Gleason Weights for Shark Text with Samples of 32, 48, 64

by

Lt. G. F. Cramer

October 1944

Abstract

The computation of Gleason weights for the three sample sizes is described, and the results tabulated. In the application to the statistical bombe, a pure ample would arise as the decipherment of a single cipher letter in all the occurrences in the message. Hence this letter must be omitted in the language count from which the weights (for a 25-letter alphabet) are computed. It was found that this made only a trivial difference in the weights except when the cipher letter is E.

Source: ENIGMA Series Volume 9, Article 5
RIP 601, Box 169, 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.

5.

GLEASON WEIGHTS FOR SHARK TEXT WITH SAMPLES OF 32, 48, 64.

Since Gleason Weights depend upon the size of the sample and also upon the relative frequencies of occurence of the various plain text letters, it is necessary to make separate calculations for different samplesizes and for different kinds of plain text. This report contains the results of these calculations for Shark plain text and for sample-sizes 32, 48, and 64.

A set of Gleason Weights consists of a set of numbers W_0, W_1 W_2, W_3 . . . . . calculated by use of the formula*

(1) $$W_k = \text{Log}_{10} \left\{ \frac{1}{c} \sum_{(=)}^{c} \left(\frac{(c-1) f_i}{1 - f_i} \right)^K \left(c \frac{(1 - f_i)}{c-1} \right)^S \right\}$$

where c is the number of distinct letters or cells involved, s is the sample-size and f_i is the probability that a letter of plain text will fall into the ith cell.

The application we had in mind when these weights were computed was one in which we would choose 32 identical cipher letters sprinkled through a Shark message, assume a stecker for these letters, and then decipher them through machines set at proper intervals with respect to each other. The 32 output letters would be "monic" if the assumed stecker and machine positions were correct, but would fall at random into 25 cells in any other case. Hence, in all of this work one output letter is impossible and c = 25. In fact, we need to compute separate weights for each of the omitted letters which would arise from various stecker assumptions. Strictly speaking, the w_k and the f_i should be written as w_{kj} and f_{ij} where the subscript j is determined by the omitted letter. It was found to be sufficiently accurate for our purposes to group the letters into 8 bands as follows: E, N, RSUITA, HLOFCD, MBJGKXWZV, Y, P, and Q.

The values of f_{ij} are obtained from a frequency count of a large number of plain text letters reduced to 1000 letters. The f_{ij} is the probability that, when the $j$the letter is the omitted letter then the output letter will be the ith letter of the alphabet.

Table I shows the actual frequencies F_i of the various plain text letters per 1000 letters and the assumed frequencies F_i which were obtained by taking means of the F_is included in the band in question.

*See memo on "Gleason Weights for Monicity"

Table II contains the values of the various f_{ij}'s. For example, if the input letter is in band 4 and the output letter whose probability is desired is in band 5, then $f_{ij} = .0209$.

The f_{ij} was computed by means of the relations

$$f_{ij} = \frac{f_i}{1000 - f_j}, \text{ when } i \neq j \text{ and } f_{ij} = 0 \text{ when } i = j$$

In the above example, this becomes $.0209 = \frac{20.2}{1000-34.5}$

In the case of the 32 sample, formula (1) becomes

$$w_K = \text{Log}_{10}\left\{\frac{1}{25}\sum_{i=1}^{26} X_i^K y_i^{32}\right\} \text{ where } X_i = \frac{24 f_i}{1-f_i} \text{ and } y_i = \frac{25}{24}(1-f_i)$$

It is no longer necessary to carry the double subscripts on the w_{kj} and f_{ij}. We use the upper limit 26 rather than 25 on the summation because, although one of the f_{ij}'s is zero, we do not specify which one it is. The $\sum$ really covers only 25 non-zero terms. In the above formula, the f_i's corresponding to letters lying in the same band are all alike and we make use of this fact to shorten the computation by combining like terms.

As an example, let us calculate w_1 for the RSUITA band. In this case, $X_1 = 4.149$, $X_2 = 2.493$, $X_3 = X_4 = \ldots = X_7 = 1.647$, $X_8 = .0000$ $X_9 = X_{10} = \ldots = X_{14} = .9144$, $X_{15} = X_{16} = \ldots X_{23} = .5248$, $X_{24} = .2792$, $X_{25} = .2082$, and $X_{26} = .0819$. The corresponding values of Y_i are: $Y_1 = .8881$, $Y_2 = .9436$, $Y_3 = Y_4 = \ldots = Y_7 = .9748$, $Y_9 = Y_{10} = \ldots = Y_{14} = 1.0034$, $Y_{15} = Y_{16} = \ldots = Y_{23} = 1.0194$, $Y_{24} = 1.0297$, $Y_{25} = 1.0327$, and $Y_{26} = 1.0381$. From these we obtain the value $\sum X_i^K y_i^{32} = 20.5367$. Dividing this by 25 and taking the log to base 10, we obtain $w_1 = 9.9146-10 = -.0854$.

The calculations of the w_k's for the 8 bands and for k = 0, 1, 2, 3, ..., 13 were carried out and the results were entered in a table which will be called Table III, but which is not included in this report. This preliminary table had -.1459 as its "most negative" entry, so all the entries were increased by .1459 in order to avoid negative numbers. Then, since it seemed desirable to have weights which are small whole numbers, each new entry was multiplied by 100, divided by 7, and rounded off to the nearest integer. These final weights w_k' were entered in Table IV which is included in the report. An inspection of the final table shows that it would have been reasonable to lump together all bands except the first and use the same weights for all cases with any omitted or input letter except E.

To illustrate the application of the weights, let us consider a case in which 32 Q's occurred in a Shark cipher text and let us suppose that we have assumed correctly that Q is steckered to A. The resulting decipherments at two different settings yielded two letter-distributions of out-put letters, one "right" and one "wrong", having the following numbers of blanks, ones, twos, threes, etc.:

	0	1	2	3	4	5	6	7
"Right" Case	13	4	4	2	1	2		
"Wrong" Case	8	11	4	0	2	1		

We are now ready to apply the previously computed Gleason Weights. We use the "A" weights which are the ones in band 3 of Table IV. The total weights are as follows:

"Right" wt. = 12(4) + 4(1) + 4(0) + 2(1) + 1(4) + 2(9) = 76
"Wrong" wt. = 7(4) + 11(1) + 4(0) + 0(1) + 2(4) + 1(9) = 56

The number of blanks was reduced by 1 before applying the weights, since the crash principle required that the "A" cell be vacant. To distinguish between "right" and "wrong" samples we would need a threshold such that relatively few wrong samples would exceed it and relatively many of the right samples would exceed it. There is to be another report which will deal with the threshold problem.

Computations of the sort already described lead to the following tables V and VI for samples of 48 and 64 letters respectively. Bands 4, 5, 6, 7 were omitted because they must be between bands 3 and 8.

TABLE I

Band	1	2	3						4					
Letters	E	N	R	S	U	I	T	A	H	L	O	F	C	D
F_i	138.5	88.5	66.8	65.0	60.8	59.0	55.1	54.8	41.0	39.0	33.8	31.7	31.4	30.2
F_i	138.5	88.5	60.3						34.5					

Band	5									6	7	8
Letters	M	B	J	G	K	X	W	Z	V	Y	P	Q
F_i	25.8	24.8	24.1	21.5	21.2	17.8	15.8	15.7	15.1	10.9	8.1	3.2
F_i	20.2									10.9	8.1	3.2

TABLE II

Values of f_{ij}

4	5	6	7	8	Band in which j falls
.1434	.1413	.1400	.1396	.1389	
.0916	.0903	.0894	.0892	.0887	
.0624	.0615	.0609	.0607	.0604	
.0357	.0352	.0348	.0347	.0346	
.0209	.0206	.0204	.0203	.0202	
.0112	.0111	.0000	.0109	.0109	
.0084	.0083	.0082	.0000	.0081	
.0033	.0033	.0032	.0032	.0000	

Gleason Wts. for Sample of 32 Letters

TABLE IV

Band	w'_0	w'_1	w'_2	w'_3	w'_4	w'_5	w'_6	w'_7	w'_8	w'_9	w'_{10}	w'_{11}	w'_{12}	w'_{13}
1	4	1	1	2	5	8	12	17	22	27	33	39	45	51
2	4	1	0	1	4	8	14	22	30	39	48	57	66	75
3	4	1	0	1	4	9	15	22	30	39	47	56	65	73
4	4	1	0	1	4	9	15	22	30	38	46	55	63	72
5	4	1	0	1	4	9	15	22	29	38	46	54	63	71
6	4	1	0	1	4	9	14	21	29	37	45	53	62	70
7	4	1	0	1	4	9	14	21	29	37	45	53	62	70
8	4	1	0	1	4	9	14	21	29	37	45	53	61	70

Table III was a preliminary table which is not included in the report.

TABLE V

Gleason Wts. for Sample of 48 Letters.

Band	w'_0	w'_1	w'_2	w'_3	w'_4	w'_5	w'_6	w'_7	w'_8	w'_9	w'_{10}	w'_{11}	w'_{12}	w'_{13}	w'_{14}	w'_{15}	w'_{16}
1	7	2	1	1	2	5	9	13	18	22	28	33	39	45	51	57	63
2	6	2	0	0	2	4	8	13	20	28	36	45	54	63	72	81	90
3	6	2	0	0	1	4	9	14	21	28	36	45	53	62	71	80	89
8	6	2	0	0	2	4	8	14	20	27	35	43	51	60	68	76	85

*Band 4, 5, 6 and 7 were omitted because they must lie between bands 3 and 8.

61 - 6 E

TABLE VI

Gleason Wts. for Sample of 64 Letters

Band	w'_0	w'_1	w'_2	w'_3	w'_4	w'_5	w'_6	w'_7	w'_8	w'_9	w'_{10}	w'_{11}	w'_{12}	w'_{13}	w'_{14}	w'_{15}	w'_{16}
1	8	4	1	0	1	3	6	9	14	18	23	28	34	39	45	51	not computed
2	9	4	1	0	1	2	5	8	13	19	26	34	43	52	61	70	77
3	9	4	1	0	0	2	5	9	14	20	27	35	43	51	60	69	
8	8	4	1	0	1	2	5	9	13	20	26	33	41	49	57	66	

*Bands 4, 5, 6 and 7 were omitted because they must lie between bands 3 and 8.