a keyword mixed sequence. The keyword determines the width of the matrix that is used. The keyword is written as the first row of a matrix and the rest of the sequence is written beneath it, taking as many rows as necessary. The transposition mixed sequence is then produced by extracting the columns of the matrix from left to right.

Keyword- **ARTILLERY**

Keyword mixed sequence in matrix:

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

Resulting sequence:

ABKURCMVTDNWIFOXLGPZEHQYJS

Keyword- **MORTAR**

Keyword mixed sequence in matrix:

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

Resulting sequence:

MBGLUZOCHNVRDIPWTEJQXAFKSY

b. The numerically keyed columnar transposition mixed sequence differs from the simple columnar only in the way it is extracted from the matrix. Instead of extracting the columns left to right, the order of the columns is determined by a numerical key based on the keyword. After constructing the matrix, the letters in the keyword are numbered alphabetically. The columns are then extracted according to the resulting numerical key.

Keyword- **CALIFORNIA**

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

Resulting sequence:

ADQZCBPYFHUIGTLESNMXOJVRKW

Keyword- **VERMONT**

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

Resulting sequence:

EBJWMDLYNGQOFPZRCKXTHSVAIU

c. Route transposition sequences are formed by any other systematic way of entering sequences into a matrix and extracting them from a matrix. They can be based on standard or keyword mixed sequences. The samples in Figure 4-l show some of the common routes that can be used. The last two omit the letter J for the convenience of a square matrix.

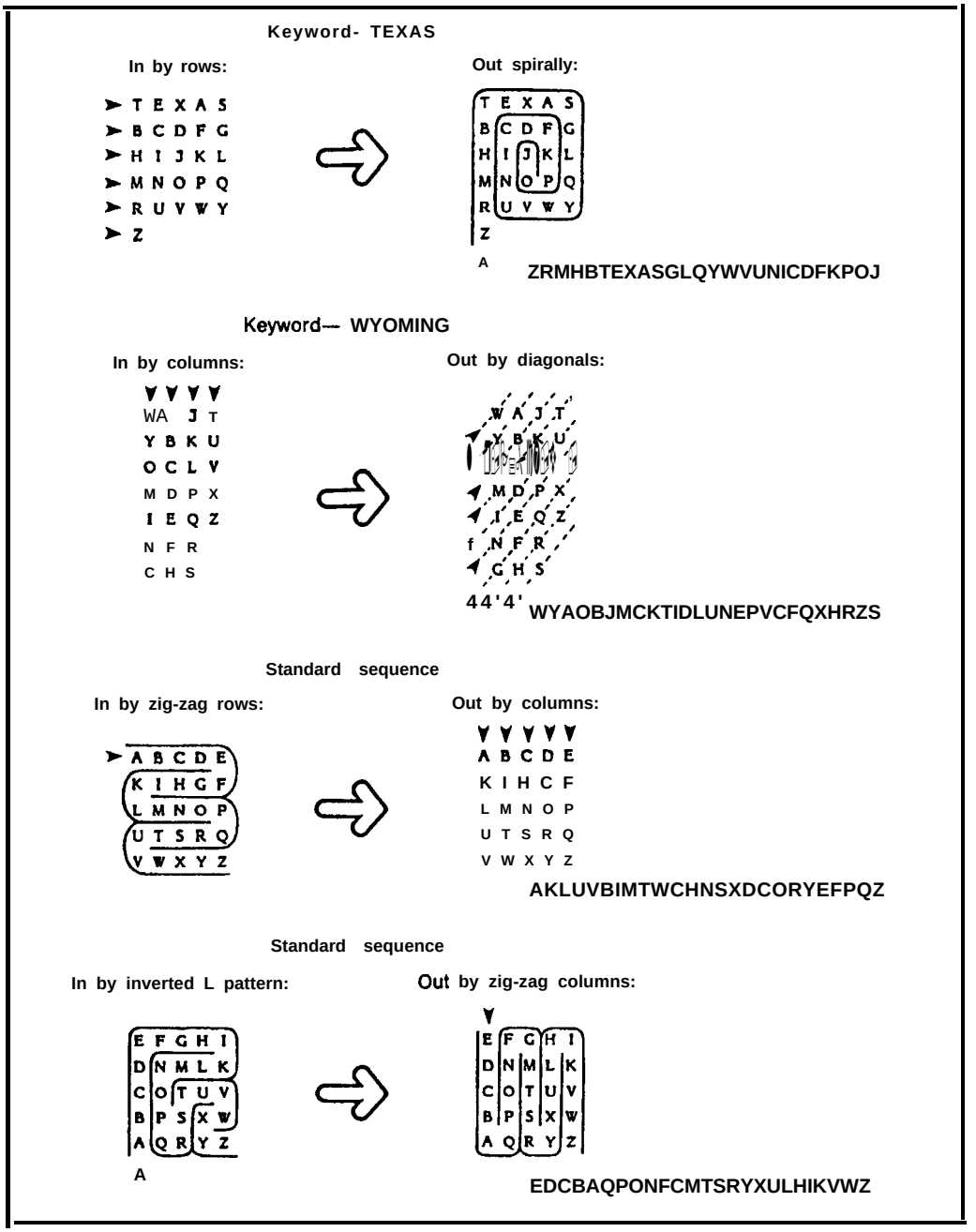


Figure 4-1. Route transposition.

4-4. Decimation Mixed Sequences

Decimation mixed sequences are produced from a standard or keyword mixed sequence by counting off letters at a regular interval.

a. As an example, consider decimating a standard sequence at an interval of 3. The new sequence begins with the first letter of the basic sequence, in this case, A. The second letter of the new sequence is the third letter that follows from the basic sequence, D. Every third letter is selected until the end of the basic sequence is reached.

Basic sequence:

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

Resulting decimated sequence:

A D C J M P S V Y . . .

The count then continues as if the sequence were written in a circle. The next letter after Y, skipping Z and A, is B. The complete resulting sequence is shown below.

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

b. The interval should have no common factors with the length of the sequence. Since any even number has a common factor of 2 with 26, only odd numbers are selected with 26 letter sequences. Intervals with common factors are not selected, because the count will return to the starting point again before all the letters are used. The interval should also be less than half the length of the sequence, because larger numbers will just duplicate in reverse order the sequence produced by a smaller number. An interval of 23, for example would produce the same sequence as an interval of 3, but in the reverse order. For a 26 letter sequence, the only usable intervals are 3, 5, 7, 9, and 11. By counting either left to right or right to left, all the basic decimated sequences can be produced.

c. Study of this method of decimation is particularly significant, because the solution of some types of polyalphabetic ciphers can yield sequences in a decimated order instead of the original order.

d. An alternate method of decimation is occasionally encountered. In the alternate method, each letter is crossed off as it is selected and that letter is not counted again. The restrictions on intervals do not apply to this method, because the starting letter can never be reached again. This method is used less, because it is subject to mistakes in the counting process that are hard to detect and correct.

4-5. Types of Mixed Cipher Alphabets

As mentioned at the beginning of this section, a mixed alphabet is any alphabet that uses one or more mixed sequences. The simplest types are those which use a standard sequence in one component and a mixed sequence in the other. These are the easiest for a cryptanalyst to reconstruct. Next in order of difficulty are those in which the same mixed sequence is used in the plain and cipher components. Most difficult are those in which two different mixed sequences are used. The next section shows how to recover each of these types of alphabets.

Section II
Recovery of Mixed Cipher Alphabets

4-6. Alphabet and Plaintext Recovery

Although this manual separates the techniques of alphabet recovery from plaintext recovery, the two processes will usually occur simultaneously, each supporting the other. When an orderly structure is found in an alphabet as individual letters are recovered, the orderly structure often helps make more plaintext recoveries. The techniques explained in this section will be used in the next section.

a. You usually begin reconstruction by recording recoveries in the form of an enciphering alphabet. An enciphering alphabet is one in which the plaintext component is arranged in A through Z order. Ciphertext letters are written in the cipher component paired with their plaintext equivalents in the plain component. The plaintext can be either the top or bottom letters, but whichever you select, you should follow it consistently in the alphabet as well as the cryptogram. Inconsistency leads to errors. In this manual, plaintext is placed above ciphertext.

b. A deciphering alphabet is one in which the ciphertext is written in A through Z order. Rearranging the alphabet into deciphering order is sometimes helpful in alphabet recovery.

c. Whenever systematically mixed alphabets are used, you should attempt to recover the systems and keys in use. The same sequences are often reused, either at different alignments of the same alphabet or in combination with other sequences. The solution can be reached much quicker when you recognize and take advantage of previous recoveries.

4-7. Reconstruction of Alphabets With One Standard Sequence

Whenever one of the two sequences is a standard sequence, recovery of the system used to produce the other sequence is made much easier.

a. The easiest type to recognize is the keyword mixed sequence. Any keyword mixed sequence has two parts-the keyword and the alphabetic progression. If you find that recovered letters are falling in alphabetic progression consistently in a portion of the sequence, it is probably a keyword mixed sequence. In this case, you can narrow down the possibilities of unrecovered letters. Consider the following partially recovered alphabet.

p: 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: s Z V T H D F G I

(1) The letters DFGI appear to be part of the alphabet section of the cipher sequence. The alphabetic progression continues at the left with the letters S and Z. All the other recovered letters appear to be part of the keyword. Between the H and the D there is room for only two of the letters at the beginning of the alphabet-A, B, and C. At least one of these must be in the keyword, leaving the other two as probable equivalents of plaintext P and Q. Similarly, there is space for only three letters between S and Z. T and V already appear, so the spaces must be filled by three of the four letters, U, W, X, and Y. Given these limitations, recovery of more plaintext is likely. Continuing the example, consider that plaintext C, F, L, P, W, and Y are recovered next.

p: 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: s X Z L V 0 T H B DFGI K P

(2) These recoveries enable several more probable letters to be placed by alphabetical progression.

p: 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: S X Y Z L V 0 T H B C D F C I J K P

(3) At this point, we can see that A and E must be in the keyword, because there is no room for them in the alphabetic progression. U or W must be in the keyword, because there is only room for one of them between S and X, and V is already placed. Similarly, M or N and Q or R must be in the keyword. Q is unlikely, even though U is available to pair with it. Placing Q and U anywhere in the blanks in the keyword suggests nothing further. R must be in the keyword, then.

(4) The letter after L in the keyword must certainly be a vowel or the keyword would be unpronounceable, and that vowel represents plaintext G. With the possibilities narrowed down this far, you might be able to spot the keyword

without referring back to the cryptogram that produced the partially recovered alphabet. The complete alphabet looks like this.

p: 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 t
c: 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

b. Recovery of decimated sequences is a straightforward process of trying out intervals. Just as a decimated sequence is produced by counting at a regular interval, the original sequence can be recovered by counting at a regular interval, too. A partially recovered alphabet with a suspected decimated sequence in the cipher component could look like this example.

p: 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: N . . . D . . . X . . F . W H . . M V . . . K . . .

(1) To determine if this is a decimated sequence, various intervals can be tried. The recovered letters suggest one obvious possibility. The letters V, W, and X all appear among the recovered letters. If they were in order in the base sequence used to generate the decimated sequence, they should reveal the interval. The interval from V to W and from W to X is -5 in each case. A trial decimation at -5, beginning with V produces the following sequence.

VWX. ..H.D . . . . . N..F..KM . . . .

(2) This sequence of letters appears to be a keyword mixed sequence. The keyword appears after the VWX and alphabetic progression resumes with the F and the KM. Once you recognize this structure, you can use it to assist in further plaintext recoveries just as in the first example shown in paragraph 4-7a. The original basic sequence used to produce the decimated sequence is shown below.

RHODEISLANBCFCJKMPQTUVWXYZ

c. Simple transposition mixed sequences often resemble decimated sequences. You will often see a regular spacing of adjacent low frequency letters, just as we saw VWX in the previous example. This is not caused by a decimation interval, but by the regular length of columns separating the letters. Recovery of the generation method of transposition mixed sequences is accomplished by rebuilding the original matrix.

p: 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: u F O V P X K Y I R Z G D T E

The almost regular spacing of the letters V, X, Y, and Z resembles a decimated sequence, but the interval is not constant. This almost, but not quite, regular spacing is an indication of simple columnar transposition. The letters V, X, Y, and Z are probably the bottom letters in their columns of the original matrix. W, which has not been recovered, probably occurs in the keyword, because there does not appear to be room for a column ending with W. Analysis of this type of sequence proceeds by rebuilding the columns. Placing the letters V, X, Y, and Z in sequence with their preceding letters as their columns, produces this partial result.

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
U . F 0 V/. . P X/. . K . Y/I . . R Z/C D . T . E .

U		.	I
.	.	.	.
F	.	K	.
O	P	.	R
V	X	Y	Z

Now the initial reconstruction appears successful. The rows above VXYZ also show alphabetic progression developmg. Q can be inserted in the next to last row with confidence. The next step is to place the rest of the letters into columns that would continue the structure in a logical way. A little trial and error will show that the columns before the V column end with T and U. The U was not the top of the V column, but the bottom of the preceding column.

p: 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: U/. F 0 V/. . P X/. . K Q Y/I . . R Z/G D . T/. E .

				.	I
G	.	.	.	.	.
D	E	F	.	K	.
.	.	O	P	Q	R
T	U	V	X	Y	Z

The longer columns belong on the left. Shifting these columns produces this result.

.	I	G	.	.	.
.	.	D	E	F	.
K	.	.	.	O	P
Q	R	T	U	V	X
Y	Z				

The matrix is now in its original form. L, M, and N can be placed between K and 0. Either H or J can be inserted between F and K and the remaining letter belongs in the keyword in the top row. S and W are in the keyword, because they are missing from the alphabetical progression. That leaves A, B, or C for the remaming letter of the keyword, with the other two on the second row. Since only one vowel has been found in the keyword up until now, A probably belongs in the keyword with B and C filling the blanks in the second row. Trial placements of A, S, and W together in the first row blanks, together with either H or J in the remaining space leads to the conclusion of JIGSAW as the keyword.

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

d. The recovery of numerically keyed columnar transposition sequences is the same as for simple columnar transposition, except the columns are not in order in the sequence. The next example shows the recovery of this kind of transposition mixed sequence.

p: 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: X M D B Z P . T Y . . S U I R W . C O V J . L . H .

This problem is again best approached through the end of alphabet letters. V, W, X, Y, and Z have all been recovered, and they make a good starting point. V, W, X, Y, and Z are placed in a row with their preceding letters above them in columns.

p: a b c d c f g h i j k l m n o p q r s t u v w x y z
c: X/M D B Z/P . T Y . . S U I R W/. C O V/J . L . H .

.	U	.	P	M
C	I	H	.	D
O	R	.	T	B
V	W	X	Y	Z

This time no alphabetic progression appears, even if we consider that one or two of the columns might be misplaced. In this case, the next thing to consider is that the sequence may be reversed. Selecting the letters to the right of V, W, X, Y, and Z instead of the left produces the following example.

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
X M D B/Z P . T/Y . . S U I R/W . C O/V J . L . H .

L	O	B	S	T
.	C	D	.	.
J	.	M	.	P
V	W	X	Y	Z

This setup is clearly correct. Next, we add the two short remaining segments.

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
X M D B/Z P . T/Y . . S/U I R/W . C O/V J . L/. H ./

		L	O	B	S	T
.	R	.	C	D	.	.
H	I	J	.	M	.	P
.	U	V	W	X	Y	Z

Moving the short columns to the right and filling in the missing letters produces the following matrix.

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

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

The final step is to recover the numerical key. If normal methods are used, it should be produced by the keyword and should show the actual order in which the columns were extracted. Numbering the letters in the keyword in alphabetical order and comparing them with the cipher sequence in the alphabet confirms that this method was used. Since the sequence was reversed, the order of columns in the cipher sequence appears in right to left order beginning with the cipher letter B.

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

3 4 1 6 7 2 5

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

e. One type of transposition sequence remains to be considered. When a route transposition process is used, the solution is to try to reconstruct the original routes. In examining attempts to solve the matrix by rebuilding columns, be alert to entry routes other than by rows. Look for spirals, diagonals, and alternate horizontals or verticals. If rebuilding the columns produces no results, consider rebuilding spiral, diagonal, or alternate row or column routes. This manual does not show examples of these approaches, but if you encounter this situation, approach it logically and try various approaches until one succeeds. The techniques of solving route transposition ciphers explained later in this manual will help in this process.

f. Each of the preceding examples was approached as if we knew, perhaps from past history, what types of sequences were used. We assumed that the plain component was a standard sequence, and the cipher sequence could then be readily reconstructed by itself. It is common, in approaching a cryptanalytic problem, to assume the simplest case and only to move on to more complex possibilities when the simplest case must be rejected. A great deal of time can be wasted by assuming something is more complicated than it is.

g. The next simplest case is where the cipher sequence is a standard sequence and the plain sequence is mixed. When reconstruction attempts fail because you started with an enciphering alphabet, rearranging the alphabet into a deciphering alphabet may yield results. Once rearranged, the solution is approached just as we did in the above examples. Look for short alphabet progression to indicate keyword mixed sequences. If that is not found, see if a decimation was used. If decimation was not used, try reconstructing the columns of a columnar transposition. Remember to try forward and reversed sequences.

h. If none of these approaches yields results, either with an enciphering alphabet or a deciphering alphabet, other approaches are called for. Either there are two mixed sequences, a more complex process was used, or random sequences were used.

4-8. Reconstruction of Alphabets With Two Mixed Sequences

Recovering alphabet structure when both sequences are mixed is more difficult than the previous examples. You are much less apt to be successful with only partial recoveries. Where the alphabet could be reconstructed during the solution of the plaintext in the previous examples, reconstruction of an alphabet with two mixed sequences must usually wait for the full solution of the plaintext. The examples in this section will begin with a fully recovered, but not reconstructed, alphabet.

a. The easiest type to recover with two mixed sequences occurs when both sequences are keyword mixed, as in the next example.

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

p: i f n j l q k s t u v w x y z g o m p h e r a b c d
c: 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

Enciphering and deciphering forms of the same alphabet are shown. The underlined portions show substantial alphabetic progression in both, which is typical of alphabets with keyword mixed sequences. A transposition or decimation would not

produce such an obvious progression. The underlined portions in both alphabets are probably in their original form. The remaining plain-cipher pairs are out of order. Your task is to reconstruct the original order. The usual approach at this point is to try to extend the alphabetic progression outward from the obvious progression. In this case, the enciphering alphabet shows two long alphabetic strings of cipher letters, HIJKLMNO and WXYZ, which must have some or all of the letters PQRSTUV in between. Similarly, the deciphering alphabet shows plaintext strings ABCD and STUVWXYZ, and some or all of the letters EFGHIJKLMNOPQR must be in between. Suppose the cipher letters PQRSTUV belong in exactly that order. If that is the case, then the plaintext letters GOMPHER must also be in the right order, preceding ABCD. We expect to find the keyword immediately before the beginning of the alphabetic sequence. GOMPHER, while not a recognizable word may be close to it. If we try GOMPHER as a keyword, then the remaming letters must be in alphabetical order. Adjusting the alphabet so GOMPHER is a trial keyword will produce this arrangement.

p: f i j k l n q s t u v w x y z g o m p h e r a b c d
c: B A D G E C F H I J K L M N O P Q R S T U V W X Y Z

Now the cipher sequence shows a recognizable word, BADGE, but the solution is incomplete. If we move the M-R pair so that plaintext M fits in alphabetic order instead of the keyword, we see the following alphabet.

p: f i j k l m n q s t u v w x y z g o p h e r a b c d
c: B A D G E R C F H I J K L M N O P Q S T U V W X Y Z

This rearrangement is the original sequence of the alphabet.

b. When transposed or decimated sequences are used in the alphabet, the solution is much more difficult. The alphabetic progression used in the previous example is not available to assist with reconstruction. A solution is still possible in many cases, however. When both sequences are the same sequence in the same direction, the alphabet can often be recovered quite readily.

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

(1) Reconstruction begins with a process called chaining. Use the plain-cipher pairs to create a 26 letter chain by linking the cipher letter of each pair to the pair with the same plaintext letter. Any pair can be used as the starting point. Beginning with the plaintext A-ciphertext L pair (abbreviated Ap-Lc) next find plaintext L. Plaintext L equals ciphertext W (Lp-WC), producing a partial

chain of ALW. Continuing with Wp-Bc, the chain is extended to ALWB. Continue adding links to the chain until you return to the original letter A. The complete chain is shown below.

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

(2) Since we were able to produce a 26 letter chain, there is a strong indication that the same sequence was used in both components. With different sequences, the chances of producing such a chain are very low. Unrelated sequences will almost always return to the starting point before using all 26 letters. The alphabet in paragraph 4-8a, for example, produces separate 23 and 3 letter chains.

(3) The sequence produced by chaining an alphabet with two identical sequences in the same direction will always either be the original sequence or a decimation of the original sequence. This narrows the possibilities for the original sequence down to six. The chained sequence and its five possible decimations are listed below.

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

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

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

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

Decimation 9:
A M U L Z J W H V B S Y Q E O K I R G T F X D P C N

Decimation 11:
A H O X U B I P Z Y G N W E F M V K D L S R C J Q T

(4) If the original sequence was a decimated sequence, the basic keyword or standard sequence used to generate the decimated sequence would be one of the above, Smce none of them are either standard or keyword mixed, the original sequence was probably transposed. Approaching each sequence above with transposition in mind, the letters V, W, X, Y, and Z have been underlined in each, searching for a basis to rebuild the columns. The last sequence (decimation 11) yields the following matrix.

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

(5) When the same sequence is used in the same direction in both components of the alphabet, a 26 letter chain will only be produced half of the time. When the two sequences are staggered by an odd number of letters, a 26 letter chain results. When the two sequences are staggered by an even number of letters, two separate 13 letter chains result. These can sometimes be recovered, too, but the solution is more difficult.

c. The chaining technique can also be used with alphabets with different sequences in the two components if they are reused at different alignments. Consider the next two alphabets, recovered at different times on the same day.

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

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

(1) To test if the same alphabet was used, chain the cipher sequences against each other. In the example, chain A of the first to T of the second, T of the first to N of the second, and so on. This produces the following chain.

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

(2) This confirms that the two alphabets used the same sequences at different alignments. If chaining produced anything but one 26 letter sequence or two 13 letter sequences, they are not the same alphabet.

(3) Write all possible decimations, as before.

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

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

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

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

Decimation 9:
A L Y T K F N X S J D O W R I C M V Q H B E Z U P G

Decimation 11:
A X I E Y J M U K O Q C N R B L S C Z T D V P F W H

(4) The decimation of 7 produces a sequence that almost looks as if it were the original. This can happen when the decimation interval and the column length of a transposed sequence are the same except for one long column. The correct sequence is a decimation of 9 read in reverse.

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

The sequence used to generate the simply transposed sequence was a keyword mixed sequence based on LEMON.

(5) The plaintext component can be reconstructed now that the correct ciphertext sequence is known. We start with the decimated sequence. Since the sequence with a decimation of 9 was used in reverse to recover the keyword LEMON, we will list it in reverse.

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

Either of the two alphabets given at the start of this problem can be used to reconstruct the plaintext sequence. The first alphabet is repeated for reference.

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

We now rearrange this alphabet so that the cipher sequence is in the same order as the recovered decimated sequence.

p: e b c d f h i j k l m p q s t u v w x y z o r a n g
c: G P U Z E B H Q V M C I R W O D J S X N F K T Y L A

d. The chaining techniques introduced in this section are also used in the solution of polyalphabetic ciphers. They will be further developed in Part Four.

Section III
Solution of Monoalphabetic Unilateral Ciphers Using Mixed Cipher Alphabets

4-9. Preparation for Analysis

The first step in approaching the unsolved cryptogram is to prepare a worksheet.

a. If prepared by hand, one-fourth inch or one-fifth inch cross section paper (graph paper) should be used if possible. Hand lettering should be clearly printed in ink. The cryptogram should be triple spaced vertically to leave room for writing. If a copying machine is available and local security rules permit, the worksheet should be copied after preparation to permit a restart with a clean worksheet whenever needed.

b. Generally, you will want to prepare at least a unilateral frequency count. Other special frequency counts may be needed also, as will be explained later. If you are unsure of system identification, you may want to calculate the ϕIC. Computer support, if available, can save a lot of time at this step.

c. Next, you should scan the text searching for repeated segments of ciphertext. Underline all repeats you find of at least three letters in length. You may find it useful to underlme two letter repeats, too.

d. If you have more than one cryptogram that appears to have been enciphered with the identical system, prepare a worksheet for each. Compare peaks and troughs of frequency counts to see if they are similar. If so, look for repeats between messages as well as within messages. Repeats between messages are another indication that the identical system was used. The more repeats you find, the easier the solution will be.

e. If you are still in doubt whether two cryptograms have been enciphered by the same system, there is a simple statistical test available, similar to the phi test. The chi test or cross product test compares two frequency distributions to determine the probability that they are from the same alphabet. The frequency of each letter in one distribution is multiplied by the frequency of the same letter in the other distribution. The results of all the multiplications are added to produce the chi value. Chi is the Greek letter that looks like an X. The formula for the chi value is—

$$X = \Sigma\ (f)(f2).$$

The expectation with a random match is 1/26th of the product of the total letters of each, or—

$$Xr = .0385\ (N1)(N2).$$

With a correct match, the expected value is .0667 times the products of the total letters, or—

$$Xp = .0667\ (N1)(N2).$$

The results can also be expressed as an index of coincidence, the usual form if produced by computer support. The formula for the cross IC, as it is called is—

$$X\ IC = \frac{Xo}{Xr} = \frac{26\ \Sigma\ (f1)(f2)}{(N1)(N2)}.$$

With a correct match, the expected IC value, as with the phi text is 1.73. If you match two alphabets and the X IC is close to 1.73, the chances are that they were enciphered with the same alphabet. Figure 4-2 illustrates a completed chi test.

PROBLEM: To determine If the two frequency counts below were from cryptograms enciphered with the same alphabet.

	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	
c1:	-	3	2	6	1	13	-	3	3	-	3	-	6	2	3	3	4	1	-	-	10	-	1	-	4	1	N=69
c2:	4	2	1	6	1	7	-	4	-	1	2	-	5	1	3	4	4	3	-	1	8	-	1	1	2	-	N=61

Product:

- 6 236 1 91 - 12 - - 6 - 30 2 9 12 16 3 - - 80 - I - 6 -

Xo = Σ (f)(f2) = 6 + 2 + 36 + . . . + 8 = 315

Xr = .0385 (N1)(N2) = .0385 (69)(61) = 162

X IC = Xo/Xr = 315/162 = 1.94

The results indicate the same alphabet Was used.

Figure 4-2. Chi test.

f. As with any statistical test, you should use this as a guide only, and take all other available information into consideration, too, For example, if you find several long repeated segments of text between two cryptograms, it is probably a waste of time to calculate a chi test by hand. You already have the evidence you need to make a decision as to what approach you will use to reach a solution.

4-10. Approaches to the Solution

There are two basic approaches to the solution-the probable word method and the brute force approach. The probable word method is to try to gain a quick entry into the system by correctly assuming a portion of the plaintext. The brute force approach is to systematically narrow down the possible keys to the system and then force a solution by exhaustively trying all those possible keys. The method in the previous chapter of solving standard alphabet systems through trying all possible decipherment is a good example of the brute force approach. In practice, the solution of any given system is likely to use a combination of the two approaches.

4-11. Solution With Known Sequences - Completing the Plain Component Sequence

When the sequences used in an alphabet are known, a quick forced solution is possible.

a. Although mixed alphabets are used instead of standard ones, the solution is exactly the same as that explained in paragraph 3-7b.

(1) Set up the known alphabet at any alignment.

(2) Perform a trial decipherment (pseudotext).

(3) Using the trial decipherment as the letters at the head of the columns, generate all possible decipherment by listing the plain component sequence vertically for each column.

b. Figure 4-3 illustrates the solution of a cryptogram with known sequences using the above steps.

Solve: LIZWF QFMYK LOILX

Plain component-keyword mixed sequence based on SEA URCHIN.

Cipher component-standard sequence.

Step 1. Set up the alphabet at any alignment.

p: s e a u r c h i n b d f g j k l m o p q t v w x y z
c: 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

Step 2. Perform a trial decipherment.

p: fnzwc mcgyd fknfx
c: LIZWF QFMYK LOILX

Step 3. Complete the plain component sequence.

FNZWC MCGYD FKNFX
CBSXH
JDEY I
KFAZN
LGUSB
MJRED
OKCAF
PLHUG
QMIRJ
TONCK
VPBHL
WQDIM
XTFNO
YVGBP
ZWJDQ
SXKFT
EYLCV
AZMJW
USOKX
REPLY BYCOU RIER
CAQMZ
HUTOS
I RVPE
NCWQA
BHXTU
DIYVR

p: s e a u r c h i n b d f g j k l m o p q t v w x y t
c: 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

Plaintext: REPLY BY COURIER

Figure 4-3. Completing the plain component.

4-12. Probable Word Method

The probable word method of solution depends on your being able to correctly identify a portion of the plaintext. When you can do this, you can begin to reconstruct the keys. The partial key recoveries lead to more plaintext recoveries, and by working back and forth between keys and plaintext, you can complete the solution. There are many ways in which you can identify plaintext. The more you know about the senders of enciphered traffic and the situation in which it was sent, the more likely you are to be able to assume plaintext correctly.

a. Stereotypes. Military organizations tend to do things in standard ways. Rules for message formats are likely to be used. Standard forms are likely to be used for recurring needs. When you learn enough about the sender's standard ways of doing things, you can use those standards. Standard formats are most likely to be found in message beginnings and endings. Messages are likely to begin with addressees, message subjects, security classifications, and references to other messages. Messages are likely to end with signatures or unit identifications. These stereotypes are bad security practices, but difficult to avoid.

(1) Consider the following example of a message where stereotypes can be used to achieve a quick solution. The previous message from the same sender, already recovered, began, TWO *PART* MESSAGE PART ONE. The text gave the itinerary of a visiting team of officers from an allied country, but was incomplete. A mixed alphabet was used with the previous message, but it has changed with the new message.

ZZZZZ NSHIX LNFOM MXKOI XLNNS HNOXF STDDR OIXLN XNMTU NOOGN

ETLNV EHPLM YVEOD TZHIN OLLDA HGOMZ HFFXG RTGKX ZZZZZ

(2) The first and last groups (ZZZZZ) are obviously not part of the text of the message. They are probably indicators of some kmd.

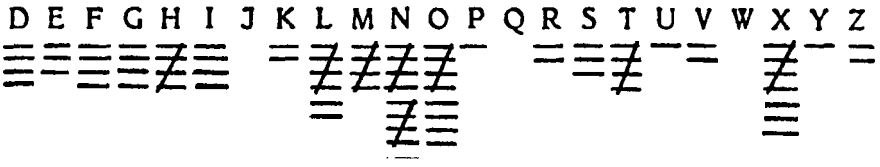
(3) We begin by preparing the following worksheet with a frequency count and underlined repeats. The indicator groups are not included in the frequency count.

NSHIX LNFOM MXKOI XLNNS HNOXF
STDDR OIXLN XNMTU NOOGN ETLNV
EHPLM YVEOD TZHIN OLLDA HGOMZ
HFFXG RTGKX

p: 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 t

c :

c: 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) If this is a follow-on to the message that began, *TWO PART MESSAGE PART ONE*, we would assume that it would begin TWO PART MESSAGE PART TWO. The underlined repeats are positioned perfectly for the repeated words TWO and *PART*, so the assumption seems well borne out.

(5) Next, we enter the assumed text in the message and the alphabet. Using those recovered values throughout the message produces the text shown below.

t w o p a r **t m e s s a g e p** a r t t w o t e a m
N S H I X L N F O M M X K O I X L N N S H N O X F

W e p a r t a t s t **e t e e t r t**
S T D D R O I X L N X N M T U **N O O G N E T L N V**

o r s e o p t e r r o e s
E H P L M Y V E O D T Z H I N O L L D A H G O M Z

o m m a g a
H F F X G **R T G K X**

p: 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: x 0 K F H I L M N **S**

(6) From the recovered ciphertext letters, it appears that the cipher sequence is keyword mixed. On that basis, ciphertext G and J are placed in alphabetical order.

t w o p a r t m e s **s a g e p** a r t t w o t e a m
N S H I X L N F O M M X K O I X L N N S H N O X F

W e p a r t **a t s t e e n t r t**
S T D D R O I X L N X N M T U N O O C N **E T L N V**

o r s e o p t e r r o n e s
E H P L M Y V E O D T Z H I N O L L D A H G O M Z

o m m a n n g a
H F F X G R T C K X

p: 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: x 0 K F G H I J L M N **S**

(7) Several possibilities for additional plaintext appear in the message with these additions. You may see other possibilities but for illustration, we will add the letters for the word COMMANDING appearing at the end of the message.

```
twopa rtmes sagep arttw oteam
NSHIX LNFOM MXKOI XLNNS HNOXF

wi depart atsi teent irt
STDDR OIXLN XNMTU NOOGN ETLNV

 o rs e icopt err onesc
EHPLM YVEOD TZHIN OLLDA HGOMZ

omman dinga
HFFXC RTGKX

p: abcdefghijklmnopqrstuvwxyz
c: X ZRO K T FGHI JLMN S
```

(8) Additional placements are possible. Ciphertext Y belongs between X and Z. P and Q fit between N and S. U, V, and W fit between Sand X. The first word on the second line appears to be WILL. The phrase SIXTEEN THIRTY HOURS appears.

```
twopa rtmes sagep arttw oteam
NSHIX LNFOM MXKOI XLNNS HNOXF

willd epart atsix teent hirty
STDDR OIXLN XNMTU NOOCN ETLNV

hours byhel icopt errl onesc
EHPLM YVEOD TZHIN OLLDA HGOMZ

omman dinga
HFFXG RTCKX

p: abcdefghijklmnopqrstuvwxyz
c: XYZRO KET DFGHIJLMNPQSUVW
```

Only the ciphertext letters A, B, and C remain to be placed. Of those, only A is used in the text, and it appears to be part of the commander's name. If C is placed as part of the keyword ROCKET and A and B placed in alphabetical order, the commander s name becomes *R L JONES.* The plaintext is TWO *PART MESSAGE PART TWO TEAM WILL DEPART AT SIXTEEN* THIRTY HOURS BY HELICOPTER *R L* JONES *COMMANDING.* The complete alphabet is shown below.

```
p: abcdefghijklmnopqrstuvwxyz
c: XYZROCKETABDFGHIJLMNPQSUVW
```

b. **Exploitation of Numbers.** Not all cryptograms will include such stereotyped beginnings and endings. Without these stereotypes, repeated words in the text offer another possible point of entry. Spelled out numbers are often easy to recognize when they repeat in messages, as shown in the next example.

H W B N F W A Z A O U R R W L W W Z M U O J R N E

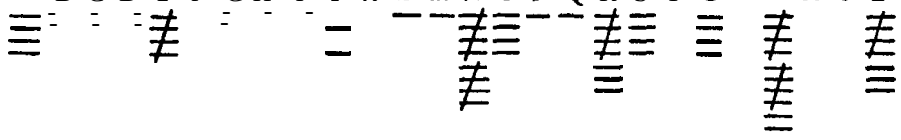
J Y I S J R J O Q W E U D R C W R S Z N N P W A Z

R C W E N B N O K F C N Z W E U D R S Z N N C N Z

W S W A Z E X X X X

p: 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:

c: 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) The pattern of consecutive short three- to five-letter repeats is characteristic of numbers. Numbers tend to occur with each other in such things as grid coordinates, times, and quantities. In the above example, the repeated RSZNN must be THREE, the only five letter number to end in a double letter. We begin by placing THREE in the alphabet and entering other occurrences of the same letters.

e r t t r t e
H W B N F W A Z A O U R R W L W W Z M U O J R N E

h t t t h r c e r
J Y I S J R J O Q W E U D R C W R S Z N N P W A Z

t e e e r t h r e e e r
R C W E N B N O K F G N Z W E U D R S Z N N G N Z

h r
W S W A Z E X X X X

p: 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: N S Z R

(2) The recovered letters suggest additional numbers. RCW, which begins with plaintext T must be *TWO*. GNZW, which includes ER as the middle two letters must be ZERO. EUD, which has no letters in common with THREE, TWO, or ZERO, can only be SIX.

```
  o e o r      i t t o   0 0 r i   t e s
H W B N F W A Z A O U R R W L  W W Z M U  O J R N E

      h t     o s i x t w o t h r e e o r
J Y I S J R J O Q W E U D R C W R S Z N N P W A Z

t w o s e       e   z e r o s i x t h r e e z e r
R C W E N B N O K F G N Z W E U D R S Z N N G N Z

o h o r s
W S W A Z E X X X X

p: 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:     N   s u     W   Z E R   C D   G
```

(3) Several more possibilities can be placed at this point. Ciphertext F can be placed between D and G in the cipher sequence as the alphabetical structure begins to appear. The last word of the message is apparently *HOURS*, needing only the U to complete it. The partially repeated *FOUR* can be seen at the end of line two, and SEVEN follows TWO on the third line.

```
  0 v e y o u r u n i t t o  o o o r ' ı n t e s
H W B N F W A Z A O U R R W L W W Z M U O J R N E

      h t n o s i x t w o t h r e e f o u r
J Y I S J R J O Q W E U D R C W R S Z N N P W A Z

t w o s e v v e n y z e r o s i x t h r e e z e r
R C W E N B N O K F G N Z W E U D R S Z N N G N Z

o h o u r s
W S W A Z E X X X X

p: 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:     NP   SU     o w   Z E R A B C D F G
```

(4) The first word is MOVE, Q can be placed between P and S in the cipher sequence. The word BY completes the third line. With ciphertext K placed from the word BY, ciphertext L and M can also be placed.

```
movey ourun ittoc oordi ntes
HWBNF WAZAO URRWL WWZMU OJRNE

    h  t ngo sixtw othre efour
JYISJ RJOQW EUDRC WRSZN NPWAZ

twose venby zeros ixthr eezer
RCWEN BNOKF GNZWE UDRSZ NNGNZ

ohour s
WSWAZ EXXXX

p: abcdefghijklmnopqrstuvwxyz
c:   KLMNPQSU     HOW     ZERABCDFG
```

(5) *COORDINATES* online one provides the plaintext letter A as ciphertext J. With J placed in the alphabet, the letter I must be in the keyword, along with T, which will not fit in the alphabetic progression. The keyword is therefore HOWITZER. The complete plaintext is *MOVE YOUR UNIT TO COOR*-DINATES ALPHA *TANGO* SIX TWO THREE FOUR TWO SEVEN BY ZERO SIX THREE ZERO HOURS.

c. Word Patterns. When neither stereotypical beginnings and endings nor repeated numbers provide a point of entry, repeated words can often be recognized by their patterns of repeated letters.

(1) Such words as ENEMY, ATTACK, and DIVISION have repeated letter patterns that make them easy to recognize. They are even easier to recognize when the words are repeated in the text. Underlining the repeats gives an indication of where the words begin and end. For example, ATTACK and BATTALION have the same pattern of repeated letters. If the ciphertext OGGORF is repeated in the text, it is much more likely to be ATTACK than a portion of the word BATTALION. It could also be EFFECT, ATTAIN, or a number of other possibilities.

(2) In the case where two or more words have identical patterns, such as ATTACK and EFFECT, letter frequencies can help to decide between the possibilities. If the letters O and F of OGGORF are high frequency letters and the rest are fairly low, it is more likely to be EFFECT than ATTACK. If all the letters are high in frequency, ATTAIN is likely.

(3) Tables have been compiled of common pattern words for various languages to assist in analysis. Table D-3 in Appendix D of this manual provides an English

language word pattern table. Word patterns are also called *idiomorphs.* There is a formal procedure for recording word patterns, which is followed in the table. When you find a pattern word repeated in a cryptogram, you can follow the same procedure to record the pattern and then look it up in the table. The procedure is this—

- Find the first repeated letter in the pattern, and designate all occurrences of that character with the letter A.

G R F L Y M F P A R P Z
A A

- Continue lettering alphabetically from left to right, making sure that each new character gets the next letter of the alphabet and each repeated character gets the same letter.

C R F L Y M F P A R P Z
A B C D B A

- Stop lettering when the last occurrence of the last repeated character is reached. In the example, P is the last occurrence of the last repeated character. The final character Z is not lettered.

C R F L Y M F P A R P Z
A B C D E B F C A F

- Designate any characters before and after the pattern characters with dashes to show the length of the word.

C R F L Y M F P A R P Z
- A B C D E B F G A F -

(4) To use the pattern, refer to Appendix D, Table D-3. The patterns are in alphabetical order beginning on page D-19. The pattern ABCDEBFGAF is located on page D-34. The only word listed for this pattern is *HEADQUARTERS.* The extra letters at the beginning and end of the pattern, designated by the dashes, fit HEADQUARTERS perfectly.

(5) The use of word patterns to solve a cryptogram is shown in the next example.

X C G X F S E A L L K Q I A V **X G J Q M U N A H D**

P V W M Q W C U T U M M U E T U M V A V I A V B A

F A V A C **Z U R F M U N N M U** X W N G D M Q Q N A

H C E U N **G U C Z U P M M Q I A T Q V G** E A L L N

C Q X M D **Q X W X G** C X F S N G U C W A B A N A U

V F U T T X V W E A **L L T U B** Q R U M E X M W R M

U T F M U N N M U X W N C E U R A B Q V **A V Q G U**

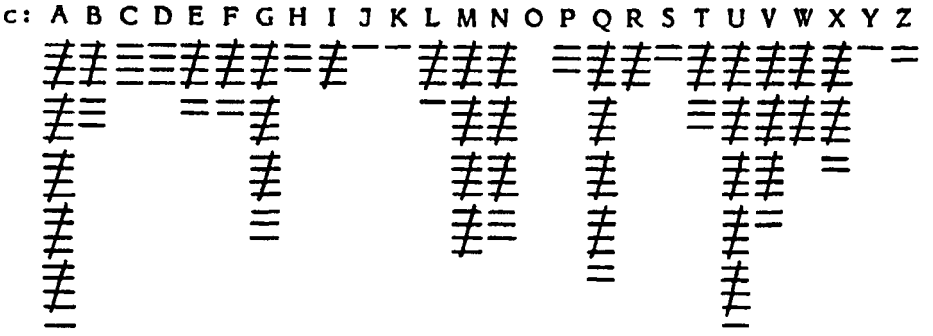
M U X W Y P V F C A **U V Q A I** D G N Q B Q V N A H

N G U C U V Q R A B Q M Q I A T Q V C A N W A B A

N A U V M Q N Q M B Q X X X X

p: 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 t

c:



(6) The cryptogram shows all repeats longer than three letters. There are a number of shorter repeats, too, which will be used if necessary. We begin the analysis by deriving the word patterns for the longer repeats. The pattern and possible words from Appendix D for each repeat are shown below.

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

(7) *CROSSROADS* is the only choice for the second patten. There is an extra letter at the end of the repeat, but that may have been caused accidentally by a repeated first letter of the next word in each case. Using *CROSSROADS* as a trial starting point, we compare common letters with the other repeats. From *CROSSROADS,* we see that cipher M equates to plaintext R, for example. Examining the possible choices for the MQIATQVG repeat, only REGIMENT is consistent with the Rp-Mc pair. Similarly, the Op-Uc and Dp-Wc pairs of *CROSSROADS* are consistent with *DIVISION* for the WABANAUV repeat and no others. The common plaintext N and I between REGIMENT and DIVISION also equate to the same cipher letters (V and A) giving further evidence that we are on the right track. Using the common letters between CROSSROADS, REGIMENT, and DIVISION with the XGGXFS possibilities shows that either ATTACH or ATTACK is consistent with the first three. We now place the letters of CROSSROADS, REGIMENT, and *DIVISION* in the alphabet and cryptogram.

```
a t tac i ● gin at er os l
X C C X F S E A L L K Q I A V XGJQM U N A H D

 n d r e d t o m o rro m o r n i n ginvi
P V W M Q WGUTUMMUET U M V A V IAVBA

c i n i t o crossro a d s t reesi
F A V A G ZURFMUNNMU X W N G D M Q Q N A

 t o s t o o r r e q i m e n t i s
HGEUNGUCZUPMMQI A T Q V G EALLN

 e a r e a d a t tac s to divlsio
C Q X M D Q X W X G G X F S N G U C W A B A N A U

n c o m m and i m o v e e or aa r d r
VFUTTX V W E A L L T U B QRUMEXMWRM

o m c r0 s s r o a dst o i v e n i n e t o
UTFMU N N M U X WNGEU R A B Q V A V Q G U

r o a d n c t i oneig t s e v ensi
M U X W Y P V F G A U V Q A I D G N Q B Q V N A H

st ô one i v v eregimentsidivi
NGUCUVQRAB Q M Q I A T Q V G A N W A B A

s i o n r e s e r v e
N A U V M Q N Q M B Q X X X X

p: 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: X FWQ I A T V U MNG B
```

(8) With this start, you should be able to see many more possible plaintext words in the text. TO*MORROW*, *VICINITY*, and R*OAD JUNCTION* all appear with

only one or two letters missing. Many spelled out numbers also appear. The repeated NGUC is STOP, a common stereotype used in telegraphic text in place of a period. EALL is WILL. XGGXFS must be *ATTACK*. The completed plaintext is—

"ATTACK WILL BEGIN AT ZERO SIX HUNDRED TOMORROW MORNING IN VICINITY OF CROSSROADS THREE SIX TWO STOP YOUR REGIMENT WILL SPEARHEAD ATTACK STOP DIVISION COMMAND WILL MOVE FORWARD FROM CROSSROADS TWO FIVE NINE TO ROAD JUNCTION EIGHT SEVEN SIX STOP ONE FIVE REGIMENT IS DIVISION RESERVE."

(9) Use of word patterns is a powerful tool to gain entry into a cryptogram. It will not always work out as easily as the example shown here. Repeated letters do not always represent repeated words. Many words that are used in messages will not be found in the word pattern tables, particularly proper names. Be alert to the patterns of repeated letters in names you would expect to find in message traffic. If you can recognize the pattern of a word, it does not have to be in the tables to use it,

4-13. Vowel-Consonant Relationships

When you can successfully discover plaintext words in a cryptogram, the solution usually comes quickly. Sometimes you will encounter a cryptogram in which you can find no basis to assume plaintext. You can find no stereotypes, no usable numbers, and no repeated pattern words. In these cases, you can use the characteristics of the language itself to determine individual letters.

a. **Language Characteristics.** Languages which use an alphabet to spell out words phonetically produce exploitable letter relationships. To make words pronounceable, vowels and consonants tend to alternate. We do not expect to find many consonants or many vowels consecutively. In cases where they do, the possibilities are limited to pronounceable combinations. Exploitation of these letter relationships begins by determining which letters are consonants and which are vowels.

(1) Vowels tend to occur next to consonants. Consonants tend to occur next to vowels. Each contacts the other more readily than it contacts its own type.

(2) Since there are more consonants than vowels in English, vowels tend to contact more different letters than consonants do. A vowel will commonly contact a lot of different consonants, whereas a consonant will tend to contact the smaller number of vowels. By studying which letters contact each other and how many different contacts each letter has, we can sort ciphertext letters into vowels and consonants fairly reliably.

(3) To make use of these vowel-consonant relationships, we use a special kind of frequency count which charts contacts as well as frequencies.

b. **Trilateral Frequency Count.** The trilateral frequency count is used to record, for each letter in a cryptogram, the letter that precedes it and the letter that follows it. Figure 4-4 shows a cryptogram and its trilateral frequency count. The pairs of letters appearing in the column below each letter of the alphabet are the preceding and following letters for each occurrence. For example, the YG that appears below the letter A shows that the first A in the cryptogram occurred as part of the segment YAG. Refer to the cryptogram itself, and you will see that the segment YAG occurs in the second group of the message. Two numbers appear above each letter of the alphabet. The top figure is the frequency of that letter, which is the same as the number of pairs of letters in the column below it. The second number is the number of different letters the basic letter contacts. This type of frequency distribution and its supporting contact information take some time to prepare by hand, but they can lead to the solution when other methods fail.

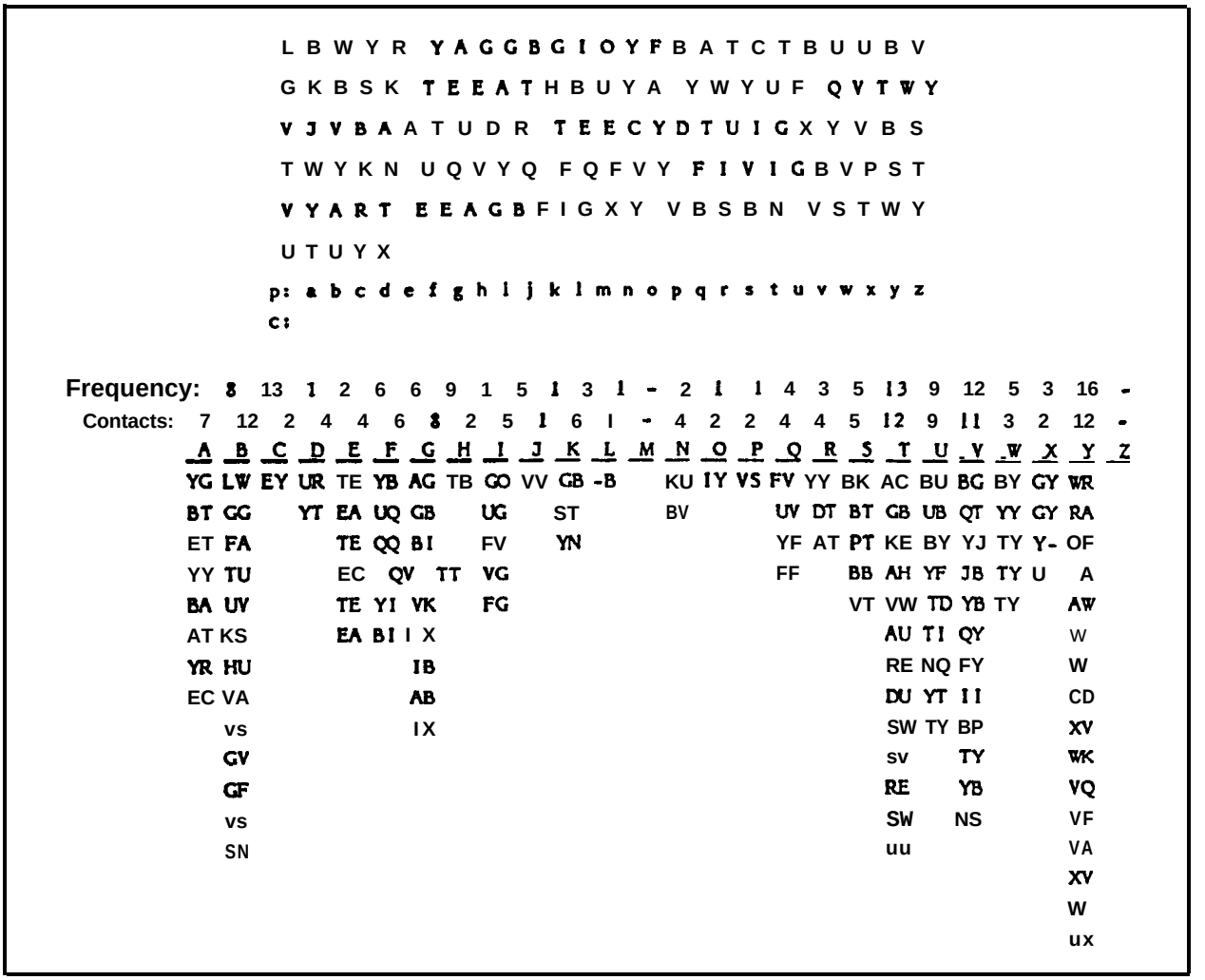


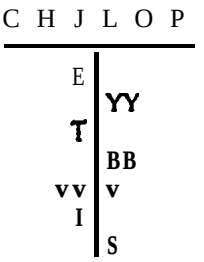
Figure 4-4. Triliteral frequency distribution.

(1) The contact information is used to determine which ciphertext letters are vowels and which are consonants. More often than not, the highest frequency plaintext letter is a vowel, even when E is not the highest frequency letter. An even more reliable indicator is the number of contacts. The letter that contacts the most different letters will usually be a vowel. In the example in Figure 4-4, ciphertext Y is likely to be a vowel for both reasons. The letters that Y contacts most frequently are likely to be consonants.

(2) In cases where there are several letters all about the same frequency and no letter stands out as a likely vowel, we can begin our approach through likely consonants instead. All or most of the lowest frequency letters should be consonants. The letters they contact most frequently are likely to be vowels.

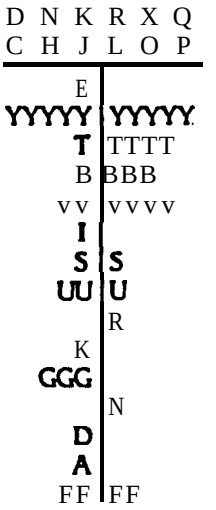
(3) We can use either a likely vowel or the set of likely low frequency consonants as our starting point. Whichever we start with, we will use both as the problem develops. The object is to separate the consonants and vowels by plotting the contacts of each m separate vowel and consonant line charts.

(4) For our example, we will pick the low frequency consonants as the starting point. The process begins by charting the contacts of the lowest frequency letters. We will begin with the letters that only occurred once in Figure 4-4-C H, J, L, 0, and P. Draw a horizontal line two to three inches long and write the selected letters above it. Draw a vertical line several inches from the center of the horizontal line producing a T-shaped figure. This is the consonant line. The contacts are charted on the line with the first letters of each pair to the left and the second to the right. Each new contact letter is charted on a new row. With the contacts for C, H, J, L, 0, and P charted, the consonant line appears below.



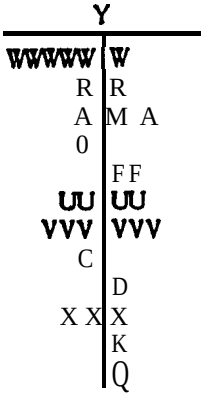
(5) Continue adding the lowest frequency letters one frequency group at a time. We first placed those with a frequency of one. Next add those with a frequency of two. Continue with those with a frequency of three and so on. Stop when the next frequency would represent more than 20 percent of the total. Going any further raises the chance too high of including a vowel that would bias the chart. If a vowel occurs only once or twice and is included, its influence will be small. If it occurs five or six times and we include it, it could lead to wrong follow-on

decisions on vowels and consonants. In our example, there are 130 letters. We want to keep our sample below 20 percent, or not more than 26 letters altogether. On this basis, we can add the frequencies of 2, 3, and 4, but not 5.



(6) The consonant line now shows that the low frequency consonants contact the ciphertext letter Y more than any other letter. The probability is very high that this is a vowel. It is tempting to select the letter V as a vowel, but it is better to proceed one letter at a time at this point.

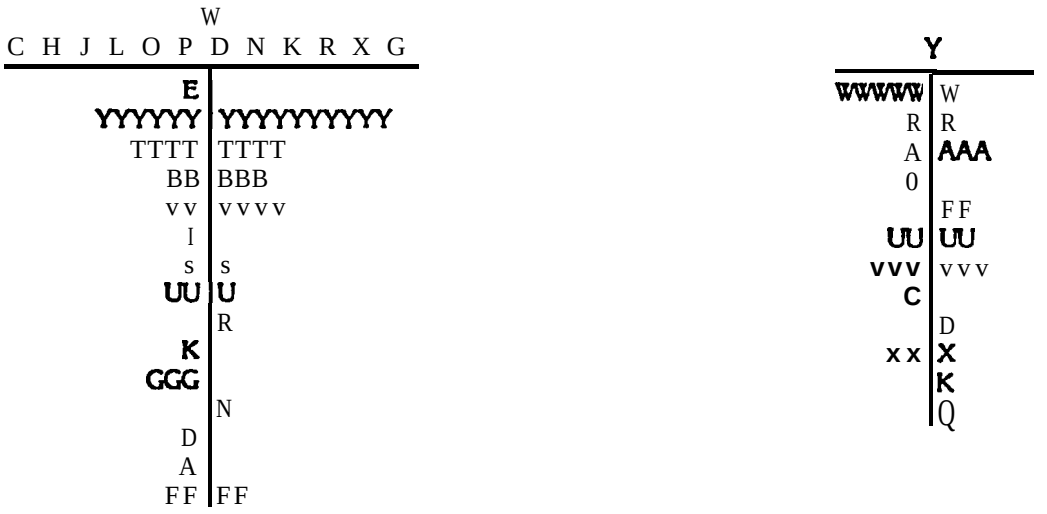
(7) Using the letter Y and its contacts, we next begin construction of a vowel line. It is charted exactly the same as the consonant line chart. The vowel line including just the letter Y's contacts is shown below.



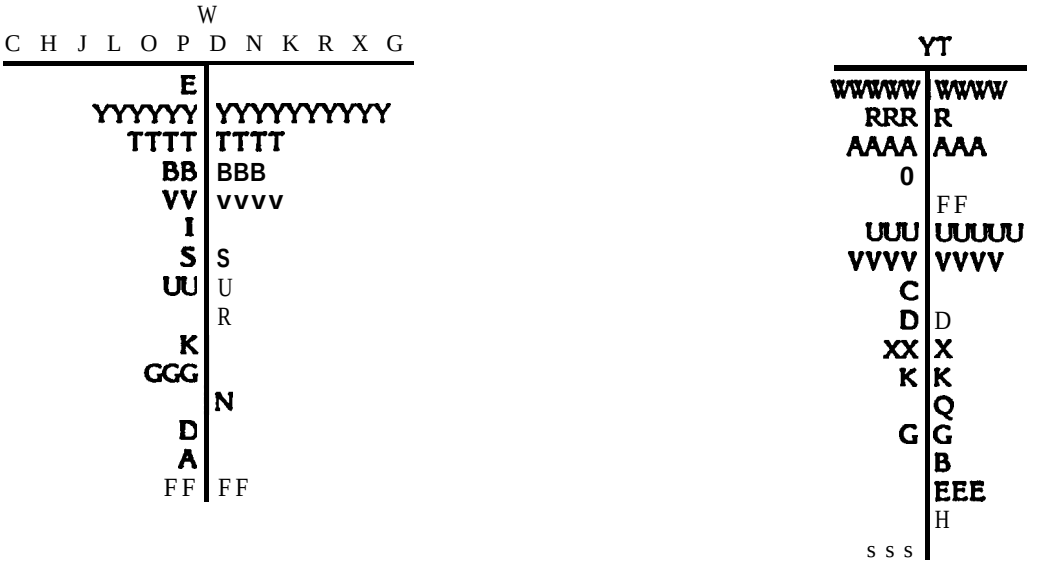
(8) The vowel line shows us we were correct in not initially accepting the letter V as a vowel. It contacts the low frequency consonants quite readily, but it also contacts a vowel readily. It may be a consonant such as R, L, or N which easily

combines with other consonants. We will not try to place V in either line at this point.

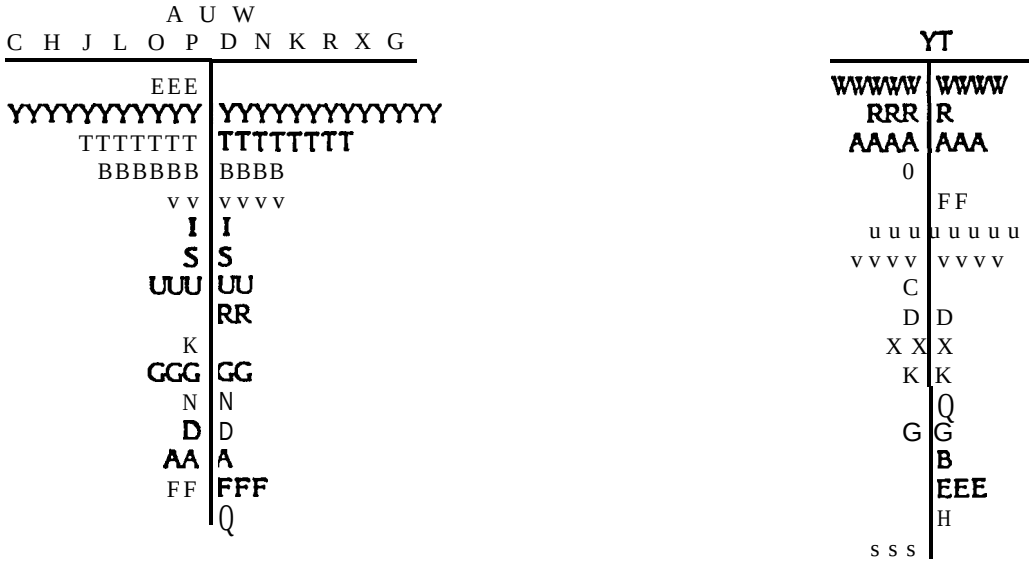
(9) The letter W contacts Y six times and is a likely consonant. We will continue by going back to the consonant line and adding W.



(10) The letter T now appears as a strong candidate for a vowel. It is second only to Y in consonant contacts so far, and just as importantly, it does not contact the already selected vowel at all. We add T and its contacts to the vowel line.



(11) The vowel line shows A and U as likely consonants. Adding these letters to the consonant line produces the next diagram.



(12) B appears to be a vowel. This is reinforced by the letters BUUB in the first line of the text. If U was correctly selected as a consonant, B is probably a vowel on the basis of this letter pattern. It is a good idea at this point to return to the text and underline all the recovered vowels.

L B W Y R Y A C C B G I O Y F B A T G T B U U B V

G K B S K T E E A T H B U Y A Y W Y U F Q V T W Y

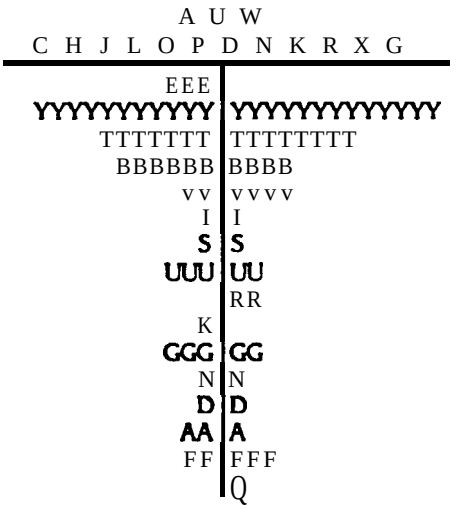
V J V B A A T U D R T E E C Y D T U I G X Y V B S

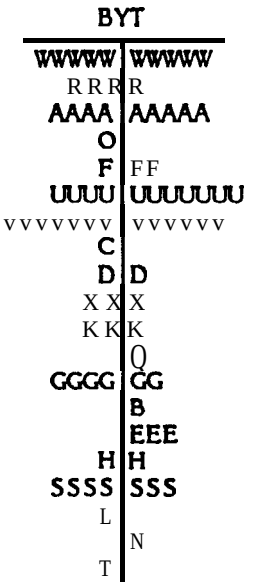
T W Y K N U Q V Y Q F Q F V Y F I V I G B V P S T

V Y A R T E E A G B F I G X Y V B S B N V S T W Y

U T U Y X

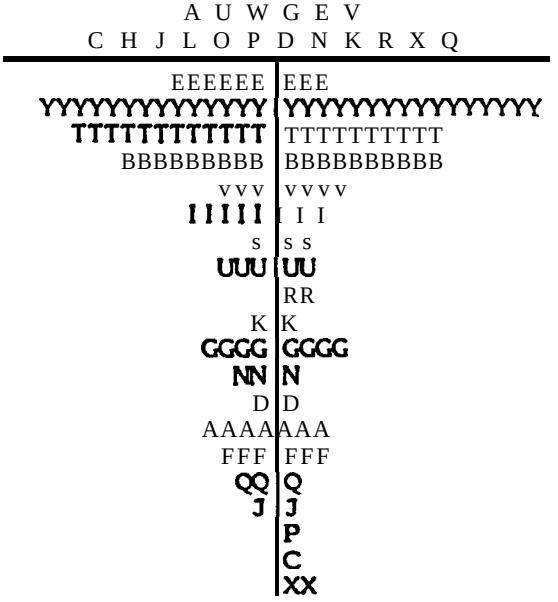
p: a b c d e f g h i j k l m n o p q r s t u v w x y
c :

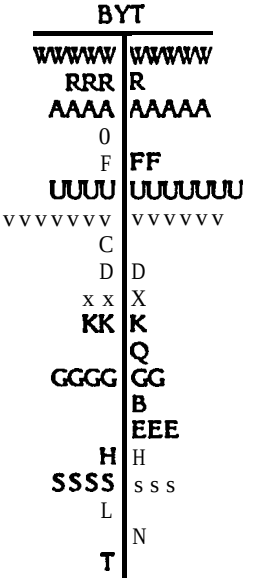




(13) Examination of the vowel-consonant patterns in the text confirms additional consonants. Double letters preceding or following the vowel are very unlikely to be vowels. We can then assign ciphertext E and Gas consonants. The GGBG segment on the first line could not all be vowels. EE occurs three times in the text following a vowel.

(14) V appears to be a consonant from the number of contacts in the vowel line, and its appearance between vowels in the segments YVB and TVY confirm it as a consonant. Placing G, E, and V in the consonant line produces this diagram.





(15) The letters F, I, and S remain unidentified. At least one of these is likely to be a vowel, since four of the letters are expected to be vowels and we have only identified three so far. Comparing the appearance of F, I, and S in the vowel and consonant lines, we see that the letter I is the best candidate for a vowel. The letter I does not appear on the vowel line at all, whereas, F and S directly contact a number of the recovered vowels. We now underline I in the text and add it to the vowel line.

L B W Y R Y A G G B G I O Y F B A T G T B U U B V

G K B S K T E E A T H B U Y A Y W Y U F Q V T W Y

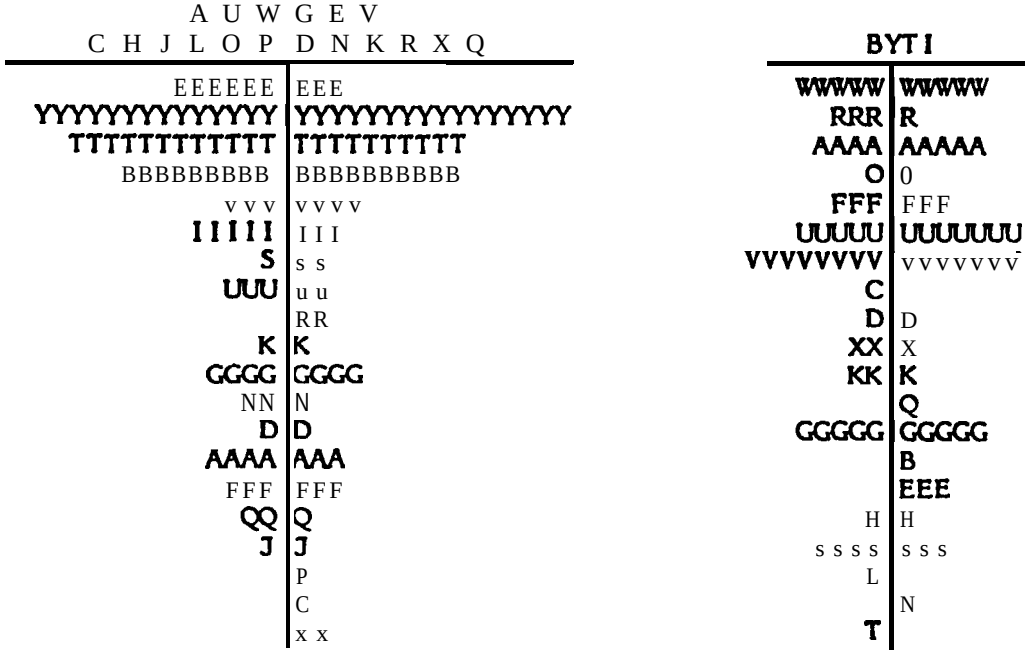
V J V B A A T U D R T E E C Y D T U I G X Y V B S

T W Y K N U Q V Y Q F Q F V Y F I V I G B V P S T

V Y A R T E E A G B F I G X Y V B S B N V S T W Y

U T U Y X

p: 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 :



(16) There are a number of directions you can take at this point. No single example can demonstrate them all. Some of the approaches that can be tried are—

- To analyze vowel combinations to determine individual vowels.

- To search for the plaintext consonants N and H. These two letters have typical patterns of contact with consonants and vowels. N tends to follow vowels and precede consonants. H tends to follow consonants and precede vowels. In some cryptograms these features will be very evident in the vowel and consonant line diagrams. In others, they will not stand out at all.
- To recover double letters by frequency analysis. Plaintext LL is the most frequent double consonant. EE and 00 are the most frequent double vowels.
- To recover common word endings such as -ING and -TION, which often appear as repeats even when complete words do not repeat.

(17) We will use several of these approaches to complete the solution of the sample problem. First, one vowel combination appears in the cryptogram, the ciphertext TB as part of the segment TGTBU. Referring to the two-letter frequency data in Appendix A, page A-2, the most frequent vowel combinations are EE, IO, OU, and EA. TB is not EE, because it is not a double letter. It is likely to be one of the other three. IO is particularly significant, because it is usually part of a -TION combination when it appears. The letters G and U, which precede and follow BT in the text, are high frequency consonants and support the -TION possibility. The letter T occurs again before G, which would produce -ITION, a very good letter combination.

(18) If TGTBU is -ITION, the letter U may appear with the typical pattern of plaintext N. Examining the occurrence of U in the vowel and consonant lines, we see that U follows vowels more often than it precedes them. It also precedes consonants more often than it follows. The differences are slight, but they help to confirm the initial assumption.

(19) Ciphertext EE occurs three times. This is likely to be plaintext LL. Each time it is preceded by ciphertext T, which we have tentatively identified as the plaintext I. ILL is another good combination that appears as part of many common words such as HILL and WILL.

(20) Y is the most common letter, and it is a vowel. While we would not usually begin analysis by assuming the most common vowel is E, our tentative identification of I and 0 make this much more likely. If Yc is Ep, then the remaining high frequency vowel, Ic, is probably Ap.

(21) Placing all the tentative recoveries in the cryptogram produces the next example.

```
  o  e  e  t t o  t a  e  o  i t i  o n n o
L B W Y R  Y A G G B  G I O Y F  B A T G T  B U U B V

t  o  i l l  i  0 n e  e  e n  i  e
G K B S K  T E E A T  H B U Y A  Y W Y U F Q V T W Y

   o  in  i l l  e  i n a t  e  o
V J V B A A T U D R  T E E C Y  D T U I G  X Y V B S

i  e  n  e  e  a  a t o  i
T W Y K N  U Q V Y Q  F Q F V Y  F I V I G  B V P S T

  e  i l l  t o  a t  e  o  o  i  e
V Y A R T  E E A G B  F I G X Y  V B S B N  V S T W Y

n i n e
U T U Y X

p: 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 t
c: I  Y  T  E  U B  G
```

(22) With the assumed letters filled in, two numbers stand out. ONE appears in the second line, and *NINE* appears in the last line. Since numbers tend to occur with each other, our next objective is to try to place additional numbers adjacent to these two. If we try SEVEN after ONE because of the -E-EN pattern, it leads to the recovery of SIX before ONE and *FIVE* before NINE.

(23) All of the high frequency plaintext letters except R are now recovered. Vc is the obvious candidate for Rp due to its high frequency and appearance in the text.

(24) Placing plaintext S, V, X, F, and R reveals this text.

```
  o v e  e s t t o  t a  e  o s i t i  o n n o r
L B W Y R  Y A G G B  G I O Y F  B A T G T  B U U B V

t  o f  i l l s i  x o n e s  e v e n  r i v e
G K B S K  T E E A T  H B U Y A  Y W Y U F  Q V T W Y

r  r o s  s i n  i l l  e  i n a t  e r o f
V J V B A  A T U D R  T E E C Y  D T U I G  X Y V B S

i v e  n  r e  r e  a r a t  o r  f i
T W Y K N  U Q V Y Q  F Q F V Y  F I V I G  B V P S T

r e s  i  l l s t o  a t  e  r o f o  r f i v e
V Y A R T  E E A G B  F I G X Y  V B S B N  V S T W Y

n i n e
U T U Y X

p: 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 t
c: I  Y S  T  E  U B  VAG  W  H
```

(25) Many possibilities for plaintext appear now. ZERO, *POSITION*, *RIVER CROSSING*, *PREPARATORY*, and *FOUR* can all be seen upon close examination.

```
m o v e w  e s t t o  t a k e p  o s i  t i o n n o r
L B W Y R  Y A G G B  G I O Y F  B A T G T B U U B V

t h o f h  i l l s i x o n e s  e v e n p  d r i v e
G K B S K  T E E A T H B U Y A  Y W Y U F  Q V T W Y

r c r 0 s  s i n g w  i l l b e g i n a t z e r o f
V J V B A  A T U D R  T E E C Y  D T U I C  X Y V B S

i v e h u  n d r c d  p d p r e  p a r a t o r y f i
T W Y K N  U Q V Y Q  F Q F V Y  F I V I C  B V P S T

r e s w i  l l s t o p a t z e  r o f o u  r f i v e
V Y A R T  E E A G B  F I G X Y  V B S B N  V S T W Y

n i n e
U T U Y X
```

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

(26) Analysis of the cipher sequence shows it to be a simply transposed keyword mixed sequence, which identifies Jp as Zc and Qp as MC.

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

Software Development

Software Development April 1996 v4 n4 p70(4)

Analysis and reporting tools for C++.

(Gimpel Software's C-Vision 4.0 and Western Wares' CC-Rider 5.1f C++ code analysis tools)

Author
Dlugosz, John

Abstract
Gimpel Software's $239 C-Vision 4.0 and Western Wares' $549 CC-Rider 5.1f C++ code analysis tools are both cost-effective and useful products for documenting and analysing substantial bodies of code. C-Vision for extended DOS contains six diverse tools, which are C-Lines, C-Comment, C-Xref, C-Format, C-Tree and C-Tdump. All six tools are command-line driven, separately executable and are capable of working with each other. C-Vision ships on only one disk, and can be customized during the initial installation. CC-Rider comes on seven disks and includes API, graphical tools, demonstration and text tool installation programs. CC-Rider also includes an educational tool that assists users in learning the browser functions. Drawbacks for CC-Rider include lengthy installation procedures and a confusing analyzer.

Type
Software Review
Evaluation

Company
Gimpel Software
Western Wares

Product
C-Vision 4.0 (Programming utility)
CC-RIDER 5.1f (Programming utility)

Topic
Software Multiproduct Review
Application Development Software
Debugging/Testing Software

Record #
18 072 358

CHAPTER 5

MONOALPHABETIC MULTILITERAL SUBSTITUTION SYSTEMS

Section I
Characteristics and Types

5-1. Characteristics of Multilateral Systems

As explained in Chapter 3, monoalphabetic unilateral systems are those in which the ciphertext unit is always one character long. Multilateral systems are those in which the ciphertext unit is more than one character in length. The ciphertext characters may be letters, numbers, or special characters.

a. Security of Multilateral Systems. By using more than one character of ciphertext for each character of plaintext, encipherment is no longer limited to the same number of different cipher units as there are plaintext units. Although there is still only one alphabet used in multilateral systems, the alphabet can have more than one ciphertext value for each plaintext value. These variant ciphertext values provide increased security. Additionally, the plaintext component of alphabets can be expanded easily to include numbers, punctuation, and common syllables as well as the basic 26 letters. When used, the variation in encipherment and the reduced spelling of numbers, punctuation, and common syllables minimize the exact weaknesses that we used in Chapter 4 to break into unilateral systems.

b. Advantages and Disadvantages. The increased security possible with variant multilateral systems is the major advantage. The major disadvantage is that by substituting more than one character of ciphertext for each plaintext value, the length of messages and resulting transmission times are increased. A second disadvantage is that more training and discipline are required to take advantage of the increased security. If training and discipline are inadequate, the security advantages are lost easily.

5-2. Types of Multilateral Systems

Multiliteral systems are further categorized by the type of substitution used. The major types are—

- Biliteral systems, which replace each plaintext value with two letters of ciphertext.
- Dinomic systems, which replace each plaintext value with two numbers of ciphertext.
- Trilateral and trinomic systems, which replace each plaintext value with three letters or numbers of ciphertext.
- Monome-dinome systems, which replace plaintext values with one number for some values and two numbers for other values.
- Biliteral with variants and dinomic with variants systems, which provide more than one ciphertext value for each plaintext value.
- Syllabary squares, which may be biliteral or dinomic, and which include syllables as well as single characters as plaintext values.

5-3. Cryptography of Multilateral Systems

The cryptography of each type of multilateral system, including some of the odd variations is illustrated in the following paragraphs. Most of these systems are coordinate matrix systems in which the plaintext values are found inside a rectangular matrix and the ciphertext values consist of the row and column coordinates of the matrix.

a. Simple Biliterals and Dinomics. The simplest multilateral systems use no variation. They typically use a small rectangular matrix large enough to contain the letters of the alphabet and any other characters the system designer wants to use as plaintext values.

(1) The plaintext values are the internals of the matrix. They may be entered alphabetically, follow a systematic sequence, or they may be random. They may be entered in rows, m columns, or by any other route.

(2) The row and column coordinates are the externals. Conventionally, the row coordinates are placed at the left outside the matrix, and the column coordinates are placed at the top. As with the internals, the coordinates may be selected randomly or produced systematically.

(3) A ciphertext value is created by finding the plaintext value inside the matrix and then combining the coordinate of the row with the coordinate of the column for that plaintext value. Either can be placed first, although placing the row coordinate before the column coordinate is more common.

(4) Five by five is a common size for a simple system (Figure 5-l). The 26 letters are fitted into the 25 positions in the matrix by combining two letters. The usual combinations are I and J or U and V. It is up to the deciphering cryptographer to determine which of the two is the correct value. There are few, if any, words in common usage in which good words can be formed using either letter of the I/J or U/V combinations. Other common sizes are 6 by 6 (which gives room for the 10 digits), 4 by 7, and 3 by 10. Many other sizes are possible.

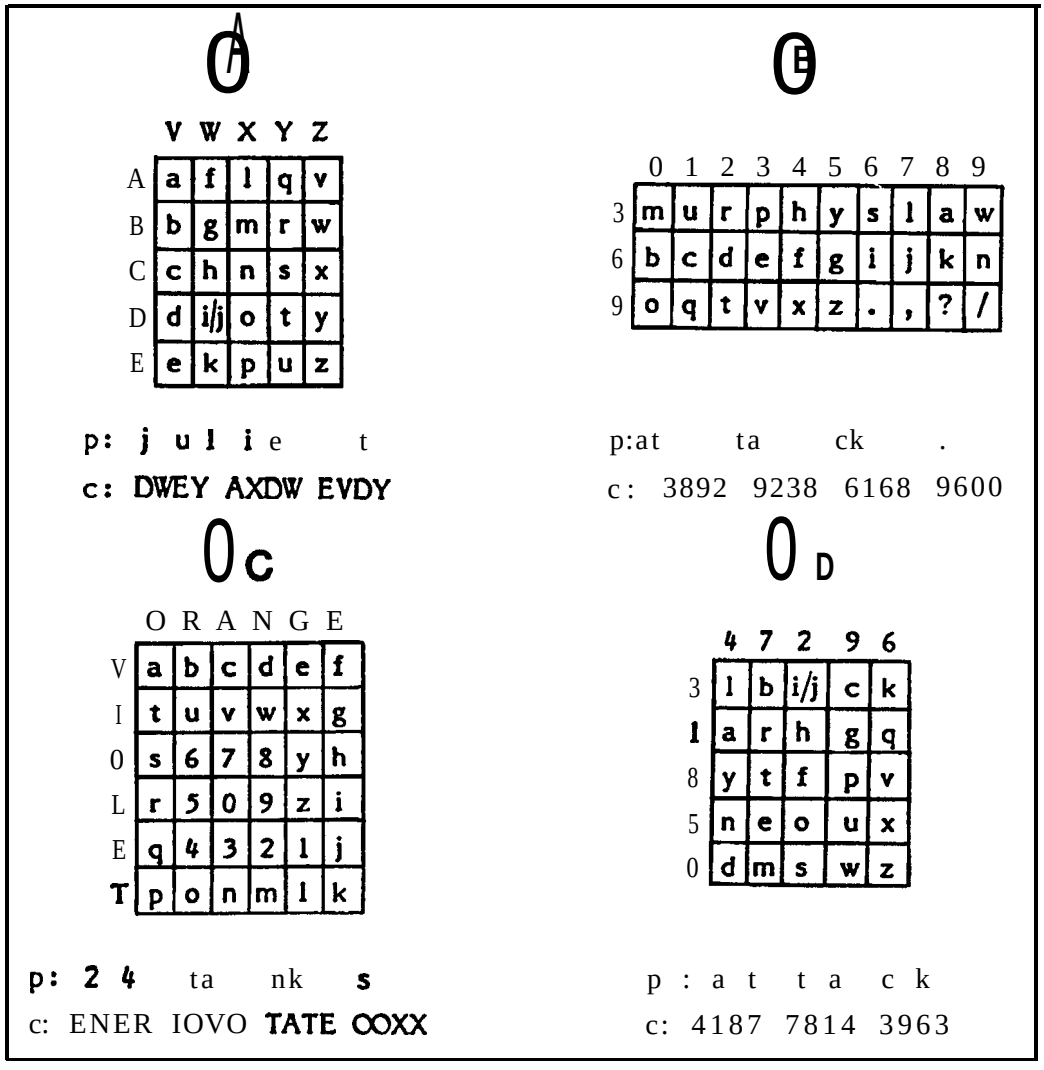


Figure 5-1. Biliteral and dinomic matrices.

(5) Example A in Figure 5-1 is a simple 5 by 5 matrix with I and J in the same plaintext cell of the square. The coordinates and the sequence within are in alphabetic order.

(6) Example B is a simple 3 by 10 matrix with orderly coordinates and a keyword mixed sequence inscribed within. The four extra cells are used for punctuation marks.

(7) Example C is a 6 by 6 matrix with a spiral alphabetic sequence followed in the spiral with the 10 digits. The coordinates in this case are related words.

(8) Example D is a 5 by 5 matrix with numeric coordinates. The plaintext sequence is keyword mixed entered diagonally. In this case, there is deliberately no repetition between the row and column coordinates. This allows the coordinates to be read either in row-column order or in column-row order without any ambiguity, as in the sample enciphered text. This is unusual, but you should be alert to such possibilities.

b. Triliterals and Trinomics. Trilateral and trinomic systems are essentially the same as biliteral and dinomic systems. The difference is that either the row coordinates or the column coordinates consist of two characters instead of one, creating a three-for-one substitution. Such systems offer no real advantage except to provide a slightly different challenge to the cryptanalyst, and have the distinct disadvantage of tripling the length of messages. They are easily recognized, and offer no increase in security.

	L	M	N	O	P
	V	W	X	Y	Z
A	a	f	l	q	v
B	b	g	m	r	w
C	c	h	n	s	x
D	d	i/j	o	t	y
E	e	k	p	u	z

	0	1	2	3	4	5	6	7	8	9
13	m	u	r	p	h	y	s	l	a	w
26	b	c	d	e	f	g	i	j	k	n
39	o	q	t	v	x	z	.	,	?	/

p: j u l i e t

c: DMW EOY ANX DMW ELV DOY

p: a t t a c k

c: 138 392 392 138 261 268

c. Monome-Dinomes. Monome-dinomes are coordinate matrix systems constructed so that one row has no coordinate. The values from that row are enciphered with the column coordinate only. This means that some ciphertext values are two characters in length (dinomes) and others are only one (monomes). If the values used as row

coordinates are also used as column coordinates, no plaintext values are placed in the monome row under those repeated column coordinates. The blanking of cells in the monome row is shown in the example below.

	1	2	3	4	5	6	7	8	9	0
	h	e	x	a	-	-	d	c	i	m
5	l	b	f	g	j	k	n	o	p	q
6	r	s	t	u	v	w	y	z	.	,

p:	e	n	e	m	y		a	t	t	a	c	k	i	n	g
c:	2	572		0	67		4	63	63	4	8	56	9	57	54

Resulting message:

257206746363485695754000O

(1) If the cells corresponding to the row coordinates in the monome row are not blanked, the deciphering cryptographer will have difficulty. Decipherment proceeds left to right, and when a 5 or a 6 is encountered in the matrix shown, it will always be a row coordinate or combine with a preceding row coordinate. It will never stand alone as a monome. If the 5 and 6 cells were not blanked, the deciphering cryptogrpher could not tell if a 5 or 6 were a monome or the beginning of a dinome. The cryptographer would have to rely on context to figure out which was intended, and that could lead to errors.

(2) The additional examples of monome-dinomes shown below demonstrate the various ways they can be constructed. The last example (top of page 5-5) is a monome-dinome-trinome.

	7	0	4	8	5	1	3	9	2	6
	w	i	l	d	-	c	a	t	-	-
6	b	e	f	g	h	j	k	m	n	o
2	p	q	r	s	u	v	x	y	z	.
5	0	1	2	3	4	5	6	7	8	9

	2	4	6	8	0
-	t	e	n	o	r
1	c	b	x	a	s
3	d	f	g	h	i
5	p	m	l	k	j
7	q	u	v	w	y
9	z	.	,	;	:

	1	2	3	4	5	6	7	8	9	0
	-	-	r	a	m	c	h	i	p	s
1	b	d	e	f	g	j	k	l	n	o
23	q	t	u	v	w	x	y	z	.	0

p: r e q u e st h e l p
c: 3 13 231 233 13 0 232 7 13 18 9

Resulting message:

3132312331302327131890000

d. Variant Systems. Variants in a multiliteral system allow plaintext characters to be enciphered in more than one way. Variants can be external or internal.

(1) External variant systems have a choice of coordinates. Either row coordinates or column coordinates or both can have variants. Examples A and B in Figure 5-2 provide two ways to encipher every letter.

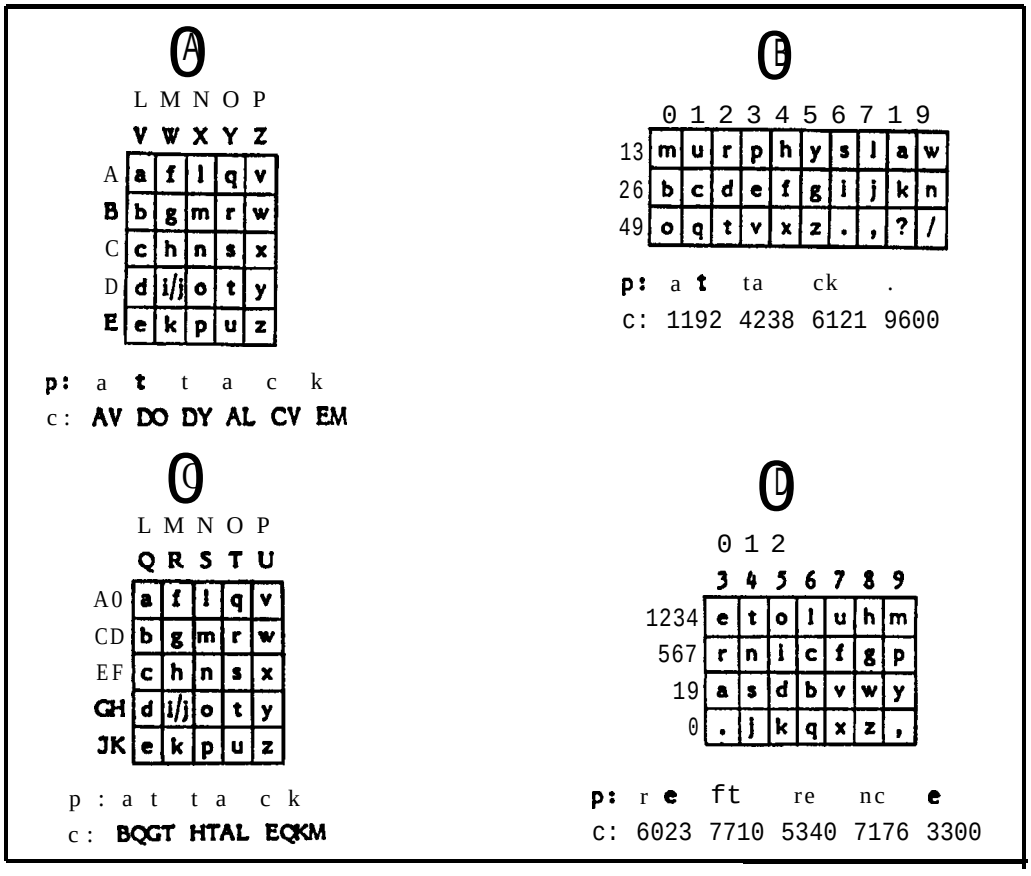


Figure 5-2. External variant systems.

Example C provides four ways to encipher every letter. Example D was constructed to provide the most variants for the most common letters. The letters E, T, and 0 can all be enciphered in eight different ways. R, N, and I can be enciphered in six different ways. A, S, D, L, U, H, and M can be enciphered in four different ways. Q, X, Z, and the comma can only be enciphered one way. When any of the systems are conscientiously used, repeated words in the text will not produce repeated ciphertext segments.

(2) Internal variant systems use larger matrices to provide variants inside the matrix. Each common plaintext letter appears more than once. Here are two examples of internal variant systems.

	3	0	2	8	6	5	1	4	7	9
7	e	e	e	e	t	t	o	n	i	s
3	e	e	e	t	t	o	r	i	a	d
9	e	e	t	t	o	r	i	a	d	u
1	e	e	t	o	r	n	a	d	u	f
6	e	t	o	r	n	a	d	u	c	m
4	t	o	r	n	a	s	u	c	m	p
8	o	r	n	a	s	l	c	y	g	w
2	r	n	i	s	l	h	y	g	v	k
5	n	i	s	l	h	f	b	v	j	x
0	i	s	l	h	f	b	p	w	q	z

	K	L	M	N	O	P	Q	R	S	T
A	l	u	c	k	y	c	h	a	r	m
B	o	b	j	e	c	t	i	o	n	s
C	g	o	l	d	r	e	c	o	r	d
D	a	f	f	e	c	t	i	o	n	s
E	r	a	p	s	e	s	s	i	o	n
F	i	n	c	e	n	d	i	a	r	y
G	t	r	i	v	i	a	q	u	i	z
H	h	e	a	v	y	m	e	t	a	l
I	m	a	s	t	e	r	w	o	r	k
J	s	i	x	t	y	s	e	v	e	n

The first example above places the letters in the matrix according to their expected frequency in plaintext. If their use is well balanced, all letters in the square will be used with about the same frequency. The second square achieves the same effect by using 10 words or phrases in the rows, which use all the letters, The first letters of the column spell out an eleventh word-logarithms.

e. Syllabary Squares. Another type of internal variant system is the syllabary square. This type includes common syllables as well as single letters. When these are used, the same square may be used for a period, changing the coordinates more frequently than the square itself.

	6	0	4	3	8	1	7	5	9	2
8	a	1	ad	al	an	and	as	at	b	2
4	c	3	ce	co	d	4	da	de	di	e
3	5	ea		ec	ed	ee	ei	el	en	ent	er
7	es	et	f	6	fi	fo	g	7	h	8
2	hi	ht	i	9	in	ing	io	ir	is	it
Oj		0	00	k	l	la	le	11	m	ma
5	n	nd		ne	ng	ni	nt	o	on	or	ou
9	p	q	r	ra	re	ri	ro	rs	rt	s
1	se	si	st	t	ta	te	th	ti	tion	t o
6	tw	ty	u	ur	v	ve	w	x	y	z

p: r ei n fo r ce m ent s
c: 94 31 56 71 94 44 09 35 13 92

p: re in f or ce m ent s
c: 98 28 74 59 44 09 39 92

The two sample encipherments of *REINFORCEMENTS* show that a syllabary square suppresses repeats in ciphertext just as single letter variant systems do. It also has the advantage of producing shorter text than single letter multilateral systems.

f. Sum Checks. It is very easy for errors to occur when messages are transmitted and received, whatever means of transmission are used. Because of this, some users introduce an error detection feature into traffic known as sum checking.

(1) In its simplest form, a sum-check digit is added to every pair of digits in numeric messages. The digit is produced by adding the pair of digits to produce the

third. If the result is larger than 9, only the second digit is used, dropping the 10's digit, for example 8 plus 9 equals 7 instead of 17. This is also known as modulo 10 arithmetic.

Ciphertext: 42 63 55 47 22 89

Ciphertext with sum check: 426 639 550 471 224 897

(2) Whenever the first two digits do not add up to the third, the receiving cryptographer is alerted that an error has occurred. The cryptographer then tries to figure out the correct digit from context or by assuming that two of the digits are correct and determining what the third should be.

(3) There are many variations on the simple system of sum checking described here. Sometimes the sum-check digit will be placed first or second in each resulting group of three. Sometimes a sum check will be applied to a larger group than two numbers. Sometimes a different rule of arithmetic will be used, such as adding the sum-check digit so that the resulting three always add to the same total. Sometimes a more complex system will be used that provides enough information to resolve many errors as well as detect them, particularly when computers are used in data and text transmissions.

(4) Computer produced sum checks can be used with any characters, not just numbers. Computer produced sum checks will normally be invisible to the user, as they are automatically stripped out when a message is received. They may or may not be invisible to the cryptanalyst. Recovery of computer produced sum checks is well beyond the scope of this text, but you should be alert to their existence.

Section II
Analysis of Simple Multilateral Systems

5-4. Techniques of Analysis

The first steps in solving any multilateral system are to identify the system and establish the coordinates. It makes little difference whether the system uses numbers or letters for coordinates. The techniques are the same in either case. Once the system is identified and the coordinates set up, a solution of the simpler systems is the same as with unilateral systems. Variant systems require additional steps. Each type is considered in the following paragraphs.

5-5. Identification of Simple Biliteral and Dinomic Systems

Simple biliteral and dinomic systems are very easy to recognize and solve.

a. First, the two-for-one nature of the system will usually be apparent. The message will be even in length. The majority of repeated segments will be even in length, although when an adjacent row or column coordinate is the same, a repeat may appear odd in length. The distance between repeats, counted from the first letter of one to the first letter of the next, will be even in length.

b. Second, unless the identical letters or numbers are used for row and column coordinates, there will be limitation by position. One set will appear in the row coordinate position, and the other set will appear in the column coordinate position. Even in the case where all coordinates are different and either the row or column coordinate character may be placed first, each pair will be limited to one from one set and one from the other. If you do not recognize it right away, charting contacts will make it obvious.

c. For systems with letters as coordinates, not more than half the alphabet will be used as coordinates. This severe limitation in letters used is the most obvious characteristic, since only very short unilateral messages are ever that limited. A phi index of coincidence will reflect that limitation, always appearing much higher than expected for a unilateral system.

d. Dinomic systems, since they are limited to the 10 digits anyway, are not quite as obvious. Simple systems should still show positional limitation, however,

5-6. Sample Solution of a Dinomic System

The next problem shows the steps in solution of a sample dinomic system. These steps apply equally to biliteral systems.

2023 2029 6224 6322 2144 4420 6362 4924 6529 2769
2043 2123 2227 4627 6521 2221 2723 6527 2349 2144
4481 8287 2423 4349 2144 4485 8089 6522 2746 2421
6365 2263 2142 2027 2324 6322 2144 4420 6362 4627
6521 2221 2723 6560 2144 4441 2047 2123 2422 6680

6666 6522 2746 4263 2069 2122 6425 2729 2924 2343
2123 4700

a. The most obvious thing about this cryptogram is that every pair of numbers begins with 2, 4, 6, or 8. The final pair begins with 0, but since it appears nowhere else, it is probably a filler. This suggests that we are dealing with a matrix with four rows.

b. Scanning the second digit of every pair, we see that there is some limitation in the column position, also. All digits are used except 8. The matrix appears to have nine columns, although it is possible that a column for 8 exists, but no values from it were used. Four by nine is a reasonable size for a matrix.

c. Next, we check for repeats and underline them. We also prepare a dinomic frequency count by setting up a 4 by 9 matrix and checking off each dinome that appears.

2023 2029 6224 6322 2144 4420 6362 4924 6529 2769
2043 2123 2227 4627 6521 2221 2723 6527 2349 2144
4481 8287 2423 **4349** 2144 4485 **8089** 6522 2746 **2421**
6365 2263 2142 2027 2324 6322 2144 4420 6362 4627
6521 2221 2723 6560 2144 4441 2047 2123 2422 6680

6666 6522 2746 4263 2069 2122 6425 2729 2924 2343
2123 4700

	1	2	3	4	5	6	7	9	0
2	15	10	10	7	1		11	4	8
4	1	2	3	10		4	2	3	

d. The two longer repeats both include patterns of repeated values, Word patterns can be constructed on repeated dinomes just as they were for repeated single letters. The word patterns for the two longer repeats are shown below.

- A B C D D E A -
24 63 22 21 44 44 20 63 62
A R T I L L E R Y

- A B C D C A E B
46 27 65 21 22 21 27 23 65
P O S I T I O N S

e. The word pattern lists in Appendix D show only one possibility for each pattern as shown. The two are consistent with each other. Using these recoveries, we can set up a matrix and place the values in it and the cryptogram.

```
e n e  y a r t i l  l e r y  a s  o
2023 2029 6224 6322 2144 4420 6362 4924 6529 2769

e  i n t o p o s i t i o n s o n  i I
2043 2123 2227 4627 6521 2221 2723 6527 2349 2144

1     a n     i l  1     s t o p a i
4481 8287 2423 4349 2144 4485 8089 6522 2746 2421

r s t r i e o n a r t
6365 2263 2142 2027 2324 6322 2144 i l 441 l e 630 r y 46p o

s i t i o n s  i l l  e i n a t
6521 2221 2723 6560 2144 4441 2047 2123 2422 6680

  s t o p  r e  i t  o  a n
6666 6522 2746 4263 2069 2122 6425 2729 2924 2343

i n
2123 4700
```

	1	2	3	4	5	6	7	9	0
2	i	t	n	a			o		e
4				l		p			
6		y	r		s				
8									

f. The plaintext words ENEMY and *AIRSTRIKE* are now obvious. Placing the M from ENEMY shows *COMMANDING* at the end of the message. Most of the remaining plaintext letters are easily recovered.

```
                                                l e r y
e n  e m  ya  rt  i  I   4420 6362 h a  s m  o v
2023 2029 6224 6322 2144          4924 6529 2769

e d i  n  to  po  s i  t i  on  so  nh  i l
2043 2123 2227 4627 6521 2221 2723 6527 2349 2144

l      a  n  d  h i l  l                s t o p a i
4481 8287 2423 4349 2144 4485 8089 6522 2746 2421

r s  t r     e o n a  r t
6365 2263 2142 ik2027 2324 6322 2144 il 441 e 6362 ry po 4627

s i  t i  o n  s w                        a t
6521 2221 2723 6560 2144 il 4441 lb 2047 2123 in 2422 6680

     s t  o p  k r  e v  i t   c  o m  m a  n d
6666 6522 2746 4263 2069 2122 6425 2729 2924 2343

i n  g
2123 4700
```

	1	2	3	4	5	6	7	9	0
2	i	t	n	a	c		o	m	e
4	b	k	d	l		p	g	h	
6		y	r		s			v	w
8									

g. The letters in the second row precede all the letters in the third row alphabetically. This suggests an alphabetic structure, although the columns are clearly not in the correct order. The first row probably contains a keyword. If we rearrange the columns so the letters in the second and third rows fall in alphabetical order, we see the next structure.

	1	3	5	7	9	0	2	4	6
2	i	n	c	o	m	e	t	a	
4	b	d		g	h		k	l	p
6		r	s		v	w	y		
8									

h. The plaintext letters area keyword mixed sequence based on INCOME TAX. After placing the remaining letters, there are still 10 blank cells in the matrix. Seven of them are used in the cryptogram, and they cluster together in segments of three or four dinomes. They show the typical pattern of numbers. In particular, the four

plaintext values of groups 50 and 51 of the message indicate time, and 66 is probably a 0. More likely than not, the remaining numbers fill the bottom row of the matrix in numerical order, but these recoveries cannot be confirmed without more information, If hill numbers could be compared to known numbers from an enemy map sheet, we could accept the values with more confidence. At this point, we are reasonably confident of the letter arrangement and the number 0, but the remaining numbers are only a possibility. However, if this were a current real life situation and the enemy referred to by the text is our own forces, we would certainly consider reporting the likelihood of air strikes on our artillery positions.

5-7. Analysis of Monome-Dinome Systems

The characteristics of biliteral and dinomic systems that stand out most are the divisibility by two and the positional limitation that makes it easy to determine matrix coordinates. By changing the length of the plaintext unit from character to character, monome-dinome systems avoid both of these characteristics. In their place, however, the frequency of the numbers (or occasionally, letters) used as row coordinates tends to be higher than the other coordinates. Choosing the highest frequency numbers as row coordinates gives a starting point to reconstruct a monome-dinome system. Consider the next example.

80796	78009	60720	51187	33812
07960	76059	69730	71070	99089
60905	96070	62050	09109	13866
96058	24710	81059	69740	79610
90591	19787	16833	07389	70805
00019	60509	07055	05458	57950
19196	97407	96960	72051	18733
81207	06910	70390	56545	35399
95205	00030	08204		

Numbers:	1	2	3	4	5	6	7	8	9	0
Frequency:	19	8	13	6	22	20	25	16	33	53

a. Repeats are underlined and the number frequencies are shown in the example. A dinomic system can be ruled out, because the repeats are an odd interval apart. The distance between the repeats is 153 characters, counting from the first character of one to the first character of the next. A three-for-one substitution is possible from the position of the repeats, but no patterns or positional limitations appear when divided into threes. The very high frequency of the numbers 0 and 9 in relation to

the other numbers suggests that the system is monome-dinome. The most likely row coordinates are 0 and 9. Other row coordinates are possible, but at this point it is best to start with the most likely candidates only.

b. Begin by breaking the message into monomes and dinomes using only the 0 and 9 as row coordinates. Mark off the divisions in pencil, keeping in mind that some changes may be required later. Start with the first character of the message and work through in order to the end, marking off the monomes and dinomes. Whenever the first character after a division is a 0 or 9, include it with the next character, If it is any other character, leave it as a monome.

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

c. With the divisions in place, we can try a word pattern on the long repeat.

```
96 07 2 05 1 1 8 7 3 3 8 1 2 07
- A  B C  D D E F C C E D B A
R  E  C O  N N A I S S A N C E
```

d. We next set up a monome-dinome matrix with row coordinates 0 and 9 and include the recovered letters. Shown below is the partially recovered matrix and the cryptogram with all letters from *RECONNAISSANCE* placed in the plaintext and the matrix.

ae r ia r e c 0 n n a i s s a n c
8/0 7/9 6/ 7/8/0 0/9 6/0 7/2/0 5/1/1/8/7/ 3/3/8/1/2/

e r e o r s t n e a r
0 7/9 6/0 7/6/0 5/9 6/9 7/3/0 7/1/0 7/0 9/9 0/8/9

o r e c o n s a
6/0 9/0 5/ 9 6/0 7/0 6/2/0 5/0 0/9 1/O 9/ 1/3/8/6/6/

r o a c i n n o r e r n
9 6/0 5/8/ 2/4/7/1/0 8/1/0 5/9 6/9 7/4/0 7/9 6/1/0

o n ä ı n ass e sa o
9/0 5/9 1/ 1/9 7/8/7/ 1/6/8/3/3/ 0 7/3/8/9 7/0 8/0 5/

r o e o o a i
0 o/o 1/9 6/0 5/0 9/ 0 7/0 5/5/ 0 5/4/5/8/ 5/7/9 5/0

r e r r e c o n naiss
1/9 1/9 6/ 9 7/4/0 7/ 9 6/9 6/0 7/2/0 5/1/ 1/8/7/3/3/

anc t e s s
8/1/2/0 7/ 0 6/9 1/O 7/0 3/9 0/ 5/6/5/4/5/ 3/5/3/9 9/

c o a c
9 5/2/0 5/ 0 o/o 3/0 0/8/2/0 4

	1	2	3	4	5	6	7	8	9	0
-	n	c	s				i	a		
0					o		e			
9						r				

e. These reveries suggest additional plaintext, particularly the message beginning *AERIAL RECONNAISSANCE REPORTS ENEMY*. Placing these new values leads to additional recoveries.

a e r i a l r e c o n n a i s s a n c
8/0 7/9 6/ 7/8/0 0/9 6/0 7/2/0 5/1/1/8/7/ 3/3/8/1/2/

e r e p o r t s e n e m y a r
0 7/9 6/0 7/6/0 5/9 6/9 7/3/0 7/1/0 7/0 9/9 0/8/9

m o r e d co I u m n s a p p
6/0 9/0 5/ 9 6/0 7/0 6/2/0 5/0 0/9 1/0 9/ 1/3/8/6/6/

r o a chin g n o r t h e r n m
9 6/0 5/8/ 2/4/7/1/O 8/1/0 5/9 6/9 7/4/0 7/9 6/1/0

o u n t ai n pass e s a t g o
9/0 5/9 1/ 1/9 7/8/7/ 1/6/8/3/3/ 0 7/3/8/9 7/0 8/0 5/

1 f r o m e o o a i f
0 o/o 1/9 6/0 5/0 9/ 0 7/0 5/5/ 0 5/4/5/8/ 5/7/ 9 5/0

u r t h e r r e c 0 n n a i s s
1/9 1/9 6/ 9 7/4/0 7/ 9 6/9 6/0 7/2/0 5/1/ 1/8/7/3/3/

a n c e d u e b y s s
8/1/2/0 7/ 0 6/9 1/0 7/0 3/9 0/ 5/6/5/4/5/ 3/5/3/9 9/

o l b l ac k
9 5/s/o 5/ 0 0/0 3/0 0/8/2/0 4

	1	2	3	4	5	6	7	8	9	0
-	n	c	s	h		p	i	a	-	-
0	f		b	k	o	d	e	g	m	l
9	u					r	t			y

f. Several things remain to be done to complete the solution. The columns can be rearranged to recover a keyword in the top row and alphabetical progression in the next two rows. Additionally, there are two unrecovered segments of text. Both of them include a number of 5s, and the preceding text in each case suggests numbers. The solution is that there is another row in the matrix with the 5 as its coordinate. It was not used enough to select from frequency alone, but once enough text was recovered, the structure can be seen. The added row includes the numbers. The complete solution appears in the next example, with the recovery of specific numbers only tentative.

```
a e r  i a l  r  e c o nnai  ssanc
8/0 7/9 6/  7/8/0 0/9  6/0 7/2/0  5/1/1/8/7/  3/3/8/1/2/

 e  r  t  p  o  r  t  s  e  n  e  m  y  a  r
0 7/9 6/0  7/6/0 5/9  6/9 7/3/0  7/1/0 7/0  9/9 0/8/9

 m  o  r  e  d  c  o  l  u  m  nsapp
6/0 9/0 5/  9 6/0 7/0  6/2/0 5/0  0/9 1/0 9/  1/3/8/6/6/

 r  o  a  c h i n  g  n  o  r  t h e  r n m
9 6/0 5/8/  2/4/7/1/0  8/1/0 5/9  6/9 7/4/0  7/9 6/1/0

 o  u  n  t  a i  npass  e s a  t  g  o
9/0 5/9 1/  1/9 7/8/7/  1/6/8/3/3/  0 7/3/8/9  7/0 8/0 5/

 l  f  r  o  m  e  o  7  6  4  2  .  f
0 0/0 1/9  6/0 5/0 9/  0 7/0 5/5  0/5 4/5 8/  5 7/9 5/0

 u  r  t  h  e  r  r  e  c  o  n  naiss
1/9 1/9 6/  9 7/4/0 7/  9 6/9 6/0  7/2/0 5/1/  1/8/7/3/3/

 d  u  b  y  l  6  0  z
8/1/2/0 7/  0 6/9 1/0 e 7/0 3/9 0/  5 6/5 4/5 0 3/5 3/9 9

 .  c  o  l  b  l  a  c  k
9 5/2/0 5/  0 0/0 3/0  0/8/2/0 4
```

	3	6	7	1	8	2	4	0	9	5
-	s	p	i	n	a	c	h	-	-	-
0	b	d	e	f	g	j	k	l	m	o
9	q	r	t	u	v	w	x	y	z	.
5	0	1	2	3	4	5	6	7	8	9

5-8. Application of Vowel-Consonant Relationships to Multiliterals

Vowel-consonant relationship solutions can be applied to multiliterals, too. As long as you can determine the coordinates of the matrix, you can set up a dummy matrix with any sequence of characters inside as a pseudoplain component. You then reduce the cryptogram to unilateral terms by deciphering with the dummy matrix. Next, solve the resulting unilateral cryptogram using any of the techniques learned with unilateral systems, including the use of trilateral frequency counts and the vowel and consonant lines.

5-9. Solution of Trilateral and Trinomic Systems

Trilateral and trinomic systems are solved in exactly the same way as biliterals and dinomics. The systems are identified by the tendency of messages to break into groups of three instead of groups of two. With simple triliterals and trinomics, positional limitation is even more evident than it is for biliterals and dinomics. Look for a limited set of pairs of characters as either the first pair of characters or the last pair of characters in every three, Once these are found, set up your coordinates and solve as before.

Section III
Analysis of Variant Multilateral Systems

5-10. Identification of Variant Systems

As with any coordinate system, analysis of variant multilateral systems begins with determination of the coordinates. If the product of the row and column coordinates is 50 or more, the system is almost certainly a variant system of some kind.

5-10. Analysis of External Variant Systems - Frequency Matching

External variant systems are generally easier to solve than internal variant systems. Frequency counts can usually be used to determine which coordinates combine with each other on the same row or column, whenever the text is long enough to give a good representative sample, as shown in the next problem.

IIUC RAPC OIPU IANU NMDR NIRI ISIU AIII PSPR AUUN
AMDG ANPG URDU IMMA PRAU MROU RIIM NAMO ICDN UUUA
UIOM ARAA AIII DSMI RRNO MMPU RGUR UNDS NIIA RMMA
PSUC UONM IOAR RADU PUPG OCIA PUMO RCMM MCDR ROIA
SORI ACNM UNRI IMII SMRA ANNA SRNM ROMI NONR RAUC

RIPN SADG AUPR IONA DUUU MRIA OGNR RAIR WIA RGNI
MOPO RAMM WI_I DRPS MIAR MOAC DGUA URAC NISR NOIG
DSSI RORM MINO MURU MMAI DOUA PGRR USXX

	A	C	G	I	M	N	O	R	S	U
A	1	3		3	1	2		3		3
D			3			1	1	3	3	3
I	6	1	1	5	3		2	1	1	1
M	3	1		4	4		4	2		2
N	3			4	4		4	2		1
0		1	1	1	1					1
P		1	3			1	1	3	3	4
R	6	1	2	5	2		3	2		1
S	1			1	1		1	2		
U	3	3		1		3	1	3	1	2

a. The cryptogram used 10 different letters as row coordinates and 10 different letters as column coordinates. Using these coordinates, a digraphic frequency count has been completed as shown. For example, the letter I is paired with itself five times, so the number 5 appears in the matrix at the point where the row and column of I intersect.

b. Examining the frequency count, we can see that there are good frequency pattern matches between certain rows and certain columns. For example, the I row and the R row are nearly identical. Similarly, the A column and the I column are nearly identical. Carrying this process further, we can match the row pairs, AU, DP, IR, MN, and OS. The column pairs are AI, CN, GS, MO, and RU. At this point, we have no idea in what order the coordinate pairs belong or which letter in each pair comes first or if it even matters which letter comes first. We have enough information, however, to reduce the cryptogram to unilateral terms.

c. To reduce the cryptogram to unilateral terms, we set up a matrix with the combined coordinates and write any sequence of letters within it, for example, A through Y.

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

K B	KG	U J	KT	S J	PK	MO	AK	H J	EB
I IUC	RAPC	OIPU	IANU	NMDR	NIRI	ISIU	AI II	PSPR	AUUN
DH	BH	EJ	NP	J E	T Y	KN	PS	LG	EA
AMDG	ANPG	URDU	IMMA	PRAU	MROU	RIIM	NAM3	ICDN	UUUA
AX	EA	AK	HP	OS	S J	ME	BH	PK	NP
UIOM	ARM	AI I	I DSMI	RRNO	MMPU	RGUR	UNDS	NI IA	RMMA
HB	DS	NE	K J	J H	VK	JS	LS	Q J	NK
PSUC	UONM	IOAR	RAW	PUPC	OCIA	PUMO	RCMM	MCDR	ROTA
XK	BS	BK	NK	XK	BP	YS	NP	ST	KB
SORI	ACNM	UNRI	IMII	SMRA	ANN4	SRNM	ROMI	NONR	RAUC
KG	UH	EJ	NP	JE	TK	WT	KO	PK	MP
RIPN	SADG	AUPR	IONA	DUUU	MRIA	OGNR	RAIR	MAIA	RGNI
SI	KS	TK	J H	PE	SB	HA	EB	PY	SM
MOPO	RAMM	MUI	I DRPS	MIAR	MOAC	DGUA	URAC	NISR	NOIG
HU	NN	PS	TO	SA	IA	HO	C		
DSSI	RORM	MINO	MURU	MMAI	DOUA	PGRR	USXX		

d. We see that repeats appear in the pseudotext that results from our trial decipherment. The repeats that were suppressed by the variants are now visible with the variants combined. The recovery of the plaintext is like any of the previous problems. When we recover the plaintext and enter the recovered values in the matrix in place of the trial sequence, we reach the solution shown below.

	A	C	G	M	R
	I	N	S	O	U
A U	l	n	k	g	i
D P	-	m	a	b	r
IR	e	f	d	s	c
MN	t	u	-	o	p
OS	y	z	x	v	w

```
en  em  yr  ep  or  te  dc  le  ar  in
KB  KG  UJ  KT  SJ  PK  MO  AK  HJ  EB
IIUC RAPC OIPU IANU NMDR NIRI ISIU AIII PSPR AUUN

ga  na  ir  st  ri  pw  es  to  fm  il
DH  BH  EJ  NP  JE  TY  KN  PS  LC  EA
AMDG ANPG URDU IMMA PRAU MROU RIIM NAMO ICDN UUUA

lv  il  le  at  co  or  di  na  te  st
AX  EA  AK  HP  OS  SJ  ME  BH  PK  NP
UIOM ARAA AI I I DSMI RRNO MMPU RGUR UNDS NI IA RMMA

an  go  si  er  ra  ze  to  fo  ur  se
HB  DS  NE  KJ  JH  VK  JS  LS  QJ  NK
PSUC UONM IOAR RADU PUPG OCIA PUMO RCMM MCDR ROIA

ve  no  ne  se  ve  nt  wo  st  op  en
XK  BS  BK  NK  XK  BP  YS  NP  ST  KB
SORI ACNM UNRI IMII SMRA ANNA SRNM ROMI NONR RAUC

em  ya  ir  st  ri  pe  xp  ec  te  dt
KG  UH  EJ  NP  JE  TK  WT  KO  PK  MP
RIPN SADG AUPR IONA DUUU MRIA OGNR RAIR MAIA RGNI

ob  eo  pe  ra  ti  on  a I  in  tw  od
SI  KS  TK  JH  PE  SB  HA  EB  PY  SM
MOPO RAMM MUII DRPS MIAR MOAC DGUA URAC NISR NOIG

ay  ss  to  pc  0 l  bl  ac  k
HU  NN  PS  TO  SA  IA  HO  C
DSSI RORM MINO MURU MMAI DOUA PGRR USXX
```

e. With the plaintext values filled into the matrix, we can see in what order the rows and columns belong. Starting with the last row of the internals, we rearrange the columns of the matrix in alphabetic order.

	M	R	G	A	C
	O	U	S	I	N
AU	g	i	k	l	n
DP	b	r	a	-	m
IR	s	c	d	e	f
MN	o	p	-	t	u
OS	v	w	x	y	z

The first row of the internals should follow alphabetically after the third row-scdef, gikln.

	M	R	G	A	C
	O	U	S	I	N
DP	b	r	a	-	m
IR	s	c	d	e	f
AU	g	i	k	l	n
MN	o	p	-	t	u
OS	v	w	x	y	z

f. All that remains is to fill in the missing letters H, J, and Q in the plaintext sequence, and to try to recognize how the coordinates were constructed. As mentioned earlier, it is common practice to couple I with J or U with V when using a 5 by 5 matrix. Since J did not appear in the plaintext, we may assume it occupies an alphabetical position within the I block. The Q clearly belongs between the P and T leaving the H in the top row. The plaintext keyword is BRAHMS (the classical composer). With that as a clue, the letters in the coordinates are shifted to their correct positions, revealing the keywords PIANO, DRUMS, MUSIC, and ORGAN.

	M	U	S	I	C
	O	R	G	A	N
PD	b	r	a	h	m
IR	s	c	d	e	f
AU	g	i/j	k	l	n
NM	o	p	q	t	u
OS	v	w	x	y	z

5-12. Analysis of Variants - Isologs

Two or more encrypted messages with different encrypted text, but the same underlying plaintext are called isologs. When isologs are encountered, your job is much easier. Isologs are particularly useful in solving variant multilateral systems, either external or internal.

a. Isologs can be recognized by one or more of these characteristics—

- Identical message lengths.
- Similar characteristics in the text, such as repeated segments or characters occurring in the same position in each message.

- External indications, such as identical times of file or identical message numbers included in the header for each message. Normally, no two different messages from the same sender receive the same file time or message number. When you see the same time of file on the same date originating from the same unit, the messages are likely to be isologs.

b. Two messages that showed the same time of file in the message header appear in Figure 5-3.

Message 1:

XLNH GVDV NZRH DKXH AMNV RPGZ XMNK DZGP XVDH QHNB
QCFH DVRP GLFP DSAZ RHFB GKNZ DBFL DLGH RSFH QKRB
TSDP QVNK DZFP DKQP QMAC NBRL RPRK NSRV NBFL FBNP
DBLM FZGV ACRK TCTH XPTM AHNL NMRM DBFS FHRH NCRZ
XCFV NBRL FPTS DHGK NKDZ FHNV

Message 2:

GYQB EDAD QTOW ATZM OPFT GSAY OTFD ZDKW KYZY VSQD
EWOS ATGW KTGS FMKP OWFS LTQT ZDEM ARVS ERGW LDFW
OYZB LTFT ZTOS FDVW EWOH QDLR GSZS AMQS QTLM FWQY
ZDGH AWET GPZW GTQM ZRGD EPFM EYKM QTLM GSGW LBAS
OTQW ZTER GWGB QBED ADZD OSAT

Figure 5-3. Isolog example.

c. Each message shows positional limitations. Message 1 has the letters ADFGLNQRTX in the row coordinate position and BCHKLMPSVZ in the column coordinate position. Message 2 has AEFGKLOQVZ in the row coordinate position and BDHMPRSTWY in the column coordinate position. The two messages are not encrypted in the same system, but they appear to be isologs.

d. The initial step in solving these isologs is to see what values equate to each other in the two messages. Pick one of the most frequent digraphs in either message as a starting point. For example, FH occurs four tunes in the first message. A frequency count, while not strictly necessary, may be helpful in spotting the most common values. The digraphs that occur in the same positions in message 2 as FH in message 1 are OS, GW, GS, and another OS.

e. The next step is to find each of the digraphs in message 2 that equated to FH from message 1. The letters OS, GW, and GS in message 2 and the digraphs in the same position in message 1 are underlined in Figure 5-3.

f. We now see that RH, RP, FP, and FH in message 1 equate to GS, GW, and OS in message 2. A check of the new values in message 1 adds the additional digraph OW in message 2, completing the equations for that set. It appears that R and F are variant row coordinates and P and H are variant column coordinates in message 1. Similarly, the message 2 variants are G and 0 on the rows and W and S on the columns.

g. Continue the process by picking additional repeated values. Complete the equations for each, working back and forth between the two messages, just as we did for the initial digraph FH. Continue until all coordinates have been combined, or you run out of digraphs to compare. You can set up a plot to keep track of the equations as shown in the next example.

Row	Column	Message 1	Message 2	Row	Column
FR	HP	RH RP FP FH	GS GW OS OW	GO	SW
DN	BZ	DZ NZ DB NB	QD QT ZT ZD	QZ	DT
	KV	NV DV DK NK	FD FT AD AT	AF	
GQ		QK QV GK CA'	ED ET LT LD	EL	
AL	CM	AM LM AC	OH GP GH OP		HP
TX	LS	XL TS	GB OY GY		BY
		GP GH QP QH	VS VWKW	KV	
		DS DL NS NL	FMAMAR		MR

h. Other combinations could have been selected than the ones shown, but these are sufficient to show all the variants in both matrices. From this point, either message can be reduced to unilateral terms and solved. Then the recovered plaintext can be applied to the other message to complete the recovery of the second matrix. Note that if the same matrix was used in both messages, the similarity should be quickly recognized and the solution accomplished more easily. The next paragraph shows the simpler technique when the same matrix is used.

5-13. Solution Using Isologous Segments

Segments of ciphertext which have the same underlying plaintext are known as isologous segments. A technique similar to the one used in isolog solution can be used any time repeated plaintext can be identified. This is likely to occur with repeated beginnings and endings to messages or with long repeated words and phrases.

a. Recognizing repeated plaintext in variant systems requires painstaking inspection of the ciphertext. Computer indexes of repeated plaintext, which show repeated text on consecutive lines along with the preceding and following text makes repeats

easier to recognize. In any long plaintext repeat, some of the ciphertext digraphs or dinomes are likely to repeat. Other ciphertext digraphs or dinomes are likely to show common row or column coordinates. Pairs with neither row nor column coordinates in common will generally be in the minority. Therefore, although a lot of trial and error may be involved, the longer repeated plaintext segments can often be identified. Consider the two message beginnings shown below.

Message 1:

3469 8489 2469 1420 8957 7238 2311 8840 9626 6269
1429 1622 8924 ...

Message 2:

7338 7127 6934 ...

b. The similarities of the text make it quite clear that the underlying plaintext is the same in both cases, and the same matrix is used for both. Proceeding on the assumption that the plaintext and matrix are the same, it is easy to match the remaining values to determine the variants. For example, from the first dinome in each message, 3 and 4 are column variants. From the second dinome in each message, 8 and 9 are column variants. All the variants can be combined from this short example, and the remainder of the solution is routine.

5-14. Analysis of Internal Variant Systems

Internal variant systems are generally more difficult to solve than external variant systems. With no coordinates to combine, frequency counts do not provide immediate clues to variants. Similarly, isologous segments are harder to recognize. Some characters are likely to repeat in isologous segments with internal variant systems, but the partial repeats caused by common row or column coordinates are much less likely to occur. Still, given enough messages from a single system to produce repeats; given operator carelessness in encryption; or given stereotyped traffic, these systems can readily be solved, too. Once a plaintext entry is formed, the remainder of a solution is not difficult. When you find isologs or isologous segments, you can equate ciphertext values just as was demonstrated in the internal variant examples. The only difference is that you do not combine coordinates through this process, but instead find all cells in the matrix that have the same plaintext value.

5-15. Analysis of Syllabary Squares

Syllabary squares are closely related to small code charts, and the solution of both types of systems is similar. The analysis of syllabary squares produces some distinct differences.

a. Isologs or isologous segments are not necessarily the same length in each case. The encipherment examples below are repeated from paragraph 5-3e.

	6	0	4	3	8	1	7	5	9	2
8	a	1	ad	al	an	and	as	at	b	2
4	c	3	ce	co	d	4	da	de	di	e
3	5	ea	ec	ed	ee	ei	el	en	ent	er
7	es	et	f	6	fi	fo	g	7	h	8
2	hi	ht	i	9	in	ing	io	ir	is	it
0	h	0	00	k	l	la	le	ll	m	ma
5	n	nd	ne	ng	ni	nt	o	on	or	ou
9	p	q	r	ra	re	ri	ro	rs	rt	s
1	se	si	st	t	ta	te	th	ti	tion	to
6	tw	ty	u	ur	v	ve	w	x	y	z

p: r ei n fo r ce m ent s
c: 94 31 56 71 94 44 09 35 13 92

p: re in f or ce m ent s
c: 98 28 74 59 44 09 39 92

b. Isologous segments can often still be recognized by the plaintext values which have no variation. In the example, there is only one way to encipher the letters M and S. When REINFORCEMENTS is enciphered, the ciphertext equivalents of M and S will always be the same. Other values are likely to begin with the same row coordinate, since syllables beginning with the same letter are likely to be on the same row, such as the R and the RE. Still others will have a possible variation, but the variation will not be used. The repeated CE syllable in both segments is an example of this. As a result of all these considerations, isologous segments are often recognizable and provide a point of entry to the system.

c. Solution of syllabary spelling will be further explained in Part Six, Analysis of Code Systems.

PART THREE

Polygraphic Substitution Systems

CHAPTER 6

CHARACTERISTICS OF POLYGRAPHIC SUBSTITUTION SYSTEMS

Section I

Characteristics of Polygraphic Encipherment

6-1. Types of Polygraphic Systems

As first explained in Part One, polygraphic cipher systems are those in which the plaintext units are consistently more than one letter long. The most common type is digraphic substitution, which replaces two letters of plaintext with two letters of ciphertext. There are also such systems as trigraphic and tetragraphic substitution. The larger types are rare, and awkward to use in military applications, so they are not included in this manual.

6-2. Digraphic System Characteristics

The simplest type of digraphic substitution, if not the simplest type to construct, uses a 26 by 26 matrix with plaintext values as coordinates to two-letter ciphertext values within the table. A sample of a digraphic substitution matrix is shown in Table 6-1.

Table 6-I. Digraphic substitution matrix.

	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	WZ	IY	NX	CW	HV	EU	SR	TQ	RP	AO	BN	DM	FL	GK	JJ	KI	LH	MF	OD	PC	QB	UT	VG	XA	YE	ZS
b	IZ	NY	CX	HW	EV	SU	TR	RQ	AP	BO	DN	FM	GL	JK	KJ	LI	MH	OF	PD	QC	UB	VT	XG	YA	ZE	WS
c	NZ	CY	HX	EW	SV	TU	RR	AQ	BP	DO	FN	GM	JL	KK	LJ	MI	OH	PF	QD	UC	VB	XT	YG	ZA	WE	IS
d	CZ	HY	EX	SW	TV	RU	AR	BQ	DP	FO	GN	JM	KL	LK	MJ	OI	PH	QF	UD	VC	XB	YT	ZG	WA	IE	NS
e	HZ	EY	SX	TW	RV	AU	BR	DQ	FP	GO	JN	KM	LL	MK	OJ	PI	QH	UF	VD	XC	YB	ZT	WG	IA	NE	CS
f	EZ	SY	TX	RW	AV	BU	DR	FQ	GP	JO	KN	LM	ML	OK	PJ	QI	U-I	VF	XII	YC	ZB	WT	IG	NA	CE	HS
g	SZ	TY	RX	AW	BV	DU	FR	GQ	JP	KO	LN	MM	OL	PK	QJ	UI	VH	XF	YD	ZC	WB	IT	NG	CA	HE	ES
h	TZ	RY	AX	BW	DV	FU	GR	JQ	KP	LO	MN	OM	PL	QK	UJ	VI	XH	YF	ZD	WC	IB	NT	CG	HA	EE	SS
i	RZ	AY	BX	DW	FV	GU	JR	KQ	LP	MO	ON	PM	QL	UK	VJ	XI	YH	ZF	WD	IC	NB	CT	HG	EA	SE	TS
j	AZ	BY	DX	FW	GV	JU	KR	LQ	MP	OO	PN	QM	UL	VK	XJ	YI	ZH	WF	ID	NC	CB	HT	EG	SA	TE	RS
k	BZ	DY	FX	GW	JV	KU	LR	MQ	OP	PO	QN	UM	VL	XK	YJ	ZI	WH	IF	ND	CC	HB	ET	SG	TA	RE	AS
l	DZ	FY	GX	JW	KV	LU	MR	OQ	PP	QO	UN	VM	XL	YK	ZJ	WI	IH	NF	CD	HC	EB	ST	TG	RA	AE	BS
m	FZ	GY	JX	KW	LV	MU	OR	PQ	QP	UO	VN	XM	YL	ZK	WJ	II	NH	CF	HD	EC	SB	TT	RG	AA	BE	DS
n	GZ	JY	KX	LW	MV	OU	PR	QQ	UP	VO	XN	YM	ZL	WK	IJ	NI	CH	HF	ED	SC	TB	RT	AG	BA	DE	FS
o	JZ	KY	LX	MW	OV	PU	QR	UQ	VP	XO	YN	ZM	WL	IK	NJ	CI	HH	EF	SD	TC	RB	AT	BG	DA	FE	GS
p	KZ	LY	MX	OW	PV	QU	UR	VQ	XP	YO	ZN	WM	IL	NK	CJ	HI	EH	SF	TD	RC	AB	BT	DG	FA	GE	JS
q	LZ	MY	OX	PW	QV	UU	VR	XQ	YP	ZO	WN	IM	NL	CK	HJ	EI	SH	TF	RD	AC	BB	DT	FG	GA	JE	KS
r	MZ	OY	PX	QW	UV	VU	XR	YQ	ZP	WO	IN	NM	CL	HK	EJ	SI	TH	RF	AD	BC	DB	FT	GG	JA	KE	LS
s	OZ	PY	QX	UW	VV	XU	YR	ZQ	WP	IO	NN	CM	HL	EK	SJ	TI	RH	AF	BD	DC	FB	GT	JG	KA	LE	MS
t	PZ	QY	UX	VW	XV	YU	ZR	WQ	IP	NO	CN	HM	EL	SK	TJ	RI	AH	BF	DD	FC	GB	JT	KG	LA	ME	OS
u	QZ	UY	VX	XW	YV	ZU	WR	IQ	NP	CO	HN	EM	SL	TK	RJ	AI	BH	DF	FD	GC	JB	KT	LG	MA	DE	PS
v	UZ	VY	XX	YW	ZV	WU	IR	NQ	CP	HO	EN	SM	TL	RK	AJ	BI	DH	FF	GD	JC	KB	LT	MG	OA	PE	QS
w	VZ	XY	YX	ZW	WV	IU	NR	CQ	HP	EO	SN	TM	RL	AK	BJ	DI	FH	GF	JD	KC	LB	MT	OG	PA	QE	US
x	XZ	YY	ZX	WW	IV	NU	CR	HQ	EP	SO	TN	RM	AL	BK	DJ	FI	GH	JF	KD	LC	MB	OT	PG	QA	UE	VS
y	YZ	ZY	WX	IW	NV	CU	HR	EQ	SP	TO	RN	AM	BL	DK	FJ	GI	JH	KF	LD	MC	OB	PT	QG	UA	VE	XS
z	ZZ	WY	IX	NW	CV	HU	ER	SQ	TP	RO	AN	BM	DL	FK	GJ	JI	KH	LF	MD	OC	PB	QT	UG	VA	XE	YS

p: at ta ck at da wn

c: PC PZ FN PC CZ AK

a. As the example shows, with any digraphic system, repeated plaintext digraphs can cause a ciphertext repeat. Repeated single letters do not cause ciphertextrepeats, Digraphic systems suppress individual letter frequencies, but show normal frequency patterns for pairs of letters. Since there are 676 possible digraphs in the English language, many more groups of text are needed for digraphic frequencies to be very useful as a direct aid to analysis.

b. Repeated plaintext words and phrases cause ciphertext repeats only when they begin in the same odd or even position. If both occurrences of a plamtext repeat begin in the odd position or both begin in the even position, the ciphertext repeats. If one occurrence is in an odd position and one is in an even position, they will produce different ciphertext. As a result, nearly half of all plaintext repeats are suppressed. This is shown in these three alternate examples, all enciphered from Table 6-1.

```
   at ze ro fo ur ze ro ze co st op
   PC CV EJ PJ DF CV EJ CV EJ DC CI

-a tz er of ou rz er oz tr os to p-
-- OS UF PU RB LS UF GS UF SD TJ --

-a tz er ot hr et ze ro ze ro st op
-- OS UF TC YF RV CV EJ CV EJ DC CI
```

c. In the first example, all three *ZEROs* produce a repeat when they all begin in the even position. In the second example, they all begin in the odd position, and only the portions of the three ZEROS that appear as complete digraphs (the ERs) produce a repeat. In the third example, the two ZEROs that begin in the even position produce repeats, but the first ZERO, which begins in the odd position, does not.

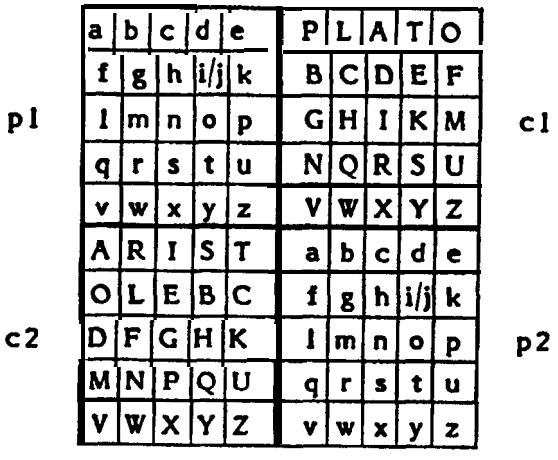
d. The suppression of individual letter frequencies and a significant portion of plaintext repeats means that digraphic systems are considerably more secure than unilateral systems and most multiliterals.

6-3. **Four-Square System**

Large table digraphics are awkward systems for military usage. In their place, there are several much more convenient small matrix digraphic systems available with about the same degree of security. The first of these is the four-square.

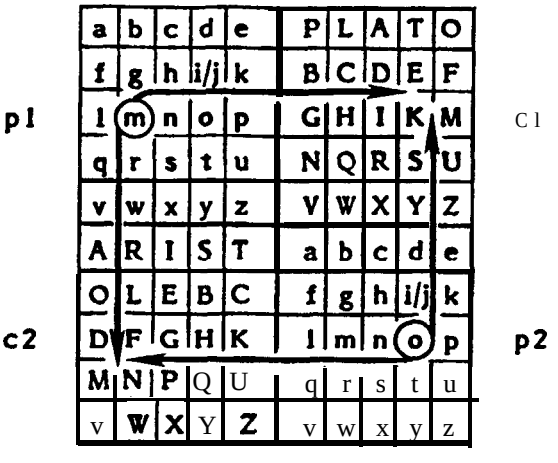
a. The four-square consists of four 5 by 5 matrices in a square. The two plaintext letters and the two ciphertext letters of each encipherment each use a different

square. The squares marked pl and p2 usually, but not always, contain standard sequences. The two squares marked c 1 and c2 can include any mixed sequence.

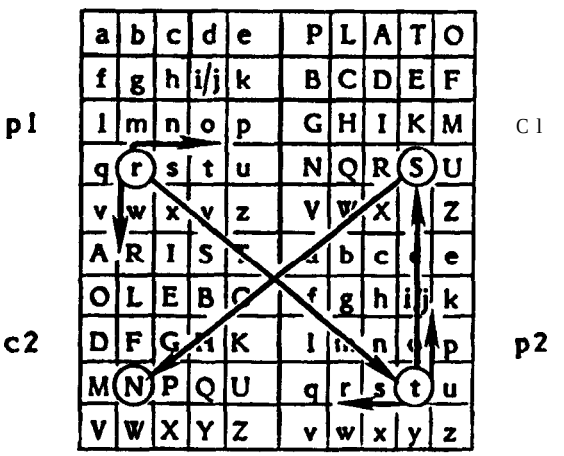


p: mo rt ar fi rt
c: K F S N L M E O U R

b. Encipherment or decipherment follows a rectangular pattern. Whether enciphering or deciphering, the letters of the digraphs are located in the appropriately labeled squares. These letters form diagonally opposite corners of a rectangle. The equivalents, plaintext or ciphertext, are the remaining corners of the same rectangle. For example, plaintext MO determines the rectangle outlined in the square below. Plaintext M determines the upper row and the left column of the rectangle. Plaintext 0 determines the bottom row and the right column of the rectangle. The ciphertext equivalent, KF, is then found in the remaining comers in the appropriately labeled squares.



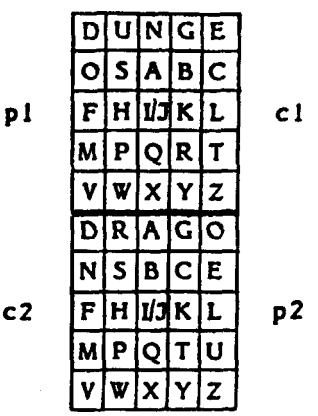
c. For a second example, to encipher RT, R is located in the pl square, and T is located in the p2 square. The ciphertext equivalent of RT is found in the remaining corners of the rectangle prescribed by RT. The first ciphertext letter, S, is found in the cl square in the plaintext T column and the plaintext R row. The second ciphertext letter, N, is found in the c2 square at the intersection of the plaintext R column and the T row. Tracing the letters from pl to p2 to cl to c2 is shown below.



d. Decipherment is handled in exactly the same way, except that the ciphertext letters in the cl and c2 squares determine the rectangle by which the plaintext letters are found.

6-4. Vertical Two-Square

The two types of two-squares are simpler than the four-square system. The first is the vertical two-square, which uses two 5 by 5 matrices one on top of the other. Normally both squares contain mixed sequences.



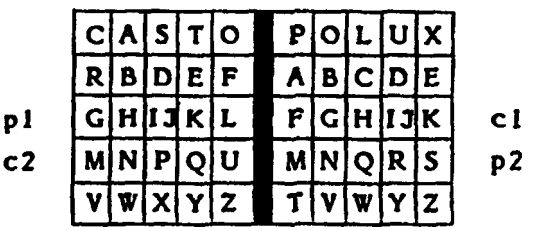
p: al lq ui et on th ew es te rn fr on tx
c: CJ IU NH GU ON PL UZ UE TE MC HD ON QZ

a. The rectangular rule used with the four-square is used with the two-square, also. Whenever the letters to be enciphered are in the same column, however, the letters become their own equivalents. The encipherment of ON and TE in the example illustrates this.

b. The case where the plaintext letters remain unchanged in the ciphertext is called a transparency. A weakness of this system is that in the long run, about 20 percent of the digraphs in a cryptogram will be transparencies. This is enough to give away more plaintext in many cases and enable a speedy solution.

6-5. Horizontal Two-Square

The second kind of two-square is the horizontal two-square, like the vertical, it uses two 5 by 5 matrices.



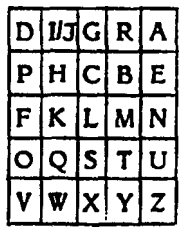
p: we ha vc no ty et be gu nt of ig ht
c: ZB FB ZR NA UY AY EB JC MW PL GI F W

a. The rectangular rule again applies. In the horizontal two-square, values on the same row are replaced with the same letters in the reverse order. This is illustrated by the encipherment of the plaintext letters be and *ig* in the example.

b. Digraphs in ciphertext which are the same as the plaintext in reverse, are called reverse transparencies. Like the direct transparencies of the vertical two-square, they occur in the long run in about 20 percent of the digraphs. They severely weaken the security of the system.

6-6. Playfair Cipher

The Playfair cipher is the most common digraphic system. *Playfair* is always capitalized, because it was named for a Lord Mayfair of England. It is the simplest of systems to construct, using only a 5 by 5 matrix, yet it is more secure than uniliterals and most multiliterals. The rules of encipherment and decipherment are a little more complex than the previous digraphic systems. Sizes other than 5 by 5 are occasionally used.



p: th es ho th ea rd to un dt he wo rl dx
c : QB CU PQ QB NE AJ DT ZU RO CP VQ GM GV

a. The first rule of encipherment and decipherment is the familiar rectangular rule. This applies any time the two letters to be enciphered or deciphered are not in the same row or column. The first four digraphs in the example follow this rule. One additional step must be remembered. In tracing the encipherment or decipherment in the matrix, always move vertically from the second letter to the third letter. For example, to encipher TH, locate the T and the H and move vertically from the H to the letter that is m the same column as the H and the same row as the T. Following this rule, TH is enciphered as QB, not BQ. Similarly, to decipher CU, locate the C and the U, move vertically from the U to find the first plaintext letter E and then the second plaintext letter S.

b. When the two letters to be enciphered or deciphered are in the same row, follow the rule, encipher *right*, *decipher left*. To encipher or decipher, pick the letter to the right or left of each letter of the given digraph, as appropriate. In the example, the plaintext letters R and D are in the same row. They are enciphered with the letters immediately to the right of each letter, producing ciphertext AJ (or AI). If a letter to be enciphered is at the right edge, as in the encipherment of HE, the next letter to the right of the right edge is considered to be the letter in the same row at the far left. The letter to the right of E is P. Similarly, if deciphering, the letter to the left of the left edge is the letter at the far right in the same row. The letter to the left of F is N. Each row is treated as if it were written in a circle with the first letter of a row immediately following the last letter.

c. When the two letters to be enciphered or deciphered are in the same column, use the rule *encipher* below, *decipher above.* To encipher EA in the example, the letters below E and A are N and E respectively. To decipher ZU, the letters above Z and U are U and N respectively. As with the rows, columns are treated as if they were written in a circle. The letter after the bottom letter in a column is the top letter; the letter before the top letter is the bottom letter.

d. The rules *encipher right, decipher left* and encipher below, decipher above produce the acronyms ERDL and EBDA. For many analysts, it is convenient to memorize these pronounceable acronyms to remember the rules.

e. The rectangular rule and the row and column rules take care of all possible cases except double letters. In the Playfair system, there is no rule for enciphering or deciphering a double letter in the same digraph. When double letters are encountered in plaintext in the same digraph, the cryptographer must break up the double letters with a null letter, such as inserting an X between them. As a result, double letters will never be encountered in the ciphertext, except in error. This is only true of the Playfair system. Four-squares and two-squares can handle double letters without any problem.

Section II
Identification of Polygraphic Substitution

6-7. General Digraphic Characteristics

Certain identifying characteristics are common to all digraphic systems. Other characteristics appear only with specific systems.

a. Message lengths, repeats, and distances between repeats are likely to be even in length in all digraphic systems because the basic unit is two-letters. Furthermore, the systems which use 5 by 5 matrices will often only use 25 letters, omitting either the I or the J in ciphertext. In some cases, these values will be used alternately just to ensure use of all letters.

b. Digraphic systems are most often mistaken for biliteral with variant systems, because both exhibit ciphertext which breaks into units of two and both can use most letters. The key distinction to look for between biliterals and digraphics is the complete absence of any positional limitation (paragraph 5-5b) in digraphic systems.

c. Two-square systems stand out because of the director reverse transparencies. Scan the text for the presence of good plaintext digraphs, either direct or reversed, to identify two-square systems. Direct transparencies indicate vertical two-squares; reversed transparencies indicate horizontal two-squares.

d. If no double letters are present in a digraphic, it is probably a Playfair system.

e. Monographic frequency counts for digraphic systems are not as flat as random text and not as rough as plaintext or unilateral systems. They generally fall in between the two. The monographic phi test can be used to confirm this, if necessary.

6-8. Digraphic Frequency Counts

There are several types of frequency counts you can take for working with digraphic systems.

a. The most common way to take a digraphic count is to break the text into digraphs and count those digraphs. For example, given text ABCDE FGHIJ . . . , you would normally break it as AB, CD, EF, GH, IJ, . . , . There are two other ways to take a digraphic count, however. If you are unsure whether there may be indicator groups or null letters at the beginning, you may not know where to begin breaking the text into digraphs. As a comparison, you can skip the first character and begin separating the text into digraphs beginning with the second character. This will produce a cornpletely different set of digraphs than the usual method: A, BC, DE, FG, HI, J . . . . The third way to produce a digraphic count is to combine the two methods to count all possible digraphs. In this case, you would count AB, BC, CD, DE, EF, FG, GH, HI, IJ, ... Unless you have a reason to want an alternate method, stick to the first method.

b. There are two ways to record your count on paper. One is to make a 26 by 26 square on graph paper, and mark the digraphs in the appropriate cells. The other way, useful with short cryptograms, is to write the letters A through Z horizontally, and mark the digraphs by putting the second letter of each digraph under the first letter of the digraph in the A through Z sequence. Then by scanning the columns under each letter for repeated letters, you can readily spot repeated digraphs. This method takes much less space than a 26 by 26 square and gives you the same information

6-9. Digraphic Coincidence Tests

The phi test and phi index of coincidence can be calculated for digraphic frequency counts as well as monographic.

a. The digraphic phi test is calculated in essentially the same way as the monographic test. In the monographic phi test, 1 out of 26 comparisons in random text was expected to be a coincidence for a probability of 0.0385. In the digraphic phi test, 1 out of 676 comparisons is expected to be a coincidence for a probability of 0.0015. The

probability of a coincidence in plaintext is 0.0069 instead of 0.0667. Thus, the formulas for the digraphic phi test are—

$$2\,\phi p = 0.0069\ N\ (N - 1).$$

$$2\,\phi r = 0.0015\ N\ (N - 1).$$

$$2\,\phi o = \Sigma f\ (f - 1).$$

$$2\,\Delta IC = \frac{676\ \Sigma f\ (f - 1)}{N\ (N - 1)} = \frac{2\,\phi o}{2\,\phi r}.$$

N is the total number of digraphs counted.
The frequency of each repeated digraph Is f.

b. As discussed in the first part of this chapter, digraphic ciphertext frequencies will occur with the same numbers as plaintext frequencies when digraphic systems are used. If the digraphic ϕo is close to ϕp but the monographic ϕo is low, the system is likely to be a digraphic system. If you are using the index of coincidence form of the test, the expected 2 AIC is 4.6. The results are much more variable than the monographic test, because of the large number of different elements counted, but it can still be used as a guide. As with any statistical test, the results should not be used by themselves, but used along with all other available information.

6-10. Examples of System Identification

Three messages in unknown systems follow to show the process that leads to system identification. Repeats are underlined, monographic and digraphic frequency counts are shown, and monographic and digraphic ICs are calculated for each. The three messages were all sent from the same headquarters to subordinate elements, and all contained a common message serial number in their header.

a. Message texts and data.

Message 1:

TVCX XSWM WZWV JEVH HCJS IUZZ TVKP VYUY JWTZ CUIK
XCEI SVJC XIUT IDDI ETWM IWHH ISWC TIXP ZTVK RIKU
IKCU ISDV UHVM IRPC WUTU CJZK VUTV JTNI XMIB VYUZ
JVTW EIZT VKEC JEIX CCXX XICM IZEV HHCK CZZI ZEVH
HCCJ SYJJ IEIZ ZCUP HISW ECXX UVEI SYUI ZZTV KKIJ

AUII J

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	19	3	11	0	0	10	28	13	11	0	5	1	0	4	0	2	8	13	15	18	10	11	5	16

Total letters = 205

Monographic IC = 1.74

	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	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	1	0	0	0	0	0	0	2	1	0	1	0	0	0	0	0	0	0	2	0	0	1	0	1
D	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E	0	0	2	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	2	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	1	0	1	1	0	0	0	1	1	2	0	0	0	0	0	0	1	2	0	1	0	1	1	0	2
J	0	0	1	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0
K	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	2	0	
T	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	
U	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	1	1
V	0	0	0	0	0	0	0	2	0	0	2	0	1	0	0	0	0	0	0	0	1	0	0	0	2	0
W	0	0	1	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	1	0	0	0	1
X	0	0	1	0	0	0	0	0	2	0	1	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Z	0	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2

Total digraphs = 102

Digraphic IC = 3.41

Message 2:

NPEC MISY DQQR PATH GFTS LYUV DNPR RWIP SPDR AGYL
RKBE FIPO EGLY RFCZ AFFP SYLE KZLF SDFN LRVI NPOC
CRYL NCYL FMPT HTYA IWES TNNE VARP TNPO OZLR YAOW
I PAV PNUE A INP XKGV EFCE EGKY RLGS AI BP KZCF NCUV
IAUA THGF GVSI PVRA EFUV AGYI LFSD EBKR TPEF SIYL

WDN PRLA VNYL ARXX

A	B	C	D	E	F	C	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
15	3	5	6	13	14	12	3	12	0	6	14	2	13	5	18	2	15	10	8	6	11	3	3	13	4

Total letters = 216

Monographic IC = 1.26

	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	0	0	0	0	0	1	2	0	2	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
B	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
D	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	1	0	0	0	0	0	0	0	0
E	0	1	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
I	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	0	0	0
J	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	2
L	1	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2	0
M	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	1	0	1	2	1	0	1	1	0	1	0	1	0	1	0	~	0	~	0	~	0	~	0	~	0
O	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
P	1	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	2	0	1	0	1	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
R	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0
S	0	0	0	2	0	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0
T	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0	1	0	0	1	0	0	0	0	0	0	0
U	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0
V	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Y	2	0	0	0	0	0	0	0	1	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Total digraphs = 108

Digraphic IC = 5.38

Message 3:

GMGH NGMO RWOG GOEG HWMM HOHR GLNM GEGG HDND HADD
OONL MFRM GFER MLEE GEYO NANW GAGW GFRF YDYL DOMA
MRYG YFOW ODGR HLNG RWDW YAGM OOOL OAOW NFHM GOAD
DOGW GDHG DWDG HOYD GMOO OWAR MMHM GERL NEOO RANL
DWRL NDNA DOOG DLHR YLHG HEED OWYR ERNG HWYA HFYL

YGGL RFML GRYA HFHE GAGM EOOW RWAG DOOM GRNW NLMF
HLEH GFGO YMOW RMHF GERA NMYD HAYF OORW NGYD MWRO
MODW NDEG DOMM YMHR GGHD YDMA NGMF RMDW MMNF HEHD
GHND YGGL ODYW GAHL OONF OWRF MMYG YAAE HDDO DDHW
YMNG MORL YLGE YFDW DGNO NAOO MFRM HMGR RAOE DOGL

DRNL OWDO HAXX

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
23	0	0	40	20	19	54	32	0	0	0	22	40	26	50	0	0	31	0	0	0	0	27	2	26	0

Total letters = 412
Monographic IC = 2.16

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

Total digraphs = 206
Digraphic IC = 8.90

b. Different analysts might approach the identification of the systems used in these messages in different ways, but here is one example of how the systems can be identified.

(1) Although the messages all carry the same message serial number, which is usually a sign of isologs, the messages are all different lengths. If they are isologs, they are not enciphered in the same system.

(2) A comparison of monographic frequency counts confirms that they are in different systems. The highs and lows in each frequency count are too different for any possibility of repeated use of the identical system.

(3) The ICs give a different picture in each. Message 1 has monographic and digraphic ICs consistent with plaintext or a unilateral system. The digraphic IC of 3.41 is slightly below the expected 4.6, but it is within acceptable limits. Message 2 shows a low monographic IC of 1.26, but the digraphic IC of 5.38 is also well within plaintext limits. This is typical of digraphic systems. Message 3 is quite high in both monographic and digraphic ICs.

(4) Messages 1 and 2 use nearly all letters. Message 3, which is twice as long as message 1, uses only 14 different letters. The high ICs and the limited letter usage are consistent with a biliteral with variants system. A close inspection of the digraphic frequency count will show rows and columns with very similar patterns, suggestmg external variants that can be combined. Different letters are used in the row position than those used in the column position. This positional limitation confirms the identification of a biliteral with variants system.

(5) Message 1 has the most repeated text, which is consistent with a unilateral system. Message 2 has only a few repeats and message 3 has only short and fragmentary repeats. In message 3, the fragmented repeat on lines 7 and 10 are in the identical relative position in message 2 as the ZTVK repeat in lines 2 and 5 of message 1. This similarity strongly confirms that the two messages are isologs.

(6) The identifications of the systems in messages 1 and 3 are clear at this point, but message 2 still needs to be clarified. The underlined repeats in message 2 are in the same relative position as in message 1, if you adjust for the slightly increased length of the message. Only some of the repeats from message 1 appear in message 2, however. This is consistent with a digraphic system, which will only show repeats that begin in the same even or odd position.

(7) In message 2, a check of the long diagonal from the AA position to the ZZ position of the digraphic frequency count shows that the only double letter that appeared was the filler XX at the end of the message. The Playfair is the only

digraphic system which will not show double letters. Finally, because the Playfair cannot encipher double letters, all double letters that occur in digraphs must be broken up by the insertion of null letters. This characteristic explains how it can be an isolog, but appear slightly longer. The three messages are all clearly isologs, and the systems are confidently identified, lacking only the final solution for full confirmation. Solution techniques for each of the major digraphic system types are explained in the next chapter.

CHAPTER 7

SOLUTION OF POLYGRAPHIC SUBSTITUTION SYSTEMS

Section I
Analysis of Four-Square and Two-Square Ciphers

7-1. Identification of Plaintext

Recovery of any digraphic system is largely dependent on the ability to correctly identify or assume plaintext. As with any system, isologs and stereotyped messages can help a great deal. Pattern words can also be of assistance. With unilateral systems, patterns of repeated letters provided an assist. With digraphic systems, patterns of repeated digraphs can do the same thing. Appendix D, beginning on page D-38, includes several types of word pattern tables. The first type, listed on pages D-38 and D-39 shows patterns applicable to any digraphic system. The means of representing digraphic patterns are simpler than those for unilateral patterns. The patterns identify the repeated digraph in a word or phrase by the letters AB in each case, and nonrepeating digraphs are just represented by dashes. Here are a few examples that show how the patterns are formed.

DE CO DE
AB -- AB

PO ST PO NE
AB - - AB --

MA IN TA IN IN C-
-- AB -- AB AB --

-M AI NT AI N-
-- AB - - AB --

7-2. Solution of Regular Four-Squares

Regular four-square ciphers, in which the plaintext squares are in A through Z order, are slightly easier to solve than the type with all mixed squares.

a. With the known plaintext squares, an additional type of word pattern can be used. Since the plaintext locations are fixed, certain words will always produce single letter ciphertext repeats. The word MI LI TA RY, for example, will always produce a repeated ciphertext letter in the first and third cipher position. When MI LI TA RY is enciphered by the matrix shown in paragraph 6-3, it produces KL KO NS SW. Four-square word patterns are shown on pages D-43 through D-47. The patterns are represented by the repeated letters only, placing A, C, E, and soon in the first letter positions of digraphs, and B, D, F, and so on in the second letter positions. Repeats between different positions are ignored. Following these rules, a few examples of four-square word patterns appear below.

re qu es te d-
UR UM AU US OY
A- A- -- A- --

el em en ts
PK LK AK RQ
-B -B -B --

qu ar te rm as te r-
UM LM US QF Ah4 US RW
AB -B AD -- -B AD --

b. Identifying the four-square from other digraphic systems is largely a matter of elimination. It will include double letters, unlike the Playfair. It will not include a high proportion of good plaintext digraphs or reversed plaintext digraphs like the two-squares. There is no ready clue to tell whether a four-square is a regular one or not, but it is often easiest to assume the simplest case for a start and only consider more complicated construction when the simple case fails to produce a solution.

c. To demonstrate the use of four-square word patterns and recovery of the system, consider the cryptogram shown below.

TATO UTOD HIDM FIPK ROFM HRVH BMAH NHKM UNAN ZMRO
SKHH RQBX FSYF KQNS QFAT KQUY SMQP SMNT MYRO RYDM
F I PK ROFM IQLT TYSQ RYRV FEDC ATGR RHTO AOTD QP

d. The underlined repeats give a chance to try a four-square word pattern as an entry to the cryptogram.

DM FI PK RO FM
-B A- -- -- AB

The only word with this pattern in Appendix D is INFORMATION. Placing *INFORMATION* in the text, and beginning reconstruction of a regular matrix produces the next example.

in form atio n
TATO UTOD HIDM FIPK ROFM HRVH BMAH NHKM WAN ZMRO

in
SKHH RQBX FSYF KQNS QFAT KQUY SMQP SMUT MYRO RYDM

form atio n
FIPK RON IQLT TYSQ RVRV FEDC ATGR RHTO AOTD QP

PI

a	b	c	d	e				R	
f	g	h	i/j	k			D	F	
l	m	n	o	p					
q	r	s	t	u		P			
v	w	x	y	z					
					a	b	c	d	e
					f	g	h	i/j	k
I	K		M		l	m	n	o	p
O					q	r	s	t	u
					v	w	x	y	z

H I c l

c2

p2

R

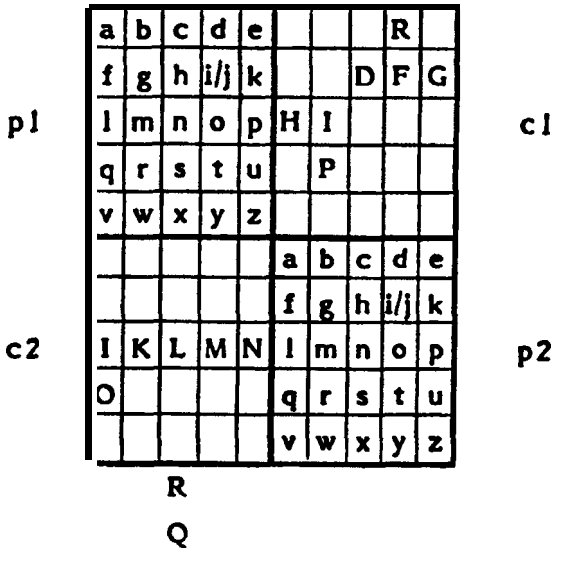
Q

e. The recovered values have been placed in the matrix, and the alphabetic construction is apparent. Additionally, four values have been placed outside the matrix for the moment as suggested by the plaintext Ns at the end of INFORMATION. H and I must be in the same row as plaintext N. R and Q must be in the same column. Several additions can now be made from the alphabetic construction. L and N fit in the third row of the c2 matrix. Further, if H and I are in the third row of the cl matrix, then they must be the first two letters on that row and G is the last letter of the second row. Placing all of these in the matrix and using the partially recovered matrix to decipher as much plaintext as possible produces the next example.

llin form atio n at
TATO UTOD HIDM FIPK ROFM HRVH BMAH NHKM WAN ZMRO

c at in
SKHH RQBX FSYF KQNS QFAT KQUY SMQP SMNT MYRO RYDM

form atio n h
FIPK ROFM IQLT TYSQ RYRV FEDC ATGR RHTO AOTD QP



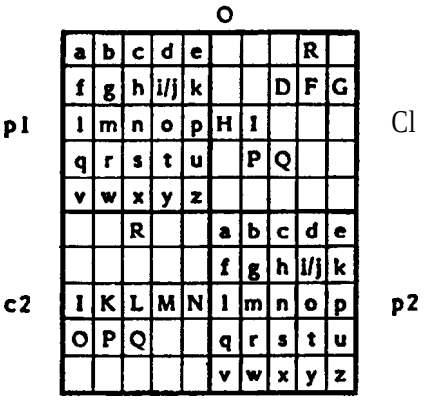
f. Next, suppose that Q in the cl matrix is in the keyword. If so, the U would normally be with it. There are not enough letters left in the alphabet after the P in the cl matrix to put both Q and U at the beginning, so Q is almost certainly right after the P.

g. We can be fairly confident of the recoveries up to this point. A number of possibilities present themselves, but as they are only possibilities, the work should be done lightly in pencil. We can next try placing the Q and R in the c2 matrix. The Q is more likely to be in the sequence than the keyword, so we will tentatively place it in the fourth row and R in the first row. We can place P in the fourth row, also, before Q. Another possibility is to place plaintext A on line one of the message, forming the word *ALL* before INFORMATION.

a llin form atio na at
TATO UTOD HIDM FIPK ROFM HRVH BMAH NHKM UNAN ZMRO

ct rs at in
SKHH RQBX FSYF KQNS QFAT KQUY SMQP SMNT MYRO RYDM

form atio nr he rs
F I PK ROFM IQLT TYSQ RYRV FEDC ATGR RHTO AOTD QP

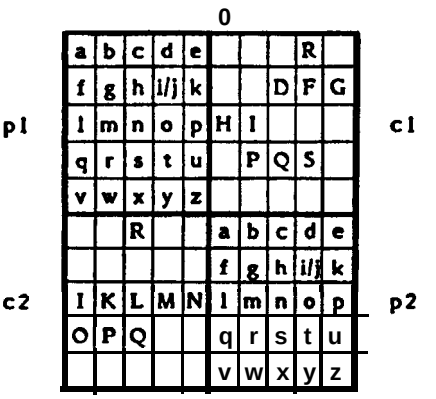


h. Next consider the plaintext RS on line two. It must certainly be preceded by a vowel, therefore, the ciphertext digraph SM must produce a vowel m the p2 position. The only vowel in the same row in the p2 matrix as the ciphertext M in the c2 matrix is plaintext 0. S must be in the fourth column of the cl matrix above the plaintext 0. The only logical place for the S is on the fourth row. Adding the S and entering the values increases our solution as shown in the next example.

a llin form atio na at
TATO UTOD HIDM FIPK ROFM HRVH BMAH NHKM UNAN ZMRO

ct tors to at in
SKHH RQBX FSYF KQNS QFAT KQUY SMQP SMNT MYRO RYDM

form atio nr st he rs
FIPK ROFM IQLT TYSQ RYRV FEDC ATGR RHTO AOTD QP



i. These additions suggest several possibilities. STOP may appear in the middle of line 2. REQUEST may be the word after *INFORMATION* on line 3. Placing these values produces good alphabetical progression in the matrix and many more plaintext possibilities.

```
  qu     a  l l in  f o r m  a t i o  n a          on          at
TATO UTOD HIDM FIPK ROFM HRVH BMAH NHKM WAN ZMRO

ro   ct   i t   nsou      ns   tots topu pdat edin
SKHH RQBX FSYF KQNS QFAT KQUY SMQP SMNT MYRO RYDM

f o r m  a t i o  n r e q  u e s t  edby          qu    e rs
FIPK ROFM IQLT TYSQ RYRV FEDC AT& RHTO AOTD QP
```

p1

a	b	c	d	e	L			R	
f	g	h	i/j	k			D	F	G
l	m	n	o	p	H	I	K	M	N
q	r	s	t	u	O	P	Q	S	T
v	w	x	y	z					

c1

		R		Y	a	b	c	d	e
					f	g	h	i/j	k
I	K	L	M	N	l	m	n	o	p
O	P	Q	S	T	q	r	s	t	u
U	V	W	X	Z	v	w	x	y	z

c2 p2

j. From here, the solution is routine. *REQUEST* is the first word. HEADQUARTERS is the last word. These values in turn fill in enough blanks in the matrix to recognize the keywords and complete the solution. The keywords are LAUREL and HARDY.

7-3. Solution of Mixed Four-Squares

Slightly different techniques must be used when standard sequences are not used in the p1 and p2 squares. The specific four-square word patterns of Appendix D, pages D-43 through D-47 no longer apply, although the general digraphic patterns that precede them on pages D-38 and D-39 are still applicable. Generally, because the matrix construction is less orderly, more text must be known or assumed to successfully complete the solution. The problem that follows shows how the solution can be approached with mixed squares.

FMFE FMPX ZPYX IYYP GGME TXGS YGGB YLF I HAGB YLMK
MRGH YRFM BYYP MMBQ YMHD MHLN MNOS YPVI DMXH RPGL
MNSO QLMP CBYL VGQI QLYX KTZG HEEM GBKM FLYK PHMA
SREE GDMK DEBG TTEB IXCN VINI SOSC HHIG THHM OQPO
TGKI VGQI PMXR CPGH YRSE PLMN LNMN ACVC OCCO KPWC

PKIP PCSU GHYR FKSC YGXX

a. The above cryptogram has been identified as a four-square. Previous messages from the same headquarters have been signed by ADAMS or MILLER. The repeated segments in the text suggest several possibilities for plaintext.

(1) The AB -- AB pattern at the beginning fits the common stereotype REFERENCE.

(2) The repeated GBYL segments appear to be numbers, and the number of characters is exactly right to fit in the expanded stereotype REFERENCE YOUR MESSAGE *NUM*BER, before the numbers. To add to this, recent messages from the addressee have been numbered in the mid 4500s. FOUR FIVE *FOUR* is probably the text of the first three numbers.

(3) GHYR occurs at good sentence length intervals and is probably STOP.

(4) These possibilities give enough values to begin reconstructing the matrix.

b. If you assume that standard 1 and p2 squares were used, entering the values in the matrix produces conflicts. ! he squares must be mixed. To recover a mixed four-square, divide a sheet of cross-section paper into four areas, representing the four squares. The areas cannot initially be limited to 5 by 5 squares, il though eventually the recovered values will condense into that size. Proceed by entering each plaintext and ciphertext pair of digraphs into the appropriate areas, maintaining the rectangular relationship. Start new rows and columns for each pair entered unless there are one or more values in common with previous entries. The entries for the first seven pairs are shown in the next diagram.

reft rtnc eyou rmts sage numb trfo urfi vtfo ur
FMFE FMPX ZPYX IYYP GGME TXGS YGGB YLFI HAGB YLMK

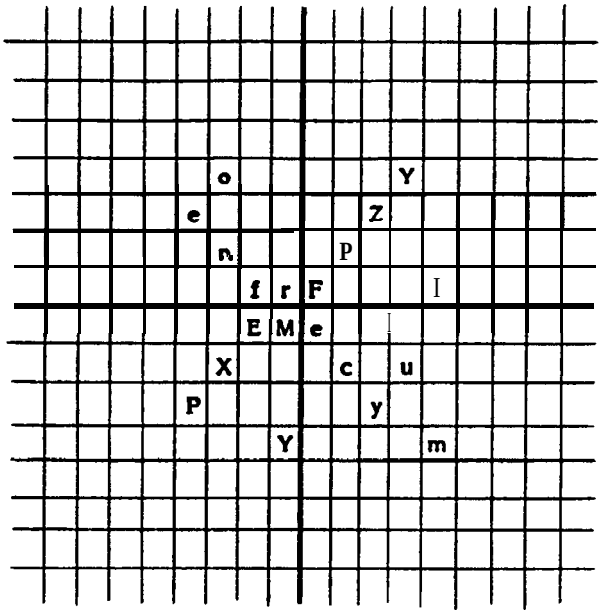
st op
MRGH YRFM BYYP MMBQ YMHD MHLN MNOS YPVI DMXH RPGL

four
MNSO QLMP CBYL VGQI QLYX KTZG HEEM GBKM FLYK PHMA

SREE GDMK DEBG TTEB IXCN VINI SOSC HHIG THHM OQFO

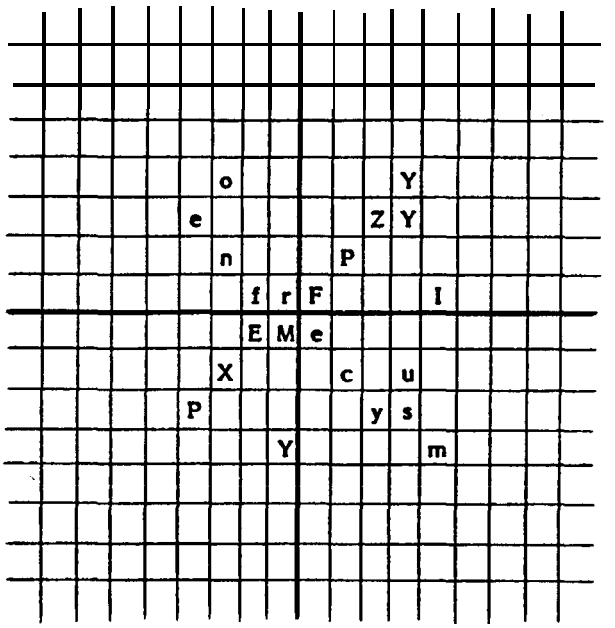
st op
TGKI VGQI PMXR CPGH YRSE PLMN LNMN ACVC OCOO KPWC

stop
PKIP PCSU GHYR FKSC YGXX



c. The first digraph pair entered was plaintext re equalling cipher-text FM, appearing in the inner corners of the four areas. We will use the notation re=FM to represent such pairs from here on with the plaintext in lower case. The next pair, fe=FE was placed on the same row as the first pair because of the common letters with the first pair. The entries continue, placing the letters on new rows and columns except when previously used values occur. The eighth pair, es=YP, presents a new situation Plaintext e and ciphertext Y are already on different rows. The new pair shows

that these two rows should be combined. The diagram below shows the entry before combining the rows. The rows are combined by writing the plaintext o of the first row in the same position on the second row.



d. When all entries have been made and all rows and columns combined wherever possible, the diagram appears as shown below. All plaintext that can be deciphered from the partially recovered matrix is also filled in.

rtft rtnc tyou rmts sage numb trfo urfi vtfo ur
FMFE FMPX ZPYX IYYP GGME TXGS YGGB YLFI HAGB YLMK

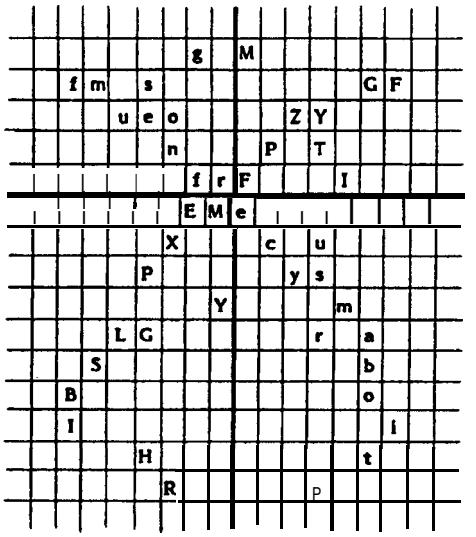
st oprc es es a
MRGH YRFM BYYP h & Q YMHD MHLN MNOS YPVI DMXH RPGL

four ou e fo
MNSO QLMP CBYL VGQI QLYX KTZG HEEM GBKM FLYK PHMA

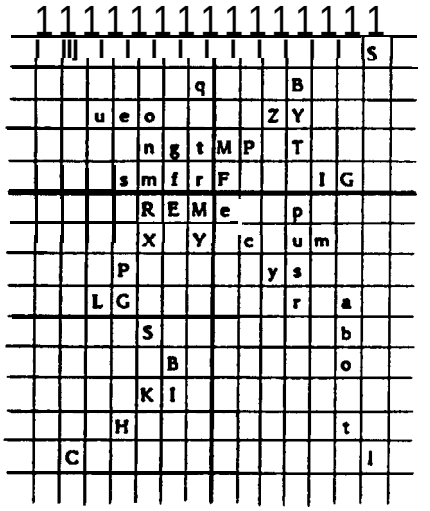
SREE GDMK DEBG TTEB IXCN VINI SOSC HHIG THHM OQPO

r st op
TGKI VGQI PMXR CPGH YRSE PLMN LNMN ACVC OCOO KPWC

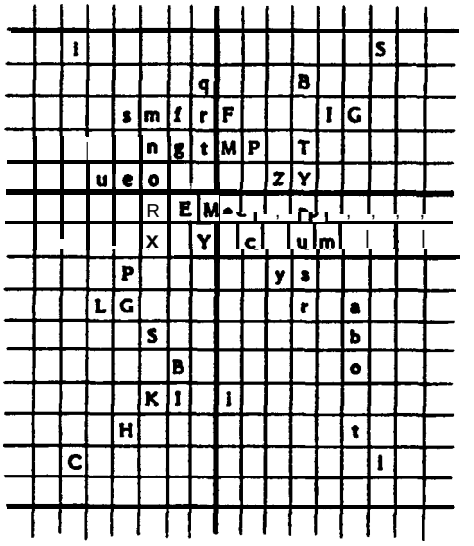
stop er
PKIP PCSU GHYR FKSC YGXX



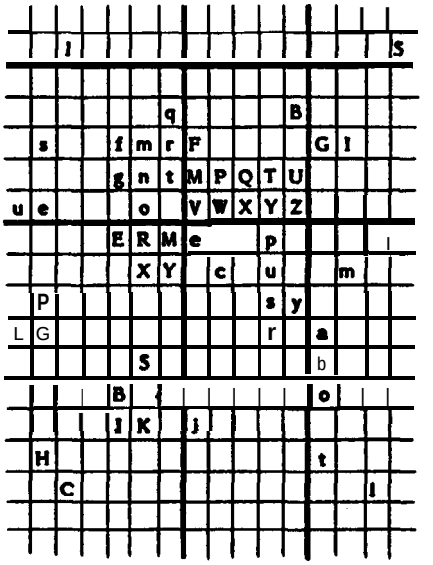
e. More plaintext can be added at this point. The four-letter number after FOUR FIVE FOUR must be NINE, because ZERO will not fit properly in the matrix. The word beginning at the end of the first line is probably *REQUEST*, and the sender is MILLER, not ADAMS. When these recoveries are added to the matrix, there are enough recoveries to see the basic structure of the four-square.



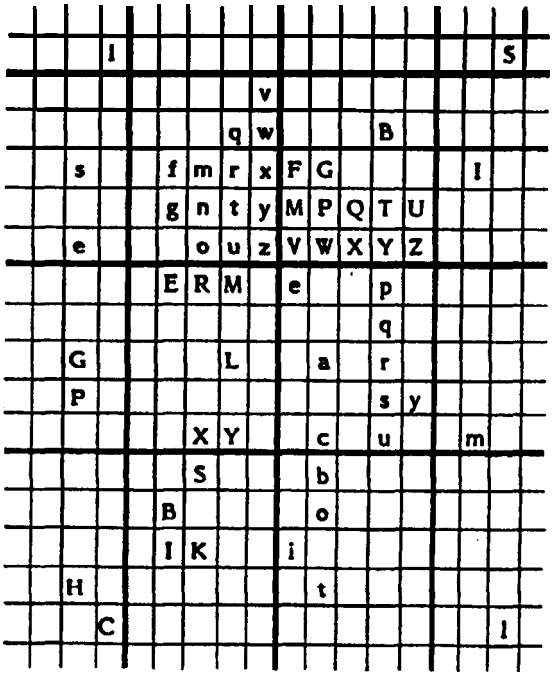
f. Each area shows signs of alphabetic progression. The upper right area shows partial rows with the letters FGI, MPT, and YZ. The lower left has rows with IK and XY. The upper left has columns with fg, mno, and qrt. The lower right has a column with prsu in it. These patterns suggest that the plaintext squares (upper left and lower right) use sequences entered by columns and the ciphertext squares use sequences entered by rows. With this in mind, the rows and columns can be rearranged. The most obvious place to start is to rearrange the rows so that the partial sequences FGI, MPT, and YZ are the last three rows in the upper squares.



g. Moving these three rows put the letters mno and fg in the correct order in the upper left area. The row before these. three rows also appears to be correct1y placed. Now examine the column arrangement. In the upper right area, the Y and Z are probably in the last two columns in the original matrix, With the T placed directly above the Y, there are just enough spaces to fill in UVWX between the T and the YZ on the bottom two rows. Then, with the U appearing in the alphabetical progression, the Q is probably the missing letter on the fourth row. The complete fourth row can be placed in MPQTU order. Similar1y, in the upper left area, the fg, mno, and qrt columns are probably the second, third, and fourth columns of that matrix. We can now rearrange the columns so the first five columns on each side of the center line reflect the original order.



h. The rearranged matrix suggests many more possibilities. In the upper left area, uvwxyz can be filled in as was done with the upper right. In the upper right, the G can be moved next to the F, combining two columns. Rows can be rearranged in the lower areas. Examining the lower right area, the fourth column must include the q by the same logic as was used in the upper right area. The correct order is pqrsu.



i. All the rows and columns outside the 5 by 5 squares can be systematically placed in the squares by following the alphabetical order. Fully combined, the four-square appears below.

			p	v					
		l	q	w	S			B	
s	f	m	r	x	F	G	I	K	L
	g	n	t	y	M	P	Q	T	U
e		o	u	z	V	W	X	Y	Z
H	E	R	M		e	t		p	v
	B	C	D	F	I	o		q	w
G	I	K	L		i	a		r	x
P	Q	S	T	U		b		s	y
V	W	X	Y	Z		c	m	u	z

j. The remaining values are easily recovered by using this matrix to fill in more plaintext in the cryptogram. The additional plamtext will suggest still more plaintext, which can be used to complete the four-square.

7-4. Solution of Two-Square Ciphers

The solution of two-square ciphers, either horizontal or vertical, is similar to the solution of a mixed four-square, only much simpler. The worksheet is divided into two areas by a vertical or horizontal line, as appropriate, instead of four. Plaintext is much easier to recognize because of the transparencies that occur. Matrix reconstruction proceeds, like the four-square, by entering digraph pairs in their rectangular relationship, except for transparencies, which are plotted in the same row or column. New values are plotted in new rows and columns, unless one or more values are in common with previous plots, as with the four-square. As recovery proceeds, working back and forth between the matrix and the text, the two-squares can be combined and condensed to the original form, like the four-square.

Section II
Analysis of Playfair Ciphers

7-5. Security of Playfair Ciphers

Breaking into Playfair ciphers is similar to the solution of mixed four-squares in some respects and very different in others.

a. The Playfair shares the rectangular principle of encipherment with four-squares and two-squares, but it is complicated further by the EBDA and ERDL rules. When recoveries are plotted, every possible rule must be considered, not just the rectangular rule.

b. Recognition of plaintext is aided by another type of word pattern that occurs with Playfair only. Whenever a plaintext digraph is repeated in reverse order, the ciphertext appears in reverse order, too. This does not happen with four-squares and two-squares. It occurs whichever rule of decipherment is used. The word DEFENDED, for example, has a Playfair word pattern of AB -- -BA, the same as DEPARTED, RECEIVER, and a number of others. Playfair word patterns are listed in Appendix D, pages D-40 through D-42. The general digraphic word patterns of pages D-38 and D-39 can also be used.

7-6. Reconstruction of Playf air Ciphers

To illustrate the analysis of Playfair ciphers and the reconstruction of the Playfair matrix, consider the following message. This message was sent from a brigade headquarters to three subordinate battalions.

DT BV VF GO OG MV CQ IH NS MN VI FC IK FK NX KH UB GK AV LH
CA CF WC YC IA VM PB CI FK CA GV UH NC BX OV LY NU CQ ED GO
OG MV CQ VW OV UB QH CM CM QM UO BX OV YG DH HB KR CY OG MV
CQ IH NS NS QR EX I U GO OG OE GO XK AV DT CB XK AV XK AV YV
TQ RH OC NS NB CS LG FN RH GO CV MX VM St FU CM GO XK AV KT
GH KT GH DT CB YV TQ

a. Initial plaintext recoveries are fairly easy with this message.

(1) The XK AV repeats on line four strongly suggest ZE *RO* with another four digit letter group in between them. The numbers are most likely to be a spelled out time.

(2) YV TQ, appearing after the time and at the end of the message, is probably *ST OP.*

(3) The series of four letter repeats beginning with ZE RO at the end of line five and continuing on line six before the final ST OP is probably another time.

(4) The repeat GO OG MV CQ has a number of possibilities in Appendix D, but in the context in which the message was sent, it is most likely to be *B AT TA LI ON.*

(5) If BATTALION is correct, then the partial repeat beginning at the end of line three represents the plaintext *TA LI* ON. This is again part of the word BATTALION, but the word started out as an even letter division with the digraph *BA.* TT, the next digraph, is impossible with the Playfair system, so a null must have been inserted, probably TX. With the addition of the null, the remainder of the word is divided into digraphs, as before, to produce the partial repeat.

(6) The ciphertext in the middle of line four, GO OG OE GO, which deciphers as *AT TA* -- *AT* using the common values from *B AT TA LI ON,* is probably *AT TA CK AT.*

b. These plaintext recoveries give more than enough information to reconstruct the original Playfair matrix. The trickiest step in matrix reconstruction is to pick the best starting point. As every possibility for the matrix is plotted, it can get very

complicated. Careful selection of what values to place first can reduce the complexity a great deal. The cryptogram is repeated below with all recovered values filled in to assist in finding the best starting point.

```
          b  at  ta  li  on
DT BV VF GO OG MV CQ IH NS MN VI FC IK FK NX KH UB GK AV LH

                    ll                                  on  b  at
CA CF WC YC IA VM PB CI FK CA CV UH NC BX OV LY NU CQ ED GO

ta  li  on                                          ba  tx  ta  l  i
OG MV CQ VW OV UB QH CM CM QM UO BX OV YG DH HB KR CY OG MV

on                          at  ta  ck  at  ze  ro      ze  ro  ze  ro  st
CQ IH NS NS QR EX IU GQ OG QE GO XK AV DT CB XK AV XK AV YV

op                                          ll          at  ze  ro
TQ RH oc NS NB GS LG FN RH GO cv MX VM si FU CM GO XK AV KT

          St  op
GH KT GH DT CB YV TQ
```

(1) Usually the best starting point, if available, is to select a digraph pair where there is a letter in common between the plaintext and ciphertext digraphs. These only occur when adjacent rows or columns are involved, using the ERDL or EBDA rules respectively. This problem does not have any recovered digraph pairs with a common letter, so another starting point must be found.

(2) The next best starting point is to find two digraph pairs with at least two letters in common between the two pairs. The ro=AV and at=GO pairs share the As and Os in common. Other pairs are also possible.

(3) The reconstruction begins by taking one of the selected pairs and plotting each possibility for it. All three rules must be considered. The three separate plots that follow show the result of plotting ro=AV for the rectangular rule, ERDL, and EBDA in turn.

Rectangular rule:

```
R  A
v  0
```

ERDL:

```
RA  o v
```

EBDA:

```
R
A
0
V
```

(4) The positioning of the letters is arbitrary. In the rectangular plot, we do not know that R is to the left of A or above V. We do not know how many rows and columns occur between the characters. We only know that the four letters form

a rectangle if that is the correct rule. In the ERDL plot, we do not know that RA is to the left of OV or if there is a column in between the pairs or not. Similarly, in the EBDA plot, we do not know that RA comes above OV or if there is a row in between, The spaces and placements are unknown until the reconstruction has proceeded further.

(5) The next step is to add our second pair to the first plots. Again, we have to consider all three rules as we add the second pair. With three possible rules for each pair, there could be as many as nine different possible plots after two pairs if we did not select some letters in common to limit the possibilities.

(6) Consider first, the addition of at=GO to the rectangular plot of the first pair.

R	A	G
V	0	T

(7) ERDL cannot be used with the second pair, since we have already placed A and 0 in separate rows. To use ERDL, they must be in the same row.

(8) When EBDA is applied to the at=GO pair and linked to the ro=AV rectangular plot, the plot looks like this.

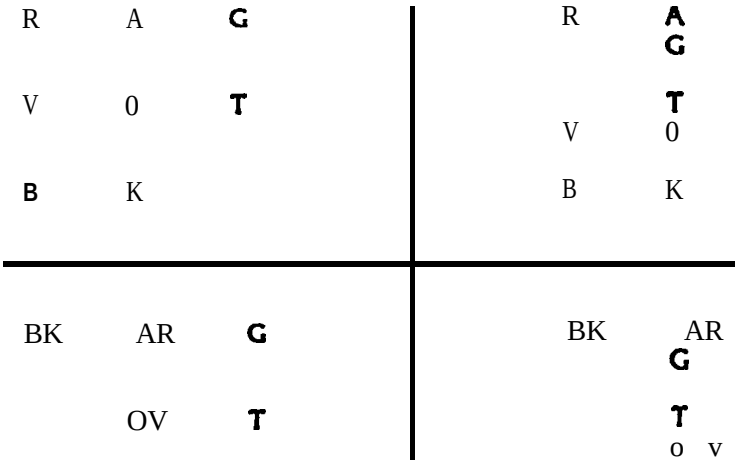
R	A
	G
	T
V	0

(9) When we try to link at=GO to the ERDL plot for ro=AV, it cannot be done. With A and 0 in the same row, the rectangular plot and the EBDA plot cannot be applied properly. If we try to plot ERDL for at=GO, it results in six different letters on the same row, which is not possible in a normal Playfair. Therefore, we can cross out or erase the ERDL plot for ro=AV.

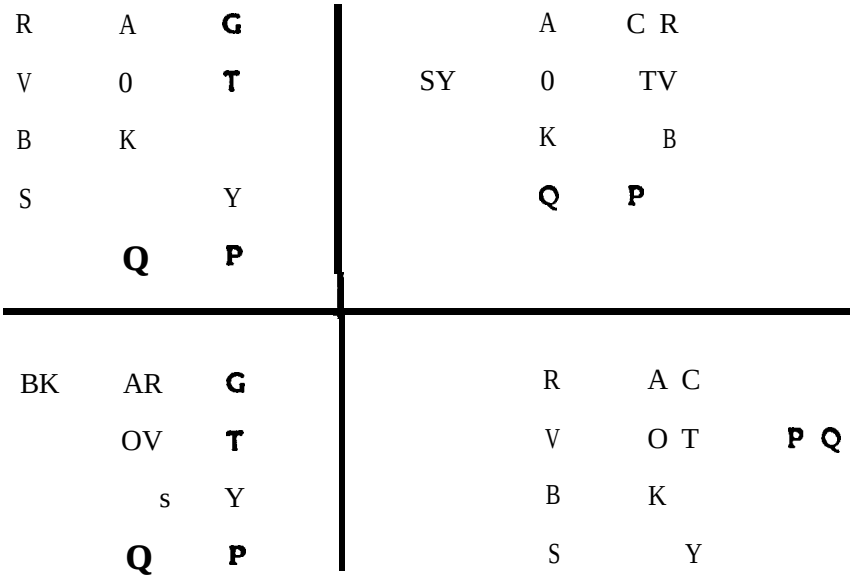
(10) We next plot all possible rules for at=GO with the EBDA plot for ro=AV. The rectangular rule is the only possibility. ERDL for at=GO is impossible, because we have already placed A and 0 in the same column. EBDA is impossible, because it would place six different letters in the same column.

R	A	G		R			R	
					2		A	G
V	0	T			T		0	T
				V	0		V	

(11) The next step is to again pick a digraph pair with at least two letters in common with the letters already plotted. The most obvious possibility is the ba=KR on line three. Following the same approach as we did with the second pair, we find four possibilities this time.



(12) Both st=YV and op=TQ have two letters in common with the recovered diagrams. Checking all possibilities for each of these produces the next four diagrams.



(13) Various approaches can be used to further build the possible diagrams. One approach is to try to recover more text. The repeated KT GH is certain to be a spelled out number. If we try to decipher KT using all of our trial diagrams, all

but the third one produce plaintext -0. The third diagram produces G-. From these results, we can rule out the third diagram, since no number has a G in the first position. The number *FOUR* is the only likely plaintext with 0 in the second position, We add fo=KT to the three remaining digrams and then try to fit ur=GH. In each case, only the ERDL rule will apply. The last of the three remaining diagrams is also eliminated, since ur=GH cannot be plotted. We are left with these possibilities.

RH	A	UC				A U G R H	
V	0	T		S Y		0	T V
B	K	F				K	F B
S		Y				*Q*	P
	Q	P					

(14) The second diagram above is impossible, since there is no way to fit the SY so that it aligns with the row above it. We are finally down to a single diagram, and with careful selection of digraph pairs to plot, we can keep it to a single diagram. Next we will plot on=CQ, tx=CY, and ze=XK.

RH	A	UC	
v	0	T	c
B	K	F	E
s	z	Y	x
	Q	P	N

(15) The X, Y, and Z on the fourth line clearly belong in sequence.

R	H	U	C	A
V		C	T	O
B		E	F	K
S		X	Y	Z
		N	P	Q

(16) The partially reconstructed matrix can now be used to add substantially more plaintext in the message.

b at ta li on x et ef a re af ro
DT BV VF GO OG MV CQ IH NS MN VI FC IK FK NX KH UB GK AV LH

ou te xt il f ef ou rt hr es to on b at
CA CF WC YC IA VM PB CI FK CA GV UH NC BX OV LY NU CQ ED GO

ta li on to re a ac es to r ba tx ta li
OG MV CQ VW OV UB QH CM CM QM UO BX OV YG DH HB KR CY OG MV

on x x a at ta ck at ze ro ve ze ro ze ro st
CQ IH NS NS QR EX IU GO OG OE GO XK AV DT CB XK AV XK AV YV

op ar t x e ry ep ar at o il eg at ze ro fo
TQ RH OC NS NB GS LG FN RH GO CV MX VM SL FU CM GO XK AV KT

ur fo ur v e st op
GH KT GH DT CB YV TQ

(17) DT CB is clearly FIVE. The word on line five, after op=TQ is AR TI LX LE RY. The second row includes the numbers -F IV EF 0U RT HR EX E-. These additions are placed in the matrix.

R H U C A

B D E F K
L N P Q

S X Y Z
v I C T O

(18) The missing M and W are easily placed alphabetically. The rows are placed in correct order by shifting the last row to the top and placing the remaining rows alphabetically. The keyword is VICTOR HUGO.

(19) To solve Playfair systems like this, it is important to remember to try all possibilities and to keep the work as simple as possible. It is very easy to overlook possible arrangements, so work very carefully. Always look for the digraph pairs with the least possibilities to plot to keep the work from getting very complex. If the square appears to be alphabetical in construction, use the alphabeticity to help you put rows and columns in the correct order whenever you can.

CHAPTER 8

PERIODIC POLYALPHABETIC SUBSTITUTION SYSTEMS

Section I
Characteristics of Periodic Systems

8-1. Types of Polyalphabetic Systems

All the substitution systems explained up to this point are monoalphabetic systems. Whether they deal with one letter at a time or several, whether they have one cipher equivalent for each plaintext letter or more than one, they are still systems with only one alphabet. The constant feature that makes a system monoalphabetic is that a given ciphertext value always translates into the same plaintext value. In polyalphabetic systems, a given ciphertext value changes its plaintext meaning.

a. Most polyalphabetic systems are monographic; they encipher a single letter at a time. Polygraphic polyalphabetics are possible, but have little practical military value.

b. A typical polyalphabetic system will use from 2 to 26 different alphabets. Polyalphabetic systems which repeat the same set of alphabets over and over again in the same sequence are known as periodic systems. Polyalphabetic systems which do not keep repeating the same alphabets in the same order are known as aperiodic systems. Periodic systems, because of their regular repeating keys, are generally less secure than aperiodic systems. Aperiodic systems, on the other hand, are generally more difficult to use, unless the encipherment is done automatically by a cipher machine or computer.

c. The classic types of polyalphabetic systems use a set of alphabets, such as the 26 alphabets pictured in Figure 8-l. Figure 8-1, known as a Vigenere square, includes all possible alignments of a direct standard alphabet. Mixed alphabets can also be used in such a square. If all 26 alphabets are used, any letter can equal any other letter. There are necessarily three elements to the encryption process with polyalphabetic ciphers, which the square and the accompanying examples illustrate. The plaintext letters are listed across the top of the square. The cipher equivalents are found in the 26 sequences below. The final element is the key that designates which alphabet is used at any given time. The key letter is found on the

left side of the square. The first example in Figure 8-l shows the use of a repeating key based on a keyword. Since the same key is repeated over and over again, the resulting system is periodic. The second example uses a nonrepeating key based on a quotation. Since this key does not repeat, it is an aperiodic system. Note that the reuse of the same alphabets does not constitute a repeating key. For the system to be classified as periodic, the same alphabets must be reused over and over again in the same sequence.

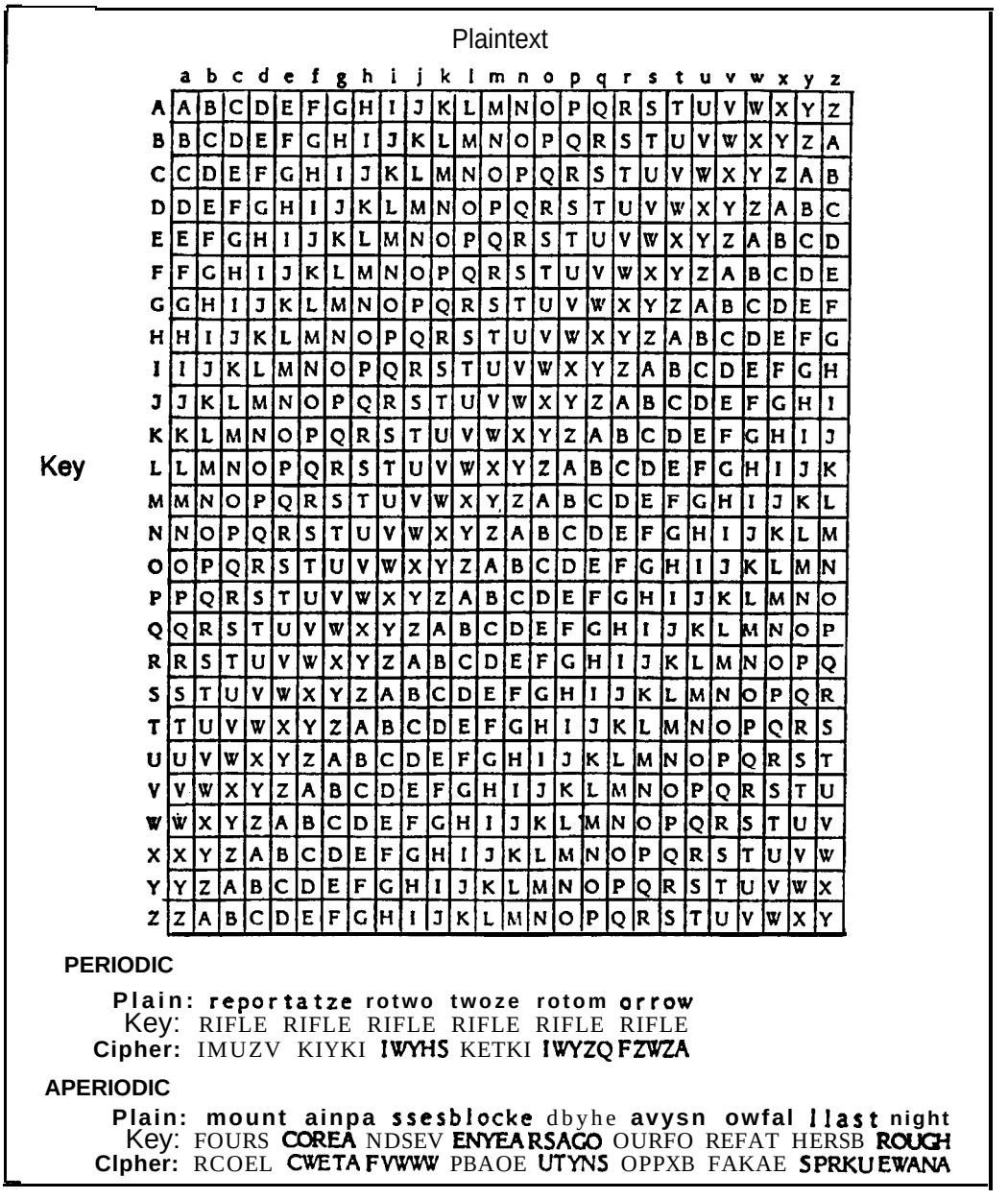


Figure 8-1. Use of Vigenere square.

d. Another way to picture the same system as the first example in Figure 8-1 is shown below. In this case, instead of using the complete alphabet square, only the alphabets actually used are shown. These alphabets are used repeatedly to produce the same results. In this example, the key is expressed in terms of the number of the cipher sequence used, instead of by the repeating key letters.

p:	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	t
C1:	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
c2:	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
c3:	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
C4:	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	C	H	I	J	K
c5:	E	F	C	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D

Plaintext: repor tatze rotwo twoze rotom orrow
Key: 12345 12345 12345 12345 12345 12345
Ciphertext: IMUZV K IYKI IWYHS KETKI IWYZQ FZWZA

e. Another type of polyalphabetic system does not use multiple alphabets in the classic sense, but instead enciphersa message in a single alphabet. Then it applies either a repeating key or nonrepeating key to the first encipherment to create a polyalphabetic. One method of applying a polyalphabetic key to a monoalphabetic encipherment is to use a numeric system and arithmetically add a key to it. For example, here is a dinomic system, which has been further enciphered by a repeating numeric additive. The first encipherment is labeled I, for Intermediate cipher, and the second encipherment is labeled C. The 8-digit repeating key is labeled K. Modulo 10 arithmetic is used (paragraph 5-3f(1)).

	0	1	2	3	4	5	6	7	8	9
3	m	u	r	p	h	y	s	l	a	w
6	b	c	d	e	f	g	i	j	k	n
9	o	q	t	v	x	z	.	,	?	/

p:	at	ta	ck	at	ze	ro	ni	ne	hu	nd	re	d.
I:	3892	9238	6168	3892	9563	3290	6966	6963	3431	6962	3263	6296
K:	4209	9336	4209	9336	4209	9336	4209	9336	4209	9336	4209	9336
c:	7091	8564	0367	2128	3762	2526	0165	5299	7630	5298	7462	5522

f. Another approach to applying a polyalphabetic key begins with the built-in encoding system used by teleprinters or computers. Paragraph 8-2 shows examples of these.

8-2. Machine Based Polyalphabetics

When text is sent electronically by radio or wire, some form of coding must be used. The earliest system of coding for electronic transmission was Morse code, which is still used widely today. When teleprinters took their place in communications, a new

binary type of coding system was devised, which can be handled by machine more easily than Morse code can. Anybinary coding system uses only two characters, which can be represented electronicllly as a signal pulse or no signal pulse, high voltage or low voltage, or one frequency or another frequency. Which of these approaches is used depends on the equipment in use and is not our concern here. We are concerned with how the two binary characters, whatever their electronic origin, are combined to represent alphabetic, numeric, and special characters, and how they may further be encrypted. Various notations have been used to represent the two binary characters-Xs and Os, 1s and Os, ts and -s, or Ms (for marks) and Ss (for spaces). We will use 1s and 0s in this text, but you should be aware that you may see other notations elsewhere, particularly in older literature.

a. The Baudot Code. Teleprinter systems generally use a 5-digit binary code known originally as the Baudot code. There are 32 possible combinations of 5 digits, which are not enough for the letters, numbers, and printer control characters needed for communications. The number of possible characters is approximately doubled by the use of upper and lower shift characters, similar to the shift key on a typewriter, giving all characters two alternate meanings except the shift characters themselves and the space character. There are still not enough characters for upper and lower case letters, so all traffic passed by such teleprinter systems use capital letters only. The standard international teleprmter code is shown in Figure 8-2. Each dot represents a 1 and each space represents a 0. Other codes are also used besides the one shown.

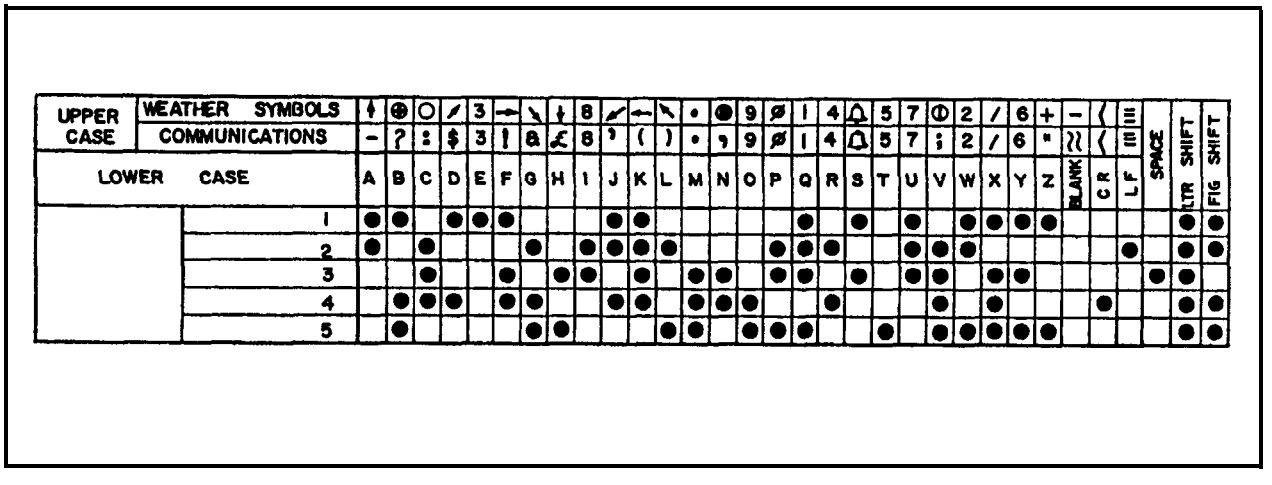


Figure 8-2. international teleprinter code.

The binary digits themselves are known as bauds-a term derived from the Baudot code. The terminology has carried over into modern computer. systems as well. Polyalphabetic keys, also in 5-digit binary form, are easily applied to coded text

electronically by baud addition. An example of this process is shown below. Although other rules are also possible, the addition of key and plaintext bauds is usual1y accomplished by the rule, Like *values* sum to 0; *unlikes sum to 1. (In* computer logic, this would be called an exclusive OR, or XOR operation.)

Plaintext:	**e**	**n**	**e**	**m**	**y**
Bauded piain:	10000	00110	10000	00111	10101
Key:	01010	11010	10100	01110	10110
Baudedcipher:	11010	11100	00100	01001	00011
Ciphertext:	**J**	**U**	**(space)**	**L**	**0**

One advantage of this rule of addition is that adding the same key to the ciphertext produces the plaintext again.

b. Computer Codes. Communications between computers use more than 5 digits. Typical computer codes use either 7- or 8-binary digits (bits), giving a range of 128 characters or 256 characters. These permit upper and lower case letters, a full range of punctuation marks and special characters, and a number of codes to control printers and communications devices as well. With the 8-bit, 256 character set, graphics may also be enabled to permit transmitting pictures as well as text. The most common standard for the first 128 characters, whether 7-bit or 8-bit, is the American standard code for information interchange (ASCII) standard, which you can find in many cornputer manuals. Encipherment and decipherment can be accomplished in 7- and 8-bit operation just as was shown for 5-digit teleprinter operations. The more complex systems are far beyond the scope of this manual, but simple repeating key systems can be solved using the techniques discussed here. One problem that computer codes present is that less than half of the possible 7-bit characters are letters and numbers, and many of them stand for printer control codes that do not print out as characters normally. Working with binary numbers themselves is unwieldy, but any 7- or 8-bit value can be represented by two hexadecimal (base 16) arithmetic digits. Hexadecimal arithmetic is not explained here, but explanations are available in many computer manuals and texts, if needed. Hexadecimal and binary numbers are also explained in Army Correspondence Course Program Subcourse SA0709.

Section II
Identifying Periodic Systems

8-3. Analysis of Repeated Ciphertext

Polyalphabetic systems normally have very flat frequency counts. The phi IC is normally close to the random expectation of 1.00. Since other systems, including

variant multiliterals and aperiodic systems, also can produce flat frequency counts, this is not enough to identify a system as periodic. The key to identifying a system as periodic is to recognize through repeated ciphertext that a repeating key is used.

a. Repeated ciphertext can occur in two ways. Whenever the same plaintext is enciphered by the same keys, the ciphertext will also repeat. Such repeats are called causal repeats. The second way that ciphertext can repeat is by pure chance. Different plaintext enciphered with different keys will sometimes produce short ciphertext repeats. Causal repeats are much more likely to occur than accidental repeats, particularly if they are longer than two or three characters. The example below, repeated from Section I, shows how causal repeats occur.

Plaintext: **repor tatze rotwo twoze rotom orrow**
Key: **12345 12345 12345 12345 12345 12345**
Ciphertext: **IMUZV KIYKI IWYHS KETKI IWYZQ FZWZA**

The plaintext words *ZERO* and *TWO* both occur twice. The repeated *ZEROs* lined up with the same alphabets, producing a ciphertext repeat. The repeated *TWOs* lined up with different alphabets and did not produce a ciphertext repeat.

b. Whenever causal repeats occur, the distance between them must be a multiple of the period length. In the example above, the two *ZEROs* occurred 10 letters apart. Note that the distances are counted from the first letter of one repeat to, but not including, the first letter of the second repeat. If the distance was not a multiple of the period five, the ciphertext repeat would not have occurred.

c. The distance between causal repeats is a multiple of the period length. Given a cryptogram of unknown period that includes ciphertext repeats, the period can be determined, or at least narrowed down, by analyzing the distances between repeats. The period must be a factor of the distance. The factors of a number are all the numbers which divide evenly into that number. When there is more than one repeat, the period must be a common factor of all such distances. For example, if a cryptogram has repeats that are 28, 35, and 42 letters apart, the only number that evenly divides all the distances is 7. The period must be 7. Utility tables showing common factor numbers are in Appendix E.

d. Here is a more complex example. Suppose a cryptogram suspected of being periodic includes the following repeats.

Repeat	Distance
CXKLRYPDL	**84**
ZBHHNST	**90**
XTVTB	36
SRM	35

The next step after determining the distances is to list the factors for each repeat, as shown below.

Repeat	Distance	Factors
GXKLRYPDL	90	2, 3, 4, 6, 7, 8, 12
ZBHHNST		2, 3, 5, 6, 9, 10
XYVTN	36	2, 3, 4, 6, 9, 12
SRM	35	5, 7

No numbers evenly divide the distances between all the repeats. In such cases, either the system was not a periodic system, or one or more of the repeats is accidental. In this problem, the SRM repeat is probably accidental, because it is the shortest. Discarding the SRM repeat from consideration, the remaining repeats all have common factors of 2, 3, and 6. Where more than one factor is possible, it is generally safest to assume the largest. If the period is actually 3, for example, it will reveal itself by repeated alphabets as the cryptogram is solved.

8-4. Analysis by Frequency Counts

Periodic systems can be identified even when there are no repeated words in the text. Causal single-letter ciphertext repeats will still occur and significantly outnumber the accidental single-letter repeats.

a. To find the causal single-letter repeats, take frequency counts for each alphabet according to its position in the suspected repeating cycle. If the period is incorrect, the separate frequency counts will remain flat. If the period is correct, the separate frequency counts will be as rough as plaintext on the average. Recognizing when a count is rough or flat is difficult by eye, particularly with anything but very long cryptograms, but the phi test performed on each separate alphabet gives a reliable indication. Taking separate frequency counts by position for each suspected period and then calculating phi tests on each is a laborious and time-consuming process by hand. It can be done when necessary, but it is best performed by computer support. Figures 8-3, 8-4, and 8-5 show computer generated output for suspected periods of 6, 7, and 8 for the following cryptogram.

LPADW GUGHG ETZHV KSRQS ACNPJ GHTHH QCKGS CHHRB HMDIH HMCJM
EXEVH LVPQS OCHPK MZYBZ SMMPF TLBGF KRAEA FBMHQ IXSZC PGAQT
KPLPS GXIVX BCFRI TSTGF SPYNS SNTAL SIOSC MJRMI ZSICF RQTUV
HLVPQ SOCHP KQFDW SFRAK MILRG GECAU HFEGN YXXZO GLGMZ DUHUC
XGRIL SARZQ FDWBB PSRUD UGJGD JNTWF BTABQ SVBGF WRDPP BFRGN

A	B	CDE	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	TUVWXYZ
10	11	11	8	6	13	20	17	9	5	7	9	11	6	4	14	10	13	19	10	7	7	5	7	3	8

TOTAL LETTERS = 250 MONOGRAPHIC IC = 1.098474

b. The average ICs for each period in Figure 8-3 and 8-4 are flat, The average IC for a period of 8 in Figure 8-5 is much higher than the other two. This clearly shows that the period of 8 is more likely correct than periods of 6 and 7.

c. The computer program used to generate these examples is listed in Appendix F. It is written in GW BASIC, and is readily adaptable to many different computers.

PERIOD = 6:

```
A B C D E F C H I J K L M N O P Q R S T U V W X Y Z
2 0 0 1 0 4 4 4 2 3 0 3 2 1 3 0 3 0 2 0 2 1 2 1 0 2

TOTAL LETTERS = 42                IC = 1.117306

A B C D E F C H I J K L M N O P Q R S T U V W X Y Z
1 0 4 2 0 4 4 2 2 0 1 2 2 0 0 4 0 4 4 2 0 0 0 3 1 0

TOTAL LETTERS = 42                IC = 1.358885

A B C D E F C H I J K L M N O P Q R S T U V W X Y Z
4 4 1 2 3 1 1 5 0 0 2 0 1 0 0 0 2 1 3 0 1 4 2 1 1 3

TOTAL LETTERS = 42                IC = 1.238095

A B C D E F C H I J K L M N O P Q R S T U V W X ' Y Z
0 3 2 3 0 0 6 1 1 1 1 1 2 2 1 5 0 2 2 6 0 0 0 1 0 2

TOTAL LETTERS = 42                IC = 1.570267

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

TOTAL LETTERS = 41                IC = 1.014634

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

TOTAL LETTERS = 41                IC = 1.236585
```

Figure 8-3. Frequencies, period 6.

PERIOD = 7:

A B C D E F C H I J K L M N O P Q R S T U V W X Y Z
2 2 2 2 1 3 4 1 1 0 0 1 1 1 1 1 3 1 2 2 1 2 0 1 0 1

TOTAL LETTERS = 36 **IC = .784127**

A B C D E F C H I J K L M N O P Q R S T U V W X Y Z
1 0 1 1 2 4 3 4 1 1 3 1 1 1 0 2 0 1 4 2 2 0 0 1 0 0

TOTAL LETTERS = 36 **IC = 1.155556**

A B C D E F C H I J K L M N O P Q R S T U V W X Y Z
2 2 1 2 0 1 1 2 3 0 1 1 4 2 1 2 2 1 2 0 1 1 2 1 0 ~ 1

TOTAL LETTERS = 36 **IC = .7428572**

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

TOTAL LETTERS = 36 **IC = .8666667**

A B C D E F C H I J K L M N O P Q R S T U V W X Y Z
0 2 0 1 0 2 2 2 0 0 1 3 2 2 1 1 1 3 4 1 1 1 2 3 1 0

TOTAL LETTERS = 36 **IC = .9079365**

A B C D E F C H I J K L M N O P Q R S T U V W X Y Z
1 2 2 0 0 2 4 4 1 1 0 0 1 0 0 4 1 2 2 2 0 2 0 1 0 3

TOTAL LETTERS = 35 **IC = 1.22353**

A B C D E F C H I J K L M N O P Q R S T U V W X Y Z
2 2 0 1 1 1 4 2 2 1 1 2 1 0 1 2 2 2 4 1 1 0 0 0 0 2

TOTAL LETTERS = 35 **IC = .9176471**

Figure 8-4. Frequencies, period 7.

PERIOD = 8:

A B C D E F C H I J K L M N O P Q R S T U V W X Y Z
3 0 0 1 1 2 3 3 0 3 4 1 1 1 0 0 0 0 4 0 0 2 1 1 0 1

TOTAL LETTERS = 32 **IC = 1.382903**

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 1 0 0 0 5 0 0 1 1 1 4 1 0 5 3 7 0 . 0 0 0 0 0 1 0

TOTAL LETTERS = 32 **IC = 2.820968**

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

TOTAL LETTERS = 31 **IC = 1.585592**

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
0 0 1 5 1 0 0 0 6 0 0 0 0 1 0 1 0 0 4 7 1 0 0 3 1 0

TOTAL LETTERS = 31 **IC = 3.075299**

A B C D E F C H I J K L M N O P Q R S T U V W X Y Z
2 1 0 0 1 0 0 4 0 0 1 3 0 0 3 1 0 0 3 0 3 2 4 0 0 3

TOTAL LETTERS = 31 **IC = 1.621505**

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

TOTAL LETTERS = 31 **IC = 1.958989**

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

TOTAL LETTERS = 31 **IC = 1.453764**

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
0 1 4 0 0 3 5 0 0 0 1 3 3 0 0 6 0 3 0 2 0 0 0 0 0 0

TOTAL LETTERS = 31 **IC = 2.460215**

Figure 8-5. Frequencies, period 8.

CHAPTER 9

SOLUTION OF PERIODIC POLYALPHABETIC SYSTEMS

Section I
Systems Using Standard Cipher Alphabets

9-1. Approaches to Solution

When standard alphabets are used with monoalphabetic systems, three approaches are possible. The simplest occurs when text can be immediately identified. Identification of only two or three letters in a standard unilateral alphabet is sufficient to reconstruct and confirm the entire alphabet. The other two methods, where text is not readily identifiable, are to match frequency patterns to the normal A through Z pattern and to generate all possible solutions. All three of these methods also apply to standard alphabet periodic polyalphabetics.

9-2. Solution by Probable Word Method

When the alphabets in a periodic system are known or suspected to be standard, the identification of one plaintext word is usually enough to recover the whole system. The period must be identified first, as explained in the previous chapter, either by analysis of repeat intervals or by the phi test. Then when a word is recognized from repeats or stereotypes, the alphabets can be written and tried throughout the cryptogram. If they produce good plaintext throughout, the problem is solved.

EIYMB EKVWO YBTOE ILMFK CRRAK WJWBZ ELUYO NZUZF ZNTIH YMZXT
IMSWG WRRPC HFGNV ZQALN QCNGJ VBFSQ RVFPO ENISI CIMHJ SJDBT
ALSDI CSOGH ZYAWW JCEQE MRCFY KIIXC SERRE RGZPB RMJDC IMRHZ
SFZXT TWQHW YHVAG UYDUS QPGJD BTSGZ JYAGK KARXQ MJE

Repeats	Distance	Factors
ZXT	105	3, 5, 7
CIM	54	3, 6, 9
JBDT	77	7, 11

Factor analysis does not show us a clearcut period length, but if we select the four letter repeat as the most likely causal repeat, 7 appears to be the correct period. If we also try *STOP* as the four letter repeat, it gives us the following text and alphabets.

re nais cer e t sen smov he av idge ing
EIYMB EKVWO YBTOE ILMFK CRRAK WJWBZ ELUYO NZUZF ZNTIH YMZXT

e p men owar ud dy erso ofb a r svi stop
IMSWG WRRPC HFGNV ZQALN QCNGJ VBFSQ RVFPO ENISI CIMHJ SJDBT

my po ions ngr i h ave nhea yr ei rced
ALSDI CSOGH ZYAWW JCEQE MRCFY KIIXC SERRE RGZPB RMJDC IMRHZ

ing p f our htho st op sonc and i
SFZXT TWQHW YHVAG UYDUS QPGJD BTSGZ JYAGK KARXQ MJE

p:	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
C1:	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
C2:	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
C3:																										
C4:																										
C5:																										
C6:	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
C7:	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

From the partial plaintext that this produces, *STOP* is clearly correct. Such words as *RECONNAISSANCE, HEAVY,* and *REINFORCED* are apparent, any one of which will complete the solution. For another type of probable word approach, applicable to periodics or aperiodic, see paragraph 10-3c on crib dragging.

9-3. Solution by Frequency Matching

With monoalphabetic systems using standard alphabets, the solution was very easy whenever a message was long enough to give a recognizable pattern. The characteristic pattern of highs and lows of a standard sequence cannot be easily concealed. The same technique applies to polyalphabetic systems, although messages necessarily must be longer to produce a recognizable pattern for each separate alphabet.

FNPDM GJRMF FTFFZ IQKTC LGHAS EOSIM PVLZF LJEWU WTEAH EOZUA
NBHNJ SXFFT JNRGR KOEXP GZSEY XHNFS EZAGU EORHZ XOMRH ZBLTF
BYQDT DAKEI LKSIP UYKSX BTERQ QTWPI SAOSF TQKTS QLZVE EYVAE
JSNFB IFNEI OZJNR RFSPR TEHNJ ROJSI UOCZB GQPLI STUAE KSSQT
EFXUJ NFGKO UHLZF HPRYV TUSCP JDJSE BLSYU IXDSJ JAEVF KJNQF

FIFMP EHYQD

a. Factor analysisshows common factors of three and six for all repeat intervals. Based, on this, a frequency count for six alphabets is produced, as listed in Figure 9-1. If the period were actually three, the first and fourth, the second and fifth, and the third and sixth frequency counts would be similar. This is clearly not the case, so the period is confirmed as six.

```
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
0 2 0 0 3 5 0 0 0 10 0 0 0 0 2 4 4 0 4 3 6 0 0 1 0 0
TOTAL LETTERS = 44          IC = 2.638478

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 0 2 2 7 1 0 1 2 0 0 1 1 6 2 1 0 4 5 2 1 1 1 0 0 0
TOTAL LETTERS = 44          IC = 1.731501

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
0 1 1 0 6 2 0 1 0 0 0 5 2 2 3 4 2 2 0 3 0 0 1 3 4 1
TOTAL LETTERS = 43          IC = 1.468439

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
0 1 0 3 3 6 3 0 4 2 3 2 0 0 0 1 0 4 1 1 1 3 0 1 0 4
TOTAL LETTERS = 43          IC = 1.439646

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
3 3 0 1 0 7 1 7 1 0 1 0 1 0 2 0 3 1 8 1 0 0 1 1 0 1
TOTAL LETTERS = 43          IC = 2.303433

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 0 0 0 1 0 3 1 4 2 5 1 1 3 1 0 1 0 2 4 2 1 0 1 3 5
TOTAL LETTERS = 43          IC = 1.295681
```

Figure 9-1. Periodic frequencies.

b. The easiest patterns to match are generally those with the highest ICs. The first, second, and fifth alphabets have the highest ICs, and all can be matched fairly easily. In the first, plaintext A equals ciphertext B. In the second, plaintext A equals ciphertext A, and in the fifth, plaintext A equals ciphertext O. Other alphabets can be matched, too, but using these as an example, the partially reconstructed text is shown below.

```
en  y  ir   r ef   c sc   t r   e  u  ov  r  le   i da   a ta
FNPDM GJRMF FTFFZ IQKTC LGHAS  EOSIM PVLZF LJEWU WTEAH EOZUA

 t i s  r  in  d  ne   s re    t es   m t e  t  wo  t  al
NBHNJ SXFFT JNRGR KOEXP GZSEY XHNFS EZAGU EORHZ XOMRH ZBLTF

n  pd  m di   e o u  e  at  c   sw  e ns   c ss   l d e  m
BYQDT DAKEI LKSIP UYKSX BTERQ QTWPI SAOSF TQKTS QLZVE EYVAE

is  n  en   a in   r or   t i   r  e  to  n  pp   e ta   e pt
JSNFB IFNEI OZJNR RFSPR TEHNJ ROJSI UOCZB GQPLI STUAE KSSQT

 j i n  w  th  r  or   f rc    p re   e t i  e  ia  r  in
EFXUJ NFGKO UHLZF HPRYV TUSCP JDJSE BLSYU IXDSJ JAEVF KJNQF

r em   t pd
FIFMP EHYQD
```

c. The letter combinations produced by the three recovered alphabets are consistent with good plaintext. Expanded plaintext can be recognized in many places. The first word is *ENEMY* for example. Filling in added plaintext is a surer and quicker means of completing the solution at this point than trying to match more alphabets. Here is the complete solution.

```
enemy airbo rnefo rcesc aptur  edbug ovair field indaw natta
FNPDM GJRMF FTFFZ IQKTC LGHAS  EOSIM PVLZF LJEWU WTEAH EOZUA

ckthi smorn ingpd enemy stren  gthes timat edatt wobat talio
NBHNJ SXFFT JNRGR KOEXP GZSEY XHNFS EZAGU EORHZ XOMRH ZBLTF

nspdi mmedi ateco unter attac  kswer eunsu ccess fulpd enemy
BYQDT DAKEI LKSIP UYKSX BTERQ QTWPI SAOSF TQKTS QLZVE EYVAE

iscon centr ating armor inthi  rdsec torin appar entat tempt
JSNFB IFNEI OZJNR RFSPR TEHNJ ROJSI UOCZB GQPLI STUAE KSSQT

tojoi nupwi thair borne force  spdre quest immed iater einfo
EFXUJ NFGKO UHLZF HPRYV TUSCP JDJSE BLSYU IXDSJ JAEVF KJNQF

rceme ntspd
FIFMP EHYQD
```

p:	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
C1:	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
C2:	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
C3:	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
C4:	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
C5:	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
C6:	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

9-4. Solution by the Generatrix Method

With standard alphabets or any known alphabets, the method of completing the plain component can be used. This method, when applied to periodic systems, is commonly called the generatrix method. The advantage of this method over frequency matching is that it will work even with fairly short cryptograms. Just as with a monoalphabetic system (see paragraph 4-11), the first step is a trial decryption at any alphabet alignment, followed by listing the plain component sequence vertically underneath each letter of the trial decryption. Whenever the plain and cipher sequences are identical and in the same direction, no trial decryption is necessary. The key difference with periodic systems is that the process must be applied to the letters of each alphabet separately. Plaintext will not be immediately obvious when you look at the generated lines of letters from only a single alphabet, so selection must be initially based on letter frequencies and probabilities rather than recognizable text. The process is illustrated with the following cryptogram enciphered with direct standard alphabets.

QNMZC TAAED FASRR TITYI UGPGW QVMAX TRMRM ZHMNZ KFQEI RIOUX XAAGR UGPG

a. The cryptogram has a period of five, which can be confirmed either through periodic-phi tests or factor analysis of all the repeats, including two letter repeats, which are not underlined.

b. The most obvious step to try is to substitute *STOP* for the four letter repeat. It does not produce plaintext elsewhere, however. More powerful methods of solution are required.

c. The cryptogram can be readily solved by the generatrix method. The first step is to separate the letters produced by each alphabet. The letters from each of the five alphabets are listed separately below. Notice that if you read all the first letters, it produces the first group of the cryptogram. The second letters produce the second group and so on.

QTFTUQTZKRXU NAAIGVRHFIAG MASTPMMMQOAP ZERYGARNEUGG CDRIWXMZIXR

d. No trial decryption is required, because the same sequence is expected for both the plain and cipher components. Therefore, the next step is to complete the plain component sequence for each letter grouping. This is illustrated in Figure 9-2.

QTFTUQTZKRXU		NAAIGVRHFIAG		MASTPMMMQOAP		ZERYGARNEUGG		CIRIWXMZIXR	
296962902836	62	888855876885	84	688966662886	79	098658889655	77	78885360838	64
RUGUVRUALSYV		OBBJHWSIGJBH		NBTUQNNNRPBQ		AFSZHBSOFVHH		DJSJXYNAJYS	
865658687865	78	844175885147	62	849628888642	73	868074886577	74	71813688168	57
SVHVWSVBMTZW		PCCKIXTJHKCI		OCUVROOOSQCR		BGTAICTPGWII		EKTKYZOBKZT	
857558546905	67	677283917278	67	876588888278	83	459887965588	82	92926084209	51
TWIWXTWCNUAX		QDDLJYUKILDJ		PDVWSPPPTRDS		CHUBJDUQHXJJ		FLULZAPCLAU	
958539578683	76	277716628771	61	675586669878	81	776417627311	52	67670867786	68
UXJXYUXDOVBY		REEMKZVLJMEK		QEWXTQQQUSET		DIVCKEVRIYKK		GMVMABQDMBV	
631366378546	58	899620571692	64	295392226899	66	785729588622	69	56568427645	58
VYKYZVYEPWCZ		SFFNLAWMKNFL		RFXYURRRVTFU		EJWDLFWSJZLL		HNWNBCRENCW	
562605696570	57	866878562867	77	863668885966	79	915776581077	63	78584789875	76
WZLZAWZFQXDA		TGGOMBXNLOGM		SGYZVSSSWUGV		FKXEMGXTKAMM		IOXOCDSFODX	
507085062378	51	955864387856	74	856058885655	69	623965392866	65	88387786873	73
XAMABXAGRYEB		UHHPNCYOMPHN		THZAWTTTXVHW		GLYFNHYULBNN		JPYPDETGPEY	
386843858694	72	677687686678	82	970859993575	76	576687667488	78	16667995696	70
YBNBCYBHSZFC		VIIQODZPNQIO		UIABXUUUYWIX		HMZGOIZVMCOO		KQZQEFUHQFZ	
648476478067	67	588287068288	70	688436666583	69	760588056788	68	22029667260	42
ZCOCDZCITAGD		WJJRPEAQORJP		VJBCYVVVZXJY		INAHPJAWNDPP		LRARFGVIRGA	
078770789857	73	511869828816	63	514765550316	48	888761858766	78	78886558858	76
ADPDEADJUBHE		XKKSQFBRPSKQ		WKCDZWWAYKZ		JOBIQKBXOEQQ		MSBSGHWJSHB	
876798716479	79	322826486822	53	527705558620	52	184822438922	53	68485751874	63
BEQEFBEKVCIF		YLLTRGCSQTLR		XLDEAXXXBZLA		KPCJRLCYPFRR		NTCTHIXKTIC	
492964925786	71	677985782978	83	377983334078	62	267187766688	72	89797832987	77
CFRFGCFLWDJG		ZMMJSHDTRUMS		YMEFBYYYCAMB		LQDKSMDZQGSS		OUDUIJYLUJD	
768657675715	70	066687798668	77	669646667864	74	727286702588	62	86768167617	63
DGSGHDGMXEKH		ANNVTIEUSVNT		ZNFGCZZZDBNC		MRELTNEARHTT		PVEVJKZMVKE	
758577563927	71	888598968589	91	086570007487	52	689798988799	97	65951206529	50
EHTHIEHNYFLI		BOOWUJFVTWOU		AOGHDAAAECOD		NSFMUOFBSIUU		QWFWKLANWLF	
979789786678	91	488561659586	71	885778889787	90	886668648866	80	25652788576	61
FIUIJFIOZGMJ		CPPXVKGWUXPV		BPHIEBBBFDPE		OTGNVPGCTJVV		RXGXLMBOXMG	
686816880561	63	766352556365	59	467894446769	74	895856579155	73	83537648365	58
GJVJKGJPAHNK		DQQYWLHXVYQW		CQIJFCCCGEQF		PUHOWQHDUKWW		SYHYMNCPYNH	
515125168782	51	722657735625	57	728167775926	67	667852776255	66	86766876687	75
HKWKLHKQBIOL		ERRZXMIYWZRX		DRJKGDDDHFRG		QVIPXRIEVLXX		TZIZNODQZOI	
725277224887	61	988036865083	64	781257777685	70	258638895733	67	90808872088	58
ILXLMILRCJPM		FSSAYNJZXASY		ESKLHEEEIGSH		RWJQYSJFWMYY		UAJAOPERAPJ	
873768787166	74	688868102886	70	982779998587	88	851268165666	60	68188698861	69
JMYMNJMSDKQN		GTTBZOKAYBTZ		FTLMIFFFJHTI		SXKRZTKGXNZZ		VBKBPQFSBQK	
166681687228	61	599408286490	64	697686661798	79	832809253800	48	54246268422	45
KNZNOKNTELRO		HUUCAPLBZCUA		GUMNJGGGKIUJ		TYLSAULHYOAA		WCLCQRGTCRL	
280882899788	77	766786740768	72	566815552861	58	967886776888	88	57772859787	72
LQAOPLOUFMSP		IVVDBQMCADVB		HVNOKHHHLJVK		UZMTBVMIZPBB		XDMDRSHUDSM	
788867866686	84	855742678754	68	758827777152	66	606945680644	58	37678876786	73
MPBPQMPVGNTQ		JWWECRNDBEWC		IWOPLIIIMKWL		VANUCWNJAQCC		YENESTIVETN	
664626655892	65	155978874957	75	858678886257	78	588675818277	72	69898985998	88
NQCQRNQWHOUR		KXXFDSOECFXD		JXPQMJJJNLXM		WBOVDXOKBRDD		ZFOFTUJWFUO	
827288257868	71	233678897637	69	136261118736	45	548573824877	68	06869615668	61
ORDRSORXIPVS		LYYGETPFDGYE		KYQRNKKKOMYN		XCPWEYPLCSEE		AGPGUVKXGVP	
887888838658	85	766599667569	81	262882228668	60	376596677899	82	85656523556	56
PSESTPSYJQWT		MZZHFUQGEHZF		LZRSOLLLPNZO		YDQXFZQMDTFF		BHQHVWLYHWQ	
689896861259	77	600766259706	54	708887776808	74	672360267966	60	47275576752	57

Figure 9-2. Generatrix method.

e. To aid in selection of the most likely generated letter sequences, numeric probability data has been added to each line of the listing. The numbers listed below each letter are assigned on the basis of logarithmic weights of the letter probabilities. To the right of each group of logarithmic weights is the sum of the weights for that group. Using this kind of weighting lets us determine the relative probabilities of each line by adding the weights for each letter. The weights in Figure 9-2 have been added according to the log weights shown in Table 9-1.

Table 9-1. Logarithmic weights of letter probabilities.

Letter:	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
Log weight:	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

f. The listing in Figure 9-2 was computer generated. When this work must be done manually, it is easier to generate the sequences without the probability data. Then scan the generated rows for each alphabet to visually select those with the most high frequency letters. Finally, if necessary, the probability data can be added only for the selected rows.

g. Only rarely will the correct rows consist entirely of those with the highest totals. Normally, you will have to try different combinations of the high probability rows until you find the correct match. The best place to start is with those rows that stand out the most from others in the same alphabet groups. In the illustrated problem shown below, alphabets four and five provide the most likely starting point. In each case, the sum of the log weights for one row are well above any others. These are listed below, superimposed above each other with room for the other three alphabets to be added.

1:
2:
3:
4: **MRELTNEARHTT 97**
5: **YENESTIVETN 88**

h. As the rows are superimposed, the plaintext will appear vertically. The next step is to see which high probability rows from other alphabets will fit well with the starting pair. Trying both of the two highest probability rows for alphabet three produces the next two possibilities.

1:				
2:				
3:	AOGHDAAAECOD	90	ESKLHEEEIGSH	88
4:	MRELTNEARHTT	97	MRELTNEARHTT	97
5:	YENESTIVETN	88	YENESTIVETN	88

i. Reading the plaintext vertically, the grouping on the right is better than the one on the left. The DTS sequence in the left grouping is unlikely, and all the letter combinations on the right are acceptable. Furthermore, the EMY combination at the beginning of the right grouping suggests *ENEMY*. The letter sequences for the first two alphabets which begin with E and N respectively are both high probability sequences. The complete solution is shown below.

1:	EHTHIEHNYFLI	91
2:	NAAIGVRHFIAG	84
3:	ESKLHEEEIGSH	88
4:	MRELTNEARHTT	97
5:	YENESTIVETN	88

"ENEMY HAS RETAKEN HILL EIGHT SEVEN THREE IN HEAVY FIREFIGHT LAST NIGHT"

Section II

Systems Using Mixed Alphabets With Known Sequences

9-5. Approaches to Solution

When mixed sequences are used in periodic systems, a variety of different techniques can be used to solve them. When the plain and cipher sequences are known, the same techniques used with standard alphabets can be used, adapted to the known sequences. When one or both of the sequences are unknown, new techniques must be used. Each situation is a little different. The major paragraphs of this section deal with each situation: both sequences are known, the ciphertext sequence is known, or the plaintext sequence is known. Techniques for solving periodics when neither sequence is known are covered in the next section.

9-6. Solving Periodics With Known Mixed Sequences

Exactly the same techniques that were used with standard alphabets can be used with any known mixed sequences.

a. Successful assumption of plaintext allows you to directly reconstruct the cipher alphabets, as before.

b. The generatrix method works, making sure that a trial decryption is first performed with the sequences set at any alignment. All possible letter combinations are then generated by completing the plain component sequence, as before. The key points to remember are to perform the trial decryption and to use the plain component as the generatrix sequence, not a standard sequence.

c. Frequency matching also works, but there are some differences in its application. Frequency counts must be arranged in the cipher sequence order, not in standard order. The pattern that the frequency counts are matched to must be adjusted to the order of the known plain component. Rearrange the patterns of peaks and troughs to fit the plain component. For example, shown below is the pattern for a standard plain sequence and the pattern that results if a keyword mixed sequence based on POLYALPHABETIC is used as the plain component.

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

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

The new pattern resulting from the mixed plaintext sequence is just as easy to match frequency counts to as the more familiar standard pattern. If it should prove difficult to match by eye alone, there is also a statistical test, called the chi test, which can be used to aid the matching process. Paragraph 9-7 demonstrates the use of the chi test.

9-7. Solving Periodics With Known Cipher Sequences

The technique of frequency matching can be used any time the cipher sequence is known, whether or not the plain sequence is also known. When the plain sequence is known, the frequency patterns of the cipher sequences are best matched to the expected plain pattern as explained in paragraph 9-6. When the plain sequence is unknown, the frequency patterns of the cipher sequences can be matched to each other. In either case, the key is that the known cipher sequence allows the frequency count to be arranged in the order of the original cipher sequence. The following problem

demonstrates frequency matching with a known cipher component sequence. The cipher component sequence in the problem in Figure 9-3 is a keyword mixed sequence based on NORWAY.

```
MZTNK XLBTQ JVMQF WQTIX JJBTF  OCMEF HMHBM KTDPO IZYGR NJDHF
IEKAD AAPID NRBUF IYMET HDOPL  WLOID AQYEF KCWDF TPFAH MAUBR
HCWYQ JJMVR SLSBD HTTPO FDMQF  JLLNQ FEOIH QQYUQ KCLPO GLBQX
JJHBL WLQVF JDKNI JMTHF TCOVZ  ORHAD KCWDF XZWXF IPWCO XHWZP
KEOUF IJTPZ FAUUP HCYRF MDMTE  TRKDF MRWCO HMCNH TVGUL KRK

N O R W A Y B C D E F G H I J K L M P Q S T U V X Z
2 2 0 3 2 0 0 0 0 3 1 6 5 7 6 0 4 0 1 1 4 0 0 3 0
TOTAL LETTERS = 50          IC = 1.804082

N O R W A Y B C D E F G H I J K L M P Q S T U V X Z
0 0 5 0 3 1 0 7 4 3 0 0 1 0 5 0 6 3 2 3 0 2 0 2 0 3
TOTAL LETTERS = 50          IC = 1.697959

N O R W A Y B C D E F G H I J K L M P Q S T U V X Z
0 5 0 7 0 4 4 1 2 0 1 1 3 0 0 4 2 6 1 1 1 5 2 0 0 0
TOTAL LETTERS = 50          IC = 1.697959

N O R W A Y B C D E F G H I J K L M P Q S T U V X Z
4 0 1 0 3 1 4 2 3 3 0 1 2 4 0 0 0 0 5 3 0 3 5 3 1 1
TOTAL LETTERS = 49          IC = 1.282313

N O R W A Y B C D E  F G H I J K L M P Q S T U V X Z
0 5 3 0 0 0 0 0 5 1 15 0 3 1 0 1 3 1 2 4 0 1 0 0 2 2
TOTAL LETTERS = 49          IC = 3.161565
```

Figure 9-3. Known cipher components.

a. Examination of the frequency patterns in Figure 9-3 shows that they do not match the usual standard sequence-pattern. This means that the plain component sequence was not a standard sequence.

b. If the cipher sequences can be correctly matched against each other, the cryptogram can then be reduced to monoalphabetic terms and solved easily.

c. Figure 9-4 is a portion of a computer listing that matches the frequency count of the cipher letters of the first alphabet with the frequency count of second alphabet letters at every possible alignment. The alignments are evaluated by the chi test. In the chi test, each pair of frequencies for an alignment is multiplied. The products of all the pairs are totaled to produce the chi value for that alignment. Figure 9-5 shows the computation carried out for the first alignment. The chi test is also called the cross-product test.

MATCHING ALPHABET 1 AND ALPHABET 2

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

MATCH 1 : 70

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

MATCH 2 : 102

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

MATCH 3 : 128

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

MATCH 4 : 90

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

MATCH 5 : 172

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

MATCH 6 : 78

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

MATCH 7 : 103

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

MATCH 8 : 88

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

MATCH 9 : 64

Figure 9-4. Chi test computer extract.

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

0 +0 +0 +0 +6 +0 +0 +0 +0 +0 +0 +0 +6 +0+35 +0 +0+12 +0 +3 +0 +8 +0 +0 +0 +0

Figure 9-5. Computation of chi value.

d. Figure 9-6 shows the highest chi values for each match of the first alphabet with the other four alphabets. For all matches except the fourth alphabet, the chi values were clearly the highest. Two matches are shown for the fourth alphabet, because the difference between the two values is not significant. Either match could be the correct one.

MATCHING ALPHABET 1 AND ALPHABET 2

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

MATCH 5 : 172

MATCHING ALPHABET 1 AND ALPHABET 3

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

MATCH 18 : 170

MATCHING ALPHABET 1 AND ALPHABET 4

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

MATCH 10 : 134

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

MATCH 19 : 132

MATCHING ALPHABET 1 AND ALPHABET 5

N	O	R	W	A	Y	B	C	D	E	F	G	H	I	J	K	L	M	P	Q	S	T	U	V	X	Z
2	2	0	3	2	0	0	0	0	0	3	1	6	5	7	6	0	4	0	1	1	4	0	0	3	0
X	Z	N	O	R	W	A	Y	B	C	D	E	F	G	H	I	J	K	L	M	P	Q	S	T	U	V
2	2	0	5	3	0	0	0	0	0	5	1	15	0	3	1	0	1	3	1	2	4	0	1	0	0

MATCH 25 : 185

Figure 9-6. Best matches.

e. To resolve which of the two matches with the fourth alphabet is correct, the highest chi values for matches between the second and fourth and the third and fourth alphabets have also been determined. These are shown in Figure 9-7.

MATCHING ALPHABET 1 AND ALPHABET 4

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

MATCH 10 : 134

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

MATCH 19 : 132

MATCHING ALPHABET 2 AND ALPHABET 4

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

MATCH 15 : 132

MATCHING ALPHABET 3 AND ALPHABET 4

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

MATCH 2 : 141

Figure 9-7. Matches with the fourth alphabet.

f. The matches of alphabet four with alphabets two and three clarify which of the matches with the first alphabet was correct. This becomes apparent when we set up the other four alphabets.

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

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

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

4:

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

g. The match of N of the first alphabet with P of the fourth alphabetic correct. The second alphabet and third alphabet matches confirm this.

h. The next step in the solution is to reduce the cryptogram to monoalphabetic terms using the matches just determined. An A through Z sequence is arbitrarily used for the plain component, and the message is decrypted just as if it were the original.

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

```
rveir ympdv otabm dpeva okpdm  bdarm mnvot prrad nvote akrum
MZTNK XLBTQ JVMQF WQTIX JJBTF  OCMEF HMHBM KTDPO IZYGR NJDHF

nfymk eabvk aypem nbarx mekas  dmkvk eporm pdmqm votmo rafoe
IEKAD AAPID NRBUF IYMET HDOPL  WLOID AQYEF KCWDF TPFAH MAUBR

mdmnv okafe umdok mread keabm  omziv kfkvo tpoev pdzad lmpba
HCWYQ JJMVR SLSBD HTTPO FDMQF  JLLNQ FEOIH QQYUQ KCLPO GLBQX

okvos dmcfm oeyip oneum vdkfb  byvmk pdmqm yvmgm nompd yimhu
JJHBL WLQVF JDKNI JMTHF TCOVZ  ORHAD KCWDF XZWXF IPWCO XHWZP

pfkem nkeab kafeu mdokm readl  vyyqm rympd mnqio vtues pyy
KEOUF IJTPZ FAUUP HCYRF MDMTE  TRKDF MRWCO HMCNH TVGUL KRK
```

i. Reduced to monoalphabetic terms, many more repeats in the text that were suppressed by the multiple alphabets now appear. The solution is completed the same as any other monoalphabetic system.

9-8. Solving Periodics With Known Plaintext Sequences by Direct Symmetry

When the plaintext sequence is known, but not the ciphertext sequence, a solution technique known as direct symmetry is possible. Direct symmetry depends on the probable word method for the initial entry into the cryptogram. It makes use of the fact that the columns can be reconstructed in their original order as recoveries are made. Consider the next example, which uses a standard plaintext sequence.

```
MBNFQ ZLHQV ERNMS EXWFJ MBUFU  LWZIA LBSMK CFXKN WSNZW TREQA
XWHRN ACTKP EVBZJ PREZB TCZWH  TKTDN LBWAU PRZOQ KFEIW KBSRD
EVRWA MBIHO MBNFQ ZLHQV ERNMB  IVZIN MVCHR MXXRD EXDFU NLWGV
ITUCG JBUFW ALWML KFSLL IFQRX  YVIHE JKAHO
```

a. The period is five. The 14 letter repeat is probably *RECONNAISSANCE.*

```
recon naiss ance  a  o  re o          e e          c   n s
MBNFQ ZLHQV ERNMS EXWFJ MBUFU  LWZIA LBSMK CFXKN WSNZW TREQA

  i         a     n                   e     n          e
XWHRN ACTKP EVBZJ PREZB TCZWH  TKTDN LBWAU PRZOQ KFEIW KBSRD

a    re    recon naiss ance          r    r    a  o  a
EVRWA MBIHO MBNFQ ZLHQV ERNMB  IVZIN MVCHR MXXRD EXDFU NLWGV

      e o   a e
ITUCG JBUFW ALWML KFSLL IFQRX  YVIHE JKAHO
```

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
E													Z				M								
L				B									R												
		N					H																		
				M										F				Q							
													Q					V							

b. With recovered letters filled in, we can see that the beginning phrase is the stereotype, *RECONNAISSANCE PATROL REPORTS.*

```
recon naiss ancep atrol repor  ts    te e          c    n s
MBNFQ ZLHQV ERNMS EXWFJ MBUFU  LWZIA LBSMK CFXKN WSNZW TREQA

 si         a   l  n                  ter r  n  n        e
XWHRN ACTKP EVBZJ PREZB TCZWH  TKTDN LBWAU PRZOQ KFEIW KBSRD

a    re    recon naiss ance          r     rt    at or  ar s
EVRWA MBIHO MBNFQ ZLHQV ERNMB  IVZIN MVCHR MXXRD EXDFU NLWGV

  p   epo   are
ITUCG JBUFW ALWML KFSLL IFQRX  YVIHE JKAHO
```

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
E													Z				M		L						
L				B									R					W	X						
		N					H								U		W								
				M										F				Q							
											J		Q		S		U	V							

c. With a known plain component, the columns are in their original order. This means that the partially reconstructed cipher sequences are also in the right order. Each cipher sequence is the same sequence, and whatever one row reveals about the spacing of letters can be transferred to other rows as well. For example, in the second row, X follows immediately after W. X can then be placed after W in row three. Similarly, all common letters can be placed by carefully counting the intervals and placing the same letters at the same intervals in each row. Here is what the matrix looks like after all such values are placed.

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
E	F	H	J		Q	R	S		U	V	W	X	Z				M		L			N	B		
L			N	B			E	F	H	J		Q	R	S		U	V	W	X	Z				M	
		N	B			E	F	H	J		Q	R	S		U	V	W	X	Z				M		L
Z				M		L			N	B			E	F	H	J		Q	R	S		U	V	W	X
	L			N	B			E	F	H	J		Q	R	S		U	V	W	X	Z				M

d. Filling all the new values into the text reveals many more possibilities. Completion of the solution is routine from this point.

```
recon naiss ancep atrol repor  tst   tene   is e locat  ngs
MBNFQ ZLHQV ERNMS EXWFJ MBUFU  LWZIA LBSMK CFXKN WSNZW TREQA

msite       ardal  ngaf   tyk      e ter r  nt n  ig t  ent
XWHRN ACTKP EVBZJ PREZB TCZWH  TKTDN LBWAU PRZOQ KFEIW KBSRD

army  re p  recon naiss ancef   rt e rr po rtst  at or war s
EVRWA MBIHO MBNFQ ZLHQV ERNMB  IVZIN MVCHR MXXRD EXDFU NLWGV

  p   depot  areb  ingb  iltu   r pi d  p
ITUCG JBUFW ALWML KFSLL IFQRX  YVIHE JKAHO
```

e. The direct symmetry technique can also be used as an alternate method when the cipher sequence is the known sequence. The matrix can be inverted, placing the cipher sequence on the top of the matrix and the plaintext equivalents inside in separate rows for each alphabet. Each row will be the plaintext sequence in the correct order. Horizontal intervals recovered in one row can then be duplicated in each sequence just as was demonstrated above for cipher sequence recovery. Unlike the technique of frequency matching, it depends on successful plaintext assumptions, however. It is not as powerful a method of solution, but if plaintext can be readily identified, it may be the quickest way to solve a cryptogram.

Section III

Solving Periodics With Unknown Sequences

9-9. Solving Periodics by Indirect Symmetry

When neither the plaintext nor the ciphertext sequence is known, the matrix cannot be initially recovered with sequences in the correct order. Frequency matching cannot be used, either. However, some of the interval relationships are preserved even when the columns are not placed in the correct order, and these interval relationships can be exploited to aid in matrix recovery.

a. To illustrate how interval relationships are preserved, consider the following two matrices. The first is the matrix in its original form. The second is the same matrix, rearranged with the plain component in A through Z order. This is the form in which you will normally recover a matrix with unknown sequences until enough is known to rearrange the columns in the correct order.

c	l	a	r	i	n	e	t	b	d	f	g	h	j	k	m	o	p	q	s	u	v	w	x	y	z
B	C	D	F	G	I	J	K	L	M	Q	R	T	U	V	W	Y	Z	S	A	X	O	P	H	N	E
M	Q	R	T	U	V	W	Y	Z	S	A	X	O	P	H	N	E	B	C	D	F	G	I	J	K	L
R	T	U	V	W	Y	Z	S	A	X	O	P	H	N	E	B	C	D	F	G	I	J	K	L	M	Q
K	L	M	Q	R	T	U	V	W	Y	Z	S	A	X	O	P	H	N	E	B	C	D	F	G	I	J

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

b. The key principle to understand when working with ananalyst's matrix, like the second one above, is that every pair of columns and every pair of rows represents an interval in the original matrix. To illustrate this, look at the plaintext A column and the plaintext G column in the bottom matrix. The letters D and R appear in the first cipher sequence. If you count the distance between the D and R in the original (top) matrix, you see that the interval is nine. Similarly, the interval for the other pairs in the two columns, R and X, U and P, and M and S, are also nine. For any two columns that you compare, the horizontal interval between the letters in each alphabet will be the same. The interval will not always be nine, of course. It depends on which two columns you are comparing. The point is that between any pairs in the same row in the same two columns, the interval will be the same.

c. Next compare the letters in the first cipher sequence and the second in the bottom matrix. In the first column, the letters D and R appear, which we already noted are nine letters apart horizontally in the original matrix. The letters R and X appear in

another column in the first and second sequences, as do U and P, and M and S. The first and second cipher sequences are an interval of nine apart. Whichever pair of letters you look at in the first and second cipher sequences, they are nine apart in the original cipher sequence. Each pair of cipher sequences represents a different interval. For example, the interval between the first and third cipher sequence is eleven. The interval between the first and fourth is seven. The interval between the second and third is two, and so on.

d. There are a number of ways in which we can use an understanding of these interval relationships to help solve a polyalphabetic cryptogram. The use of interval relationships where sequences are unknown and columns are out of order is called indirect symmetry. This contrasts with the earlier situation with known sequences and columns in the correct order, where we used direct symmetry to aid in the solution.

e. To put indirect symmetry to use, consider the following example. Initial recoveries in a polyalphabetic system have produced the following information.

a	b	c	d	e	f	g	h	i	j	...
R	.	.	.	T	.	.	.	M	.	...
M	.	.	.	F	.	.	.	.	.	...
T	.	.	.	.	.	M	.	.	.	...

f. In comparing the plaintext A and E columns, we see that the letters R and T and the letters M and F are the same interval apart. We do not know what the interval is, but we know it is the same in each case.

g. The same interval appears when we compare the first and third cipher sequences, where R and T appear in the first column. Since we know the interval will be the same for any pair of letters between the first and third sequences, and we know M and F have the same interval as R and T, we can add the letter F in the plaintext I column in the third sequence under the letter M.

h. Any time we can establish an interval relationship for two pairs in a rectangular pattern as above, and can find three of the four letters, also in a rectangular pattern elsewhere, we can add the fourth letter to complete the pattern. The pairs must be read in the same direction in each case. Notice that we cannot add F in the plaintext G column in the first sequence. The interval from the first to the third sequence is not the same as the interval from the third to the first.

i. Matching pairs are usually found by reading horizontally in one case, and vertically with one letter in common in the second case, as in the above example. Matching relationships may be found anywhere in matrix, however, and are not restricted to

cases with one letter in common. You can find most such matching pairs by examining every column in which you have recovered at least three letters. For each letter in the column, look for a match with letters on the same row that are the same as one of the other letters in the column. When you find such letters, check for every possible complete rectangular relationship, and see if you can find the same relationship with one letter missing elsewhere. Often the addition of one or two letters is all you need to recognize more plaintext in the cryptogram and complete a solution.

j. If you have reason to believe that the plaintext sequence is the same as the cipher sequences, you can use the plaintext sequence in establishing interval relationships, too. All the techniques that apply to the ciphertext sequences apply to the plaintext sequence as well, when it is the same sequence.

9-10. Extended Application of Indirect Symmetry

Indirect symmetry can be used in other ways, too. For example, when enough letters have been recovered, you can list all the pairs of letters between each pair of sequences, and develop partial decimated chains of letters for each, as was explained in paragraph 4-8 with monoalphabetic substitution. These partial chains from different alphabet combinations can then be combined together geometrically to recover the original sequence. This technique is illustrated in the following indirect symmetry problem.

```
refer encey ourme ssage numbe  reigh teigh tthre esixs top
SMHPT ZZOPH KRION FJTYN WRSFN  SMKYZ JMKYZ JNPVN ZJKRX JOFSB

JMILM JMPPM VEVST JMIZK CTWFN  SMWEY LNBKG KKRET VHMSG ZJIEL

          si xthre eeigh tfour  fours evens top
ZOGSJ RMBZV ANPVN ZMKYZ JCRCT  EOVVX ZWBLX JOFOA TMEXB PUBGA

          o nesev enzer ozero  hours
YBWPG ZYXJA WMNPF ZZJPT KFBVA  IOVVX HOSOM KZBZV AZRIN YUBV
```

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

a. Through recognition of the stereotyped beginnings and the use of many numbers, the text shown has been recovered, and the recovered values filled into the matrix.

More values can be filled into the text, but we will first concentrate on the application of indirect symmetry.

b. To recover additional values through indirect symmetry, examine each column with more than two recovered letters in it. Beginning with the fifth column, take each letter in turn, and scan the same row as the selected letter for letters that are the same as those in the column. The first letter, Z, has no letters in common in its row with the letters M, B, P, and N.

c. For the second letter, M, the common letter Z does appear in its row. Having found a common letter, examine each rectangular relationship that exists between the two columns. We first see that Z and W have the same interval as M and Z. Links with this common letter will not add any more values, however.

d. The next rectangular relationship shows that P and L have the same interval as M and Z. Reading M and Z vertically, we look for P or L on the same rows as the M and Z to complete the relationship. We find neither P in the second row nor L in the first row. If either occurred, we could fill in the other. The letters can be written in a column off to the side for future use.

e. Having observed all relationships from the column with the common letter Z, we look for another column with a common letter on the M row. B and P do not occur except in our added column. The letter N does occur in the second row, however. Examining relationships in the N column, we see that Z and J have the same interval as M and N reading horizontally. With that established, we read M and N vertically and look for Z in the second row or J in the last row. This time we find Z in the second row. We can add J in the last row in the same column with Z to complete the rectangular relationship.

f. Continuing this process, all the letters shown in bold print can be added to the matrix without making any new plaintext recoveries.

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

g. It would be easy at this point to return to plaintext recovery to complete the solution, but another technique can be used to recover the original cipher sequences and rebuild the matrix. This technique involves listing all links that result by matching each cipher sequence with every other cipher sequence. Sequence 1 is matched with

sequences 2, 3, 4, and 5, in turn. Then sequence 2 is matched with 3, 4, and 5; sequence 3 is matched with 4 and 5; and sequence 4 is matched with 5. If the plaintext sequence were the same as the ciphertext sequence, it would only have been necessary to match the plaintext with each cipher sequence to get all combinations. When all links have been plotted and combined into partial chains wherever possible, the results are shown below.

1-2: ECHKOR TWZM FJN IL AS
1-3: EHOV TZB SIP WM KR FN
1-4: FZP WL KE SV JM OC TI AR
1-5: OSTFX IZN ER WJ KA CV
2-3: CHKORV WZMB FJN LP AS
2-4: AOE NMP SRC FWI JZL
2-5: LZJX CRS MN OA WF
3-4: XFTSO NZIVC BP ML RE JW
3-5: HRA BNX PZF KVS MJ IT
4-5: PN LJ EA VT CS IF ZX

h. Each set of partial chains represents a decimation of the original sequence. Sometimes, you will be fortunate at this point to find that one of the partial chains directly represents the original sequence (decimation one). When this happens, the original sequence is the obvious starting point. It does not occur in this example, so the best technique is usually to select a set with one of the longer chains as a starting point and relate all other sequence combinations to it. Notice that the chains produced by sequences 1-2 and by sequences 2-3 are obviously produced by the same interval, since many of the partial chains are identical. They make a good starting point for this problem. Begin by listing each chain fragment on paper, horizontally. Write the separate chains in different rows so they will not run into each other.

E C H K O R V
T W Z M B
F J N
I L P
A S

i. The next step is to relate other chains to the existing plot. By examining the intervals or patterns that letters from other chains have in relation to the starting chains, they can be added by following the same rule. For example, the 1-3 combination can

be added by observing that it will fit the starting chains by skipping every other letter. This will also enable linking the fifth fragment, AS, with the fourth. After adding all the 1-3 chains, the plot looks like this example.

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

j. Next, search for another combination that can be added to the plot. The 3-4 combination links by counting backwards every fifth letter, as shown by the V and C of the NZIVC chain. This ties all the chain fragments together into one longer chain. When all combinations are added, each by their own rule, it results in almost complete recovery.

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

k. This technique is known as linear chaining. Sometimes you will be unable to combine the fragments into one long chain. When all intervals are even, you will always end with two separate 13-letter chains, which may be combined by trial and error or by figuring out the structure of the original matrix. A second technique, called geometric chaining, which could have been applied here also, is explained in paragraph 9-11.

l. Continuing, the chain above must be a decimation of the original sequence. Since V, W, and X are spaced consistently nine apart, trying a decimation of 9 produces the next sequence.

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

m. With G missing from alphabetical progression, the sequence is keyword mixed, based on GAMES. We can now return to the polyalphabetic matrix and rearrange the columns using the GAMES sequence on each cipher row.

o	a	.	u	.	b	.	v	y	.	n	.	m	e	.	x	p	f	r	z	i	g	s	c	h	t
K	L	N	O	P	Q	R	T	U	V	W	X	Y	Z	G	A	M	E	S	B	C	D	F	H	I	J
O	P	Q	R	T	U	V	W	X	Y	Z	G	A	M	E	S	B	C	D	F	H	I	J	K	L	N
R	T	U	V	W	X	Y	Z	G	A	M	E	S	B	C	D	F	H	I	J	K	L	N	O	P	Q
E	S	B	C	D	F	H	I	J	K	L	N	O	P	Q	R	T	U	V	W	X	Y	Z	G	A	M
A	M	E	S	B	C	D	F	H	I	J	K	L	N	O	P	Q	R	T	U	V	W	X	Y	Z	G

n. The unused letters can be determined by returning to the plaintext and deciphering the rest of the message. The plaintext sequence turns out to be a simple transposition mixed sequence based on OLYMPIC. The repeating key is KOREA.

o. The approach shown to solving this problem is not necessarily the way in which you would solve it in actual practice. It would probably be more effective to return to the plaintext earlier than was done in this example. This approach was selected to show the variety of indirect symmetry techniques that can be used, not necessarily because it would yield the quickest solution.

9-11. Solution of Isologs

Whenever isologs are encountered between periodic messages with different period lengths, it is possible to recover the original cipher sequences without any initial plaintext recovery. The cryptograms can then be reduced to monoalphabetic terms and quickly solved. Two different techniques may be used, depending on whether the same alphabets or different alphabets are used in the isologs.

a. When isologous cryptograms use the same alphabets with different repeating keys, the cipher sequences can be recovered by the indirect symmetry process. Take the following two messages, for example.

Message 1:

AOPDY JBFKW ATILB XCTKZ KIKVN SHUAJ COWLA PDBRU KRXAT WALBZ
ZVYZZ YRNCI FPPOJ OBYJQ SESQK SPGUK XIKVW AVUCW MYTXY ZCYZB
PHBJE SCWXC TKZKV PKN

(period 3)

Message 2:

DCFHC SBOHH BOENY GMGKB HQOQF FIXHS CVURB KKWUX UEXEQ HBFHP
SYCCZ NZSFZ MDFST WBNFB VNXEB VYDUS VQOQR TMXMI MNQJR VJOSE
YQBQC CFSAX KODTV WHS

(period 4)

(1) To solve the isologs, the two messages are first superimposed with the alphabets numbered for each.

```
1: AOPDY JBFKW ATILB XCTKZ KIKVN   SHUAJ COWLA PDBRU KRXAT WALBZ
   12312 31231 23123 12312 31231   23123 12312 31231 23123 12312
2: DCFHC SBOHH BOENY GMGKB HQOQF   FIXHS CVURB KKWUX UEXEQ HBFHP
   12341 23412 34123 41234 12341   23412 34123 41234 12341 23412

1: ZVYZZ YRNCI FPPOJ OBYJQ SESQK   SPGUK XIKVW AVUCW MYTXY ZCYZB
   31231 23123 12312 31231 23123   12312 31231 23123 12312 31231
2: SYCCZ NZSFZ MDFST WBNFB VNXEB   VYDUS VQOQR TMXMI MNQJR VJOSE
   34123 41234 12341 23412 34123   41234 12341 23412 34123 41234

1: PHBJE SCWXC TKZKV PKN
   23123 12312 31231 231
2: YQBQC CFSAX KODTV WHS
   12341 23412 34123 412
```

(2) With periods of 3 and 4, there are 12 different ways in which the alphabets of the first are matched to the alphabets of the second. These begin with the first alphabet of message 1 matched with the first alphabet of message 2 and continue through alphabet 3 matched with alphabet 4. After these 12 matches, the cycle of matches starts over again. For other periods, the number of different alphabet matches is the least common multiple of the two period lengths. The least common multiple of 6 and 4 is 12. The least common multiple of 6 and 9 is 18. For periods of 8 and 9, 72 different alphabet matches are required.

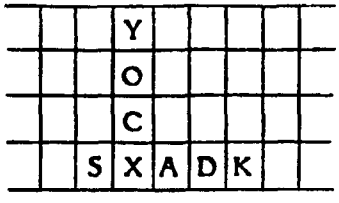
(3) Analysis continues by plotting the links for each alphabet pair. For example, the first link is A1=D1, the second link is O2=C2, and the third link is P3=F3. The next example shows all links plotted and combined into partial chains.

```
1-1: SXADK IE NFM BH WR CJ
2-2: YOCX LN SF BW ZPD QE AT
3-3: TKBY PF HI RU ZS VM
1-4: KOSVY UXG DH BE
2-1: PYCM AH KU JT ZD
3-2: KTGD OWI JS RE ZC HQ
1-3: BB KK (all links the same)
2-4: FOV ZB AE YN KS JQ PW
3-1: KH WU TQ RZ JF XV EC
1-2: IQB NSC WH LR XJ
2-3: AB KO CF SV YR
3-4: IZVQ TO PK LF EN WS
```

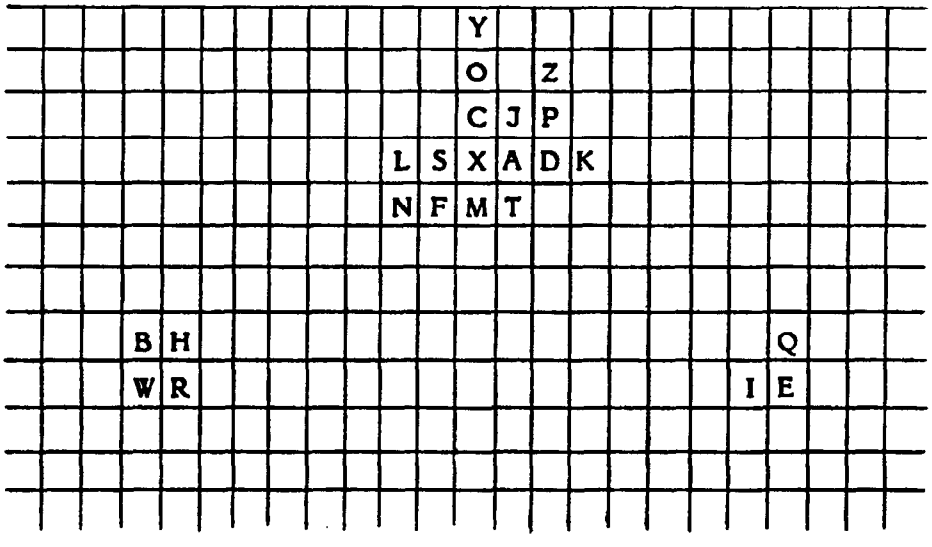
(4) The 1-3 plot shows that the same alphabets were used in both these positions.

(5) The partial chains can be combined into one long chain by a process of geometric chaining. Geometric chaining will often produce results when linear chaining is not effective. Geometric chaining is plotted horizontally and vertically, instead of in one straight line. Relationships between alphabet matches can be discovered more readily with this method.

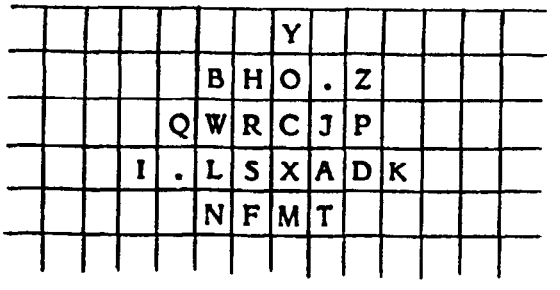
(6) Geometric chaining begins, as with linear chaining, by selecting one alphabet match to plot horizontally. We can select the 1-1 match for its 5-letter chain as a starting point. Next, select a second alphabet match to intersect it plotted vertically. For our example, we will use the 2-2 match, producing the following initial plot.



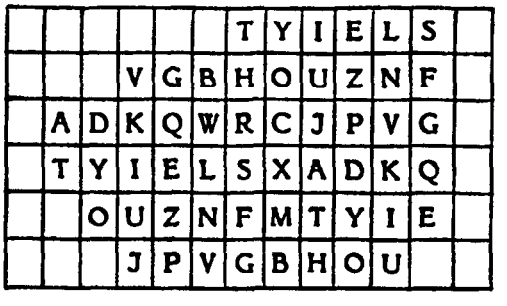
(7) To this initial plot, we add as many other fragments from the 1-1 and 2-2 matches as we can at this time. We can also set up plots separated from these for each one that cannot be linked to it.



(8) The next step is to find another alphabet match that can easily be added to the plot. For example, the 1-2 match proceeds in the diagram along a lower left to upper right diagonal, as shown by the NSC and XJ fragments. All the 1-2 fragments can be added by the same diagonal rule. This ties in the separate plots from above, also.



(9) Each additional alphabet combination can be added to the plot now. In many cases, you may see different possibilities for rules. For example, the 3-4 match can be seen to proceed by an up 3, left 1 rule, as shown by the TO link. A simpler equivalent is to plot by the upper left to lower right diagonal, as shown by the PK link. The simplest way to describe the 3-3 match is up 1, right 2, as shown by the TK or BY links. This is similar to a knight's move in chess. When all matches are plotted, they produce this diagram.



(10) The rows can easily be extended into one 26-letter chain at this point, but if alphabetic progression can be spotted by any other rule, it can be used instead. For example, starting with the V in the upper left part of the diagram, VWXY appears by a descending knight's move. Continuing from the Y that repeats near the left side, the sequence can be extended further. The complete sequence appears below.

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

(11) Using the new recovered sequence and the relationships between the alphabets of messages 1 and 2, the matrices for both messages can be set up. Using the first cipher sequence for message 1, all the cipher sequences for message 2 can be lined up with it using the links already plotted. Here is how the message 2 alphabets line up with alphabet one. The first 1-1, 1-2, 1-3, and 1-4 links from the isologs are shown in bold print to demonstrate how they were lined up.

C1: G R A I N B C D E F H J K L M O P Q S T U V W X Y Z
C2:
C3:

C1: B C D E F H J K L M O P Q S T U V W X Y Z G R A I N
C2: M O P Q S T U V W X Y Z G R A I N B C D E F H J K L
C3: G R A I N B C D E F H J K L M O P Q S T U V W X Y Z
C4: I N B C D E F H J K L M O P Q S T U V W X Y Z G R A

(12) Similarly, the alphabets in the first matrix can be completed by plotting the relationships between the second message and the first. The solution then becomes a matter of reducing them to monoalphabetic terms.

(13) In cases where the two periods have a common factor, the sequences can still be recovered, but they cannot be fully aligned. In this case, the chi test can be used to match the sequences by frequencies, if necessary, once the sequences are known.

b. A different technique must be used if different alphabets are used between the isologs, not just different repeating keys. For example, consider the next two messages.

Message 1:

AUUJB NFMOI AXCQD LHXPE OCPZD XMZAN HUGQV OIAZZ POPAA FOZUY
OQEOX BRDHA MVUUO SFBNW XJXWO XVEZP IPHYM WODOT CMOTU CTUPT
UOYRO SBBMP CMMXA ATYAN

(period 3)

Message 2:

ZCIPY RZXLG ZXSNP CNLNH LQDZU FXALR SIGIH MQTCA GTNMQ TCZGG
ZYZTG GORIB NDISF YZGUB KGKEZ IMDJS HLIYN EZKFF XXLOG CYCSG
KTHJL VTINA ORDLW MPDZK

(period 4)

(1) The sequences are different in the two messages, and they cannot be directly chained together. If you listed the links resulting from the two messages using the previous technique, they would lead nowhere and contradictions would quickly develop. The cipher sequences of each must be kept separate.

(2) The method of recovering the cipher sequences when they are different is to set up periodic matrices one over the other, as shown below. Message 1 and message 2 equivalents are then plotted in the correct sequence for each in the same columns. Initially, this will result in more than 26 columns, but as incomplete columns are combined with each other, the matrices will collapse to the correct width. This method could be used with more than two isologs also, by superimposing as many matrices as there are isologous messages.

```
1: AUUJB NFMOI AXCQD LHXPE OCPZD  XMZAN HUGQV OIAZZ POPAA FOZUY
   12312 31231 23123 12312 31231  23123 12312 31231 23123 12312
2: ZCIPY RZXLG ZXSNP CNLNH LQDZU  FXALR SIGIH MQTCA GTNMQ TCZGG
   12341 23412 34123 41234 12341  23412 34123 41234 12341 23412

1: OQEOX BRDHA MVUUO SFBNW XJXWO  XVEZP IPHYM WODOT CMOTU CTUPT
   31231 23123 12312 31231 23123  12312 31231 23123 12312 31231
2: ZYZTG GORIB NDISF YZGUB KGKEZ  IMDJS HLIYN EZKFF XXLOG CYCSG
   34123 41234 12341 23412 34123  41234 12341 23412 34123 41234

1: UOYRO SBBMP CMMXA ATYAN
   23123 12312 31231 23123
2: KTHJL VTINA ORDLW MPDZK
   12341 23412 34123 41234
```

Message 1:

1	A			J			F			I			C		
2		U			B			M			A			Q	
3			U			N			O			X			D

Message 2:

1	Z				Y				L				S		
2		C				R				G				N	
3			I				Z				Z				P
4				P				X				X			

(3) The first three groups of each message are plotted above. Each time a previously used letter appears in the same sequence, the two columns can be combined. For example, in message 2, the Zs in the third sequence allow those two columns to be combined, and similarly, the Xs in the fourth sequence can be combined. In the next example, the complete messages are plotted and all possible columns are combined.

Message 1:

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

Message 2:

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

(4) These matrices can easily be completed by direct symmetry, remembering that the sequence in each matrix is different.

Message 1:

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

Message 2:

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

(5) Either cryptogram can now be reduced to monoalphabetic terms and solved, as before.

CHAPTER 10

APERIODIC POLYALPHABETIC CIPHERS

10-1. Simple Manual Aperiodic Systems

Chapter 9 showed that periodic polyalphabetic systems are generally more secure than monoalphabetic systems. However, the regular, repeating nature of the keys in periodic systems are a weakness that an analyst can exploit. Using factor analysis or the phi test, the analyst can readily determine how many alphabets there are and which letters are enciphered by which alphabets. Aperiodic polyalphabetic systems eliminate the regular, repeating use of alphabets so the analyst cannot easily tell which letters are enciphered by which alphabets. There area number of ways to use a limited set of alphabets but suppress their regular repetition. The following subparagraphs show the most common types of these, and briefly discuss their weaknesses and approaches to their solution. They are presented to make you aware of the possibility that such techniques can be used, but no detailed explanation of their solution is given.

a. Word Length Aperiodic. The simplest type of aperiodic changes alphabets with each word instead of each letter. The analyst cannot tell which letters are encrypted by which alphabet until the text is recovered. However, the major weakness of this system is that when repeats occur, they are likely to be word length, and plaintext word patterns show through as clearly as with monoalphabetics. When alphabets are known, the generatrix method makes the plaintext obvious.

b. Numerically Keyed Aperiodic. Another approach, similar to word-length encipherment, is to change alphabets after a number of letters, determined by a numerical key. The numerical key is often based on the repeating key. The key is generated by the same process used with a numerically keyed transposition

sequence. The letters in the repeating keyword are numbered alphabetically. Then the key determines how many letters are enciphered consecutively by each alphabet. For example, here is a short message enciphered by a numerically keyed aperiodic based on the keyword BLACK.

	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	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
5	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
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
3	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
4	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

2 5 1 3 4 2 5 1 3
en emyat t ack ingo na llfro n ts
FO PXJLE T CEM SXQY OB WWQCZ N VU

This system, while more complicated than a word-length aperiodic, allows many repeats and patterns to appear. When the alphabets are known, use of the generatrix method also quickly reveals the plaintext.

c. Interruptor Letter Aperiodic. Another approach to breaking up the cyclic nature of periodic systems is through the use of an interruptor letter. In interruptor letter systems, the alphabets are used in rotation like a periodic system, but whenever a preselected plaintext (or alternatively, ciphertext) letter is encountered, the rotation is interrupted and encipherment returns to the first alphabet. This is a more secure method than the previous two, but it can have the effect of creating repeats that would not otherwise occur. For example, if a plaintext R is used as an interruptor letter, every time REINFORCEMENTS appears in the text, encipherment from the second letter on will be identical every time. The letter after the initial R will be enciphered by the first alphabet each time because of the interruption. The same thing will happen with any word that begins with the interruptor letter. Use of a ciphertext interruptor letter instead of a plaintext letter will avoid many of these repeats, but the interruptions will generally occur much less often in such a case.

10-2. Long-Running Key Aperiodic

Much more common than the simple manual aperiodic systems described in the previous paragraph are those that use a long-running, ever changing key. These systems may be enciphered manually, by cipher machine, or by computer, as first discussed in paragraph 8-1. Figure 8-1 gave an example of using a book key where the key

letters were a quotation. A quotation, particularly from a book, provides a ready source of long-running keys, but it is relatively unsecure, because the key itself is so orderly. More often, the keys will be random or pseudorandom. The keys are applied to the plaintext using an alphabet chart like the Vigenere square in Figure 8-l. The keys may be generated by a pseudorandom, repeatable process or by a random, nonrepeatable process. Both the sending and receiving cryptographer must have a copy of the same book or pad of keys. When these are intended for single usage of the keys, the system is called a one-time pad system. Truly random one-time pad systems are absolutely unbreakable when used properly. When keys are reused, however, whether by mistake or by design, the messages with the reused keys are likely to be recoverable. Manual one-time pad systems are slow systems to use and present logistics problems for any large scale usage. The volume of keys must be at least equal to the volume of messages to be sent, When more than one communications link shares the use of copies of the same pad, careful procedures must be set up to prevent reuse of the same keys by different users.

10-3. Solution of Long-Running Key Aperiodic

The solution of messages enciphered in long-running key systems may be possible in three situations. First, the key generation process may be known in advance from prior recoveries or other sources. Second, the keys may be so orderly that they are recognizable when partially recovered, as can occur when plaintext is used as the source of keys. Third, the same sequence of keys is reused. We are primarily concerned with the third case, where keys are reused.

a. Depth **Recognition.** A reuse of long-running keys is called a depth. Messages using the same keys are called messages in depth. If the keys begin at the same point in two or more messages, the messages are in flush depth, If the keys begin at different points in two or more messages, but include reused keys for at least part of the messages, they are in offset depth. The solution of messages in depth first requires you to recognize that the depth exists.

(1) One way to recognize depth is through exploitation of indicator systems. In one-time pad systems and in many types of cipher machine or computer systems, the starting point or settings for the keys must be known by the enciphering and deciphering cryptographers. This information on the keys is often passed from cryptographer to cryptographerthrough the use of an indicator system. The first way to recognize a depth is to find two messages or transmissions with identical indicators. Identical indicators will often tip-off that a flush depth is occurring.

(2) The second way to recognize depth is to find repeated text between two or more messages. Except for short accidental repeats, repeated ciphertext will only occur when the same plaintext is enciphered with the same keys. In periodic

systems and simple manual aperiodic, this will often occur within a single message as the same keys are reused. With long-running key aperiodic, this will only occur between messages when keys are reused. If all depths are expected to be flush depths, the search for repeats is a matter of superimposing messages and looking for repeats in the same position in each message. If depths are offset, they are more difficult to find by inspection alone.

(3) The third way to recognize depth is to use a type of coincidence test known as the kappa test. Whether whole words and phrases are repeated using the same keys or not, individual characters using the same keys will occur frequently when depths are present. When two messages are matched together, letter by letter, and do not use the same keys, 1 out of 26 letters (or 3.85 percent) will randomly match. Of course, if a different alphabet is used, or if characters other than letters are also used, the expected number of matches by chance alone will be 1 out of the total number of different characters used. On the other hand, if the messages are correctly placed in depth, a letter by letter comparison (the kappa test) will produce matches about 6.67 percent of the time. Also, the results can be expressed as a kappa index of coincidence showing the ratio of observed coincidences to random expectation. As with searching for repeats, it is much easier to find flush depths than it is to find offset depths, but with computer support, messages can be matched in every possible alignment to search for depths.

(4) As an example of depth recognition, consider the three messages that follow. Each has similar indicator groups that suggest the messages may be in depth with each other. Messages 1 and 2 have identical indicators. Message 3 differs only in the last digit of the second group.

Message 1:

JJ632 0406 **HJJBW KBZGA OWSON** SRJCF **AGORU** EOGVA CNWIH CLVZX **MDSAF EMFGP** VNNNN ABJPZ TJNVL **QMGGN** TVBAP **MDODN ODMIO NOIWO** XANAC **CNLXS EMBWV** CVZYD **FTPUC TQNAW** ZUTUH **J** J632

Message 2:

JJ632 0406 FWFQA VSAIA UOSOS SHMQD YCLNO YOOQV **GNVSD BOIIG** XDRAF **GFEMM GTCZN VMYSN UHCYM** CZBPP BOVYW BLQIO **AKEXM NMNTN SODPA UNBMO QYYQS GOBMA** WSUQL **JJ632**

Message 3:

JJ632 0407 **KDHYW QOEBJ DBJGH** PYGEP **HOQNY OOISH UYMHX MGTUC EYWTG RLRKQ YKISC QNPTB** JFCRA EKZXA LLCOZ **HIKYE UJPKC SHWHN** VWAXF APEVC XJDQS FISYL SQLCY **JAGRP JJ632**

(5) There are no repeats longer than three letters between any of the three messages. Because of the identical indicators, we first try to match messages 1 and 2 at a flush depth using the kappa test, The number of matches multiplied by 26 and divided by the number of comparisons equals the kappa IC. Do not count the indicator groups in the comparisons.

1: JJ632 0406 HJJBW KBZGA OWSS SRJCF AGORU EOGVA CNWIH CLVZX
2: JJ632 0406 FWFQA VSAIA UOSOS SHMQD YGLNO YOOQV GNVSD BOIIG

1: MDSAF EMFGP VNNNN ABJPZ TJNVL QMGGN TVBAP MDODN ODMIO NOIWO
2: XDRAF GFEMM GTCZN VMYSN UHCYM CZBPP BOVYW BLQIO AKEXM NMNTN

1: XANAC CNLXS EMBWV CVZYD FTPUC TQNAW ZUTUH JJ632
2: SODPA UNBMO QYYQS GOBMA WSUQL JJ632

2 to 1: offset 0
13 matches out of 115 comparisons
Kappa IC = 2.94

(6) As shown by the kappa test, the number of matches is well above random expectation The two messages appear to be in flush depth with each other. Next we try message 3 matched with the first two at a flush depth.

1: JJ632 0406 HJJBW KBZGA OWSON SRJCF AGORU EOGVA CNWIH GLVZX
2: JJ632 0406 FWFQA VSAIA UOSOS SHMQD YGLNO YOOQV GNVSD BOIIG
3: JJ632 0407 KDHYW QOEBJ DBJGH PYGEP HOQNY OOISH UYMHX MGTUC

1: MDSAF EMFCP VNNNN ABJPZ TJNVL QMGGN TVBAP MDODN ODMIO NOIWO
2: XDRAF GFEMM GTCZN VMYSN UHCYM CZBPP BOVYW BLQIO AKEXM NMNTN
3: EYWTG RLRKQ YKISC QNPTB JFCRA EKZXA LLCOZ HIKYE UJPKC SHWHN

1: XANAC CNLXS EMBWV CVZYD FTI'UC TQNAW ZUTUH JJ632
2: SODPA UNBMO QYYQS GOBMA WSUQL JJ632
3: VWAXF APEVG XJDQS FISYL SQLCY JAGRP JJ632

3 to 1 and 2: offset 0
9 matches out of 235 comparisons
Kappa IC = 1.00

(7) The flush match of message 3 is clearly not a correct match, because of the low kappa index of coincidence. We next try offsets of 1, 2, 3, 4, and 6 letters to the right.

```
1: JJ632 0406 HJJBW KBZGA OWSON SRJCF AGORU EOGVA CNWIH GLVZX
2: JJ632 0406 FWFQA VSAIA UOSOS SHMQD YGLNO YOOQV GNVSD BOIIG
3: JJ632 0407 KDHY WQOEB JDBJG HPYGE PHOQN YOOIS HUYMH XMGTU

1: MDSAF EMFGP VNNNN ABJPZ TJNVL QMGGN TVBAP MDODN ODMIO NOIWO
2: XDRAF GFEMM GTCZN VMYSN UHCYM GZBPP BOVYW BLQIO AKEXM NMNTN
3: CEYWT GRLRK QYKI S CQNPT BJFCR AEKZX ALLCO ZHIKY EUJPK CSHWH

1: XANAC CNLXS EMBWV CVZYD FTPUC TQNAW ZUTUH JJ632
2: SODPA UNBMO QYYQS GOBMA WSUQL JJ632
3: NVWAX FAPEV GXJDQ SFISY LSQLC YJAGR PJJ63 2

3 to 1 and 2: offset 1
13 matches out of 234 comparisons
Kappa IC = 1.44

1: JJ632 0406 HJJBW KBZGA OWSON SRJCF AGORU EOGVA CNWIH CLVZX
2: JJ632 0406 FWFQA VSAIA UOSOS SHMQD YGLNO YOOQV GNVSD BOIIG
3: JJ632 0407 KDH YWQOE BJDBJ GHPYG EPHOQ NYOOI SHUYM HXMGT

1: MDSAF EMFGP VNNNN ABJPZ TJNVL QMGGN TVBAP MDODN ODMIO NOIWO
2: XDRAF GFEMM GTCZN VMYSN UHCYM GZBPP BOVYW BLQIO AKEXM NMNTN
3: UCEYW TGRLR KQYKI SCQNP TBJFC RAEKZ XALLC OZHIK YEUJP KCSHW

1: XANAC CNLXS EMBWV CVZYD FTPUC TQNAW ZUTUH JJ632
2: SODPA UNBMO QYYQS GOBMA WSUQQ 33632
3: HNVWA XFAPE VGXJD QSFIS YLSQL CYJAG RPJJ6 32

3 to 1 and 2: offset 2
8 matches out of 233 comparisons
Kappa IC = 0.89

1: JJ632 0406 HJJBW KBZGA OWSON SRJCF AGORU EOGVA CNWIH GLVZX
2: JJ632 0406 FWFQA VSAIA UOSOS SHMQD YGLNO YOOQV GNVSD BOIIG
3: JJ632 0407 KD HYWQO EBJDB JGHPY GEPHO QNYOO ISHUY MHXMG

1: MDSAF EMFGP VNNNN ABJPZ TJNVL QMGGN TVBAP MDODN ODMIO NOIWO
2: XDRAF GFEMM GTCZN VMYSN UHCYM GZBPP BOVYW BLQIO AKEXM NMNTN
3: TUCEY WTGRL RKQYK ISCQN PTBJF CRAEK ZXALL COZHI KYEUJ PKCSH

1: XANAC CNLXS EMBWV CVZYD FTPUC TQNAW ZUTUH 53632
2 : SODPA UNBMO QYYQS GOBMA WSUQL JJ632
3: WHNVW AXFAP EVGXJ DQSFI SYLSQ LCYJA GRPJJ 6 3 2

3 to 1 and 2: offset 3
6 matches out of 232 comparisons
Kappa IC = 0.67
```

```
1: 15632 0406 HJJBW KBZGA OWSON SRJCF AGORU EOGVA CNWIH GLVZX
2: JJ632 0406 FWFQA VSAIA UOSOS SHMQD YGLNO YOOQV GNVSD BOI IG
3: JJ632 0407      K DHYWQ OEBJD BJGHP YGEPH OQNYO OISHU YMHXM

1: MDSAF EMFGP VNNNN ABJPZ TJNVL QMGGN TVBAP MDODN ODMIO NOIWO
2 :  XDRAF GFEMM GTCZN VMYSN UHCYM GZBPP BOVYW BLQIO AKEXM NMNTN
3: GTUCE YWTGR LRKQY KISCQ NPTBJ FCRAE KZXAL LCOZH IKYEU JPKCS

1: XANAC CNLXS EMBWV CVZYD FTPUC TQNAW ZUTUH JJ632
2 :  SODPA UNBMO QYYQS GOBMA WSUQL JJ632
3: HWHNV WAXFA PEVGX JDQSF ISYLS QLCVJ AGRPJ 3632

3 to 1 and 2: offset 4
9 matches out of 231 comparisons
Kappa IC = 1.01
```

```
1: JJ632 0406 HJJBW KBZGA OWSON SRJCF AGORU EOGVA CNWIH GLVZX
2 :  JJ632 0 4 0 6  FWFQA VSAIA UOSOS SHMQD YGLNO YOOQV GNVSD BOI IG
3: 31632 0407      KDHYW QOEBJ DBJGH PYGEP HOQNY OOISH UYMHX

1: MDSAF EMFGP VNNNN ABJPZ TJNVL QMGGN TVBAP MDODN ODMIO NOIWO
2: XDRAF GFEMM GTCZN VMYSN UHCYM GZBPP BOVYW BLQIO AKEXM NMNTN
3: MGTUC EYWTG RLRKQ YKISC QNPTB JFCRA EKZXA LLCOZ HIKYE UJPKC

1: XANAC CNLXS EMBWV CVZYD FTPUC TQNAW ZUTUH JJ632
2:  SODPA UNBMO QYYQS GOBMA WSUQL 31632
3 :  SHWHN VWAXF APEVG XJDQS FISYL SQLCY JAGRP JJ632

3 to 1 and 2: offset 5
17 matches out of 230 comparisons
Kappa IC = 1.92
```

(8) The offset of five is clearly the best match of those tried, and the kappa index of coincidence is a good value for a correct match. The three messages are now correctly placed in depth.

b. Depth Reading. When the messages are superimposed properly, they can be solved by a process known as depth reading. With only a few messages, the process of applying the key must be known. With manual systems, standard alphabets are commonly used. With cipher machine or computer based systems, the process of baud addition is usually known or can be figured out easily. The three messages in our example use the standard alphabet Vigenere square of Figure 10-l.

Plaintext

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
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 10- 1. Vigenere square.

(1) With three messages in depth, almost any correct assumption of plaintext will lead to a quick solution. For example, trying the word REPLACEMENT as the first word of message 3 produces the results shown below.

```
                        rchte amarr i
1: JJ632 0406 HJJBW KBZGA OWSON SRJCF AGORU EOGVA CNWIH GLVZX

                        ctivt gearw i
2: JJ632 0406 FWFQA VSAIA UOSOS SHMQD YGLNO YOOQV GNVSD BOIIC

                        repla cemen t
3: JJ632 0407         KDHYW QOEBJ DBJGH PYGEP HOQNY OOISH UYMHX

Key:                  TZSNW OKSXW K

1: MDSAF EMFGP VNNNN ABJPZ TJNVL QMGGN TVBAP MDODN ODMIO NOIWO
2: XDRAF GFEMM GTCZN VMYSN UHCYM GZBPP BOVYW BLQIO AKEXM NMNTN
3: MGTUC EYWTG RLRKQ YK I SC QNPTB  JFCRA EKZXA LLCOZ HI KYE UJPKC

1: XANAC CNLXS EMBWV CVZYD FTPUC TQNAW ZUTUH JJ632
2: SODPA UNBMO QYYQS GOBMA WSUQL  JJ6 32
3: SHWHN VWAXF APEVG XJDQS FISYL  SQLCY JAGRP 3J6 3 2
```

(2) Recovering the key from the assumption of REPLACEMENT and using it to decipher the other two messages produces good segments of plaintext in each message. It is easy to build on these assumptions to recover additional plaintext. For example, assuming that the second message begins PROTECTIVE *GEAR* and that the word after TEAM in the first message is *ARRIVING* leads to additional recoveries.

```
                  resea rchte amarr iving
1: JJ632 0406 HJJBW KBZGA OWSON SRJCF AGORU EOGVA CNWIH GLVZX

                  protc ctive gearw illbe
2: JJ632 0406 FWFQA VSAIA UOSOS SHMQD YGLNO YOOQV GNVSD BOIIC

                        repla cemen tfiri
3: JJ632 0407         KDHYW QOEBJ DBJGH PYGEP HOQNY OOISH UYMHX

Key:            QFRXW TZSNW OKSXW KWBPZ

1: MDSAF EMFGP VNNNN ABJPZ TJNVL QMGGN TVBAP MDODN ODMIO NOIWO
2: XDRAF GFEMM GTCZN VMYSN UHCYM GZBPP BOVYW BLQIO AKEXM NMNTN
3: MGTUC EYWTG RLRKQ YK ISC QNPTB  JFCRA EKZXA LLCOZ HI KYE UJ PKC

1: XANAC CNLXS EMBWV CVZYD FTPUC TQNAW ZUTUH JJ632
2: SODPA UNBMO QYYQS GOBMA WSUQL JJ632
3: SHWHN VWAXF APEVG XJDQS FISYL SQLCY JAGRP JJ632
```

(3) This process of assuming text can be continued to a complete solution. Correct assumptions are easily verified. Incorrect assumptions are quickly disproved.

(4) The most difficult step is making the first correct assumption. Message beginnings are the most likely area to yield results, because they are likely to be very stereotyped. Sometimes, just trying the letters RE at the beginning of a message will be enough to suggest the text of the messages in depth with it. When message beginnings do not yield results, more powerful techniques are available.

c. Crib Dragging. When you cannot assume the beginning of a message, you can still often correctly assume a particular word that will be in a message. The assumptions can come from familiarity with previous messages, results of traffic analysis and direction finding, or other intelligence sources. Once you suspect a word is in one of two or more messages in depth, you can systematically try the word at every position, recover the keys each position would produce, and try the keys in the other message or messages to see if the keys produce more plaintext. This is a laborious process performed manually, but a sure one. Fortunately, there are some short cuts that can be used to simplify the process.

(1) Two messages in depth can generally be combined in such a way that you can skip the step of key recovery and proceed directly to checking for plaintext. With the Vigenere square of Figure 10-1, this can be accomplished by treating one message as if it were laintext, the other as ciphertext, and producing the resulting key stream, whiCR is actually a combination of the two ciphertexts. To demonstrate this process, consider the beginnings of messages 1 and 2 from the previous example. If we combine message 1 and message 2 as if they were plaintext and ciphertext respectively, it produces a combination text for the first groups of YNWPE, Message 1 letters are used as keys in the Vigenere square. Message 2 letters represent the internals of the Vigenere square. For example, key H matched agamst internal F produces plaintext Y.

Message 1: H J J B W . . .
Message 2: F W F Q A . . .
Combination: Y N W P E . . .

(2) If we now apply the correct plaintext of message 1 to the combination text using the Vigenere square, it will directly produce the plaintext of message 2. The

combination text is again found in the key letter position in the square, and the plaintext is found in the same position for each message as the original ciphertexts.

Message 1: H J J B W . . .
Message 2: F W F Q A . . .
Combination: Y N W P E . . .
Message 1: r e s e a . . .
Message 2: P r o t e . . .

(3) The combination text can be systematically used to try out a plaintext assumption in every position by a process known as crib dragging. *Crib* is a common synonym for assumption in cryptanalysts. Consider the following two messages in depth. The first message was sent by a unit undergoing an artillery barrage. It is likely that the word *ARTILLERY* will be found in the message.

Message 1: IOZHN EJBTK AKRZE STXVZ GCAVH FJRVX DQNUT
Message 2: UKMWR SDCXM HVOUS OFHUD PICDV BKUPC OEWKK

(4) The first step to trying out *ARTILLERY* in message 1 is to create the combination text. Message 1 is treated as plaintext and message 2 as ciphertext.

Message 1: IOZHN EJBTK AKRZE STXVZ GCAVH FJRVX DQNUT
Message2: UKMWR SDCXM HVOUS OFHUD PICDV BKUPC OEWKK
Combination: MWNPE OUBEC HLXVO WMKZE JGCIO WBDUF LOJQR

(5) The results of trying *ARTILLERY* in each of the first three positions are shown below.

Message 1: IOZHN EJBTK AKRZE STXVZ GCAVH FJRVX DQNUT
Message 2: UKMWR SDCXM HVOUS OFHUD PICDV BKUPC OEWKK
Combination: MWNPE OUBEC HLXVO WMKZE JGCIO WBDUF LOJQR
1: a r t i l l e r y
2: m n g x p z y s c

Combination: MWNPE OUBEC HLXVO WMKZE JGCIO WBDUF LOJQR
1: a r t i l l e r y
2: w e i m r f f v a

Combination: MWNPE OUBEC HLXVO WMKZE JGCIO WBDUF LOJQR
1: a r t i l l e r y
2: n g x w f m i t f

(6) Obviously, not one of the first three tries is the correct placement of ARTILLERY. The process can be speeded up, however, by plotting the crib vertically and the resulting text for message 2 on a descending diagonal.

```
Message 1:    IOZHN EJBTK AKRZE STXVZ GCAVH FJRVX DQNUT
Message 2:    UKMWR SDCXM HVOUS OFHUD PICDV BKUPC OEWKK
Combination:  MWNPE CUBEC HLXVO WMKZE JGCIO WBDUF LOJQR
Crib: a mwn
      r  neg
      t   gix
      i    xm w
      l     P zf
      l       tfm
      e        yfi
      r         svt
      y          ca f
```

(7) The plot above is identical in results to the three separate plots that preceded. Once this format is adopted, it is easier to write in a whole row at a time.

```
Message 1:    IOZHN EJBTK AKRZE STXVZ GCAVH  FJRVX DQNUT
Message 2:    UKMWR SDCXM HVOUS OFHUD PICDV  BKUPC OEWKK
Combination:  MWNPE OUBEC HLXVO WMKZE JGCIO  WBDUF LOJQR
Crib: a mwnpe oubec hl xvo wmkze j gcio   wb
      r  negv flsvt ycomf ndbqv axtzf     nsu
      t   gix hnuxv aeqoh pfdsx czvbh     puwn
      i    xm wcjmk ptfdw eushm rokqw     ej Icn
      l     p tfmpn swigz hxvkp urntz     hmofq w
      l       zfmpn swigz hxvkp urntz     hmofq wz
      e        yf i g lpbzs aqodi nkgms   afhyj psn
      r         svt ycomf ndbqv axtrf     nsulw cfag
      y          ca fjvtm ukixc heagm     uzbsd jmhop
```

(8) The plaintext for message 2 appears on the sixth diagonal, as highlighted above. Once the text is spotted and the crib confirmed, it becomes a matter of depth reading, as before. The worksheet can now be set up and the rest of the text recovered.

```
                  artil lery
Message 1: IOZHN E JBTK AKRZE STXVZ GCAVH FJRVX DQNUT

                  olumn spot
Message 2: UKMWR SDCXM HVOUS OFHUD PICDV BKUPC OEWKK

Key:              ESILZ PGAB
```

(9) With cipher machine and computer based systems that use baud addition, adding two messages in depth together by baud addition eliminates the key. The baud addition of the two ciphertexts is identical to the baud addition of the two original plaintexts.

(10) Whatever type of alphabet square or system of combining bauds is used, there is usually a way to combine texts in depth to eliminate the effects of the key. If you are unsure how to approach a particular type of system, test samples you create for yourself in the system to see how ciphertext can be combined to eliminate the effect of the key.

PART FIVE

Transposition Systems

CHAPTER 11

TYPES OF TRANSPOSITION SYSTEMS

11-1. Nature of Transposition

Transposition systems are fundamentally different from substitution systems. In substitution systems, plaintext values are replaced with other values. In transposition systems, plaintext values are rearranged without otherwise changing them. All the plaintext characters that were present before encipherment are still present after encipherment. Only the order of the text changes.

a. Most transposition systems rearrange text by single letters. It is possible to rearrange complete words or groups of letters rather than single letters, but these approaches are not very secure and have little practical value. Larger groups than smgle letters preserve too much recognizable plaintext.

b. Some transposition systems go through a single transposition process. These are called single transposition. Others go through two distinctly separate transposition processes. These are called double transposition.

c. Most transposition systems use a geometric process. Plaintext is written into a geometric figure, most commonly a rectangle or square, and extracted from the geometric figure by a different path than the way it was entered. When the geometric figure is a rectangle or square, and the plaintext is entered by rows and extracted by columns, it is called columnar transposition. When some route other than rows and columns is used, it is called route transposition.

d. Another category of transposition is grille transposition. There are several types of grilles, but each type uses a mask with cut out holes that is placed over the worksheet. The mask may in turn be rotated or turned over to provide different patterns when placed in different orientations. At each position, the holes lineup with different spaces on the worksheet. After writing plaintext into the holes, the mask is removed and the ciphertext extracted by rows or columns. In some variations, the plaintext may be written in rows or columns and the ciphertext extracted using the grille. These systems may be difficult to identify initially when first encountered, but once the process is recognized, the systems are generally solvable.

e. Transposition systems are easy to identify. Their frequency counts will necessarily look lust like plaintext, since the same letters are still present. There should be no repeats longer than two or three letters, except for the rare longer accidental repeat. The monographic phi will be within plaintext limits, but a digraphic phi should be lower, since repeated digraphs are broken up by transposition. Identifying which type of transposition is used is much more difficult initially, and you may have to try different possibilities until you find the particular method used or take advantage of special situations which can occur.

f. Columnar transposition systems can be exploited when keys are reused with messages of the same length. As will be explained in Chapter 13, the plaintext to messages with reused keys can often be recovered without regard to the actual method of encipherment. Once the plaintext is recovered, the method can be reconstructed.

11-1. Examples of Columnar Transposition

The most common type of transposition is columnar transposition. It is the easiest to train and use consistently.

a. Simple Columnar **Transposition.** At its simplest, columnar transposition enters the plaintext into a rectangle of a predetermined width and extracts ciphertext by columns from left to right. For example, a simple columnar transposition with a width of seven is shown below.

Plaintext: ENEMY TANKS APPROACHING HILL EIGHT SIX THREE STOP

E	N	E	M	Y	T	A
N	K	S	A	P	P	R
O	A	C	H	I	N	G
H	I	L	L	E	I	G
H	T	S	I	X	T	H
R	E	E	S	T	O	P

Ciphertext:

ENOHH RNKAI TEESC LSEMA HLISY PIEXT TPNIT OARGC HPXXX

(1) The cryptographer receiving the above message knows only that a width of 7 was originally used. The cryptographer rebuilds the matrix by determining the length of each column and writing the ciphertext back into the columns. With a width of 7 and a length of 42, each column must have 6 letters. Inscribing the ciphertext into columns from left to right recreates the original matrix, and the plaintext can be read by rows.

(2) Not all messages will come out even on the bottom row. Here is the same message with STOP omitted. The columns are not all the same length. In this case, the matrix is called an incompletely filled matrix.

E	N	E	M	Y	T	A
N	K	S	A	P	P	R
O	A	C	H	I	N	G
H	I	L	L	E	I	G
H	T	S	I	X	T	H
R	E	E				

Ciphertext:

ENOHH RNKAI TEESC LSEMA HLIYP IEXTP NITAR CCHXX

(3) The deciphering cryptographer must now perform the additional step of determining which columns will be longer than the others. With 38 letters and a given width of 7, dividing 38 by 7 produces 5 with a remainder of 3. This means that the basic column length is 5, but the first 3 columns are 1 letter longer. Sometimes, cryptographers will avoid this additional step by padding message texts so that the bottom row is always completely filled.

(4) The solution of these systems is extremely easy. The security depends on just one number, the matrix width. All you have to do to solve a message enciphered by simple columnar transposition is to try different matrix widths until you find the right one. To try each width, you just do exactly what the deciphering cryptographer does. Divide the total length by the trial width and the result and remainder will tell you the basic column length and how many longer columns there are.

(5) If you suspect that only completely filled matrices are being used, the solution is easier. You only need to test widths that evenly divide into the message length in that case. For example, with a length of 56, you would try widths of 7 and 8. If neither of these worked, you would also try 4, 14, 2, and 28 to cover all possibilities. It is better to try the possibilities closest to a perfect square before you try very tall and very wide matrices.

b. Numerically-Keyed Columnar Transposition. Numerically-keyed transposition systems are considerably more secure than simple columnar transposition. You cannot exhaust all possibilities with just a few tries as you can with the simple systems. The transposition process is similar to that used to produce transposition mixed sequences.

(1) The numerical key is commonly based on a keyword or key phrase. Unlike keywords used to produce mixed sequences, the keyword may have repeated letters in it. To produce a numerical key from a keyword with repeated letters, the repeated letters are numbered from left to right.

```
1 2 6 4 8 3 7 5
A A R D V A R K

  1     I   1 1
2 9 1 4 0 8 6 1 2 3 3 7 5
T R A N S P O S I T I O N
```

(2) As with simple columnar transposition, matrices may be completely filled or incompletely filled. In either case, the plaintext is written horizontally and the ciphertext is extracted by column in the order determined by the numerical key, The following example shows an incompletely filled matrix.

```
5 6 1 4 3 2
O R A N G E
```

R	E	Q	U	E	S
T	R	E	I	N	F
O	R	C	E	M	E
N	T	S	I	M	M
E	D	I	A	T	E
L	Y				

Ciphertext:

QECSI SFEME ENMMT UIEIA RTONE LERRT DYXXX

(3) The decipherment process for the receiving cryptographer is more complicated than with simple columnar transposition. The cryptographer must decide the column lengths, as before. With the above message, the cryptographer divides the length of the message by the length of the numerical key. In this case, 32 divided by 6 is 5 with a remainder of 2. The basic column length is 5 with two longer columns at the left. The cryptographer then sets up a matrix with the key at the top and marks the column lengths.

```
5 6 1 4 3 2
O R A N G E
```

·	·	·	·	·	·
·	·	·	·	·	·
·	·	·	·	·	·
·	·	·	·	·	·
·	·	·	·	·	·
·	·				

(4) The ciphertext is now entered by columns according to the numerical key to produce the plaintext.

(5) The solution of numerically-keyed systems is more complex than for simple columnar transposition. It is more than just trying all possibilities. The solution of numerically-keyed columnar transposition is explained in Chapter 12.

11-3. Route Transposition

There are many other ways to transpose messages than columnar transposition using squares and rectangles. The shape of the geometric figure used can be varied, and the method of inscribing and extracting text can be varied. Columnar methods are the most common in military usage, because they are the easiest to learn and use reliably, but other methods may be encountered. Some of these common methods are shown below.

a. Route transposition using other geometric figures.

(1) The rail-fence cipher is inscribed by zigzag pattern and extracted by rows.

			N					M						R					G			
		I		F				E		E				A		R			N		N	
	E				O		C				N		S				I		I				O	
R						R						T						V					W

Ciphertext: NMRGI FEEAR NNEOC NSIIO RRTVW

(2) The triangular pattern is inscribed by rows and extracted by columns.

				R				
			E	I	N			
		F	O	R	C	E		
	M	E	N	T	S	A	R	
R	I	V	I	N	G	N	O	W

Ciphertext:

RMIFE VEONI **RIRTN** NCSGE **ANROW**

b. The next examples show just some of the possibilities for route transposition using squares or rectangles. Each example is based on REINFORCEMENTS *ARRIVING* NOW to help you see how the route was entered. The route can be:

(1) Inscribed by spiral, out by columns.

R	E	I	N	F
R	R	I	V	O
A	O	W	I	R
S	N	G	N	C
T	N	E	M	E

Ciphertext:

RRAST ERONN IIWGE NVINM FORCE

(2) Inscribed by diagonals, out by alternating rows.

R	I	O	M	A
E	F	E	S	V
N	C	T	I	G
R	N	R	N	O
E	R	I	N	W

Cipher-text:

RIOMA VSEFE NCTIG ONRNR ERINW

(3) In by outward spiral, out by alternating diagonals.

N	G	N	O	W
I	R	C	E	M
V	O	R	E	E
I	F	N	I	N
R	R	A	S	T

Ciphertext:

NIGNR VIOCO WERFR RNEME IASNT

(4) In by L-pattern, out by spiral from lower right.

R	R	R	O	W
E	A	I	N	G
I	S	V	I	N
N	T	N	E	M
F	O	R	C	E

Ciphertext:

ECROF NIERR ROWGN MENTS AINIV

c. Completely filled squares or rectangles are more common with route transposition than with columnar transposition. The reason is that it is often difficult for the cryptographers to figure out how to handle an incompletely filled matrix. It is simpler in practice to completely fill each matrix than to provide rules to cover every incompletely filled situation.

d. The solution of route transposition is largely a matter of trial and error. When you suspect route transposition, see if the message length is a perfect square or if the matrix can be set up as a completely filled rectangle. Then try entering the ciphertext by different routes, and look for visible plaintext by another route.

CHAPTER 12

SOLUTION OF NUMERICALLY-KEYED COLUMNAR TRANSPOSITION CIPHERS

12-1. Completely Filled Matrices - Determining Matrix Size

When completely filled matrices are known or suspected, the first step in their solution is to determine the matrix size. As discussed in Chapter 11 for simple columnar transposition, the correct width must be an even divisor of the message length. With simple columnar transposition, the correct width could be confirmed easily, because plaintext will appear on the rows when the width is correctly selected. It is not as simple with numerically-keyed transposition, Although each row will contain the letters of plaintext for that row when the width is correctly selected, the letters will be out of order. The key to recognition is the vowel count on each row. Vowels should appear consistently with fairly even counts on each row when the correct width is tried. In plaintext, vowels appear about 40 percent of the time even in small samples of text. This is necessary for text to be pronounceable. If some of the rows have too many or too few vowels, you probably have the wrong width. Consider the next cryptogram.

ERESO RIERU GRFPT TEOAE OOSNN MNIEU SDEES MTSUR FYSBW TEARC EUXRQ GXXXX

a. The cryptogram has 56 letters, assuming the final Xs are all nulls. If a completely filled matrix is suggested by past experience, then the matrix is probably either 7 or 8 letters wide. Write the cryptogram by columns into a trial matrix of each width and count the vowels in each row.

E	R	E	N	E	F	R	3
R	U	O	M	E	Y	C	3
E	G	A	N	S	S	E	3
S	R	E	I	M	B	U	3
O	F	O	U	T	W	X	3
R	P	O	E	S	T	R	2
I	T	S	S	U	E	Q	3
E	T	N	D	R	A	G	2

E	E	T	O	U	M	S	C	4
R	R	T	S	E	T	B	E	2
E	U	E	N	S	S	W	U	4
S	G	O	N	D	U	T	X	2
O	R	A	M	E	R	E	R	4
R	F	E	N	E	F	A	Q	3
I	P	O	I	S	Y	R	G	3

b. The first matrix, with a width of seven letters, has the more regular spacing of vowels. The letter Q in the first matrix also has a U on the same row, whereas the second matrix does not. The first matrix is clearly the better possibility.

12-2. Matrix Reconstruction by Anagramming

Continuing the same problem, the object now is to rearrange the columns into the original order. The rearrangement of letters to find the original plaintext order is called anagramming. You may be able to see possibilities for complete words on some of the rows, but the Q and the U on the seventh row provide the most obvious starting point. To recover the numerical key at the same time, number the columns in numerical order before starting reconstruction.

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

7	5			
R	E			
C	E			
E	S			
U	M			
X	T			
R	S			
Q	U			
G	R			

a. All the letter combinations produced by placing columns 7 and 5 together look reasonable for plaintext. At this point, you can see that the last two rows should

both be followed by vowels. Both the 1 and 6 columns end with two vowels, Here is what each looks like when added to the initial two columns.

7	5	1
R	E	E
C	E	R
E	S	E
U	M	S
X	T	O
R	S	R
Q	U	I
G	R	E

7	5	6
R	E	F
C	E	Y
E	S	S
U	M	B
X	T	W
R	S	T
Q	U	E
G	R	A

b. Both possibilities give good plaintext letter combinations, but at this point, several words are suggested in the second match. REF.. ..CE could be part of REFERENCE. XTW could be part of SIX TWO, and the UMB in that case would suggest NUMBER. With these probable words, clearly column 3 follows 756. Column 7 is the left-hand column, because the letters needed for REFERENCE, SIX, and NUMBER are on the row above in column 4. Adding columns 3 and 4 produces the next matrix.

7	5	6	3		4	
R	E	F	E	R	E	N
C	E	Y	O			M
E	S	S	A			N
U	M	B	E	R	S	I
X	T	W	O			U
R	S	T	O			E
Q	U	E	S			S
G	R	A	N			D

c. The remaining two columns are easily filled in to complete the solution.

7	5	6	3	2	1	4
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	S	I
X	T	W	O	F	O	U
R	S	T	O	P	R	E
Q	U	E	S	T	I	S
G	R	A	N	T	E	D

both be followed by vowels. Both the 1 and 6 columns end with two vowels. Here is what each looks like when added to the initial two columns.

7	5	1
R	E	E
C	E	R
E	S	E
U	M	S
X	T	O
R	S	R
Q	U	I
G	R	E

7	5	6
R	E	F
C	E	Y
E	S	S
U	M	B
X	T	W
R	S	T
Q	U	E
G	R	A

b. Both possibilities give good plaintext letter combinations, but at this point, several words are suggested in the second match. REF.. ..CE could be part of REFERENCE. XTW could be part of SIX TWO, and the UMB in that case would suggest NUMBER. With these probable words, clearly column 3 follows 756. Column 7 is the left-hand column, because the letters needed for *REFERENCE,* SIX, and NUMBER are on the row above in column 4. Adding columns 3 and 4 produces the next matrix.

7	5	6	3			4
R	E	F	E	R	E	N
C	E	Y	O			M
E	S	S	A			N
U	M	B	E	R	S	I
X	T	W	O			U
R	S	T	O			E
Q	U	E	S			S
G	R	A	N			D

c. The remaining two columns are easily filled in to complete the solution.

7	5	6	3	2	1	4
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	S	I
X	T	W	O	F	O	U
R	S	T	O	P	R	E
Q	U	E	S	T	I	S
G	R	A	N	T	E	D

12-3. Incompletely Filled Matrices - Hat Diagrams

Incompletely filled matrices are also solved by anagramming, but it is a more difficult process because you cannot initially tell which letters are on the same row with each other. If you know or can correctly assume the width of the matrix, you can limit the possibilities. Consider the next cryptogram, which is expected to have a matrix width of eight letters.

EARTR RQ-IRE TALOA OXUWA UETNE IOTAE ROCTT EROTE EAOSN GHNRD

SEDOO TELHT COEAI TONQR DIMSF EXXXX

a. With a length of 76 letters and a suspected width of 8, there must be four columns with 10 letters and four columns with 9 letters. We can show the range of letters that could be placed in each column by trying the first four columns as the longer columns and alternately, the last four columns as the long columns. The true arrangement is probably neither, but it will serve to show the possible range of first and last letters for each column.

E	T	U	R	E	D	H	N
A	A	E	O	A	S	T	Q
R	L	T	C	O	E	C	R
T	O	N	T	S	D	O	D
R	A	E	T	N	O	E	I
R	O	I	E	G	O	A	M
G	X	O	R	H	T	I	S
H	U	T	O	N	E	T	F
R	W	A	T	R	L	O	E
E	A	E	E				

E	E	W	T	R	H	E	O
A	T	A	A	O	N	L	N
R	A	U	E	T	R	H	Q
T	L	E	R	E	D	T	R
R	O	T	O	E	S	C	D
R	A	N	C	A	E	O	I
G	O	E	T	O	D	E	M
H	X	I	T	S	O	A	S
R	U	O	E	N	O	I	F
				G	T	T	E

b. These two extreme situations can be combined into a single diagram, called a hat diagram, It is constructed by using the first diagram. Next, combine the letters that the second diagram shows can precede the already listed letters by adding them to the top of the first diagram. Similarly, draw a line across the bottom of the first diagram to show the possible bottom letters. The altered first matrix is now the completed hat diagram.

1	2	3	4	5	6	7	8
				R			
			T	O	H		
		W	A	T	N	E	
	E	A	E	E	R	L	O
E	T	U	R	E	D	H	N
A	A	E	O	A	S	T	Q
R	L	T	C	O	E	C	R
T	O	N	T	S	D	O	D
R	A	E	T	N	O	E	I
R	O	I	E	G	O	A	M
G	X	O	R	H	T	I	S
H	U	T	O	N	E	T	F
R	W	A	T	R	L	O	E
E	A	E	E				

c. The completed hat diagram can now be used as a guide to show how columns may be aligned together. Its value can be seen if you try to place the Q in the text before a U. There are two Us in the cryptogram. The Q is necessarily near the top of the matrix. The U in column 2 can only be at the bottom of the matrix. The U in column 3 can only be at or near the top of the matrix. The correct U to place with the Q is now obvious. Lining up the Q in column 8 with the U from column 3 produces an initial reconstruction.

8	3	
O	W	
N	A	
Q	U	
R	E	
D	T	
I	N	
M	E	
S	I	
F	O	
E	T	

d. Next, there is an X near the bottom of the matrix in column 2. It will combine well with the SI of the first two columns to produce SIX.

8	3	2
O	W	E
N	A	T
Q	U	A
R	E	L
D	T	O
I	N	A
M	E	O
S	I	X
F	O	U
E	T	W

e. SIX is not the only number near the bottom of the matrix. FOUR and TWO are likely on the last two rows, and column 4 is available with RO near the bottom.

8	3	2	4
O	W	E	A
N	A	T	E
Q	U	A	R
R	E	L	O
D	T	O	C
I	N	A	T
M	E	O	T
S	I	X	E
F	O	U	R
E	T	W	O

f. The E after SIX suggests EIGHT. The numbers themselves suggest the word *COORDINATES*, which appears in the middle of the matrix. With these words written in, the rest of the columns can be placed.

8	3	2	4	7	5	1	6
O	W	E	A	L	T	E	R
N	A	T	E	H	E	A	D
Q	U	A	R	T	E	R	S
R	E	L	O	C	A	T	E
D	T	O	C	O	O	R	D
I	N	A	T	E	S	R	O
M	E	O	T	A	N	G	O
S	I	X	E	I	G	H	T
F	O	U	R	T	H	R	E
E	T	W	O	O	N	E	L

g. All letters are now used, but several letters appear at both the top and bottom of the matrix. The first word of the message is *ALTERNATE,* and the letters before it all appear correctly at the bottom of columns. The L at the bottom after ONE correctly appears as part of ALTERNATE at the top. Removing the duplicated letters and shifting *ALTERNATE* to begin at the left-hand column completes the solution.

4	7	5	1	6	8	3	2
A	L	T	E	R	N	A	T
E	H	E	A	D	Q	U	A
R	T	E	R	S	R	E	L
O	C	A	T	E	D	T	O
C	O	O	R	D	I	N	A
T	E	S	R	O	M	E	O
T	A	N	G	O	S	I	X
E	I	C	H	T	F	O	U
R	T	H	R	E	E	T	W
O	O	N	E				

h. This solution depended on correctly identifying the width of the matrix and the fortunate appearance of the Q and U. Without the Q and U and without any indication of the width, a great deal more trial and error would be required for a successful solution. Hat diagrams can be constructed for different possible widths, for example, and probable words searched for within the structure of the diagram. The solution is still possible in most cases, although it will often take longer than the example did. When the same keys are reused for a period, special situations can arise which make the solution much easier. The next chapter shows the techniques that can be used in these special situations.

CHAPTER 13

TRANSPOSITION SPECIAL SOLUTIONS

13-1. Special Exploitable Situations

Military forces are rarely equipped to change cryptosystem keys with every message transmitted. The logistics and management problems of providing enough different keys and controlling their use are difficult to handle. Normally, keys will be reused for a period before they are changed. With transposition systems, several special situations can arise when keys are reused that make a solution possible when the system might otherwise resist successful analysis. One of these situations arises in columnar transposition whenever similar beginnings and endings are used with the same width matrix. The keys do not have to be the same in this case as long as the width is the same. Another more general situation occurs whenever two or more different messages of the same length occur using exactly the same keys. Each of these situations is explained in the following paragraphs.

13-2. Similar Beginnings and Endings

With columnar transposition, repeated message beginnings or endings can cause an easily recognizable and exploitable situation. When the same width keys are used and the beginnings are the same, the tops of the columns in each message will consist of the same letters. When the length of the repeated beginning is several times as long as the width of the matrix, these repeated letters are easy to spot.

a. The next two messages demonstrate the techniques that can be used when similar beginnings are encountered. Repeated segments between the two messages are underlined.

Message 1:

```
ASOLI LBOAE WDLIR ACIEL NSAIR  IEDLS NDWND TQNIH UAOTL FMLIF
1             2             3             4             5
AMPES DBREU SCEPV NELOM YEODC  SHCA! TIELT MNAEE IDERA
          6             7             8
```

Message 2:

```
QNILB TSROI RRIEP LIHUE OZYAS  OLSUT ARZEO LTMUI MTQBR OAUSC
1             2             3             4             5
IEEHT RXOLI RSWBO DSERD EODPL  TIAFS EIFAE SDEEE ZT
          6             7             8
```

(1) There are eight repeated segments in each, which shows that the messages are each eight columns wide. The repeated segments are not in the same order, which shows that the two messages use different numerical keys.

(2) Message 1 has 95 letters. Dividing 8 into 95 gives 11 with a remainder of 7. This means that all but one column must have 12 letters. The distance between repeats shows that this is true. All segments have 12 letters except for the fifth segment, which has 11 letters. The fifth segment, beginning IFA, must be the right-hand column of the matrix.

(3) Message 2 has 92 letters. Four columns have 12 letters and four columns have 11 letters.

(4) All repeated segments contain three letters except for the ASOL segment. The column beginnmg ASOL is probably the left-hand column.

(5) As a result of these observations, we can place the first and last columns in each matrix, and we can separate the middle six columns into two groups of three, based on the length of the columns in message 2.

Message 1:

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

Message 2:

3	246	157	8
A	R L L	Q U E	I
S	I T I	N S O	F
O	E M R	I C D	A
L	P U S	L I P	E
S	L I W	B E L	S
U	I M B	T E T	D
T	H T O	S H I	E
A	U Q D	R T A	E
R	E B S	O R F	E
Z	O R E	I X S	Z
E	Z O R	R O E	T
O	Y A D		

(6) Completion of the solution from here is straightforward. Anagram each group of three columns in each message, and the solution is complete. The similar beginningis *ALL REQUISITIONS FOR MEDICAL.*

b. Messages with similar endings, such as a repeated signature block, show repeated segments which represent the bottoms of columns instead of the top. The solution is approached the same way, except that the text will not necessarily appear in the same columns in both messages.

13-3. Messages With the Same Length and Keys

Whenever two or more messages have the same length and are transposed with the same keys, they can be solved together. The more messages you find that are the same length and use the same keys, the easier they are to solve. This technique can be used regardless of the type of transposition system.

a. Solving messages with the same length and keys is particularly effective with columnar transposition. The next example shows how the solution can be approached. The five messages all use the same keys. Their positions have been numbered for easy reference and to aid in key recovery.

	12345	67890	1 1 1 1 1 12345	11112 67890	22222 12345
Message 1:	L P Q R Y	T T L P U	A R R S I	U E D E O	E T S R E
Message 2:	Q S N E T	B B U H B	H R S M D	R E D A A	O A E E E
Message 3:	A O E E W	O V C U C	M T N I S	F R D E R	E S O T E
Message 4:	I O O O E	O D N R N	N N P O H	T T Y C E	T T W R A
Message 5:	J N U O T	E K U F R	R C V A D	O O N N I	T A I F E

(1) The Q in message 2 in position 1 must certainly be followed by the U in position 8.

(2) Position 1 must be at the top of a column in the original matrix, since columns are extracted beginning at the top. Position 8 is also probably at the top of a column. This applies not just to message 2, but to all five messages. The position 1 column can be written next to position 8.

1	8
L	L
Q	U
A	C
I	N
J	U

(3) Position 2 must be from the second row of the matrix. If position 8 is from the top row, then position 9 must be from the second row, also. Similarly, positions

3 and 10 are from the third row. Positions 4 and 11 are from the fourth row. Positions 5 and 12 are probably from the fifth row, although these are short messages and there might not be as many as five rows.

	18	29	30	41	52
Message 1:	L L	P P	Q U	R A	Y R
Message 2:	Q U	S H	N B	E H	T R
Message 3:	A G	O U	E C	E M	W T
Message 4:	I N	O R	O N	O N	E N
Message 5:	J U	N F	U R	O R	T C

(4) Now the task is to find additional columns to add to the fragments already started. For example, the QU in message 2 should be followed by a vowel, and the most likely letter after JU in message 5 is N. There are three columns with an N in message 5, and only one of these, position 19, has a vowel in message 2. Therefore, we will add columns 19, 20, 21, 22, and 23 to our fragments.

	189	290	301	412	523
Message 1:	L L E	P P O	Q U E	R A T	Y R S
Message 2:	Q U A	S H A	N B O	E H A	T R E
Message 3:	A G E	O U R	E C E	E M S	W T O
Message 4:	I N G	O R E	O N T	O N T	E N W
Message 5:	J U N	N F I	U R T	O R A	T C !

(5) All of the fragments produce good plaintext except, possibly, the last one. QUA will usually be followed by an R. Of the two columns with an R in message 2, column 12 provides the best combinations.

	1892	2903	3014	4125	5236
Message 1:	L L E R	P P O R	Q U E S	R A T I	Y R S U
Message 2:	Q U A R	S H A S	N B O M	E H A D	T R E R
Message 3:	A G E T	O U R N	E C E I	E M S S	W T O F
Message 4:	I N G N	O R E P	O N T O	O N T H	E N W T
Message 5:	J U N C	N F I V	U R T A	O R A D	T C I O

(6) All of the matches give good plaintext, except the fifth set, which clearly does not belong now. It is easy now to see words to build on, such as *ARTILLERY*, QUARTERS or HEADQUARTERS, JUNCTION, SUPPORT, *FIVE,* and others. All of these leads are added to the completely anagrammed messages.

	1	2	1			1	1		2	1			2	1		2	1		1	2	1		2	1
	1	4	2	5	1	8	9	2	5	3	6	2	9	0	3	6	4	7	3	0	1	4	7	5	8
Message 1:	A	R	T	I	L	L	E	R	Y	S	U	P	P	O	R	T	R	E	Q	U	E	S	T	E	D
Message 2:	H	E	A	D	Q	U	A	R	T	E	R	S	H	A	S	B	E	E	N	B	O	M	B	E	D
Message 3:	M	E	S	S	A	G	E	T	W	O	F	O	U	R	N	O	T	R	E	C	E	I	V	E	D
Message 4:	N	O	T	H	I	N	G	N	E	W	T	O	R	E	P	O	R	T	O	N	T	O	D	A	Y
Message 5:	R	O	A	D	J	U	N	C	T	I	O	N	F	I	V	E	F	O	U	R	T	A	K	E	N

(7) The final step in the solution is to recover the numerical keys. Looking at the beginning, the pattern starts to repeat after seven letters, so the original matrix was seven letters wide. The numerical key, derivable by observing the order in which the columns were extracted, was 4275136.

b. The technique of solving messages of the same length and keys can be used with any transposition system. It can be used as the basis for recovery of more difficult transposition systems such as large grilles and double transposition. The cyclic attern of columnar transposition aided the solution of the example above. Given four or more messages of the same length and keys, however, the complete messages can often be anagrammed without the help of the cyclic pattern.

PART SIX

Analysis of Code Systems

CHAPTER 14

TYPES OF CODE SYSTEMS

14-1. The Nature of Code Systems

As explained in Chapter 1, the key feature that distinguishes a code from a substitution cipher is that a code will substitute for words as well as characters.

a. Codes range in size from small charts or lists on a single sheet of paper to books as large as an unabridged dictionary.

b. Plaintext values are replaced by code groups or code words. A code group or word may replace anything from a single character to a whole sentence.

c. Since codes can compress whole sentences into a small code group, not all codes are used for security purposes. Some are used for economy instead, by replacing common sentences and phrases with a single group. For example, radio operators use Q and Z signals as a brevity code. Q and Z signals are three letter code groups beginning with Q or Z that stand for common communications procedures, A single code Q or Z signal replaces sentences or phrases such as QSA, My signal strength *is* ... and ZNN, I have *nothing* now. Operators memorize the Q and Z signals that they commonly use and the result is quicker, more economical communications.

d. Some codes are used for prearranged messages only. Limited in size and purpose, a single code group may be transmitted as a signal to begin a preplanned attack, for example. Prearranged message codes are sometimes referred to as pamcodes. Prearranged message codes may also take the form of innocent communications, so that an apparently harmless message contains a secret meaning. The message, *Les sanglots longs des violons de l'automne,* a harmless sentence in French, signaled the French underground in World War II that the Allied invasion of France was to begin soon. Codes with an innocent appearance but a secret meaning are known as open codes.

e. Prearranged message codes can only be used for limited, preplanned purposes. General purpose codes which can be used for any communications are more common. All general purpose codes must include within them, a provision for spelling words that are not included in their vocabulary. Even when very large book codes are used, proper names will sometimes need to be encoded that are not in the code's vocabulary. General purpose codes thus share some of the characteristics of substitution ciphers.

f. Codes are at their weakest when they are used to spell words. Most codes are broken into through spelling. Large codes attempt to defeat this weakness by providing many variants for letters and common syllables. The letter E might be encoded by 10 different code groups in a large code, for example. Other code groups would represent common syllables with E in them like RE, ER, EN, and ENT. In this respect, codes are similar to syllabary squares, and the initial approach to analysis can be similar between syllabary squares and codes.

g. When a high degree of security is required using codes, there are two approaches to increasing the security of codes. One is to use very large book codes, since the larger the code, the more secure it is. The other is to further encipher the code to produce an enciphered code. Any of the cipher procedures discussed earlier in this manual can be used, but the most common is to use polyalphabetic encipherment. Repeating keys and long-running keys may be used. It is one way to combine the advantages of brevity with the added security of polyalphabetics, although such procedures are time-consuming to use. They cannot be used practically in rapidly changing combat situations, for example, when speed of communications is important. Large codes and enciphered codes were common earlier in this century when a high degree of security was desired. Today, with advances in electronics, cipher machine and computer based systems are more common when a high degree of security is required.

14-2. Book Codes

Codes too large to be printed on just one or two pages are called book codes. They may range from small pamphlets to large bound books.

a. The code values in book codes may consist of letters, numbers, or a combination of letters and numbers. Usually, the code groups are a constant length, but there are occasional exceptions. Code values used primarily for voice communications will sometimes consist of pronounceable words rather than regular length groupings of characters. We will refer to only code groups in the rest of this chapter and the next, but you should understand that comments about code groups also apply to code words.

b. The simplest book codes consist of a single orderly listing of code groups and their meanings. The code groups are listed in the book in alphabetical or numerical order, and their meanings are also in a logical order. This single listing is used for encoding and decoding, and is called a one-part code. The plaintext values may be strictly alphabetical in arrangement or may be separated into separate sections for words, letters and syllables, and numbers. Occasionally, they will be arranged topically with such things as units in one section, weapons systems in another, place-names in another, and so on. The key feature of one-part codes is that when the code groups are listed in order, their plaintext meanings will also be in a logical order. A sample portion of a one-part code is shown below.

CODE GROUP:	PLAINTEXT:
AA0	A
ABD	A 0
ACF	ABANDON
ADH	ABOUT
AEJ	ACC I DENT
AFL	ACTION
AGN	ACTIVE
AHP	ACTIVITY
...	...
...	...

c. The orderly structure of one-part codes makes them easy to use, but greatly reduces their security. The analyst can use the structure to narrow down possible meanings for code groups. More secure codes are randomly arranged, and are necessarily printed in two parts. One section lists the code groups in order, and it is used for decoding. The other section, containing exactly the same information, lists the plaintext values in order, and is used for encoding. This type of code is called a two-part code. Portions of the encoding and decoding sections of a two-part code are shown below. Note that one group occurs in common between the two parts.

ENCODING SECTION:		DECODING SECTION:	
KTOL	A	ABAB	RES I STANCE
YNIF	A	ABEC	SIZE
ACEJ	AB	ABID	CHEMICAL
VAUW	ABANDON ING S	ABOF	T-72
WHOD	ABILITY	ABUG	QUALITY
AOUT	ABLE	ACAH	15
LWOQ	ABLE TO	ACEJ	AB
TEER	ABOUT	ACIK	VERIFY ING S
...	...	...	...
...	...	...	...

14-3. Matrix Codes and Code Charts

Small codes can be conveniently printed in the form of a small coordinate matrix system.

a. Typically 10 by 10 or larger, matrix codes, also known as code charts, can contain letters, syllables, numbers, and a small vocabulary of words. They are very easy to

use, and communicators can be trained in their use quickly and easily. They also offer more security than most simple ciphers.

b. Code charts are easily changed from one cryptoperiod to the next by simply changing the coordinates, while retaining the same matrix.

c. They are a very close relative to the syllabary square cipher. If the syllabary square shown in Chapter 5 contained some words as well as letters, syllables, and numbers, it would be a code instead of a cipher.

d. One type of code chart places two plaintext values in each cell-an upper value and a lower value. The lower values are all words. The upper values are all numbers, letters, or syllables. Two of the cells are set aside as shift values to indicate whether to read the upper values or lower values in the code groups that follow. A sample chart of this type is shown in Figure 14-l. This example uses letters for coordinates, and has variants on each row and column. The word ARTILLERY, for example, could be encoded as TF, TI, QF, or QI. The cells MU and UU are begin and end spell indicators. The bottom values in each cell are used until a begin spell group is sent. Then the top values are used until the end spell group is used to shift back to the lower values.

	C,D	E,H	F,I	J,K	T,L	M,O	U,V	Y,G	Z,N	P,Q	X,R	W,S	B,A
M,H	$$$ Action, ive, ivity, s	$2 Addition, al	15 Advance, d, ing, s	45 After	A Aggressor, ive (ly), s	AD Air	Spell/fig. Begins	AL Airborne	AM Aircraft/ Airplane, s	AN Ammunition	AND Antiaircraft	AR Antitank	ARE Area (of)
T,Q	$5 Armor, ed	$3 Arrive, al, d, ins, s	16 Artillery	5$ Assemble, d, ins, s	AS Attack, ed, ing, s	AT Attempt, ed, ing, s	B Azimuth (in degrees)	BA Battalion, s	BE Battery, ies	BY Begin/start, ed, ings, s	C Bomb, ed, er, ing, s	CA Bridge, d, ing, s	CAN Capture, d, ing, s
K,Z	$ $4 Casualty, ies,	- - Command er, ing, s	17 Communicate, d, ing, ion, s	55 Company, ies	CE Complete, d, ing, ion, s	CH Concentrate, d, ing, ion, s	CO Contact, ed, ing, s	D Coordinate, d, ing, ion, s	DA Corps	DAY Counterattack, ed, ing, s	DE Crew, ed, es, ing	DI Defend/de- fense, s (of)	DO Delay, ed, ing, s
O,L	1 Destroy, ed, ing, s	$5 Detach, ed, ment (of), s	18 Dispose, al, d, ition, s	E Division, s	EA Dump, s	ED East (of)	EE Encounter, ed, ing, s	EN Enemy's	ENT Engineer,	ER Enlisted Man/Men	ERS Equip, ment, ped, ping	ES Escape, d, ing, s	EST Estimate, d, ing, s (at)
R,X	2 Expect, ed, ing, s (at)	$6 Fight, er, ing, s	19 Fire, d, ing, s	ET Flank, s	F Force, d, ing, s	FO Forward	FOR Friend, ly	G From	H Front, al, s	HA Fuel, s	HE Gun, s	I Has/have	IL Headquarters
S,P	3 Heavy, ily	$7 Hill, s (No.)	2$ Hold, ing, s/held	IN Hostile, ity, ities	ING Hour, s	ION How	IS Identify, ied, ies, ing, ication	IT Immediate, ly	IVE Infantry	J Inform, ation, ed, ing, s	K Install, ation, ed, ing, s	L Junction, s (of)	LA Land, ed, ing, s
W,N	4 Large	$8 Left (of)	21 Line, s (of)	LE Locate, d, ing, ion, s	LI Machine gun, s (nest)	LO Main	LY Map, ped, ping, s	M Mechanize, d	MA Message, nger, s	ME Mile, s (from), (to)	MENT Mine, d, ing, s	MI Mission, s	MY Morning
A,B	5 Mortar, s	$9 Move, d, ing, ment, s	22 Near	N Night	NA No/not/no- thing/negat	ND North (of)	NE Number, s, (of)	NI Objective, s	NO Observe, ation, d, ing, s	NOT Occupy, ied, ies, ing	NT Officer, s	O Operate, d, ing, ion, s	OF Order, ed, ing, s
C,E	6 Over	1$ Patrol, led, ling, s	23 Penetrate, d, ing, ion, s	ON Plan, ned, ning, s (to)	OR Platoon, s	OU Point, ed, ins, s	OUR Position, s	P Post, ed, ing, s	PE Prepare, d, ation, ing, s	Q Prisoner, s	QU Proceed, ed, ing, s, ure	R Radio, ed, s	RA Railway/ Railroad, s
I,G	7 Ready (for) (to)	11 Rear	25 Receive, d, ing, s/receipt	RE Reconnais- sance	RED Refer, ence, red, ring, s (to)	RES Regiment, al, s	RI Reinforce, d, ing, ment, s	RO Replace, d, ing, ment, s	RS Report, ed, ing, s	RT Request, ed, ing, s	S Require, d, ing, isition, s	SA Reserve, d, ing, s	SE Ridge, s
D,J	8 Right (of)	12 River/ Stream	3$ Road, s/ Route, s	SH Scout, ing, s	SI Section, s/ Sector, s	SO Send, ing, s/sent (to)	ST Shell, ed, ing, s	T Small/ Small arms	TA South (of)	TE Squad, s	TED Strength, s (of)/strong	TER Stop, ped, ping, s	TH Supply, ies (of)
F,V	9 Support, ed, ing, s	13 Tank, s	35 Target, s	TI Today	TION Tomorrow	TO Tonight	TR Troop, s	U Truck, s/ Vehicle, s	UN Unit, s (of)	US Until	V Urgent, cy, ly	W Vicinity (of)	WE Water
U,Y	$1 West (of)	14 What/who	4$ When	X Where	Y will	Z With	Spell/fig. Ends	Period . Withdraw, al, ing, s	Comma , Woods	Colon : Yard, s (from), (to)	Smicln ; Yesterday	Dash -- You, r	Paren ( ) Zone, s (of)

Figure 14-1. Sample code chart.

CHAPTER 15

ANALYSIS OF SYLLABARY SPELLING

15-1. Identification of Syllabary Spelling

The key to breaking into codes and syllabary ciphers is to identify and exploit syllabary spelling. If possible, try to locate instances where the same word is spelled in different ways by combining the syllables and letters in different combinations each time. This situation can be exploited fairly easily.

a. Identifying repeated syllabary spelling in syllabary squares was demonstrated in Chapter 5.

b. In codes, only certain groups represent letters and syllables, but these tend to cluster together. With code charts, if begin spell or letter shift groups are used, identifying these special purpose groups serves to point right to groups used for spelling. Often begin spell-end spell groups or letter shift-word shift groups are the highest frequency groups and tend to alternate in the text. This makes them quite easy to spot.

c. In codes where no shift groups are used, the code groups that represent letters and syllables tend to cluster together, just as code groups that represent numbers do. If necessary, computer produced indexes of code groups and the code groups they appear with will help to isolate those used for spelling.

15-2. Recovery of Syllabary Spelling

By comparing different spellings of the same word, you can often figure out which groups represent single letters and which represents syllables. Then, the groups which represent syllables can be replaced by groups that represent single letters. Reduction to single letter terms, in turn, enables recognition of word patterns. This approach to

recovery of syllabary spelling applies equally to syllabary squares, code charts, and book codes. The segments below, each of which represents the same plaintext, illustrates how spelling can be recovered.

A:	**81**	**35**	**25**	**74**	**60**	**60**	**11**	**54**	**88**	**88**	**14**	**28**
B:	**83**	**29**	**60**	**60**	**11**	**59**	**88**	**14**	**28**			
C:	**81**	**35**	**29**	**60**	**60**	**11**	**59**	**88**	**11**	**60**	**25**	**35**
D:	**83**	**25**	**76**	**60**	**11**	**59**	**88**	**14**	**25**	**35**		

a. The first three segments all include the text 60 60 preceded by two, three, or four dinomes. If we suppose that the four dinome spelling is all single letters because it is longer than the others, then the two dinomes in segment B must each represent digraphs. Segment C with its three dinomes helps to confirm this breakout.

b. Similarly, segments A and B end with 88 14 28. Segment D ends 88 14 25 35; therefore, 28 must equate to 25 35.

c. Similar comparisons show that 14 equates to 11 60, 59 equates to 54 88, and 76 equates to 4 60.

d. We now take the first segment, for example, and replace all the dinomes that equate to two other dinomes with the single letter equivalents.

Segment A:	**81**	**35**	**25**	**74**	**60**	**60**	**11**	**65**	**88**	**88**	**14**	**28**
Replacement:	**81**	**35**	**25**	**74**	**60**	**60**	**11**	**54**	**88**	**88**	**11**	**60**	**25**	**35**

e. Reduced to single letter terms, the word pattern for the replacement segment is -ABCDDEFGGEHBA. This word pattern equates to the word *RECONNAISSANCE.*

f. These recoveries can, in turn, be used to recover additional plaintext. Whether the system is a syllabary square, a code chart, or a book code, the initial entry is the hardest part. Once the first confirmed recoveries are made, follow-on recoveries are easier.

g. The example above depended on finding sufficient repeated text to reduce the segments to single letter equivalents. This will not always be possible, but it is only one of the approaches an analyst can use to aid in recovery of the system. Anything that provides clues to the plaintext can help solve the system. Information from other sources such as traffic analysis and direction finding can help. Traffic passed in

other systems may provide isologs or clear clues to the content of the text. If the code is a one-part or uses an orderly matrix, the orderliness itself is a major aid in recovering plaintext. Encoded numbers may also help.

15-3. Recovery of Numbers

Another vulnerable point of entry in syllabary squares and codes is encrypted numbers, as has been demonstrated with other systems. Numbers, whether spelled out or encrypted by direct equivalents tend to occur with each other. Grid coordinates will typically occur in groups of four or six digits. Times are usually four digits, and tend to be rounded off into multiples of 5, 10, or 15 minutes. Times always begin with 0, 1, or 2. The third digit of a time is always 5 or less. Because of these characteristics, it is often quite easy to recognize the equivalents of 0, 1, 2, 3, 4, and 5. Even when variants are used, they tend to stand out. Given these six values, others readily follow. Recovered grid coordinates, in turn, give major clues to the rest of the text. Numbers like 7.62 (millimeter), 47 (AK-47 rifle), 45 (caliber), and 72 (T-72 tank) all provide clues to surrounding text.

15-4. Recovery of Words

Initial entry into code systems is often made through the elements that are most like a cipher. Spelled out words and encoded numbers are the weakest points in a code. Once these cipher-like groups are recovered, making further recoveries depends on recognizing the meaning of code groups that represent words and phrases. Slightly different skills are required to recover the vocabulary of a code than are required for ciphers. Cipher analysis tends to be more mathematical in nature.

a. Code recovery is more related to language skills, particularly when the text is not in English. Although words can be recovered as their English equivalents, the actual foreign language words must be known to take advantage of any alphabetic structure m the code. In languages where the sentence structure varies from English, the characteristic structures must be familiar to make sense of the code.

b. Codes are less apt to be fully recovered than ciphers. Code groups cannot be recovered until they are used, and large codes may contain many groups that remain unused for a long time. Each code group must be observed in use several times before its plaintext value can be confidently assigned. Errors are very common in encrypted traffic, and a group must be reused several times just to be sure it is not in error, It also takes repeated usage, in many cases, to be sure which of several words with similar meanings represent a particular code group. Recovery of book codes may never be completed, even when most text becomes readable at an early stage.