SRH-212

ELEMENTARY CIPHER SOLUTION

NAVY DEPARTMENT

OFFICE OF CHIEF OF NAVAL OPERATIONS

CODE AND SIGNAL SECTION

1930

REVIEWER'S NOTE:

The first review of this document was conducted by personnel of the U. S. Navy. The original classified versions were retained by them and have been placed in the NSG Repository, Crane, Indiana

~~SERVICE FOR OFFICIAL USE ONLY~~

ELEMENTARY CIPHER SOLUTION

NAVY DEPARTMENT
OFFICE OF CHIEF OF NAVAL OPERATIONS
CODE AND SIGNAL SECTION
1930

ELEMENTARY CIPHER SOLUTION

- FOREWORD -

Students of Cryptanalysis have considerable difficulty in mastering the very first step - the solution of a simple substitution cipher. Many lose interest at this point and go no further. In this connection, the writer recalls his first attempt to solve a cryptogram. He wrote "E" for the highest cipher value and then stared at the cryptogram for several hours, waiting for inspiration to come to his aid. When solution was finally achieved, the letter assumed for "E" turned out to be an "O".

The solution of an elementary cipher is simple, once the technique is really understood. Text-books on Cryptanalysis have neglected some of the details of actual solution. Students have been left to work out their own methods and have wasted days or weeks re-discovering principles which can be taught in a few hours.

To a great extent, methods of solution can be reduced to a set of rules and principles and can be taught by precept and example. The technique of elementary cipher solution is illustrated in the solution of the problem which appears on the next page.

The cryptogram was submitted from an outside source and its subject matter was entirely unknown. A careful record was kept of the various steps in the actual solution. The cipher system was simple enough to be readily understood by the beginner and yet difficult enough to require sound methods of attack.

--- THE PROBLEM ---

The following cryptogram was recently submitted by a "Novice in Cryptography", who estimated that solution would require 1000 hours of work.

```
4521941806  8528012018  7582922088  0402010620  1221534327
2899957584  4547539373  4684544692  5844604353  8293473484
5459491946  5284859346  4594928691  2745935610  5360914683
1237449294  5988759385  5917018241  1253434685  2819215591
7560170174  3723915837  2801922546  9321842036  0119562894
2600754327  8419107183  5391088834  1028192575  3619304494
5560141295  3433274546  5808885811  0907941082  4546118492
9493068284  5993068284  5927344591  0810220176  5285173458
2108849675  5434948830  5330922085  8688780985  5020109697
1253209321  5391452175  1071352791  8240441143  4001451975
4344
```

The above cryptogram was published as Problem No. 2 in communication Bulletin Number Eighty-Two, October 1929. No details of the system were submitted by its inventor.

-- THE SOLUTION --

1. The first step in the solution of an unknown cipher is to analyze the cryptogram. This analysis includes:

(a) Preparation of Frequency Table and determining the general class of cipher.
(b) Searching for repetitions.
(c) Factoring (if applicable).
(d) Preparing frequency table for the individual alphabets.
(e) Preparing lists of repetitions.

In most types of numeral ciphers two digits are substituted for one letter. A count of the numbers in the cipher text was taken and is shown in the Frequency Table on page 18. The numbers from 00 to 99 were listed and a line was drawn after a number each time that number was discovered in the cryptogram.

Most numeral ciphers belong to one of three classes:

(1) Single alphabet, usually with the numbers 11 to 36.
(2) Three related alphabets, with 78 possible different cipher values and an average of 60 to 66 different values actually employed.
(3) Four related (sometimes unrelated) alphabets, with 100 possible different cipher values and an average of 80 to 88 different values actually employed. In this type of cipher an alphabet has only 25 cipher values, and J is combined with I or else U is combined with V.

A study of the Frequency Table shows that the cipher can not be either class (1) or class (2). Each block of 25 consecutive numbers (00-24:25-49:50-74:75-99) has all the appearances of being a single "alphabet". We may therefore safely assume that the cipher is class (3) and proceed with attack on that basis.

In numeral ciphers which utilize 100 numeral values, the numerals are usually arranged in normal order, but each block of 25 consecutive numbers may be set at any point relative to the sequence of letters. In such ciphers it is often possible to "line up" the numeral sequences and reduce the cipher to the equivalent of a single alphabet. However, in this frequency table the letter distribution is so irregular that the four alphabets can not be lined-up by matching frequencies.

In this type of cipher, a plain text letter may be enciphered from any one of four alphabets, selected at random. It is possible, for instance, that all the "L's" might be enciphered from the first alphabet and all the "D's" from the second, thus giving these two cipher values a frequency four times as high as normal. The alphabets are therefore likely to be somewhat different from a normal frequency table.

As soon as the individual alphabets have been segregated the Lists of Repetitions should be prepared. The List of Repeated Groups is given in Table 2 (page 18). This list also includes the "Reversible Di-graphs".

Repeated groups represent letter combinations of high frequency. Careful study of these groups will often reveal the exact values of some of the letters involved. Without repeated groups of letters, cipher solution would be impossible in the majority of instances.

2. The second step is to prepare the work sheet. The cryptogram is recopied - but the letters or symbols are arranged in blocks according to the system used. In this cryptogram, two digits represent one letter so the numbers are spaced into pairs. (In a periodic cipher the letters of cipher text would be arranged in blocks, the length of the key.) Three spaces should be left between each line to allow space enough for "marking" and entering assumptions.

Over each cipher value write its frequency (from the appropriate alphabet in the case of poly-alphabet ciphers). Underscore all repetitions and reversible digraphs. (Use table 2). Examine the text and overscore any "peculiarities" of letter distribution. (The above should be done in ink to withstand erasure).

Recording the frequencies on the work-sheet is of the greatest importance when dealing with a minimum of text. It discloses "peculiarities" of letter distribution in the text that would otherwise be overlooked. It is essential to the vowel classification. It enables the cryptanalyst to visualize the text when working out letter combinations. It often causes a single letter of very low frequency to reveal a letter of high frequency (and sometimes a whole word). It saves constant reference to the frequency tables, which interrupts the train of thought. It saves considerable time in the end. In fact, with a complex solution, it might mean the difference between success and failure.

Table 3 (appended) shows the work sheet with the above accomplished (and with the vowels classified).

Long repetitions (6 letters or more) should be blocked off from the rest of the cryptogram with heavy vertical lines. Long repetitions indicate repeated words or phrases and are of great assistance in breaking the cipher. (There are no long repetitions in this problem and the blocking-off of word lengths is not shown until a later stage).

Cross-section paper (with one-quarter inch squares) makes the best possible work sheet. A typewritten work-sheet is nearly as good. Even spacing is essential. Three spaces should be left between lines. Do not overcrowd the work-sheet. Use printed Block capitals. The use of colored pencils for marking off repetitions, etc., is a big help. A carelessly prepared work-sheet can prevent solution merely by confusing the mind of the cryptanalyst.

3. The third step is to classify the vowels and consonants. They are classified by a study of:

(1) Frequencies
(2) Spacing
(3) Letter combinations
(4) Repetitions

(1) The low frequency values are almost invariably consonants of low or medium frequency. The intermediate frequency values are usually consonants but may be vowels. They cannot be classified except as they combine with letters already classified and are the most difficult to classify. The high frequency values are either the vowels "A,E,I,O" or consonants of high frequency.

(2) It is unusual to find over two or three consonants in succession, or two consonants of low frequency in combination. Vowels usually stand alone - combinations of more than two vowels are extremely rare. A gap of six or eight letters between two known vowels indicates the need of one or more intermediate vowels.

(3) Consonants combine with vowels, most of which are of high frequency. Vowels combine with consonants, many of which are of low frequency. Letters associated with low frequency values are vowels. Letters associated with high frequency values are consonants.

(4) Of the 30 most frequent letter pairs, 22 are vowel-consonant or consonant-vowel, 5 are consonant-consonant, and 3 are vowel-vowel combinations. Repetitions in the cipher text indicate high frequency letter combinations. Therefore, the repetition of a given letter combination creates the presumption that one of the letters is a vowel and the other a consonant.

"U" is of low frequency and can be classified only by "spacing" after A,E,I and O have been classified. The vowel of 5th highest frequency in an alphabet is almost invariably a "U". It is usually impossible to classify "Y" as a vowel - partly on account of its very low frequency and partly because "Y" is sometimes a consonant.

Mark each vowel by a circle as soon as classified - both on the work sheet and the frequency table. (To save making extra tables this has already been done in Tables 1 and 3). Values identified as consonants should be marked by an overscore or some similar method.

To explain this process, all the parts of the cryptogram where "75" occurs have been copied below:

```
7  2  9  6  6          1  2  9  9  8          4  6  9  8  6
20-18-75-82-92         99-95-75-84-45         59-88-75-93-85

9  9  9  4  3          1  1  9  7  6          7  2  9  2  7
53-91-75-60-17         26-00-75-43-27         19-25-75-36-19

9  2  9  3  6          8  7  9  7  2          8  7  9  7  5
84-96-75-54-34         45-21-75-10-71         45-19-75-43-44
```

"75" is associated with enough low frequency values to prove it o be a vowel.

4. In like manner, "84" and values in the immediate vicinity are copied below:

```
 2  9  9  8  2        1  9  9  3  9        2  6  9  3  4
95-75-84-45-27       73-46-84-54-46       47-34-84-54-59

 9  2  9  6  8        8  7  9  7  2        7  6  9  7  7
46-52-84-85-93       93-21-84-20-36       43-27-84-19-10

 9  3  9  6  8        3  6  9  4  6        7  4  9  2  9
46-11-84-92-94       06-82-84-59-27       27-08-84-96-75
```

"84" is associated with high frequency letters in almost every instance. It also combines with "75", which is already classified as a vowel. "84" is a consonant.

"88" appeared to be a vowel when first examined, but careful study after several other vowels had been classified showed that it probably was a consonant.

This process is continued until all the letters have been classified. The positive identification of two or three vowels and consonants is a great aid in classifying the remaining letters. Cipher combinations like 30-33-30 are particularly valuable. If 30 is a vowel, 33 is probably a consonant and vice versa. 30 makes the better vowel - by frequencies and by spacing. Combinations like 46-84-54-46 are equally good.

The classification of vowels is by analysis alone and does not require very much skill or experience. The classification can be considered quite accurate, although it is not infallible, and most of the vowels can be classified as such.

5. At this point it is possible to establish the relationship between the four alphabets. (This is explained later on.) However, in the actual solution this relationship was purposely ignored in order to see how difficult the solution would be if the four alphabets were all entirely different and independent. This made the solution more interesting as the frequency tables were abnormal and, in some instances, more of a hindrance than a help. We will proceed on the more general method of solution, and afterwards take up the relationship between the four alphabets.

6. The fourth step is the "Breaking-in" process:-

Experience has proved that the best points of attack in a message are:

(a) The very beginning
(b) The very end
(c) Places where the letter distribution is "peculiar"

We must correctly assume the plain-text equivalents of a few cipher values. This is not guessing - "Black Magic", clairvoyance, or mental telepathy - it is merely the application of orderly reasoning.

(a) The cryptogram was composed by a novice and a novice would not know any better than to begin a message with the word "THE" - or, if he were describing his cryptogram, the words "This Cipher". These two possibilities were therefore assumed.

```
         8  7  8  2  3  6  6  8  7  2
        45-21-94-18-06-85-28-01-20-18
         T  H  E  -  -  -  -  -  -  -
    or,  T  H  I  S  C  I  P  H  E  R
```

The word "Cipher" is impossible on account of the vowel spacing. "THE" is justified by the frequencies. "THIS" is a possibility. This about exhausts the word possibilities for the beginning of this message, as we have not the slightest idea as to the subject matter.

The values for 45-21 are checked (or may be derived) by a study of letter combinations.

45 and 21 are the first two letters of a word and are consonants of very high frequency. 45-21 is repeated elsewhere in the message thus showing it to be a combination of relatively high frequency.

The only combinations of high frequency letters which could begin a word are listed below, together with the relative frequencies of initial letters.

Letter Combination	Relative Frequency	Initial Letter	Relative Frequency
TH	50	T	17
ST	20	S	5
SH	10	R	4
TR	8	N	2
SN	3		

"TH" is the "best" combination and should be assumed first.

(b) (c) Now study the end of the message. The letter distribution of the last 20 cipher values is "peculiar" for two reasons:-

(1) Values repeated in this portion of the message.
(2) So many values between "40" and "45" used.

Retain the assumption that "45-21" is "TH". Fill in these values wherever they occur at the end of the message.

NOTE: CAPITALS are used for Basic Assumptions,
small letters for New Assumptions.

$$\overset{8}{\underset{T}{\underline{45}}}-\overset{7}{\underset{H}{21}}-\overset{9}{\boxed{75}}-\overset{7}{10}-\overset{2}{\boxed{71}}-\overset{1}{35}-\overset{6}{\boxed{27}}-\overset{9}{91}-\overset{6}{\boxed{62}}-\overset{2}{\underline{40}}-\overset{5}{\underline{44}}-\overset{3}{\boxed{11}}-\overset{7}{\underline{43}}-\overset{2}{\underline{40}}-\overset{8}{\boxed{01}}-\overset{8}{\underset{T}{\underline{45}}}-\overset{7}{19}-\overset{9}{\boxed{75}}-\overset{7}{\underline{43}}-\overset{5}{\underline{44}}$$

Note that the following underscored numbers are repeated in this part of the message: - 40, 43, 44, 45 and 75.

The following letter combinations are repeated in the message:-

45-21 = 2 75-43 = 2 10-71 = 2

27-91 is a reversible digraph

$$\overset{8}{\underset{T}{\underline{45}}}-\overset{7}{\underset{H}{21}}-\overset{9}{\boxed{75}}-\overset{7}{10}-\overset{2}{\boxed{71}}-\overset{1}{35}-\overset{6}{\boxed{27}}-\overset{9}{91}-\overset{6}{\boxed{62}}-\overset{2}{\underline{40}}-\overset{5}{\underline{44}}-\overset{3}{\boxed{11}}-\overset{7}{\underline{43}}-\overset{2}{\underline{40}}-\overset{8}{\boxed{01}}-\overset{8}{\underset{T}{\underline{45}}}-\overset{7}{19}-\overset{9}{\boxed{75}}-\overset{7}{\underline{43}}-\overset{5}{\underline{44}}$$

The very last cipher value is of course a word ending. It is a high frequency consonant - probably N, R, S, or T. The next to last letter must also be an N, R, S, or T.

But consider:

$$\overset{8}{\underset{T}{45}}-\overset{7}{19}-\overset{9}{\boxed{75}}-\overset{7}{43}-\overset{5}{44}$$

1st - our alphabets.
2nd - the letter combinations.
3rd - the relative frequencies of final letters.

Out of the last 5 letters in this cryptogram we have "43", "44", and "45" - three adjacent values in the same alphabet. They must each be different values and "45" has been taken for a "T". Cancel "T" as a possibility for "43" and "44".

The possibilities for 43-44 and their relative frequencies are as follows:-

NR=1 NS=12 RN=4 RS=9 SN=3 SR=2

NR, SN and SR are impossible word endings, and R is eliminated as a possibility for "44". The relative frequencies of S and N as final letters are about equal and give no new clue as to the identity of "44".

NS and RS are each much better letter combinations and much better word endings than RN. 44 is probably "S", and 43 may be either "N" or "R".

The possibility that 44 is N should not be considered until the assumption that "44 is S" has been definitely disproved.

Enter these values in the end of the message.

8	7	9	7	2	1	6	9	6	2	5	3	7	2	8	8	7	9	7	5
45-	21-	75-	10-	71-	35-	27-	91-	82-	40-	44-	11-	43-	40-	01-	45-	19-	75-	43-	44
T	H									S		N R			T			N R	S

Now consider what can be done. Words or parts of words can be built up from these scattered letters. If the assumptions are correct we are on the way to solution - if incorrect we will soon discover it. Now the last word of a message is usually fairly long. If it were 7 letters long, the last two words could be "IN -- T -- NS", "ON -- T -- NS" or else "OR -- T -- RS" and the third word would end in "S". "O" is the better assumption for "11", but "I" is possible.

If "45-21-75" is "THE", the third word has eight letters and the word spacing is still logical. "E" is unusually good for "75". It's frequency is very high - "ENS" and "ERS" are excellent word endings - "75-43" occurs elsewhere in the message and finally it gives a "THE" where one is needed.

Just as we were satisfied that 44 was S so may we take 75 as E. If incorrect it will show up eventually and then we can make some other assumption.

Now take "19" - It appears twice in 28-19 (both consonants)
It appears once in 19-21 (and 21 = H)
It is a high frequency consonant

"T" is the most probable value, with "S" a fair second. Other values are too improbable to be considered.

"01" is hard to identify. It is probably E, A or O - possibly "I". It cannot be the same letter as 11.

"40", the first letter of our last word, is apparently a medium frequency consonant - probably D, L, C, M or P - possibly, any other consonant.

"82" is a high frequency vowel. It cannot be "E" since 75 is "E". 82 is therefore "A", "I", or "O".

It is now time to work out the letter combinations and see what these assumptions will reveal.

WORD			WORD OR TWO WORDS								WORD		WORD						
8	7	9	7	2	1	6	9	6	2	5	3	7	2	8	8	7	9	7	5
45	21	75	10	71	35	27	91	82	40	44	11	43	40	01	45	19	75	43	44
T	H	E						a	d	S	O	n	d	e	T	t	e	n	S
								i	l		i	r	l	a		~~s~~		r	
								o	~~c~~				~~c~~	(o)					
									m				m	(i)					
									p				p						
									?				?						

The following words are possibilities for the last word:

PATTENS
MATTERS
LETTERS
DOTTERS
HATTERS
FATTERS
BATTENS

PATTERS
MITTENS
LITTERS
FITTERS

POTTERS

"S" is eliminated at once as a possibility for "19", and "c" for 40. Of these words, MATTERS and LETTERS are the most logical and should be tried first.

WORD			WORD OR TWO WORDS								WORD		WORD						
8	7	9	7	2	1	6	9	6	2	5	3	7	2	8	8	7	9	7	5
45	21	75	10	71	35	27	91	82	40	44	11	43	40	01	45	19	75	43	44
T	H	E						a	m	S	O	R	m	a	T	T	E	R	S
								i											
								o											

The above suggests nothing. Furthermore when these values were filled-in elsewhere they did not look "good".

WORD			WORD OR TWO WORDS								WORD		WORD						
8	7	9	7	2	1	6	6	6	2	5	3	7	2	8	8	7	9	7	5
45	21	75	10	71	35	27	91	82	40	44	11	43	40	01	45	19	75	43	44
T	H	E						a	L	S	O	R	l	e	T	T	E	R	S
								i											
								o											

"Letters" suggests that the cryptogram may be about the cipher-system and that in turn suggests "NUMBERS" or "NUMERALS". "Numbers" does not fit. "Numerals" fits perfectly - by vowel spacing - by frequencies - by letter combinations - and by context. The cipher has been broken.

7. The fifth step is the "Mopping-up" process. Verify the assumptions, and, if the assumptions are correct, finish the solution of the cryptogram.

We have assumed values for fifteen different letters and they all seem logical. It is time now, and not until now, to fill in these values throughout the message and see if they make other good letter combinations and give skeletons of words. This has been done in Table 4 (appended). The letter combinations are excellent.

The skeleton of "HERE" and "THERE" appear, and "53" is disclosed as being "E".

```
 7  9  7  5            7  7  9  9  9
21-53-45-27         19-21-53-91-75

 H  -- R  E          T  H  -- R  E
```

The word "NUMERAL" was next discovered and three new values were added:

```
 7  2  3  9  9  4  6  6
10-71-83-53-91-08-88-34

 N  U  m  E  R  a  l  -
```

The new value "08" = "A" helped disclose the following:

```
 4  5  5  7  9  4  7  1  8  1  2  6
59-27-34-43-91-08-10-22-01-76-52-85

      a  R  R  A  N  g  E  d
```

Each new value revealed other new values and the rest of the solution was very simple. By the time twenty cipher values had been identified, the relationship between the four alphabets was disclosed. At the same time, the method of filling in the letter sequence became fairly evident. Thus in this particular instance it was possible to completely reconstruct the "Key" before completely decrypting the message. (However, solution can be completed without taking advantage of this fact).

8. The "Key" of the cipher and the translation of the message appear in Table 5 (appended). "I" and "J" had separate values, while "U" had to be doubled up with "V".

One error in coding was noted:-

There were five "U's" in the message;
"V" was used for "U" three times,
"T" was used for "U" twice.

9. The sixth and last step is to completely reconstruct the system. The cipher device consisted of five concentric discs; four of them bearing numeral sequences in normal order, and one bearing the letters of the alphabet (minus U) in mixed order.

In this particular cipher, the relationship between the four alphabets is easily established after the vowels have been classified. The frequencies are too erratic to trust but we can line up the alphabets so as to make the vowels coincide.

The alphabets are lined up properly in Table 6 (appended). "04 and 78" were not classified as vowels but with only one occurrence that can be discounted. "71", classified as a vowel, lines up with consonants, but with two occurrences that may be due to incorrect classification of "71". The other vowels line up too well to be the result of chance.

The "Consolidated Frequency Table" checks the alignment and gives the equivalent frequencies of a single mixed alphabet. With the aid of this frequency table the cryptogram can be readily solved.

If desired, the cryptogram can be converted to the cipher values of the first alphabet, a complete list of repetitions prepared, and the cryptogram solved on this basis.

10. In this particular solution, the initial assumption that "45-21" was "TH" proved correct and solution was fairly simple. The complete solution required about three hours. However in many cases the first assumptions prove wrong and new assumptions must be made. To offset this the cryptanalyst may have a good idea of the contents of the message - may be able to fit in a word or phrase by "symmetry" - and may succeed in solving the cryptogram almost by inspection. When a cipher has been "broken" it is usually apparent at once and the "verification" is often merely a formality.

PRINCIPLES INVOLVED

- PROCESSES -

1. The solution progresses through the following stages:-

Tabulation (of values)
Classification (of vowels and consonants)
Indentification (of letters)
Reconstruction (of system)

- CLASSIFICATION -

2. Whenever possible, classify the vowels and consonants before assuming values. This is most important.

- FREQUENCIES -

3. Beginners seem to have the idea that "E" must be represented by the highest cipher value. This is far from true. "E" is the highest letter more often than any other letter - that is all. The frequency table is only a guide in the identification of letters - and sometimes an unreliable guide. The real value of a frequency table comes in the preliminary analysis where we deal with the frequencies of all 26 letters of the alphabet and not of one alone. Repetitions are far more important than frequencies when it comes to the identification of letters.

- IMPORTANCE OF LETTERS -

4. All letters are about equally important.

"E" is one of the poorest letters to identify first, as it combines with so many letters that it does not help in further identifications. "E" will always be discovered without special search.

"N" is probably the most valuable letter to identify first, (and one of the easiest) on account of its frequent occurrence in "ING", "ENT", "AND" and "ION".

Do not despise the low frequency letters. A "G" may disclose an "N" or a "Q" a "U".

- FORCING THE SOLUTION -

5. Do not force the solution. The attack should always follow the line of least resistance. "Forcing" merely delays solution.

The correct technique of code solution or cipher solution is to find a weak-point in the cryptogram and then work on it until the system is torn wide open. "Peculiar" letter distribution always indicates the weak place. The beginning and end of a message are always weak. There are usually several good "points of attack" in a cryptogram.

Above all do not "force" an assumption. If an assumption does not "check", shift the point of attack. If the assumption is correct it will be verified eventually. If incorrect it will finally be disproved.

- ASSUMPTIONS -

6. Langé says "The motto of the cryptanalyst is 'Just Suppose'." Consider what words would probably or even could possibly appear - then try and fit them in.

The fundamental basis of code and cipher solution is illustrated by the following anecdote:-

> Deacon Brown's old gray mare strayed away and was eventually found by the village half-wit. When questioned as to how he found her he replied:-
>
> "Well, I says to myself, 'where would I go if I was a horse?' and I went there, and she had."

Adhere to one Basic Assumption until it is definitely proved or disproved. Do not give up an assumption too easily, but do not cling to it too long. Experience is the only guide as to the time which should be spent on a given assumption.

It is best to assume a word or two and check the letter values in a few places before filling in the assumed values throughout the cryptogram.

- LETTER COMBINATIONS -

7. Work out the letter combinations very carefully when making Basic Assumptions. Letter combinations will very often build-up a word by analysis where all other methods of assumption have failed. The "Digraphic Frequency Table" (on page 24) is most important. Beginners are too inclined to neglect this table. Check the letter combinations in some other part of the cryptogram where the same letters appear.

- SYMMETRY -

8. Apply Symmetry of Form wherever possible. This was not described in the problem as it so happened that it could not be used.

EXAMPLES: -	LVKKVKKVNNV	XDRDOD	XFXHS
	MISSISSIPPI	PANAMA	ENEMY

As far as possible, assume words or phrases with one or more letters repeated in them. Then fit them to the cipher-text by Symmetry.

- SYSTEM -

9. Always study the cipher to see:-

(1) If it follows any definite system
(2) If there is any manner in which the system can be reduced to the basis of a single alphabet
(3) If the sequence of this single alphabet follows any given system

- WORD LENGTHS -

10. A long repetition indicates a word or phrase and is particularly valuable because the word or phrase can be blocked off from the rest of the text.

The frequencies of initial and final letters are very different, and are also different from the normal frequencies. The letter combinations are more easily worked out when it is known that a single word (and not the junction of two words) is being dealt with.

Each message has one word beginning (of the first word) and one word ending (of the last word).

Each repeated word (or phrase) gives:-

3 word beginnings, and
3 word endings.

EXAMPLE:-

```
B --- E / B (repeated word or phrase) E / B ----

----- E / B (repeated word or phrase) E / B ---- E

            B - Word Beginning
            E - Word Ending
```

Letter combinations which appear in repeated words have little significance, as the word may be of frequent occurrence with the subject matter and yet of infrequent use in the language. Short letter combinations (2, 3 or 4 letters) which are repeated throughout the message have great significance as they represent letter combinations of frequent occurrence in the language, and without which the language could not be written.

Repeated words of unusual letter combinations tend to distort the frequency tables. An unusual letter combination often indicates the break between two words.

EXAMPLES:- --- NT/N --- EE/O --- S/ST --- TH/TH ---

- CONCLUSION

Three problems are given in Table 7 (appended). Their solution will prove that the principles described herein are really understood. The student should be able to solve a simple substitution cipher (single-mixed alphabet) before attempting to solve these problems.

TABLE 1 -- FREQUENCY TABLE OF THE CRYPTOGRAM.

1st Block	2nd Block	3rd Block	4th Block
00 1	25 11	50 1	75 ~~1111~~ 1111
01 ~~1111~~ 111	26 1	51	76 1
02 1	27 ~~1111~~ 1	52 11	77
03	28 1111 1	53 ~~1111~~ 1111	78 1
04 1	29	54 111	79
05	30 111	55 1	80
06 111	31	56 11	81
07 1	32	57	82 ~~1111~~ 1
08 1111	33 11	58 ~~1111~~	83 111
09 11	34 ~~1111~~ 1	59 1111	84 ~~1111~~ 1111
10 ~~1111~~ 11	35 1	60 1111	85 ~~1111~~ 1
11 111	36 11	61	86 11
12 ~~1111~~	37 111	62	87
13	38	63	88 ~~1111~~ 1
14 1	39	64	89
15	40 11	65	90
16	41	66	91 ~~1111~~ 1111
17 111	42	67	92 ~~1111~~ 1
18 11	43 ~~1111~~ 11	68	93 ~~1111~~ 111
19 ~~1111~~ 11	44 ~~1111~~	69	94 ~~1111~~ 111
20 ~~1111~~ 11	45 ~~1111~~ 111	70	95 11
21 ~~1111~~ 11	46 ~~1111~~ 1111	71 11	96 11
22 1	47 11	72	97 1
23 1	48	73 1	98
24	49 1	74 1	99 1

TABLE 2 - LIST OF REPETITIONS.

21-53-91 = 2

21-53 = 3	---------	53-91 = 3	---------
08-88 = 2	27-45 = 2	53-43 = 2	75-43 = 2
10-71 = 2	28-01 = 2	(53-91 = 3)	84-54 = 2
12-53 = 2	28-19 = 2	---------	85-28 = 2
17-01 = 2	43-27 = 2	---------	91-08 = 2
(21-53 = 3)	45-21 = 2	---------	92-20 = 2
---------	45-46 = 2	---------	92-94 = 2
			93-21 = 2

- REVERSIBLES -

-----	27-91	53-43	85-93
-----	30-33	-----	91-27
-----	33-30	-----	92-94
-----	43-53	-----	93-46
-----	45-46	-----	93-85
-----	46-45	-----	94-92
	46-93		

TABLE 3

```
 8  7  8  2  3  6  6  8  7  2  9  6  6  7  6  1  1  8  3  7  5
45-21-94-18-06-85-28-01-20-18-75-82-92-20-88-04-02-01-06-20-12

 7  9  7  6  6  1  2  9  9  8  2  9  8  1  9  9  3  9  6  5  5
21-53-43-27-28-99-95-75-84-45-47-53-94-73-46-84-54-46-92-58-44

 4  7  9  6  8  2  6  9  3  4  1  7  9  2  9  6  8  9  8  9  6
60-43-53-82-93-47-34-84-54-59-49-19-46-52-84-85-93-46-45-94-92

 2  9  6  8  8  2  7  9  4  9  9  3  5  3  5  6  8  4  6  9  8
86-91-27-45-93-56-10-53-60-91-46-83-12-37-44-92-94-59-88-75-93

 6  4  3  8  6  1  5  9  7  9  6  6  7  7  9  9  9  4  3  8  1
85-59-17-01-32-41-12-53-43-46-85-28-19-21-53-91-75-60-17-01-74

 3  1  9  5  3  6  8  6  2  9  8  7  9  7  2  8  7  2  6  8  1
37-23-91-58-37-28-01-92-25-46-93-21-84-20-36-01-19-56-28-94-26

 1  9  7  6  9  7  7  2  3  9  9  4  6  6  7  6  7  2  9  2  7
00-75-43-27-84-19-10-71-83-53-91-08-88-34-10-28-19-25-75-36-19

 3  5  8  1  4  1  5  2  6  2  6  8  9  5  4  6  5  3  2  1  8
30-44-94-55-60-14-12-95-34-33-27-45-46-58-08-88-58-11-09-07-94

 7  6  8  9  3  9  6  8  8  3  6  9  4  6  6  7  9  4  7  1  8
10-82-45-46-11-84-92-94-93-06-82-84-59-27-34-43-91-08-10-22-01

 1  2  6  3  6  5  7  4  9  2  9  3  6  8  6  3  2  3  6  7  3
76-52-85-17-34-58-21-08-84-96-75-54-34-94-88-30-33-30-92-20-83

 2  6  1  2  6  1  7  7  2  1  5  9  7  8  7  9  9  8  7  9  7
86-88-78-09-85-50-20-10-96-97-12-53-20-93-21-53-91-45-21-75-10

 2  1  6  9  6  2  5  3  7  2  8  8  7  9  7  5
71-35-27-91-82-40-44-11-43-40-01-45-19-75-43-44
```

(WORK SHEET)

TABLE 4

```
 8  7  8  2  3  6  6  8  7  2  9  6  6  7  6  1  1  8  3  7  5
45-21-94-18-06-85-28-01-20-18-75-82-92-20-88-04-02-01-06-20-12
 T  H                 E        E  A                 E

 7  9  7  6  6  1  2  9  9  8  2  9  8  1  9  9  3  9  6  5  5
21-53-43-27-28-99-95-75-84-45-47-53-94-73-46-84-54-46-92-58-44
 H     R  E           E     T                                S

 4  7  9  6  8  2  6  9  3  4  1  7  9  2  9  6  8  9  8  9  6
60-43-53-82-93-47-34-84-54-59-49-19-46-52-84-85-93-46-45-94-92
    R                             T                    T

 2  9  6  8  8  2  7  9  4  9  9  3  5  3  5  6  8  4  6  9  8
86-91-27-45-93-56-10-53-60-91-46-83-12-37-44-92-94-59-88-75-93
    R  E  T        N        R              S              E

 6  4  3  8  6  1  5  9  7  9  6  6  7  7  9  9  9  4  3  8  1
85-59-17-01-82-41-12-53-43-46-85-28-19-21-53-91-75-60-17-01-74
          E  A           R           T  H     R  E        E

 3  1  9  5  3  6  8  6  2  9  8  7  9  7  2  8  7  2  6  8  1
37-23-91-58-37-28-01-92-25-46-93-21-84-20-36-01-19-56-28-94-26
       R           E              H           E  T

 1  9  7  6  9  7  7  2  3  9  9  4  6  6  7  6  7  2  9  2  7
00-75-43-27-84-19-10-71-83-53-91-08-88-34-10-28-19-25-75-36-19
    E  R  E     T  N  U        R           N     T     E     T

 3  5  8  1  4  1  5  2  6  2  6  8  9  5  4  6  5  3  2  1  8
30-44-94-55-60-14-12-95-34-33-27-45-46-58-08-88-58-11-09-07-94
    S                          E  T                 O

 7  6  8  9  3  9  6  8  8  3  6  9  4  6  6  7  9  4  7  1  8
10-82-45-46-11-84-92-94-93-06-82-84-59-27-34-43-91-08-10-22-01
 N  A  T     O                 A        E     R  R     N     E

 1  2  6  3  6  5  7  4  9  2  9  3  6  8  6  3  2  3  6  7  3
76-52-85-17-34-58-21-08-84-96-75-54-34-94-88-30-33-30-92-20-83
                   H           E

 2  6  1  2  6  1  7  7  2  1  5  9  7  8  7  9  9 | 8  7  9 | 7
86-88-78-09-85-50-20-10-95-97-12-53-20-93-21-53-91-|15-21-75-|10
                      N                    H     R | T  H  E | N

 2  1  6  9  6  2  5 | 3  7 | 2  8  8  7  9  7  5 ||
71-35-27-91-82-40-44-|11-43-|40-01-45-19-75-43-44 ||
 U  M  E  R  A  L  S | O  R | L  E  T  T  E  R  S ||
```

(Partial solution, assumed values filled in.)

TABLE 5.

-- KEY --

C	06	32	58	80
B	07	33	59	81
A	08	34	60	82
M	09	35	61	83
N	10	36	62	84
O	11	37	63	85
P	12	38	64	86
Q	13	39	65	87
L	14	40	66	88
K	15	41	67	89
J	16	42	68	90
R	17	43	69	91
S	18	44	70	92
T	19	45	71	93
I	20	46	72	94
H	21	47	73	95
G	22	48	74	95
UV	23	49	50	97
W	24	25	51	98
F	00	26	52	99
E	01	27	53	75
D	02	28	54	76
X	03	29	55	77
Y	04	30	56	78
Z	05	31	57	79

TRANSLATION

This code is easily deciphered when the thin discs are at hand but if not it is pretty near impossible to break period There are four codes with ninety different numeral and twenty-six alphabetical combinations it can be arranged for a change daily by simply moving up either the numerals or letters.

TABLE 6

THE FOUR ALPHABETS IN CORRECT ALIGNMENT

1st		2nd		3rd		4th		CONSOLIDATED FREQUENCY TABLE
00	1	26	1	52	11	99	1	1111
01	~~1111~~ 111	27	~~1111~~ 1	53	~~1111~~ 1111	75	~~1111~~ 1111	~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~ 11
02	1	28	~~1111~~ 1	54	111	76	1	~~1111~~ ~~1111~~ 1
03		29		55	1	77		1
04	1	30	111	56	11	78	1	~~1111~~ 11
05		31		57		79		
06	111	32		58	~~1111~~	80		~~1111~~ 111
07	1	33	11	59	1111	81		~~1111~~ 11
08	1111	34	~~1111~~ 1	60	1111	82	~~1111~~ 1	~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~
09	11	35	1	61		83	111	~~1111~~ 1
10	~~1111~~ 11	36	11	62		84	~~1111~~ 1111	~~1111~~ ~~1111~~ ~~1111~~ 111
11	111	37	111	63		85	~~1111~~ 1	~~1111~~ ~~1111~~ 11
12	~~1111~~	38		64		86	11	~~1111~~ 11
13		39		65		87		
14	1	40	11	66		88	~~1111~~ 1 (?)	~~1111~~ 1111
15		41	1	67		89		1
16		42		68		90		
17	111	43	~~1111~~ 11	69		91	~~1111~~ 1111	~~1111~~ ~~1111~~ ~~1111~~ 1111
18	11	44	~~1111~~	70		92	~~1111~~ 1	~~1111~~ ~~1111~~ 111
19	~~1111~~ 11	45	~~1111~~ 111	71	11	93	~~1111~~ 111	~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~
20	~~1111~~ 11	46	~~1111~~ 1111	72		94	~~1111~~ 111	~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~ 1111
21	~~1111~~ 11	47	11	73	1	95	11	~~1111~~ ~~1111~~ 11
22	1	48		74	1	96	11	1111
23	1	49	1	50	1	97	1	1111
24		25	11	51		98		11

- TABLE 7 -

-- PROBLEMS --

TEXT - Newspaper reports during the period of preliminary discussions prior to the London Conference for the Limitation of Armaments (January 1930).

PROBLEM NO. 1

```
5481592510  5764655523  0400240416  7463810534  0370601764
6572275154  3404733881  4637516551  0358202438  5066486059
3532019351  6455151646  6519604707  1663776454  5177504717
9519511223  3863207412  2450512238  1834274260  5965906554
1622602450  6059597207  7257222023  2065346520  4759943600
4617160316  2474515524  2077240612  0310
```

PROBLEM NO. 2

```
5862932467  6401577238  8657163824  6292387043  7415744038
4074210785  2540384924  3189493911  9444063833  4916139920
2523670049  5538227457  0720881948  0658983822  6031259906
0757196725  6788569785  6840781668  2058315707  9243813943
7458139920  6182770624  4074930760  0374006173  0685078621
9714782385  6840017800  9379239379  1549782274  2591315803
5777005973  6705687906  9949602492  8865402198  3886495139
1649584438  7921680670  8831254531  0992434178  4306913858
7881673431  3941759849  9401068870  5178963159  7807589293
9386785161  4992256094  2158448542  8631670374  9638259322
4478888277  0660947439  07
```

PROBLEM NO. 3

```
2723265747  0070952778  2807755710  3829384188  6858736307
2082240110  1424089501  7088391548  6971738648  3871395795
5643760794  0038782931  3300756989  3969209187  2726412639
5907381573  4814754207  3920913983  1739697626  9123303317
3917003975  2612573915  1110590087  7828579317  7050151069
2930397156  1073311070  8839153858  3920749512  0781262578
7350424806  3544911469  0078765869  7721261843  1795147571
7187690714  1148186911  5717188621  2615287326  1043571407
6343717557  8662661529  8116107569  1110438927  2939486675
2615693507  9426394310  9090263315  1082358864  0069711435
8469302223  2648433623  2943589139  4864205801  5723299548
3662151595  5911070069  1107730541  2939429133  1066337116
5773184600  3266473626  9000640888  2026630245  7139070056
6972263130  1815332958  0843295948  6686430754  3815114869
4200623547  3607694809  2639693010  1875212326  1600230263
3326100826
```

TABLE 8 - MECHANICS OF ENGLISH.

3	B	//
6	C	𝍸 /
8	D	𝍸 ///
26	E	𝍸 𝍸 𝍸 𝍸 𝍸 /
4	F	////
3	G	///
12	H	𝍸 𝍸 //
13	I	𝍸 𝍸 ///
1	J	/
2	K	//
7	L	𝍸 //
6	M	𝍸 /
14	N	𝍸 𝍸 ////
16	O	𝍸 𝍸 𝍸 /
4	P	////
	Q	
13	R	𝍸 𝍸 ///
12	S	𝍸 𝍸 //
17	T	𝍸 𝍸 𝍸 //
6	U	𝍸 /
2	V	//
3	W	///
	X	
4	Y	////
	Z	

FIGURE 2.—(Basis of 200 letters) Literary text.

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

FIGURE 3.—(Basis of 200 letters) Telegraphic text.

Frequency of initial and final letters.

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
Initial —	9	6	6	5	2	4	2	3	3	1	1	2	4	2	10	2	-	4	5	17	2	-	7	-	3	-
Final —	1	-	-	10	17	6	4	2	-	-	1	6	1	9	4	1	-	8	9	11	1	-	1	-	8	-

Relative frequencies of the vowels.

A 19.5%   E 32.0%   I 16.7%   O 20.2%   U 8.0%   Y 3.6%

Average number of vowels per 20 letters, 8.

The following are the proportions of vowels and consonants to the total number of letters:

Vowels A E I O U Y	40.33%	40.33%
High-frequency consonants H N R S T	34.09	59.67%
Medium-frequency consonants D L C M P F W G B V	23.81	
Low-frequency consonants J K Q X Z	1.77	
Total	100.00	100.00

letters.

(Taken from Hitt's Manual.)

FIRST LETTER.

SECOND 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
A	...	1	7	10	22	3	2	26	4	2	2	7	8	11	2	9	...	13	12	9	...	2	4	1	12	...
B	5	...	...	1	2	...	...	...	1	...	...	1	1	1	2	...	...	...	2	1	3	...	...	...	1	...
C	6	...	1	1	14	2	...	...	11	...	...	...	...	11	3	...	...	2	3	1	1	...	1	...	1	...
D	6	...	...	12	30	1	...	...	2	...	...	4	...	30	1	...	...	4	1	1	1	...	1	...	3	...
E	...	11	14	16	12	2	6	33	10	2	6	18	14	12	1	7	...	36	11	12	2	16	5	...	1	1
F	3	...	...	2	8	2	1	...	2	...	...	2	1	3	25	...	...	...	3	1	1	...	...	...	1	...
G	4	...	...	1	3	...	...	...	2	...	...	...	...	11	2	...	...	3	...	...	...	...	...	...	1	...
H	1	...	11	2	4	1	4	...	...	...	...	1	...	2	1	1	...	2	10	50	...	...	3	...	2	...
I	2	1	4	12	6	5	1	12	1	...	5	9	8	12	1	3	...	12	13	22	2	3	6	...	1	1
J	...	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
K	1	...	1	...	2	...	...	...	...	...	...	...	...	...	...	...	...	2	1	...	1	...	...	...	...	...
L	14	6	2	1	6	1	1	1	6	...	...	9	...	3	0	3	...	8	2	3	5	...	...	...	...	...
M	7	...	...	3	13	2	...	2	3	...	...	...	4	1	10	...	...	4	1	1	...	...	...	...	2	...
N	38	...	...	3	23	...	2	1	31	...	3	...	2	2	30	...	...	4	3	...	11	...	2	...	...	...
O	1	1	12	4	8	8	3	12	18	2	...	4	7	9	3	7	...	13	15	22	...	2	6	1	5	...
P	2	...	...	1	8	...	...	...	1	...	...	2	4	2	3	2	...	1	8	1	4	...	...	3	1	...
Q	...	...	...	...	2	...	...	...	...	...	...	...	...	1	1	...	...	...	1	...	...	...	...	...	...	...
R	10	1	3	3	40	3	6	2	6	...	...	1	2	1	23	8	...	2	2	8	11	...	...	...	2	...
S	10	1	...	3	25	1	2	...	17	...	1	2	1	12	7	2	...	9	11	6	11	...	1	...	6	...
T	23	1	3	12	13	5	2	3	20	...	...	2	1	24	8	2	...	10	20	11	6	...	2	2	7	...
U	1	2	1	6	1	3	2	2	...	3	...	3	1	...	17	1	5	3	5	5	...	...	...	1	...	...
V	3	1	...	...	5	...	...	...	5	...	...	...	...	...	3	...	...	2	...	...	5	...	...	...	1	...
W	1	...	...	2	8	...	1	1	...	...	...	1	1	2	4	...	...	...	2	3	...	...	...	...	3	...
X	1	...	...	...	4	...	...	...	2	...	...	...	...	...	1	...	...	...	...	...	1	...	...	...	...	...
Y	3	2	...	2	4	...	1	1	...	...	...	8	1	2	...	1	...	3	1	7	...	...	...	...	...	...
Z	1	...	...	...	...	...	...	...	1	...	...	...	...	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

Most frequent digraphs.

✱ TH—50	AT—25	ST—20
ER—40	EN—25	IO—18
ON—39	ES—25	LE—18
AN—38	OF—25	IS—17
RE—36	OR—25	OU—17
✱ HE—33	NT—24	AR—16
IN—31	EA—22	AS—16
ED—30	TI—22	DE—16
ND—30	TO—22	RT—16
HA—26	IT—20	VE—16

Most frequent trigraphs.

✱ THE—89	TIO—33	EDT—27
AND—54	FOR—33	TIS—25
THA—47	NDE—31	OFT—23
ENT—39	HAS—28	STH—21
ION—36	NCE—27	MEN—20

✱ Somewhut Lower for Telegraphic Text.