DRAFT

SOLUTION OF THE JAPANESE "RED MACHINE"

The earliest form of cipher messages used by the Japanese Diplomatic Communications Net produced cipher text which conformed to the telegraphic regulations then in effect; that is, each five-letter group contained a minimum of two vowels, and was pronounceable within the meaning of the regulations. This ~~fact~~ possibly was the a determining factor in ~~which caused~~ leading the Japanese to adopt the type of cipher machine used by them in their diplomatic communications ~~since~~. The orthographical construction of the Japanese Romaji text allowed them to make consonant-for-consonant and vowel-for-vowel substitutions on their Romanized plain text to produce cipher text which conformed to the above-mentioned regulations. It is interesting to note that none of the languages of the other great powers of the world possessed this characteristic.

The first cipher messages produced by the cipher machine were encountered in 1933 in March. These messages could be readily identified by a five-figure indicator appearing as the first cipher group of the message text. This machine was used sporadically during ~~this~~ that year and only a very few intercepts were obtained from 1933. This indicator group was composed ~~of five digits~~ permutations of the digits 01234 or 56789.

TOP SECRET

the Diplomatic circuits. In the Japanese controlled Far Eastern net, however, several messages were picked up in the latter part of the year. All these messages exhibited the characteristics of consonant-for-consonant and vowel-for-vowel substitutions noted above.[1]

In March 1934 the external characteristics of the cipher text changed and it was apparent that the Japanese had departed from the above-mentioned type of substitution and had indiscriminately mixed the vowel and consonant substitutions; that is, a vowel or consonant could be represented either by a vowel or a consonant. All cipher messages from this date on possessed these same characteristics.

The amount of cipher message text transmitted per day gradually increased after this date as more and more dependence was placed on the machine and its distribution was expanded. During the Manchurian incident it was used extensively and as many as ten and fifteen messages on a single day was not an uncommon occurrence in

---

1. It is interesting to note that during this period a group of simple monoalphabetic substitution systems were in addition to the machine and code systems. In these monoalphabetic substitutions vowels were enciphered by vowels and consonants by consonants.

the Far Eastern Diplomatic Net.

Our first serious effort at solution began in October 1936. By this time considerable traffic had accumulated and it was noted that a great many messages were being sent in what we called for lack of a better name the "Five-Number System." (This term arose from the fact noted above that each message in this system could be identified by means ~~of an~~ its indicator ~~of five numbers, carried at the head of the message.~~) Extracts from these files when subjected to diagnostic study indicated the following:

1. The text was definitely a form of cipher.
2. The nature of the substitution was such that repetitions were permitted to occur.
3. A radical change took place on the first, eleventh, and twenty-first of each month.
4. The keying element produced a very long cycle which at that time appeared to be on an indeterminate length.

Once these points facts were established, it was deemed to be most advisable to go back to the earlier type of traffic and concentrate our efforts on the longest message which could be found. The message selected had been sent under date of December 21, 1933 and was slightly over two thousand characters in length. The first study made on this message was an effort to determine if there was any relation among the vowels, and if a sequence could be formed by means of a statistical study of sequent letters in alternate positions in the cipher text. For this purpose a tabulation of the vowels following each of the vowels at the interval 1-3 was made. The results of this study are given herewith.

1st letter \ 2d letter	A	E	I	O	U	Y
A	16	19	35	22	12	30
E	7	25	20	21	13	33
I	28	13	26	36	17	22
O	35	24	23	13	16	17
U	16	23	17	17	9	14
Y	22	19	20	21	23	31

The best selection of letters from the above table produced the sequence

1	.	3	.	5	.
A	.	O	.	I	.
35		35		36	

After these three letters were selected, the remaining three letters seemed to fit best in the following arrangement

1	.	3	.	5	.
U	.	Y	.	E	.
23		23		33	

This left the problem of amalgamating the two foregoing sequences to form one sequence of six letters. The characteristics of the Romaji text are such that long sounds in Japanese are formed by doubling the vowels on which the sound is based. For example, JŌHŌ and BŌKYŌ KYŌTEI become JOOHOO and BOOKYOO KYOOTEI in telegraphic plain text. A tabulation, therefore, of the sequent vowels (interval 1-2) occurring in the message should divulge which letters were sequent in the vowel sequence of the cipher system. The results of such a tabulation follows:

1st letter \ 2d letter	A	E	I	O	U	Y
A	6	9	4	0	4	4
E	5	2	8	2	3	3
I	5	6	3	1	5	8
O	4	4	0	3	13	7
U	9	1	2	6	0	1
Y	3	1	5	14	0	4

When each of the highest of the foregoing tabulations were used to build up a chain of letters the following sequence resulted:

1	2	3	4	5	6
A	U	O	Y	I	E
9	9	13	14	8	8

It will be noted that this resulting sequence corroborates the order of letters found in the two partial sequences built up from the tabulation based on the interval 1-3.

The fact that a sequence could be derived from the intervals 1-2 and 1-3 when coupled with the fact that repetitions occurred in the cipher text itself led us to believe that the fundamental nature of the cipher machine used was somewhat similar to that used in the Kryha cipher machine, but the stepping interval of the alphabetic sequence had to be more or less regular, even though it

~~TOP SECRET~~

produced a long keying period. The obvious step was therefore to make an attempt to reconstruct the consonant sequence incorporated in the machine by means of the same sort of statistical tabulation. This method was tried, *on each of several long messages*, but due to the scattering of the coincidences over twenty letters as compared to six ~~made~~ it *was* more or less impossible to produce a sequence which could be used with any degree of surety that the arrangement of the letters was correct. ~~For this reason~~ *From some of our single letter frequency counts it was suspected that the alphabetic elements of the system remained in effect for a period of ten days. The evidence available supported the theory that each month was divided into three ten day periods, the first of these periods included the dates from the 1st to the 10th of the calendar month, the second, the dates from the 11th to the 20th, and the third from the 21st to the 30th. (What arrangements were made for the 31st of the calendar month was not clear at this point.)* *Accordingly* it was thought to be ~~most~~ advisable to pick a period of ten days in which considerable traffic was available, and make an attempt by statistical means to determine the length of the keying cycle applied to the two sequences.

An examination of all the traffic available led to the conclusion that the traffic bearing ~~At this time some traffic had arrived for~~ *for* the second ten days of December 1936, ~~and~~ *held the greatest promise for confirming or denying the suspicion noted above. Accordingly* all messages bearing the dates December 11 to 20 were selected and frequency tables made for each. Numerous repetitions were found, and most ~~noteworthy of all the points noticed was~~ *it was significant* that *in each message of considerable length* six of the twenty-six letters stood out in frequency *well* above the other twenty. When these letters were weighted with their

~~TOP SECRET D~~

expected frequency of occurrence in Japanese telegraphic text, the percentage for these six corresponded almost exactly with the percentage obtained ~~for them~~ from an actual count of their occurrences in the messages for this period.

This observation led at once to the ~~obvious~~ conclusion that the cipher machine was essentially the same as that used in December 1933, and that the six vowels were replaced by a mixture of vowels and consonants taken indiscriminately from the twenty-six letters of the alphabet. The question then arose: "Can the method used for the reconstruction of the vowel sequence for the message of December 21, 1933 be applied to reconstruct the corresponding sequence of six letters used in the period of December 11 to December 20, 1936?" Inasmuch as these six letters of high frequency of occurrence in the cipher text for this period were the letters B, E, I, N, O, and X, it was thought that the characteristic tendency of the letter "O" to occur as a plain text doublet might permit the use of the method with some success. The method was therefore applied to the messages available for study, and data from all the messages were condensed into a single ~~used as a basis for the~~ tabulation~~s~~. An analysis of ~~the~~ this tabulation~~s~~ led to contradiction after contradiction, until finally attempts to reconstruct the sequence from the consolidation results were thrown away in disgust. However, it was noticed that within a single long message, tabulations on the interval 1-2 gave consistent results, but when compared with tabulations from a second long message which also gave somewhat consistent results on the same interval, the two sequences formed were the exact reverse of each other.

This last observation led to the obvious interpretation: In some messages the sequences were used in one direction, while in other messages they were used in reverse directions. Accordingly, an effort was made to segregate the messages of the period under study into two classes, one in which the sequences were used in a relative normal or direct order, and a second in which the sequences were used in a relative reverse order. Repetitions between messages gave a firm basis for classification into these two categories, and while not all the messages in the period under study could be classified on this basis, certain of them seemed to fall into each of the categories without much question. When tabulations were made on the basis of interval 1-2 for each category, the sequences reconstructed were identical, but one was the reverse of the other as was indicated in the preceding paragraph, thus ~~verifying~~ strengthening the observation that a means of reversing the sequences had been incorporated in the machine in its construction.

Up to this point nothing had been discovered which tended to disclose any clues as to the length of the keying cycle. In discussing the results of our studies with Lieutenant Wenger* of the Navy Code and Signal Section, Mr. Friedman was advised by him to make a study of the messages based on the assumption that the basic keying cycle was a result of superimposing a ~~break~~ drive-wheel similar to that of the Kryha on the two sequences. Since a machine based on two sequences had been encountered in certain Japanese Naval Attache-

* Admiral - 1952.

traffic, he thought that the Diplomatic Net might use a similar machine. This Navy machine was built around a break-wheel of 47 teeth, on which some teeth were inoperative, and which caused a jump or skip in the order of alphabet stepping which ran normally, one letter of the sequences at each encipherment of a plain-text letter. One, two, or three teeth had been found to be omitted, and the resultant jump or skip corresponding thereto consisted of one, two, or three letters in the alphabet sequences. Also, in some messages, the two sequences would be used in one direction and in others it would be reversed; In both these cases the plain-text and cipher-text sequences were not identical.

Accordingly, it was concluded that separate coincidence tests ought to be made for the six letters and for the twenty letters. Based on the product of each the numbers from 30 up to 47 with six and with twenty, the total length of the keying cycle for each. Since 31 x 6 x 20 gives a product equal to 3,720, and since we had no message of even half that length, such a coincidence test was impossible. However, if only the letters in the six sequence were considered in the coincidence test the greatest length for the cycle was product of 47 and 6, or 282. With this in mind, the message of December 21, 1933 was studied, and coincidence counts for the products of 6 with each of the numbers from 31 to 47 were made, considering only the vowels in our counts. The greatest number of counts was obtained for the interval 1 - 259, which indicated that the break-wheel had 43 effective teeth.

Insofar as the message of December 21, 1933 was concerned, ~~we~~ it could be ~~were~~ fairly positive that ~~it~~ this particular message was prepared by means of a cipher machine possessing the following characteristics:

1. The machine was similar to the Kryha in operation.
2. Instead of one sequence of twenty-six letters as in the Kryha, two sequences, one of twenty consonants and the other of six vowels were used.
3. The ~~break-wheel controlling~~ the alphabet ~~skip had~~ stepping cycle was 43 ~~effective teeth.~~ message characters in length.
4. The order of the letters in the vowel sequence (cipher) ~~had~~ as determined by interval counts ~~[illegible]~~ seemed to be AWOYIE. The order of the vowel sequence (plain) was not known but might be identical with the cipher component.

The last two points noted above led to the following proposition which could readily be tested: Can the order of the plain component for the vowel sequence be confirmed and information on the stepping cycle ascertained by attempting to decipher the vowels, assuming (1) an effective cycle of 43 characters and (2) identical sequences for the plain and cipher components of the vowels? Since the plain-text vowel combinations YUU and YOO were of high expectancy in Japanese plain-text, these combinations must be represented by certain of the three-vowel combinations appearing in the cipher text. Also, if the action of the cipher machine was similar to the Kryha, when the message was written on a width of 43 characters, the effect of the motor wheel would be progressively constant down each column with a corresponding constant columnar displacement or stepping of the alphabets. Accordingly, the first step in performing the test was to write the message on a width of 43 characters and to look for favorable spots to check the three-vowel cipher combinations appearing in the same three columns.

A work sheet on the width 43 was prepared (See Fig. ___) and it was noted that in the last three columns (columns 41, 42, and 43) two sets of three cipher-text vowel combinations appeared. The combination OOY was found in line 18 and the combination OUO at line 36. These were considered particularly favorable since by inspection it was noted that the last two letters of each, OY and UO, were sequent letters in the sequence AUOYIE which resulted from the interval studies noted above. Also both combinations of cipher began with the letter O, which it was hoped would represent plain-text Y. It was easy to test these possibilities by simply completing the cipher component and determining if the plain-text combinations YUU and YOO could be logically derived. The results of this test are shown below:

Line 18		Line 36	
	O O Y		O U O
	Y Y I		Y O Y
	I I E		I Y I
	E E A		E I G
	A A U		A E A
	U U O		U A U
	O O Y		O U O
	Y Y I		Y O Y
	I I E		I Y I
	E E A		E I E
	A A U		A E A
	U U O		U A U

It was most encouraging (in fact, it was one of those rare exciting moments in a cryptanalyst's life) to note that both plain-text possibilities YOO and YUU underlined above resulted from a simple step-by-step progression in identical positions from lines separated by a multiple of 6 (36 - 18 = 18). Obviously, the next step was to extend the test beyond these three columns in both directions, to see if further encouraging evidence could be found. A section of cipher text of line 18 and that portion of line 19 which followed, was selected because this part of the text contained a large number of vowels. The results are shown below:

Column No.	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	1	2	3	4	5	6	7	8	9	10	11	12	13
Cipher text - Line 18, 19	E	R	A	A	E	G	O	J	A	U	L	A	A	C	I	O	O	Y	G	O	O	R	I	B	Y	M	G	Q	I	H	O
	A	-	U	U	A	-	Y	-	U	O	-	U	U	-	E	Y	Y	I	-	Y	Y	-	E	-	I	-	-	-	E	-	Y
	U	-	O	O	U	-	I	-	O	Y	-	O	O	-	A	I	I	E	-	I	I	-	A	-	E	-	-	-	A	-	I
	O	-	Y	Y	O	-	E	-	Y	I	-	Y	Y	-	U	E	E	A	-	E	E	-	U	-	A	-	-	-	U	-	E
	Y	-	I	I	Y	-	A	-	I	E	-	I	I	-	O	A	A	U	-	A	A	-	O	-	U	-	-	-	O	-	A
	I	-	E	E	I	-	U	-	E	A	-	E	E	-	Y	U	U	O	-	U	U	-	Y	-	O	-	-	-	Y	-	U
	E	-	A	A	E	-	O	-	A	U	-	A	A	-	I	O	O	Y	-	O	O	-	I	-	Y	-	-	-	I	-	O
	A	-	U	U	A	-	Y	-	U	O	-	U	U	-	E	Y	Y	I	-	Y	Y	-	E	-	I	-	-	-	E	-	Y
	U	-	O	O	U	-	I	-	O	Y	-	O	O	-	A	I	I	E	-	I	I	-	A	-	E	-	-	-	A	-	I
	O	-	Y	Y	O	-	E	-	Y	I	-	Y	Y	-	U	E	E	A	-	E	E	-	U	-	A	-	-	-	U	-	E
	Y	-	I	I	Y	-	A	-	I	E	-	I	I	-	O	A	A	U	-	A	A	-	O	-	U	-	-	-	O	-	A
	I	-	E	E	I	-	U	-	E	A	-	E	E	-	Y	U	U	O	-	U	U	-	Y	-	O	-	-	-	Y	-	U
Tentative plain text	Y	-	O	U	E	-	A	-	O	O	-	E	I	-	A	Y	O	O	-	E	I	-	I	-	U				I		A

The underlining above indicates the pattern of a simple step-by-step advancement of the vowel cipher component on each side of the plain-text YOO noted above. The letters from row 18 between columns 32 and 43 looked very good as plain-text; likewise those of line 19 from columns 1 through 7 seemed very favorable. While the letters preceding column 32 were not very encouraging, it was concluded that it might be expecting too much to hope that the simple step-by-step pattern extended without interruption over so great a stretch of cipher text.

The results of the foregoing test, while not contradictory, were still not conclusive, and additional testing was needed. Accordingly, it was decided to examine other sections of the cipher text using the same technique. It was noted that, in columns 2, 3, and 4, there were two combinations of three vowels, EEA in line 8 and EIE in line 20, similar in pattern to those found in columns 41, 42, and 43. These combinations appeared on lines also separated by a multiple of 12 (20 - 8 = 12). The test for these combinations follows:

	2	3	4			2	3	4
Line 8	E	E	A		Line 20	E	I	E
	A	A	U			A	E	A
	U	U	O			U	A	U
	O	O	Y			O	U	O
	Y	Y	I			Y	O	Y
	I	I	E			I	Y	I
	E	E	A			E	I	E
	A	A	U			A	E	A
	U	U	O			U	A	U
	O	O	Y			O	U	O
	Y	Y	I			Y	O	Y
	I	I	E			I	Y	I
Tentative Plain-Text	Y	O	O			Y	U	U

The appearance of YOO and YUU in related positions added some weight to the validity of the results obtained up to this point, and it was concluded that as the next step efforts should be made to relate the overlapping texts of lines 18 and 19 with a view to ascertaining the columnar displacement of the vowel sequence.

The following modified Vigenère square was used as a basis for ascribing the alphabetic designation to the tentative plain-text:

Cipher	A	U	O	Y	I	E	
Alphabet 1	A	U	O	Y	I	E	
2	E	A	U	O	Y	I	
3	I	E	A	U	O	Y	plain
4	Y	I	E	A	U	O	
5	O	Y	I	E	A	U	
6	U	O	Y	I	E	A	

The results of the application of this square to the texts of lines 18 - 19 and lines 19 - 20 are as follows:

Column #	34	35	36	37	38	39	40	41	42	43	1	2	3	4	5	6	7	8	9	10	11	12	13
Lines 18 - 19	A	U	L	A	A	C	I	O	O	Y	G	O	O	R	I	B	Y	M	G	Q	I	H	O
Tentative Plain	O	O	-	E	I	-	A	Y	O	O	-	E	I	-	I	-	U	-	-	-	I	-	A
Vowel Alphabet #	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3
Lines 19 - 20	H	A	W	O	C	A	Z	D	A	U	T	E	I	E	H	U	F	E	I	U	D	I	L
												Y	U	U									
												3	4	5									

An examination of the relationship of the alphabets in columns 2, 3, and 4 ~~The foregoing~~ led to the conclusion that the relative columnar displacement was a multiple of six, less one, and it was quite possible that this displacement resulted from the application of a 47-tooth wheel to the alphabetic sequence. If such were indeed the case, then it appeared that all the vowels in column 34 through column 4 inclusive could be reduced to plain-text, assuming of course the stepping pattern to be constant through these columns. Obviously, the next step was to apply the alphabetic designations to these remaining letters columns of lines 19 - 20 ~~the work sheet~~ and derive the plain-text for ~~of~~ the vowels. The results of this test ~~are shown in Figure ___~~ follow:

The plain text results for the vowels seemed most satisfactory; however, the question of what to do about the consonants still had to be answered. After giving some thought as to the desirability of continuing work on the vowels, it was concluded that sooner or later the work on the consonants needed to be done. Also, it was thought that if the machine utilized a single ~~motor~~ drive wheel which was applied to two sequences, one for the vowels and the other for the consonants, it would be possible to project its effect to the consonants and test the results statistically. If a 47 tooth motor wheel was used and it caused a single step displacement of the consonant sequence for each letter enciphered, its effect would be to displace the sequences a total of 47 steps between adjacent letters reading down the column of the work sheet of Figure ___. Accordingly, starting arbitrarily with any given letter of the section between column 34 through column 7 inclusive, it would be possible to ascribe each of the consonants of this section to one of twenty alphabets.

If the foregoing assumptions were correct, the frequency of occurrence of a given cipher letter in each of the twenty positions would correspond to the plain-text expectancy of the letter it represented in each alphabet.

It would therefore be possible to recover the consonant sequence by tabulating the occurrences of each cipher letter in each of the twenty alphabets and matching these distributions against each other.