Clark Test

by

Lt. Alfred H. Clifford

22 May 1945

Abstract

This test (invented by and named after a Britisher) evaluates the merit of a story from its "confirmations" – the number of self-steckers and pairs of intersteckered menu letters. We computed the table late in 1942, but not (as it now appears) the best possible way.

Source: ENIGMA Series Volume 5, Article 4
RIP 607, Box 171, 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.

CLARK TEST

1. Introduction.

Back in the Fall of 1942, at the instigation of Mr. Howard, and under his supervision, we computed the relative merits of bombe stories due to their Stecker confirmation pattern. Suppose a story has a self Steckers, 2b letters intersteckered in b pairs, and c letters steckered to letters not on the menu. If m is the number of letters on the menu, we have of course

$$a + 2b + c = m$$

The triad of numbers a - 2b - c we call the "pattern" of the story.

The results of the computations were tabulated and mimeographed (Table 1 below). The pattern is indicated by the letters S - C - O instead of a - 2b - c. The number heading each column of Table 1 is the number of menu letters. Within each column, the ratio R is proportional to the Bayes' factor in favor of the type indicated; but the constant of proportionality varies (more or less regularly) from column to column, so one should not use the table to compare two stories from menus containing different numbers of letters. We proceed to describe briefly how the table was computed, and incidentally to point out how they could have been done more directly.

2. Probability of a Chance Story With Given Pattern.

For a menu of m letters, $26 \times 25^{m-1}$ ways of steckering the m letters are mechanically possible to each try. The number of these that survive the consistency test (reciprocity and no two letters steckered to the same letter), and have pattern a - 2b - c is:

$$\binom{m}{a} \binom{m-a}{2b} S(2b) \frac{(26-m)!}{(26-m-c)!} \qquad (1)$$

The first term represents the number of ways of selecting a of the m letters to be self-steckered. The second is the number of ways of selecting 2b letters to be intersteckered in pairs, and the third

$$S(2b) = \frac{(2b)!}{2^b b!} = 1 \times 3 \times 5 \times 7 \times \ldots \times (2b-1)$$

is the number of ways of intersteckering them. The last term is the number of ways of steckering the remaining c menu letters to c off-menu letters.

The probability of getting a wrong story, at a given try, with pattern a - 2b - c is the above expression (1) divided by $26 \times 25^{m-1}$.

3. Probability of a Jackpot Having Given Pattern.

This may be found in two apparently different ways. The total number of Steckers satisfying the 6-20 condition (6 self-stecker, 20 letters in 10 pairs) is

$$\binom{26}{6} S(20) = 1.507 \times 10^{14} \qquad (2)$$

The desired probability is the ratio of the number of admissable Steckers which give the indicated pattern on the menu to the above total number of Steckers.

This number is the expression (1) of the preceding paragraph multiplied by the number of ways of intersteckering the remaining 26 - m - c off-menu letters among themselves, with 6 - a self Steckers:

$$\binom{26-m-c}{6-a} \quad S(20-m+a-c) \qquad (3)$$

The desired probability is therefore the product of (1) and (3), divided by (2).

4. Bayes Factor in Favor of a Pattern.

This is the quicker route to the desired Bayes Factor F (m;a, 2b,c). The expression (1) cancels out, leaving F(m;a,2b,c) =

$$\frac{26 \times 25^{m-1}}{\binom{26}{6} S(20)} \binom{26-m-c}{6-a} \quad S(20-m+a-c)$$

This is proportional to the number of ways the given story can be extended to a full Stecker, with a factor of 25 for each menu letter. It is very simple to compute, and can be found as accurately as desired.

5. Alternative Derivation of Jackpot Porbability.

Instead of counting the number of Steckers giving the indicated pattern on a given menu, imagine the Stecker is known, say

$$\begin{array}{cccccccccccccccccccccccc} A & B & C & D & E & F & G & I & K & M & O & Q & S & U & W & Y \\ & & & & & & \overline{L} & \overline{N} & \overline{P} & \overline{R} & \overline{T} & \overline{V} & \overline{X} & \overline{Z} \end{array},$$

and consider that the menu may consist of any one of the $\binom{26}{m}$ sets of m letters. The number having pattern a - 2b - c is

$$\binom{6}{a}\binom{10}{b}\binom{10-b}{c}\ 2^c \qquad (4)$$

The first term is the number of ways of selecting a of the 6 self-steckers A.....F. The second is the number of ways of selecting b of the 10 pairs both letters of which are chosen. The third is the number of ways of selecting c of the remaining 10 - b pairs, of which just one letter is to be chosen, and 2^c is the number of ways of picking the letters in the C pairs.

The desired probability is (4) divided by $\binom{26}{m}$. That this is equal to (1) times (3) divided by (2), as previously derived, is a matter of simple algebra.

6. Computation of Table 1.

The way in which we computed Table 1 was by breaking down the chance answers according to pattern on a percentage basis, doing likewise for jackpots, and taking their ratio (proportion of jackpots divided by proportion of chance answers).

For the first of these, we had the original computations of (chance) stories, broken down according to values of a + 2b ("number of B's which are A's"). Within each of these categories, the breakdown according to values of a is given. It was thus a simple matter to compute the proportion of chance stories in each category a - 2b - c.

For the jackpots, we used expression (4). This had the convenience that only the first factor $\binom{6}{a}$ involves a, and hence the rest - once computed - could be used over and over for each new value of m.

Since the proportions were taken to three figures only, the rarer cases may be a bit off.

E 5 - 31

A. H. CLIFFORD
22 May 1945

TABLE I. Ratio of Probability that given Combination of Selfs and Confirmations will Arise from True Problem to Probability that Same Combination will be Produced by Chance.

Left Hand Digit = S = Number of Selfs
Middle Digit = C = Number of Confirmations
Right Hand Digit = O = Number Outside

S + C + O = Number of Columns

Abbreviations: t = 10, e = 11, w = 12, h = 13, f = 14

10		11		12		13		14		15		16		17	
SCO	R	SCO	R	SCO	R	SCO	R	SCO	R	SCO	R	SCO	R	SCO	R
109	.02	10t	.01	10e	.03	30t	.05	40t	.14	50t	.30	529	.45	629	.87
208	.23	209	.06	20t	.02	12t	.00	22t	.00	32t	.00	349	.00	449	.00
028	.03	029	.01	02t	.00	409	.27	04t	.00	14t	.00	169	.00	269	.00
307	1.3	308	.08	309	.17	229	.04	509	.33	609	.37	628	.40	089	.00
127	.40	128	.10	129	.02	049	.00	329	.12	429	.12	448	.44	548	.71
406	4.9	407	1.9	408	.75	508	.76	149	.00	249	.00	268	.00	368	.00
226	3.3	227	.99	228	.25	328	.47	608	.29	069	.00	088	.00	188	.00
046	.67	047	.11	048	.01	148	.05	428	.69	528	.71	547	1.3	647	.74
505	11.	506	4.3	507	1.7	607	.50	248	.11	348	.23	367	.43	467	.72
325	18.	326	5.9	327	1.7	427	2.1	068	.00	168	.00	187	.00	287	.00
145	7.4	146	1.8	147	.35	247	.71	527	1.7	627	.71	646	1.3	0t7	.00
604	10.	605	5.0	606	1.2	067	.03	347	1.1	447	1.4	466	2.6	566	2.1
424	70.	425	22.	426	7.0	526	5.0	167	.12	267	.23	286	.43	386	.72
244	57.	245	15.	246	3.5	346	5.0	626	1.8	087	.00	0t6	.00	1t6	.00
064	16.	065	2.9	066	.50	166	1.0	446	5.1	546	3.6	565	6.5	665	2.2
523	70.	524	53.	525	16.	625	5.0	266	1.7	366	2.4	385	4.4	485	4.3
343	250	344	81.	345	21.	445	20.	086	.08	186	.23	1t5	.43	2t5	.70
163	123	164	32.	165	6.3	265	9.7	545	12.	645	3.4	664	6.8	0w5	.00
		623	54.	624	15.	085	1.4	365	12.	465	11.	484	19.	584	11.
		443	283	444	78.	544	48.	185	2.4	285	3.5	2t4	6.6	3t4	7.2
		263	218	264	55.	364	59.	644	12.	0t5	.23	0w4	.67	1w4	.67
		083	40.	084	19.	184	18.	464	47.	564	25.	583	47.	683	11.
				543	180	643	43.	284	24.	384	25.	3t3	45.	4t3	32.
				363	345	463	240	0t4	2.9	1t4	5.2	1w3	8.7	2w3	11.
				183	105	283	160	563	105	663	20.	682	38.	0f3	.00
						0t3	20.	383	172	483	93.	4t2	140	5t2	95.
						562	309	1t3	43.	2t3	49.	2w2	70.	3w2	80.
						382	493	662	100	0w3	10.	0f2	123	1f2	15.
						1t2	206	482	513	582	189				
								2t2	359	3t2	274				
								0w2	135	1w2	84.				