

Enigma Machine – Indicator Analysis

by

Lt. Robert E. Greenwood

15 November 1943

Abstract

A thorough job was done on the sample of a day's Home Waters traffic (678 messages) sent us early in 1942 by the British as an educational problem. The British procedure is described in Article 2, Volume 8, from an operational standpoint. The present article goes more into the underlying theory.

Source: ENIGMA Series Volume 2, Article 8
RIP 604, Box 170, 370 27/22/07
NARA, RG 38, Crane Collection

Editor: Frode Weierud, © May 2009

TOP SECRET ULTRA

FOREWORD TO ENIGMA SERIES

CRYPTANALYTICAL RESEARCH PAPERS

This series consists of original memoranda written by members of the cryptanalytical research section of the U.S. Naval Communications Intelligence Staff, and by others working with the research group. A brief description of the contents of each paper is given in the Index to each volume. While an effort toward completeness has been made, the reader is referred for greater detail to the various R.I.P.'s put out by the Atlantic Operations Department, especially R.I.P. 450. There he will also find polished techniques, which appear in this Series of their original form.

The name of the author and the date of the paper are also given in the Index, which lends an historical flavor to the Series. The Editor feels that there is considerable merit in an anthology for this sort, full of original ideas both good and bad, which supplements the finished publication. It should be further emphasized that R.I.P. 450 is concerned mainly with the techniques themselves, while this Series considers the cryptanalytical or mathematical theories which underlie the techniques. On the other hand, machine research (from an engineering point of view) is not covered in this Series.

Some of the papers in this Series are expository, but most represent original work. It must always be borne in mind that we owe to the British the basic solution of the Enigma, and many of the basic subsidiary techniques, together with the underlying mechanical and mathematical theories. Much of what we call "original" is only a retracing of steps previously taken by the British, and the Editor has striven to point this out in the Index. But there is also a great deal that extends or improves British methods, and some that strikes out in new directions.

It must be pointed out that the author of a paper may be entitled to credit only for his literary toil. Our group of eight or ten men worked as a team, and an assignment of "credit" would be as difficult as it is undesirable. In this line of endeavor, a chance remark may be worth a week's work.

TOP SECRET ~~UKRAIN~~

8.

ENIGMA MACHINE INDICATOR ANALYSIS

I. Problem: To determine as much of the nature and settings of the machine as possible from an analysis of the enciphered indicator keys and messages in these keys.

II. Assumptions: It will be assumed that traffic is available for study in which:

1. The same stecker was used.
2. The same wheel order was used.
3. The same Ringstellen key was used.
4. The same Grundstellen key applies.
5. It is known that the first letter of the indicator keys determines the settings of LHW, the second letter MW and the third letter RHW.

In effect the above means that all messages must be in the same day's traffic.

III. Summary of Attack: A search of the enciphered indicator keys is made for keys having at least two of their three letters the same as possible. The corresponding messages will be then slid against one another in the hope of obtaining a good coincidence pattern between messages. Further study may reveal the identity of one or more wheels used in the machine.

IV. Basic Theory

1. Enciphered keys ABK and ABP.

We review for a moment the method of enciphering keys. The wheels are set by means of marks at the positions called for in the Grundstellen key. The first letters of all keys are enciphered mono-alphabetically at the next position, say G plus 1, since the machine moves once before encipherment, the second letters are all enciphered at position G plus 2 and the third letters are enciphered at position G plus 3. Thus, if there are two keys ABK and ABP, it is known that LHW and MW were set at the same marks because of the identities $A = A$ and $B = B$. Furthermore, the RHW was set at two different

places, since K is not the same as P.

Since the keys are enciphered, it is not possible to say that the difference in the setting is the same as the numerical positional difference between K and P in the normal alphabet. The two messages are slid against each other with the view of obtaining a good coincidence pattern in the text. Suppose that we obtain a coincidence pattern which is sufficiently outstanding to warrant the conclusion that in that position we have mono-alphabetic substitution for each pair of super-imposed letters. Suppose that this coincidence pattern occurred after a displacement of 7 letters. It is concluded that the plain equivalents of K_c and P_c are 7 letters apart.

With sufficient line-ups, we could build up a chain of such letters whose plain equivalents are correctly spaced. It then remains to take these chains, slide them against a normal alphabet and find these locations at which the reciprocal property (required of all Enigma enciphering alphabets) holds true.

For example: Suppose we have found that the plain equivalents of Q_c and K_c are also 7 letters apart then we have:

Q_c - - - - - K_c - - - - - P_c - - - - -

We do not know whether $Q_c K_c P_c$ equals $A H O$, or whether $Q_c K_c P_c$ equals $E L S$, or $M T A$, or $N U B$, etc. We do know that $Q_c K_c P_c$ is not equal to $B I P$, since P cannot go over into itself; also we know that $Q_c K_c P_c$ is not equal to $C J Q$ since Q_c equals C and P_c equals Q are not reciprocal as it required for Enigma alphabets.

If sufficient line-ups are available, it is readily seen that the enciphering alphabet for the third indicator letter can be built up following the procedure outlined above.

REMARKS: The amount of skip must be less than 26 letters if the LHW is a one-notcher (or less than 13 letters if the LHW is a double-notcher) since the MW will advance one step after a notch and the identity B equals B will no longer hold.

Assuming complete success in the identification of this alphabet, we would know that for given stecker, wheel order, Ringstellen, and for encipherment by the Grundstellen position G + 3, the following hypothetical substitution results:

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

This substitution would only hold for the wheels set in accordance with the position (G + 2), which position effects encipherment for (G + 3).

2. Enciphered keys ABK and ARY

Suppose two messages with enciphered keys ABK and ARY have been lined up to give an extraordinary coincidence pattern at a slide of 63 letters. This choice of keys requires that MW and RHW be set differently for the two messages. Then if RHW is a single-notched wheel, since 63 equals $(2 \times 26) + 11$ steps, it is known that MW has advanced two times. (If RHW is a double notch wheel, then MW has advanced 4 times.) The case of a single notched wheel affords the following information:

1. In the enciphering alphabet (G + 3) (Discussed in the previous section) the plain equivalents for K_c and Y_c are 11 letters apart.
2. In the enciphering alphabet (G + 2), the plain equivalents for B_c and R_c are two letters apart.

In accordance with previous outlined procedure we may build up chains and test by the reciprocal property and the fact that no letter represents itself. With good fortune we may determine the substitution alphabet for stage (G + 2) of the machine (i.e. the alphabet when the machine is set at the position indicated by (G + 1).

REMARK: Due to the shortness of average messages in this system, it is generally not possible to extend these tests and reasoning to the LHW and recover the enciphering alphabet for stage (G + 1).

3. Identification of RHW

We are now in a position to attempt the identification of RHW. We first endeavor to discover the number and position of the notches on the wheels used.

If we have two enciphered keys such as ABC and ABX with an extraordinary coincidence pattern at an interval greater than 13 (and less than 26) then RHW must be a single notch, else MW would have moved. If on the other hand, we find no line-ups in which the MW has remained fixed for more than 13 letters, we may assume that RHW has two notches.

Additional information may be obtained from the substitution alphabets for stages (G + 2) and G + 3). Suppose that the enciphered

TOP SECRET ~~CREAM~~

indicator keys ABL and ACR were used on two messages which show an extraordinary coincidence pattern at an interval of 7 (or any number less than 26). The difference in MW settings is due to the fact that RHW had a notch between the plain equivalents for L_c and R_c . If these are P_c and T_c , then the notch on RHW occurs between P and T. This eliminates the possibility of a two-notch wheel, since all such notches occur at M and Z. It may eliminate some of the 5 single notch wheels, indeed it does eliminate all except those for which the notch occurs between P and T.

4. With favorable line-up, MW may be identified following the general procedure above.

5. Relative Probability for Coincidence Pattern

A given coincidence pattern may arise from one of two mutually exclusive causes:

- (1) Correct alignment, same plain text (Cause C_1).
- (2) Incorrect alignment, different plain text letters (Cause C_2). (Since all Enigma substitution alphabets are reciprocal and no plain text letter may represent itself).

To determine the probability, we use Bayes' Theorem: If C_1 and C_2 are two mutually (and collectively exhaustive) causes of an event, and k_1 and k_2 are the a-priori probabilities of the occurrence of these causes, and p_1 and p_2 are the respective probabilities that the given event should follow from each of the causes, then the probability (P_i) that the event was caused by C_i is:

$$P_i = \frac{k_i p_i}{k_1 p_1 + k_2 p_2}$$

and the likelihood (L) for a correct rather than an incorrect alignment causing the event is:

$$L = \frac{k_1 p_1}{k_2 p_2}$$

Let us consider the a-priori probability that two messages with enciphered indicators ABK and ABP can be correctly aligned. Since the only difference in wheel settings is in RHW, and since there are only 26 possible settings of RHW we conclude:

(a) that there is no chance of a correct alignment at an interval of 26 or greater and

(b) that for any interval less than 26 the chance for correct alignment is $\frac{1}{26}$. It will be noted that we have tacitly assumed that RHW is a single notcher. These odds, of course, will have to be changed for two notch wheels.

However, we shall arbitrarily assign the following a-priori odds: If third letters are different, chance of a correct alignment is $\frac{1}{26}$, if second and third letters are different chance of a correct alignment is $(\frac{1}{26})^2$. It should be noted that if the second letters are different, the

possible motion of the middle wheel due to passing a notch may change the above.

We shall call this probability of correct alignment ℓ_1 then the probability of incorrect alignment is $\ell_2 = 1 - \ell_1$.

The ratio $\frac{p_1}{p_2}$ may be computed as follows: Let us assume that we are comparing an overlap of N letters where there are:

- non-coincident places
- single coincidences
- digraphic coincidences
- trigraphic coincidences
- tetragraphic coincidences

and no coincidences higher than tetragraphic (note: $a + b + 2c + 3d + 4e = N$). The incorrect alignment may be considered as random text, and the correct alignment as plain text.

Then the ratio $\frac{p_1}{p_2}$ is seen to be:

$$\begin{aligned} \frac{p_1}{p_2} &= \left(\frac{1 - \frac{IC_1}{26}}{\frac{25}{26}} \right)^a \left(\frac{\frac{IC_1}{26}}{\frac{1}{26}} \right)^b \left(\frac{\frac{IC_2}{26^2}}{\frac{1}{26^2}} \right)^c \left(\frac{\frac{IC_3}{26^3}}{\frac{1}{26^3}} \right)^d \left(\frac{\frac{IC_4}{26^4}}{\frac{1}{26^4}} \right)^e \\ &= \left(\frac{26 - IC_1}{25} \right)^a (IC_1)^b (IC_2)^c (IC_3)^d (IC_4)^e \end{aligned}$$

Hence, we obtain:

$$L = \frac{\ell_1}{1 - C_1} \left(\frac{26 - IC_1}{25} \right)^a (IC_1)^b (IC_2)^c (IC_3)^d (IC_4)^e$$

as a measure for the likelihood of a particular coincidence pattern caused by correct alignment. Due to the small value of the $\frac{1}{1-k_1}$ factor, a good coincidence pattern will be required in order to give a good (i.e. high) likelihood. In practice it has been found advisable to consider the following as minimum requirements (which are not stringent enough for extra-ordinary coincidence).

- (a) First and third or first and second letters of indicator identical -- at least a tetragraphic coincidence.
- (b) First letters identical--at least a pentagraphic coincidence.
- (c) Other cases-- at least a hexagraphic coincidence. Enough additional single, digraphic coincidences, etc., will make these coincidence patterns extraordinary.

6. Conclusion

It is presumed that mechanical means are available for determining tetragraphic coincidences. Likelihoods for a given pattern involving a tetragraphic coincidence can then be calculated, or what is better, their logarithms can be calculated. Then, depending on the nature of the two indicator key patterns (such as ABK and ABP, or ABK and ARY) a tentative decision can be made as to whether the coincidence pattern is extraordinary. If extraordinary, the various results can be collected and an effort made to determine the substitution alphabet (G + 3), the substitution alphabet (G + 2), the position of the notch or notches on RHW and hence the identity of RHW and under favorable conditions, the identity of MW.

V. Test of Theory

This theory was tested by Navy cryptanalysts in the case of 700 messages with an average length of 150 letters and found to be applicable. There is reason to believe it would work with only 400 messages, and that with luck an even smaller quantity would suffice.

Test Case

A set of 678 messages was examined, and all pairs of messages showing a tetragraphic coincidence were tabulated. Additional coincidences were then marked, and the logarithms of the likelihoods were calculated. Table I was used to tabulate these likelihoods. A list of all high likelihoods is given in Table II. Due to the fact that numerical weights were not assigned to coincidences higher than tetragraphic, certain pairs are marked as "good",

| Indicators | Slid | 26's | 1's | Log. Prob. | Overlap |
|------------|------|------|-----|------------|---------|
| ARF ARW | 7 | 0 | 7 | 4.10 | 148 |
| AJL AIB | 13 | 0 | 13 | 1.01 | 127 |
| ATW AAP | 133 | 5 | 3 | 1.11 | 162 |
| CNS CII | 20 | 0 | 20 | 1.00 | 156 |
| CII CID | 77 | 2 | 25 | 1.29 | 196 |
| EOV EKF | 215 | 8 | 7 | good | 89 |
| FCK FBO | 81 | 3 | 3 | 1.36 | 84 |
| FWD FLF | 118 | 4 | 14 | 1.16 | 90 |
| FXP FUC | 5 | 0 | 5 | 1.64 | 155 |
| GDA GDZ | 22 | 0 | 22 | 1.80 | 86 |
| GIM GSK | 161 | 6 | 5 | 1.16 | 140 |
| DPB DWT | 20 | 0 | 20 | good | 110 |
| JBY JBR | 10 | 0 | 10 | good | 153 |
| KHI KGF | 120 | 4 | 16 | good | 31 |
| KQG KQC | 2 | 0 | 2 | 1.57 | 178 |
| MNX MDS | 15 | 0 | 15 | 1.08 | 153 |
| MLN MNX | 94 | 3 | 16 | good | 86 |
| MHS MCG | 80 | 3 | 22 | good | 128 |
| MZD MDS | 3 | 0 | 3 | good | 170 |
| MTJ MEC | 51 | 1 | 25 | good | 189 |
| MLJ MLN | 21 | 0 | 21 | 1.14 | 152 |
| NYF NIP | 88 | 3 | 10 | 2.36 | 116 |
| NPL NJL | 78 | 3 | 0 | good | 94 |
| PCE PEX | 211 | 8 | 3 | good | 94 |
| PLF PTT | 97 | 3 | 19 | 1.40 | 131 |
| QSL QAQ | 64 | 2 | 12 | 1.76 | 164 |
| RNG RNV | 17 | 0 | 17 | good | 107 |
| SQV SBE | 229 | 8 | 21 | good | 64 |
| TFR TTK | 14 | 0 | 14 | good | 70 |
| UJY UPK | 79 | 3 | 1 | good | 53 |
| UJY UUS | 14 | 0 | 14 | good | 64 |
| VCY VZS | 41 | 1 | 15 | good | 71 |
| VZS VDX | 43 | 1 | 17 | good | 150 |
| VZS VEE | 292 | 11 | 6 | 1.19 | 176 |
| XKO XPT | 142 | 5 | 12 | good | 19 |
| XLG XLE | 12 | 0 | 12 | 3.30 | 156 |
| XRX XZJ | 40 | 1 | 14 | good | 160 |
| XQJ XYV | 196 | 7 | 14 | good | 125 |
| YCJ YCA | 1 | 0 | 1 | 1.06 | 195 |
| YPL YCA | 100 | 3 | 22 | 1.76 | 110 |
| YCJ YFB | 138 | 5 | 8 | 1.52 | 50 |
| YJY YIP | 34 | 1 | 8 | 1.04 | 125 |
| YEK YCA | 29 | 1 | 3 | 1.21 | 71 |

number 3 wheel with notch at V.

In like manner we may draw conclusions concerning MW.

We state (without proof) the following:

Enciphering Alphabet (G plus 2)

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

The notch in MW is located at E, hence MW is wheel number 2.

Conclusion:

It is readily seen that with 700 messages the enciphering alphabets (G + 2) and (G + 3) may be recovered, and the identity of MW and RHW be established. The location of tetragraphic repeats may be done by machine, although the marking of other coincidences will probably be done by hand. The log weights may be readily scored.

The log weights of some "incorrect" alignments seem to be rather high. It is believed that a more accurate theory will reduce these values. The discussion of a-priori probability was incomplete, also it was assumed that in the case of incorrect alignments, that samples of "random text" were being compared.

At the time that this report was written (November 1943) the original line-ups have been preserved only in part. All line-ups whose initial indicator letter was X have been preserved and these are included herewith. 2 July 1943 the traffic and folder of original line-ups were in the files of OP-20-GM5.

number 3 wheel with notch at V,
In like manner we may draw conclusions concerning MW.

We state (without proof) the following:

Enciphering Alphabet (G plus 2)

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

The notch in MW is located at E, hence MW is wheel number 2.

Conclusion:

It is readily seen that with 700 messages the enciphering alphabets (G + 2) and (G + 3) may be recovered, and the identity of MW and RHW be established. The location of tetragraphic repeats may be done by machine, although the marking of other coincidences will probably be done by hand. The log weights may be readily scored.

The log weights of some "incorrect" alignments seem to be rather high. It is believed that a more accurate theory will reduce these values. The discussion of a-priori probability was incomplete, also it was assumed that in the case of incorrect alignments, that samples of "random text" were being compared.

At the time that this report was written (November 1943) the original line-ups have been preserved only in part. All line-ups whose initial indicator letter was X have been preserved and these are included herewith. 2 July 1943 the traffic and folder of original line-ups were in the files of OP-20-GM5.

TABLE IIISUMMARY OF REPEATS

When only left-hand indicator is the same:

| <u>Length of Repeat</u> | <u>Total</u> | <u>Good</u> | <u>Bad</u> | <u>%Good</u> |
|-------------------------|--------------|-------------|------------|--------------|
| 9 | 1 | 1 | 0 | 100% |
| 8 | 1 | 1 | 0 | 100% |
| 7 | 1 | 1 | 0 | 100% |
| 6 | 5 | 3 | 2 | 60% |
| 5 | 6 | 3 | 3 | 50% |
| 4 | <u>407</u> | <u>16</u> | <u>391</u> | 04% |
| | 421 | 25 | 396 | |

When left-hand and middle indicators are the same:

| <u>Length of Repeat</u> | <u>Total</u> | <u>Good</u> | <u>Bad</u> | <u>%Good</u> |
|-------------------------|--------------|-------------|------------|--------------|
| 6 | 1 | 1 | 0 | 100% |
| 5 | 1 | 1 | 0 | 100% |
| 4 | 8 | 2 | 6 | 25% |
| 3 | <u>70</u> | <u>10</u> | <u>60</u> | 14% |
| | 80 | 14 | 66 | |

When left-hand and right-hand indicators are the same:

| <u>Length of Repeat</u> | <u>Total</u> | <u>Good</u> | <u>Bad</u> | <u>%Good</u> |
|-------------------------|--------------|-------------|------------|--------------|
| 10 | 1 | 1 | 0 | 100% |
| 3 | 5 | 2 | 3 | 40% |

APPENDIXTHEORETICAL OVERLAPS IN A DAY'S TRAFFIC

Assumed: 1. That you have seventy-five messages one hundred twenty-five letters long (9,375 total letters).

2. Possible machine positions in order total seventeen thousand.

Then: Of the 17,000 machine positions

9,767 positions will never be used

5,424 positions will be used once

1,487 positions will be used twice

322 positions will be used three times or more.

Of our total of 9,375 letters, 5,424 will be used to fill positions only used once. This leaves 3,951 letters that are used in overlaps two or more deep. This means that you could expect the same significant repeats as would occur in comparing two samples of the underlying text approximately 2,500 letters long.

Fortunately the language is highly stereotyped and has a very high percentage of numbers.

If you have 9,375 random letters, you will get from sheer chance:

Circa 2,300 trigraph repeats

91 tetragraphs

3 pentagraphs

1/8 hexagraph

15 November 1943.

TABLE III A

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
 Normal or plain cipher G C J E F X V W P S

| Letter under Position I | Examination of alphabets | Results |
|-------------------------|----------------------------------|----------|
| 1 A | Cp = Cc | Rejected |
| 2 B | Ep = Jc Np = Ec | " |
| 3 C | Cp = Gc Ep = Cc | " |
| 4 D | Dp = Gc Gp = Jc | " |
| 5 E | Ep = Gc Qp = Ec | " |
| 6 F | Hp = Cc Cp = Pc | " |
| 7 G | Gp = Gc | " |
| 8 H | Jp = Cc Kp = Jc | " |
| 9 I | Xp = Xc | " |
| 10 J | Jp = Gc Mp = Jc | " |
| 11 K | Xp = Fc Zp = Xc | " |
| 12 L | Op = Jc Jp = Sc | " |
| 13 M | Mp = Gc Gp = Wc | " |
| 14 N | Pp = Cc Cp = Xc | " |
| 15 O | Bp = Fc Fp = Vc | " |
| 16 P | Sp = Jc Np = Sc | " |
| 17 Q | Sp = Cc Op = Sc | " |
| 18 R | Rp = Gc Gp = Xc | " |
| 19 S | Sp = Gc Qp = Sc | " |
| 20 T | Tp = Gc Gp = Fc | " |
| 21 U | Sp = Sc | " |
| 22 V | Vp = Gc Mp = Vc | " |
| 23 W | Zp = Jc Jp = Fc | " |
| 24 X | Op = Vc Vp = Sc | " |
| 25 Y | Two reciprocal pairs (WS and PV) | Accepted |
| 26 Z | Bp = Cc Cp = Jc | Rejected |