	DATE OF REQUEST	SUSPENSE DATE
	25 Jan 61	10 Feb 61

FILE OR SERIAL NUMBER AND SUBJECT	From File of Special Consultant (Friedman) General Solution for the ADFGVX Cipher System Register No. 129 Serial No. 1010 ~~Confidential~~

TO	NAME AND EXTENSION OF PERSON REQUESTING FILE Mr. William Friedman LI 6-8520	ORGANIZATION, BUILDING, AND ROOM NUMBER 310 2nd. Str., SE, Wash., D. C.

RETURN TO	Mrs. Christian, AG-24, NSA, Ft. Geo. G. Meade, Md.	DATE RET'ND.	INITIAL HERE

INSTRUCTIONS: WHEN TRANSFERRING FILE TO ANOTHER PERSON, COMPLETE SELF-ADDRESSED TRANSFER COUPON BELOW, DETACH, STITCH TO BLANK LETTER-SIZE PAPER AND PLACE IN OUT-GOING MAIL SERVIC

10203

2ND TRANSFER COUPON

TO:

FILE (serial number and subject)

TRANSFERRED TO: (name and extension)

ORGANIZATION, BUILDING, AND ROOM NUMBER

DATE (sig) (ext.)

Declassified and approved for release by NSA on 05-06-2014 pursuant to E.O. 13526

REF ID:A68203

~~Confidential~~ Register No. 129

WAR DEPARTMENT
OFFICE OF THE CHIEF SIGNAL OFFICER
WASHINGTON

GENERAL SOLUTION
FOR THE
ADFGVX CIPHER SYSTEM

1010

Box 32

30 April 1959

This document is re-graded "CONFIDENTIAL" UP of DOD Directive 5200.1 dated 8 July 1957, and by authority of the Director, National Security Agency.

Paul S. Willard
Paul S. Willard
Colonel, AGC
Adjutant General

Classification changed to ~~RESTRICTED~~
By Authority of
HAROLD G. HAYES, Col., Signal Corps
Acting Chief, Army Security Agency

By MASON G. CAMPBELL, 1st Lt., SigC
1 April 1946

NO ACCOUNTING NECESSARY

REGISTRATION CANCELED
by
Authority Hqs, ASA ltr dated 27 Feb 46
2d Ind 11 Mar 46, signed:
HAROLD G. HAYES, Col., Signal Corps
Acting Chief, Army Security Agency

REF ID:A68203

~~Confidential~~

Register No. 129

WAR DEPARTMENT
OFFICE OF THE CHIEF SIGNAL OFFICER
WASHINGTON

GENERAL SOLUTION
FOR THE
ADFGVX CIPHER SYSTEM

TECHNICAL PAPER
OF THE
SIGNAL INTELLIGENCE SECTION
WAR PLANS AND TRAINING DIVISION

PRO PATRIA VIGILANS

UNITED STATES
GOVERNMENT PRINTING OFFICE
WASHINGTON : 1934

CONTENTS

GENERAL SOLUTION FOR THE ADFGVX CIPHER SYSTEM

SECTION I

INTRODUCTORY REMARKS

1. **Nature of the problem.**—During the World War, the Germans employed a combined substitution-transposition system known as the "ADFGVX cipher" because the cipher text consisted solely of the letters, A, D, F, G, V, X. At the close of the war there were three methods [1] of solution, dependent upon *special* cases involving:

(*a*) Finding two messages with similar beginnings;
(*b*) Finding two messages with similar endings;
(*c*) Finding several messages which were enciphered by means of completely filled rectangles (the "exact factor" method).

A study was undertaken, to provide a *general* method of solution not only for its historical interest but also for the technical features involved.

2. **Purpose of this paper.**—This paper is written to present the results of the study made, since it is believed that the principles involved may be applied to other cipher systems.

[1] For a detailed exposition of the principles involved, the reader is referred to *Elements of Cryptanalysis*, Signal Corps Training Pamphlet Number 3, and a paper by First Lt. J. Rives Childs, *Report on German Military Ciphers*.

SECTION II

DISCUSSION AND ILLUSTRATION OF METHOD

3. **General discussion.**—The general solution[1] herein presented is based on a study of the frequencies of the initial and final elements of the digraphs used in encipherment and presupposes that the substitution checkerboard has not been consciously manipulated with a view to making it perfectly homogeneous. (See Givierge, p. 217.) It is assumed that the reader has solved ADFGVX messages by the principle of similar beginnings and endings.

Twelve messages enciphered in the same key were submitted for examination. They are attached hereto, in appendix II.

The first step is to determine whether the transposition rectangle involves an odd or an even number of columns. The method is based on the following reasoning: If an even width is involved, then all the letters in any one column will be either initial or final letters of digraphs; if an odd width is involved, the letters in a column will be alternately initial and final letters of digraphs. Since the first letters of every message are in the same category—i.e., either initial or final—they may be combined into a single frequency table. Furthermore, since it is certain that the third, fifth, etc., letters are in the same category as the first letter, so long as they are in the same column, they may be added to the frequency table. It is, however, necessary to limit the number of letters taken from the beginning of any one message to a reasonable length of column, depending on the size of the message. It may be safely assumed that a transposition rectangle of no more than 25 columns is being used. For a similar reason, the second, fourth, etc., letters may be put into a single frequency table. If an even width is used, the two tables will be similar; if an odd width is used, the two tables will be different. The similarity or difference is easily discernible with as few as 20 or 25 letters per table.

4. **Illustration.**—The following are the tables[2] for the test messages:

First, third, etc.

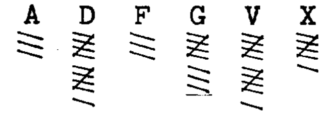

FIGURE 1

Second, fourth, etc.

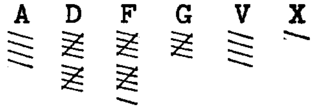

FIGURE 2

[1] This solution, obtained by Mr. F. B. Rowlett, Dr. S. Kullback, and Dr. A. Sinkov, was made without a knowledge of the remarks contained in General Givierge's *Cours de Cryptographie*, p. 209 ff. It was only after the successful completion of the solution that this reference was encountered—during the preparation of this paper. A translation of General Givierge's remarks is attached as appendix I.

[2] The portions used to make these distributions are as follows:

Message	Length	Letters taken	Message	Length	Letters taken
I	212	VDDGGGVF	VII	254	GAFGFFXFVF
II	108	VDAA	VIII	144	DGVVG
III	186	DAGAAFG	IX	182	GDDDDXV
IV	110	ADXV	X	130	DGDDF
V	202	DFXFDDVV	XI	186	VFDDVAX
VI	120	GDGF	XII	224	XFDFXVVDV

There can be no question that the two tables are dissimilar, and the indication is clear that a transposition rectangle with an odd number of columns is here the case. The outstanding difference in frequency in the cases of the letters V, X, and F in figure 1 as compared with the same letters in figure 2 sufficiently differentiates the two tables. Letters V and X are of high frequency in the *odd* positions but of low frequency in the *even* positions, whereas the letter F is of low frequency in the odd positions and of high frequency in the even positions.

Messages III and XI each contain 186 letters; if they are superimposed, appearances of V and X will most probably occur in alternate positions of one category and appearances of F in alternate positions of the other. *A reversal of this alternation definitely indicates the end of one column and the beginning of another.*

With this in mind, examine messages III and XI.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
III.	D	A	G	A	A	F	G	A	G	V	D	A	F	G	G	X	F	D	X	D	F	V	V
XI.	V	F	D	D	V	A	X	G	D	A	D	F	G	G	G	G	F	G	D	D	F	X	X

	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
III.	X	G	F	X	F	D	X	D	D	A	G	A	D	D	G	V	A	D	D	V	D	D	G
XI.	D	A	F	D	D	X	G	G	A	V	G	A	G	D	V	D	F	D	F	D	D	D	G

	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
III.	A	F	G	A	V	G	D	G	X	D	D	D	A	V	F	V	D	D	F	D	A	A	A
XI.	A	F	A	F	D	A	A	A	G	V	A	V	F	G	G	V	A	D	D	G	D	D	F

	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92
III.	A	D	X	A	G	D	X	A	G	G	D	D	A	V	G	V	F	G	D	V	F	V	D
XI.	G	F	V	D	D	A	D	F	G	A	F	D	F	V	D	D	F	V	V	V	A	D	A

	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115
III.	G	G	X	G	G	A	F	F	V	F	D	A	X	G	D	D	D	G	D	A	F	D	A
XI.	G	D	X	F	X	X	X	F	F	D	X	G	D	F	D	G	F	D	D	F	G	D	A

	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138
III.	D	G	G	A	D	D	G	D	X	A	F	V	D	F	D	X	F	V	G	D	D	V	A
XI.	G	F	A	A	G	G	A	D	X	D	G	V	D	G	A	V	G	V	D	F	D	D	F

	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161
III.	V	F	D	D	D	V	F	A	G	D	F	F	F	X	A	A	D	F	A	D	G	G	V
XI.	X	G	A	G	X	F	G	V	F	V	V	D	G	V	D	X	D	F	F	F	X	G	X

	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184
III.	F	D	A	V	D	G	X	F	V	D	A	A	V	G	D	X	F	G	G	D	D	X	G
XI.	G	X	A	G	A	G	V	G	D	V	V	X	G	F	V	D	X	D	D	X	F	V	D

	185	186
III.	D	A
XI.	D	X

Figure 3

Within the first ten letters, it is noticed that the V's and X's fall in the odd places and the F's in the even; in the next ten the F's are in the odd places and the V's and X's in the even. This reversal would indicate that column 1 of the transposition rectangle ends somewhere near the tenth letter. This same sort of reversal takes place in the third ten, but this time the break is definitely indicated. The simultaneous appearance of V and X in the sequent positions 22 and 23 indicates that 22 is the end of one column and 23 the beginning of another.[1] If columns 1 and 2 constitute 22 letters of the text, it follows that (1) each of them contains 11 letters; (2) the width of the transposition rectangle is 17 columns; and (3) the transposition rectangle contains but a single short column.

There is one other message which also contains one short column, viz, message VII of 254 letters, since (17×15)−1=254. The columns of this message contain four more letters than the corresponding columns of III and XI. Assuming, momentarily, the last column to be the short one, message VII may be added to the superposition of III and XI provided these sets of four additional letters are accounted for.[2]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
III.	D	A	G	A	A	F	G	A	G	V	D	A	F	G	G	X	F	D	X	D	F	V
XI.	V	F	D	D	V	A	X	G	D	A	D	F	G	G	G	G	F	G	D	D	F	X
VII.	G	A	F	G	F	F	X	F	V	F	G	A	G	G	X	D	X	X	D	D	F	A
								F	X	A	V								G	V	D	D

	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
III.	V	X	G	F	X	F	D	X	D	D	A	G	A	D	D	G	V	A	D	D	V	D
XI.	X	D	A	F	D	D	X	G	G	A	V	G	A	G	D	V	D	F	D	F	D	D
VII.	V	D	V	F	F	A	D	A	V	A	V	A	D	A	A	F	V	F	D	V	F	D
								F	V	G	G								X	F	X	X

	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66
III.	D	G	A	F	G	A	V	G	D	G	X	D	D	D	A	V	F	V	D	D	F	D
XI.	D	G	A	F	A	F	D	A	A	A	G	V	A	V	F	G	G	V	A	D	D	G
VII.	G	D	X	D	D	F	V	D	F	F	X	V	A	D	X	V	A	X	D	V	X	A
								D	V	F	X								F	F	V	D

	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88
III.	A	A	A	A	D	X	A	G	D	X	A	G	G	D	D	A	V	G	V	F	G	D
XI.	D	D	F	G	F	V	D	D	A	D	F	G	A	F	D	F	V	D	D	F	V	V
VII.	F	D	G	X	F	D	G	F	D	D	F	A	A	F	V	F	F	V	X	D	G	F
								V	D	V	V								D	D	V	A

	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110
III.	V	F	V	D	G	G	X	G	G	A	F	F	V	F	D	A	X	G	D	D	D	G
XI.	V	A	D	A	G	D	X	F	X	X	X	F	F	D	X	G	D	F	D	G	F	D
VII.	D	D	F	D	D	D	X	F	F	A	G	A	A	G	V	D	G	G	V	D	F	G
								F	X	F	X								G	G	X	D

FIGURE 4

[1] There is nothing of an absolute nature in this point. It is merely an indication based upon probabilities—not a conclusive proof.

[2] In figure 4 the extra four letters pertaining to the columns of message VII are shown as falling under the last letters of the columns of messages III and XI, but this is only an arbitrary placement. It is sufficient to place these extra letters in such positions as will make the first one of the series begin in an even place.

	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132
III.	D	A	F	D	A	D	G	G	A	D	D	G	D	X	A	F	V	D	F	D	X	F
XI.	D	F	G	D	A	G	F	A	A	G	G	A	D	X	D	G	V	D	G	A	V	G
VII.	F	D	F	V	A	F	F	G	F	X	G	G	D	G	G	D	D	A	V	D	X	A
								D	A	X	D								D	F	A	F

	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154
III.	V	G	D	D	V	A	V	F	D	D	D	V	F	A	G	D	F	F	F	X	A	A
XI.	V	D	F	D	D	F	X	G	A	G	X	F	G	V	F	V	V	D	G	V	D	X
VII.	V	F	X	D	D	X	V	A	G	D	V	X	D	G	X	X	D	V	F	V	F	D
								V	D	D	F								D	D	D	A

	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176
III.	D	F	A	D	G	G	V	F	D	A	V	D	G	X	F	V	D	A	A	V	G	D
XI.	D	F	F	F	X	G	X	G	X	A	G	A	G	V	G	D	V	V	X	G	F	V
VII.	A	F	D	F	X	D	X	G	D	A	A	D	V	D	D	V	A	D	D	V	D	V
								F	V	D	F								A	V	D	G

	177	178	179	180	181	182	183	184	185	186
III.	X	F	G	G	D	D	X	G	D	A
XI.	D	X	D	D	X	F	V	D	D	X
VII.	A	F	V	F	X	F	A	A	V	D
							D	F	V	D

Figure 4

Employing the same reasoning as before, it is quite evident that there is a break between positions 55 and 56. It follows, therefore, that the first five columns of the transposition rectangle are all long ones.

These five columns are of two sorts or classes which we shall designate + and −; one class comprises the odd-numbered columns, the other the even-numbered columns of the transposition rectangle. Let us see if each of the five columns can be properly classified accordingly.

For each column a frequency table of the letters in the odd and even places is made. If the table corresponds to that already made for the alternate positions of the first column, in the sequence odd → even (see figs. 1 and 2), it is designated as +. If the table is reversed, that is, if the approximation to figures 1 and 2 is closest when the sequence is even → odd, it is designated as −.

It is found that the first five columns are, respectively, + + + − +.

To help in the classification of further columns, a more mathematical procedure is desirable. The tables of figures 1 and 2 may be enlarged to correspond to the additional information that the rectangle is only 17 columns wide.[1] The letters are weighted in accordance with their

[1] The portions used to make these distributions are as follows (underlined portions indicate additional letters over those used in making figs. 1 and 2):

Message	Length	Letters taken	Message	Length	Letters taken
I	212	VDDGGGVFDFVD	VII	254	GAFGFFXFVFGFXA
II	108	VDAAVD	VIII	144	DGVVGFXG
III	186	DAGAAFGAGV	IX	182	GDDDDXVGVD
IV	110	ADXVFX	X	130	DGDDFVF
V	202	DFXFDDVVVDX	XI	186	VFDDVAXGDA
VI	120	GDGFXAG	XII	224	XFDFXVVDVDAVD

frequencies as follows:

First, third, etc.							Second, fourth, etc.					
A	D	F	G	V	X		A	D	F	G	V	X
4	14	5	11	15	10		9	15	14	8	7	2

FIGURE 5

Message V of 202 letters has two short columns (17×12−2=202), whereas the three messages dealt with thus far each have but one. That is, message V has one short column in common with messages III, VII, and XI, and one more short column that messages III, VII, and XI do not have. Can this additional short column of message V be located? Suppose column 1 is the additional short column. Then the letters which would appear in column 2 are F X F X F F F V A G F D. These letters when weighted according to figure 5 yield 77; when weighted according to figure 5 reversed yield 144. The calculation is as follows:

Distribution of letters into odd and even positions:

$\underline{\text{A}}$ D $\underset{\cancel{\equiv}}{\text{F}}$ G V X $\qquad\qquad$ A $\underline{\text{D}}$ $\underline{\text{F}}$ $\underline{\text{G}}$ $\underline{\text{V}}$ $\underline{\underline{\text{X}}}$

$$\text{Weighted Values}\begin{cases}4+25+15+14+\ 8+\ 7+\ 4 & \text{Total}=\ 77\\ 9+70+14+\ 5+11+15+20 & \text{Total}=144\end{cases}$$

In other words, column 2 for this arrangement is−. It is, however, known from what has preceded that 2 is a + column; hence, column 1 is not the additional short column of message V. Assuming 2 to be the extra short column, no such contradiction is obtained, for the calculation is as follows:

Assuming column 2 to be short, the letters of column 3 are X A V D A G F D V D G F.

Weighted frequency value for odd-even sequence (letters X V A F V G and A D G D D F)=136.

Weighted frequency value for even-odd sequence (letters A D G D D F and X V A F V G)=109.

Hence column 3 is a + column, which is consistent with the formula + + + − + for columns 1 to 5, as previously determined.

If all the foregoing reasoning is correct, and column 2 is the additional short column, it must be the next to the last column of the transposition rectangle. Since it is +, the last column must be a − one; therefore, there are nine − columns and eight + columns; it is now definitely determined that the − columns are the odd ones and the + columns the even ones.

The single short column which is common to messages III, XI, and VII is one of the columns beyond the fifth. Assuming each possibility in turn, there is obtained for the class of each column the following distributions of + and −:

Assumption	Column																	Summation of +'s and −'s
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
(1) 6th short	+	+	+	−	+	+	+	+	−	+	+	+	−	−	−	−	−	10+, 7−
(2) 7th short	+	+	+	−	+	+	−	+	−	+	+	+	−	−	−	−	−	9+, 8−
(3) 8th short	+	+	+	−	+	+	−	−	−	+	+	+	−	−	−	−	−	8+, 9−
(4) 9th short	+	+	+	−	+	+	−	−	+	+	+	+	−	−	−	−	−	9+, 8−
(5) 10th short	+	+	+	−	+	+	−	−	+	−	+	+	−	−	−	−	−	8+, 9−
(6) 11th short	+	+	+	−	+	+	−	−	+	−	−	+	−	−	−	−	−	7+, 10−
(7) 12th short	+	+	+	−	+	+	−	−	+	−	−	−	−	−	−	−	−	6+, 11−
(8) 13th short	+	+	+	−	+	+	−	−	+	−	−	−	+	−	−	−	−	7+, 10−
(9) 14th short	+	+	+	−	+	+	−	−	+	−	−	−	+	+	−	−	−	8+, 9−
(10) 15th short	+	+	+	−	+	+	−	−	+	−	−	−	+	+	+	−	−	9+, 8−
(11) 16th short	+	+	+	−	+	+	−	−	+	−	−	−	+	+	+	+	−	10+, 7−
(12) 17th short	+	+	+	−	+	+	−	−	+	−	−	−	+	+	+	+	+	11+, 6−

FIGURE 6

The correct assumption must satisfy the following conditions:

First, there must be nine minus and eight plus columns;

Second, the short column must be minus.

Only assumptions (3) and (5), in which column 8 and column 10 are short columns, satisfy these conditions.

Therefore, column 2 is followed by either column 8 or 10. Testing 2–8 for a monoalphabetic distribution, there is obtained:

	A	D	F	G	V	X
A						
D		—			—	—
F				≡	≡	
G	=	≡				
V						
X			—		=	

Testing 2–10, there is obtained:

	A	D	F	G	V	X
A						
D		—			—	—
F		=	=		—	—
G	—	—	—	—		—
V						
X		—		—	—	

Obviously the combination 2–8 is the better.

It is possible by introducing messages with additional short columns to determine more of the key; thus, it was found by using messages XII and VI that the first three columns were 16–5–7. Anagramming will yield the solution at least as rapidly.

The final solution is given by the transposition key:

16	5	7	6	9	3	14	1	13	11	17	10	4	12	15	2	8
V	I	K	I	N	G	S	C	R	O	W	N	H	O	T	E	L

The substitution key is:

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

APPENDIX I

TRANSLATION FROM "COURS DE CRYPTOGRAPHIE" BY GEN. M. GIVIERGE—PAGE 209 ff.

The most frequently encountered combinations of substitution and transposition involve transposition diagrams. But even so the number of possible variations is very great, for each plain-text letter may be represented by one or two characters and in the latter case the transposition may take place before or after the substitution and may be applied to single characters or to pairs. Besides, the substitution may be either monoalphabetic or polyalphabetic. We shall cite some of these possibilities, indicating in a summary fashion what ideas may be utilized to obtain the solution.

In general, given only one message to study, the chances of recovering the key are quite small. With a greater amount of text at hand, one may find messages of the same length, or having identical beginnings, and may avail himself of the ideas applying in such cases. It may even happen that one is in possession of the plain text of certain of the cryptograms.

Quite often the greatest difficulty is encountered in finding the number of columns in the transposition rectangle. Cryptograms having the same beginnings permit one to obtain the lengths of columns from which information one may get an idea of the kind of table used and which are the long and short columns. The examination of additional messages in the same key permits the elimination of some of the incorrect hypotheses. Sometimes one can even arrange the columns into long and short ones and then with several cryptograms of different lengths which also differ as to the number of long columns, it may happen that one can reconstruct the order of some of the columns. Other remarks, as we shall see, may help in this research.

When the substitution is monoalphabetic, letter for letter, and the transposition is effected by means of a design, the order of the two operations is immaterial; one may substitute before transposing or else substitute after the first cryptogram has been obtained by transposition.

Such a cryptogram yields a frequency table analogous to that for a monoalphabet and in most cases E can be picked out without difficulty. But the characteristic grouping is lacking for finding digraphs which would permit one to fit columns together.

If one knows the number of columns in the rectangle, an attempt may be made to guess a probable word using considerations of the type presented in the chapter on Transpositions. The frequencies of the cipher letters will serve as a guide in identifying those letters which make up the probable word.

However, this identification and the determination of the portion of the table which corresponds to the probable word are very difficult except in the case when the probable word is very characteristic either because of very frequent letters or better still because of very infrequent letters. Although the encipherment involved is quite simple, it nevertheless yields cryptograms of a high degree of security especially if the key is long and the cryptograms are short.

In order to study cryptograms built up by first substituting numbers for the letters and then transposing the individual digits, we shall first suppose that the substitution table is the one given below:

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

Each letter is enciphered by setting down the number of the column, followed by that of the row in which it is found. We shall now try to elucidate what takes place in the transposition.

After this last operation has been completed, we will have a cryptogram containing only the first five digits which give it a characteristic appearance.

It would be a good idea to begin by counting the frequency of each digit.

Let us make a remark on this point. If the frequency of each letter is entered into the enciphering table and the sums obtained for each row and column, we find:

											Total
	A	7	B	1	C	4	D	4	E	17	33
	F	1	G	1	H	0	I	7	J	0	9
	K	0	L	5	M	3	N	9	O	7	24
	P	3	Q	1	R	7	S	7	T	7	25
	U	7	V	2	X	0	Y	0	Z	0	9
Total....		18		10		14		27		31	

The 1's in the cryptogram arising from a column (as the first figures of a group), together with those resulting from a row (as the second figures of a group) have a normal frequency proportional to 33+18=51.

Similarly the remaining digits will have frequencies proportional to the figures given below:

No.	Frequency
2........	9+10=19
3........	24+14=38
4........	25+27=52
5........	9+31=40

If then the cryptogram at hand has been enciphered with a table analogous to that which we have chosen, with the letters in the same place as in our example, the frequencies of the digits will be proportional to the above numbers.

Another remark is the following: If the transposition rectangle has an even number of columns, the first line will contain an exact number of cipher digraphs; if the table has an odd number of columns, one digraph will be cut in two, the first figure being in the last column of the first line, the second figure being in the first column of the second line. In the first case (even width) each column of the table will contain only initial figures of cipher digraphs or else only final figures. In the second case (odd width) there will be a mixture in each column, the odd lines beginning with an initial digit, the even lines with a final digit.

If then in the first case we suppose the columns to be known and separated in the resulting cryptogram, the odd columns will yield frequencies based on the initial figures of digraphs, the even columns will yield frequencies based on the final figures of digraphs. Now for certain figures these frequencies are considerably different. 5 as an initial is much more frequent than as a final (31 against 9); 1 as a final is quite a bit more frequent than as an initial (33 against 18). These characteristics in the columns of the cryptogram will show up even though they are not sufficiently definite to permit the determination of the end of one column and the beginning of the next. In the case of a rectangle of an odd number of columns, this characteristic disappears, the distribution being practically uniform, but it will be discovered if one considers the digits in the even places of a column and the digits in the odd places provided the columns are sufficiently long for some law to be discovered.

It is therefore possible, given several messages, to determine the number of columns in the rectangle, at any rate after some trial.

When this has been accomplished, one may try to pair off the columns on the basis of frequency, granting that the best pair is that one which gives the greatest number of E's.

Suppose we have the following cryptogram:

```
25534   45414   51143   13441   13353   11423   13121   55135   35341

24244   12141   45311   45525   45322   55
```

By means of the 5's and the 1's, one can with a sufficient degree of approximation distinguish six columns in the following order: 1 odd, 2 even, 1 odd, 1 even, 1 odd. As the total number of letters is divisible by 6, the rectangle is completely filled and we will not have any doubt about long and short columns (a difficulty which it must be admitted will arise and will require trials to be made on a number of messages in order to permit one to make hypotheses about the beginnings and ends of columns).

Let us therefore cut the given cryptogram into six equal columns:

```
  I.........255344541451        IV.........513535341242
 II.........143134411335         V.........441214145311
III.........311423131215        VI.........455254532255
```

and try to fit the columns together. In making this trial we may suppose the odd and even columns known and will consequently try to fit together only an odd and an even column. If, however, we were not sure of the character of the columns, we would make our trials on all the columns.

Let us successively set columns II, III, and V next to I, IV, and VI:

A		B		C		D		E		F		G		H		I	
I	II	I	III	I	V	IV	II	IV	III	IV	V	VI	II	VI	III	VI	V
2	1	2	3	2	4	5	1	5	3	5	4	4	1	4	3	4	4
5	4	5	1	5	4	1	4	1	1	1	4	5	4	5	1	5	4
5	3	5	1	5	1	3	3	3	1	3	1	5	3	5	1	5	1
3	1	3	4	3	2	5	1	5	4	5	2	2	1	2	4	2	2
4	3	4	2	4	1	3	3	3	2	3	1	5	3	5	2	5	1
4	4	4	3	4	4	5	4	5	3	5	4	4	4	4	3	4	4
5	4	5	1	5	1	3	4	3	1	3	1	5	4	5	1	5	1
4	1	4	3	4	4	4	1	4	3	4	4	3	1	3	3	3	4
1	1	1	1	1	5	1	1	1	1	1	5	2	1	2	1	2	5
4	3	4	2	4	3	2	3	2	2	2	3	2	3	2	2	2	3
5	3	5	1	5	1	4	3	4	1	4	1	5	3	5	1	5	1
1	5	1	5	1	1	2	5	2	5	2	1	5	5	5	1	5	1

If we then consider the frequencies of the various digraphs, we have:

	A	B	C	D	E	F	G	H	I
11	1	1	1	1	2				
12									
13									
14				1		1			
15	1	1	1			1			
21	1					1	2	1	
22								1	1
23		1		1		1	1		1
24			1					1	
25				1	1				1
31	1				2	3	1		
32			1		1				
33					2			1	

	A	B	C	D	E	F	G	H	I
34		1		1					1
35									
41	1		1	1	1	1	1		
42		2							
43	2	2	1	1	1			2	
44	1		2			1	1		2
45									
51		4	3	2				4	5
52						1		1	
53	2					2	3		
54	2		1	1	1	2	2		1
55							1	1	

We observe in this table, in which the correct solutions are B, D, and I, that the highest frequencies are those of E (51), and that the incorrect pairings do not often present a higher frequency (the digraphs spreading out over the whole scale of numbers). It does, however, happen that the combination which corresponds to E will not exhibit this property in an incorrect pairing except for a few accidental cases the chances for which decrease as the length of column increases.

Keeping that remark in mind, let us apply it to the case in which the alphabet has been entered normally into a square but for which we do not know the number scheme adopted for the rows and columns. In other words, we do not know the key of the substitution table. We have the means of discovering E. It will be the group corresponding to the greatest frequencies obtained by coupling each column successively with all the others; remembering, of course, that one must expect errors as the result of chance.

When the columns have been coupled together in this way and E has been discovered, one has the means of obtaining new letters on the basis of frequency since each number digraph represents one letter. We already know, if the alphabet has been inserted normally, which letters are on the same row and which are in the same column with E, also which of these are frequent and which are infrequent. The determination of each new letter will give additional information about its row and its column. For example, J corresponds to a group whose first figure is the same as that for E. Among the letters having the same second figure as J only I is frequent. Therefore the row F G H I J will correspond to an infrequent letter in the E column and will have one frequent letter in it. One will also observe that on the row K L M N O, only the letter K is very infrequent. If then one row involves four frequent letters the fifth being absent, it will probably be the third and the column corresponding to the absent letter will be the first one.

Once several plain-text letters have been obtained, an attempt is made to obtain the transposition rectangle by setting the paired columns next to each other in their proper relationship. We will thus have solved the following problem: Reconstruct the transposition key and the coordinates at the side and top of the substitution square, given the alphabet which has been inscribed in the square.

If one found that the transposition key involved an odd number of letters, then it would be necessary to use the even letters in each column first with one column on the right and then with that same column on the left, after which it would be necessary to treat the odd letters in the same way. One would thus have to make four trials for each one of the preceding cases.

The reasoning, of which we have given but a short sketch above (with an attempt to stray as little as possible from general considerations without getting into problems in cryptanalysis which are beyond the elements of that science), is based upon the inequality of the frequencies of the rows and columns of the enciphering square. Thus, we were able to use the characteristic frequencies of the 5 and the 1 to draw conclusions about the make-up of the transposition diagram.

An attempt has therefore been made to construct a table in which the total frequencies of the rows and columns would be practically the same, without making the alphabet a secret or variable element and permitting one to keep it in writing with the possible risk, of course, of having it fall into the hands of the enemy. The following table satisfies that condition:

											Total
	E	17	K	0	H	0	V	2	Y	0	19
	G	1	L	5	I	7	R	7	J	0	20
	Q	0	C	4	N	9	B	1	T	7	21
	X	1	D	4	F	1	S	7	U	7	20
	Z	0	A	7	P	3	M	3	O	7	20
Total----		19		20		20		20		21	

The difficulty of cryptanalysis has been considerably increased and in particular the difference in frequencies of the 5 and 1, which was pointed out in the preceding system, no longer appears. We can, however, still attack messages based on such a procedure by making various trials based on different hypotheses as to the lengths of the columns in the rectangle unless the comparison of several messages gives more precise means of obtaining that information. The work is based on this observation that E is the only frequent letter in its row and column. When an attempt is made to couple two columns, the correct result, viz., that one which yields the greatest number of E's, will show very few other combinations using the same initial or the same final digit as E.

We see then that, in spite of its complication, a system of simple substitution and transposition may be attacked by a cryptanalyst with a fair possibility of success. The best test would be to have for study a minimum number of messages in the same key, the variable elements being the alphabet inscribed in the square and the transposition key.

APPENDIX II

Messages Used in Explanation

I

```
VDDGG GVFDF VDVVF VDGAD DAFFF
VDXFD DXDVX ADVDV FXGDF VADDG
DGDGV GDDDF XFADA VDVGD GADXV
DADAD FXAVF VDDAA VDFFD FVGDF
VDDGV DDDDA VADAF ADDXA DDGAD
FVGFV DGADV FXVXD GDDAG GDDXF
FDDXA DFGDA GXDDA VFDAF GVFVF
AFFVF AFXGF XDGVA DFVDG GAVGG
DDGDV XAXFD DX (212 letters)
```

II

```
VDAAV DDFXF XDDAX GXFXD DFXAD
VAGDD FAXDV AVDVD DFVFV FFGDG
FVAXV XAVGD VDXFD XDGAX GFGGF
VFGDF VDXAV XDDVG DDVGV AGFXF
AAAXD DXG (108 letters)
```

III

```
DAGAA FGAGV DAFGG XFDXD FVVXG
FXFDX DDAGA DDGVA DDVDD GAFGA
VGDGX DDDAV FVDDF DAAAA DXAGD
XAGGD DAVGV FGDVF VDGGX GGAFF
VFDAX GDDDG DAFDA DGGAD DGDXA
FVDFD XFVGD DVAVF DDDVF AGDFF
FXAAD FADGG VFDAV DGXFV DAAVG
DXFGG DDXGD A (186 letters)
```

IV

A D X V F X V G G V F D D V A F G A A V F D G V D

D D G D G F D V V A F G X F X F D D D D V G D A X

D A X D D D A G V F F A A D V G D F X G X G V G D

D D D A D V X V F A V D A X X D F A A F A V D V G

V D V D D A X D A A (110 letters)

V

D F X F D D V V V D X F X F X F F F V A G F D X A

V D A G F D V D G F A D A A D F D V F G D A D F V

F V F X G X D D A G D V G V F D G X X D F F G D G

X G V D D V D D F G F V G D D V F V A G X X D F V

D X A V F G A G A G A X D V D F X G V G D A D D X

A G X D A D F D G X F D G G F V G X V V G D D D A

G X V D G V D V G X D D F D D V A G A A D G D D F

D G A G D F D D D D X G V G V G G G D G X D F G F

A D (202 letters)

VI

G D G F X A G V F V D D X G X D V D D A X D A A X

F A G V G D X F F V X F A D G F F D X A A F V X F

D F X F V G D G F X F D V V X V G D F V D D V F D

F V V D V D G G V F X F G V X F F V G V D D G D D

D D G D D A V G V X G A F F X F V D D D (120 letters)

VII

G A F G F F X F V F G F X A V A G G X D X X D D F

A G V D D V D V F F A D A V A V F V G G A D A A F

V F D F V D X F X X G D X D D F V D F F X D V F X

V A D X V A X D V X A F F V D F D G X F D G F D D

F V D V V A A F V F F V X D G F D D V A D D F D D

D X F F A G F X F X A A G V D G G V D F G G G X D

F D F V A F F G F X G D A X D G D G G D D A V D X

A D F A F V F X D D X V A G D V V D D F X D G X X

```
DVFVF DDDDA AFDFX DXGDA AFVDF
DVDDV ADDVD VAVDG AFVFX FAAVD
DFVD (254 letters)
```

VIII

```
DGVVG FXGGG ADFAF VVVAX AVGGV
VDVGV VDAVG DGDGA VFDDA DDDXX
DXFVF XGVGG DGDFG GDADF DDXAV
FDDVF ADXGD ADGVA FFXAD FADXD
GFADF DDGVD VXAVA DDXFF AGDXF
FVFGF GFDFD VDXXD DGGD (144 letters)
```

IX

```
GDDDD XVGVD VDAVG FGDFV DVAVD
GFAGX AVFFG VADDD AXXAX DGADG
XAVVD GXXAA AVADA DGXDV GDDDD
GVFXA AVGGV FXDAF DGVGA FGDDF
AVVGD DVDFX DVDGF VAAGD XFDVA
ADAGD AXFVG DDDAG VAVFG XXFDD
GXFVD GGDAV DAGGF DAXDX FFVGF
AXXAD DF (182 letters)
```

X

```
DGDDF VFAVD VFDAD GFVGV GGDFV
DVVXD DFDDV GXGVD XGVGD XDGDX
FXFDX VDAAD DFXDD AFFAA FVFAG
DAAGG FAXGV XXFXA DGDFD GXGDA
DAXGV VVDAA GGVFG VAVFV AAGAX
GXDGA (130 letters)
```

XI

```
VFDDV AXGDA DFGGG GFGDD FXXDA
FDDXG GAVGA GDVDF DFDDD GAFAF
DAAAG VAVFG GVADD GDDFG FVDDA
```

DFGAF DFVDD FVVVA DAGDX FXXXF

FDXGD FDGFD DFGDA GFAAG GADXD

GVDGA VGVDF DDFXG AGXFG VFVVD

GVDXD FFFXG XGXAG AGVGD VVXGF

VDXDD XFVDD X (186 letters)

XII

XFDFX VVDVD AVDAD VFAGD GVADD

FDAAD XADFV GVDGF XFGDV FVDDD

DGDVV AVVVF ADDAX AVFVA DAXDV

GDDFA XDDGX GVFXA VXVFD GDXDF

DVXAD VAVAV GVDDD AFDFA DVFFV

VGDAG FXDDF ADVXV DFXFF VVGFX

XGFVA VFAGG DAVVD XDXGD DVVAD

DDAGA AGXFG DDDGV FGFVG VXGVF

DFFDA ADVDD XGDFD DVDDG AFGD

(224 letters)

○