MOST SECRET

COUNTS

Some of us have been thinking about the possibilities of a somewhat new method of attack. This note summarises the conclusions that have been reached so far. A good deal of experiment and thought is required before we embark on any scheme of mechanisation. I hope that everyone in the M.R. will think about the theory and conduct experiments when they can spare the time. The special case of Brown has not been considered very carefully. A preliminary talk with Keene suggests that mechanisation will be possible when we have decided what we want. Some notes on the function will be produced shortly.

It is assumed that we have a good idea of the ringstellung and at least one cilli. Each possible assumption of wheel order and ringstellung gives a "place" on the machine for the cilli message or messages. The identification of the right place and the discovery of the stecker is based on the peculiarities of the distribution of letters in German clear text, and in particular on the frequency of E, N, X and the rarity of J, Y, C.

Tapes :-

In a particular place we are in the habit of trying out a set of stecker assumptions by decoding the message with only these stecker and looking for signs of German clear text. If we assume the stecker of m letters and the message contains M letters, we expect $(m/26)^2 \times M$ letters in the decode. This process is known as "running a tape."

Counts :-

Suppose that we have a Type X machine with two stecker boards for input and output. Suppose that the input stecker board is plugged up to the assumed set of stecker and the output stecker board is self-steckered. If we now decode the message we shall expect $(m/26)M$ letters in the decode, which I will call the "count letters". If we are in the correct place, these count letters will be the stecker of the corresponding letters in the true decode, and should therefore have a language distribution. Suppose finally that the X-machine, instead of printing out the count letters, tells us that the letters A, B,.....Z occur n_1, n_2, n_3, ...n_{26} times. This process will be known as taking a count.

The test :-

For each count we can calculate the value of the function

$$\chi = \frac{N(N-1)}{\sum n(n-1)}, \text{ where } N = \sum n$$

When the count letters are random, we should expect $\chi = 26$, but when they have a language distribution we should expect $\chi = 15$.
This gives a test which is very powerful for large values on N and surprisingly efficient even for values of N as low as 25.

General principles :-

The exact method of attack will depend on the peculiarities of the problem, such as the number of cillies, the lengths of the messages involved, the length of chain given by the cillies, the occurrence of the letters E, N, X, J, Y, C in the cillies, the type of traffic, and the number of places to be tried. The main idea is to take counts on sets of stecker suggested by the cillies, to select favourable counts by means of the χ test, to examine these further,

and to run a tape as soon as a good story has been obtained. Whenever the assumed set of stecker includes a stecker of one of the letters E, N, X, J, Y, or C, the assumption can be discarded if the count shows an unreasonable number of occurrences of the assumed stecker of this letter.

First stage :-

Consider the most unfavourable case of one cilli giving six separate letters of no particular merit. For example the cilli

TMO → RFV

gives three two-chains

T-R, M-F, O-V.

It is probably best to start by assuming that one of these six letters is self-steckered. In any place T/T implies a stecker of R, and provided the message is long enough, we can do a test on the count obtained from this assumption. This can be done for all six self-stecker and for all places. The best counts would be selected and combined with the counts obtained from the 26 pairs of stecker assumptions derived from one of the other two chains. If the message is short it will be necessary to combine the assumption T/T with the 26 pairs of assumptions derived from M-F or from O-V before taking a count.

Second stage :-

After this preliminary stage we shall be left with a number of sets of stecker assumptions which give good counts and need further investigation. We then begin to guess the stecker of E, N, X, J, Y, C and try out our guesses by comparing the counts that they give with the basic count given by the original assumption. At this stage two points are helpful.

(I) When a single stecker assumption gives a count in which a particular letter occurs more frequently than would be expected, we have some evidence that the stecker assumption is consistent with the letter being the stecker of E, N, or X.

(ii) When a single stecker assumption gives a count in which a particular letter occurs we have evidence that the stecker assumption is not consistent with the letter being steckered to J, Y, or C.

These two points are of course only extreme cases of the statement that each true stecker should give a count which fits in with the language distribution.

When a sufficiently long story has been obtained, a tape can be run. At this stage, it will sometimes be helpful to see how many letters are likely to be self-steckered. If this number is not much greater than six, it will probably be a good thing to add them to the story when running a tape.

Failure :-

If we are unlucky, and no one of T, R, M, F, O, V is self steckered it is likely that two of these letters will be steckered to each other, and we can try the 15 assumptions. Twelve of these should be quite easy to test, because T/M for instance implies stecker of R and F. The remaining three, T/R, M/F and O/V will be tiresome.

Alternative for difficult cases :-

When the message is short, or when the distribution of letters in the text is unfavourable, there may be a grave danger of missing the right place even if one of the six letters is self-steckered. This is partly due to the fact that a self-stecker assumption is only to be expected to produce half as many count letters as an assumption that two different letters are steckered. It may therefore be better to take counts on all 26 pairs of stecker given by each of the three pairings T-R, M-F, O-V. This involves thirteen times as much work, but gives us three chances of spotting the right place. Of course, it may be enough to use only two, or even one of the pairings. It is likely that one pairing will offer a more favourable chance than the others. For example M and F may occur more frequently in the text than the other four letters, so stecker assumptions based on the M-F pairing will produce more count letters than those based on the other pairings.

More favourable cases :-

It seems hardly worth discussing methods for dealing with cases in which we have three chains, or two cillies, or even a female. It looks as if most of the fun will be in the modification of the method to take advantage of special features. But when the material is approaching the standard that we have till now regarded as necessary for hand methods, the idea of testing a story by taking a count and applying the χ test instead of simply running a tape in the old way will make a very great difference. In fact we badly need the hypothetical X-machine.

Taking counts :-

When taking a count on a hypothetical X-machine, I suggested that the output should be self-steckered. It might be better to have the assumed stecker in the output with the rest of the output letters self-steckered and to type out the decode as it is being counted. This might reveal the correct story by a lucky combination of letters.

Methods at present available :-

For each place the alphabets in the positions of the various letters of the message can be obtained from the X machines. From a message of length M we obtain 26 M letter triads $\alpha\,\theta\,\phi$ such that $\theta\,\phi$ is a letter pairing at a machine position in which the letter α occurs in the text. Thus the triad $\alpha\,\theta\,\phi$ means that ϕ is a counts letter corresponding to the stecker assumption α / θ. Of course the triads occur in pairs for $\alpha\,\theta\,\phi$ implies $\alpha\,\phi\,\theta$.

The triads can be entered on half a Foss sheet. For each triad $\alpha\,\theta\,\phi$ the letter ϕ is written in the $\alpha\,\theta$ square or in the $\theta\,\alpha$ square according as $\alpha \geq \theta$ or $\alpha \leq \theta$. The resulting analysis will be called the “count triangle”. Each square of the count triangle represents a stecker assumption and contains the count letters corresponding to that stecker assumption. To take a count on any set of stecker assumptions we record the count letters in the corresponding squares by dots in appropriate places in a row of a small Foss sheet.

A further analysis that is useful at the second stage can be made in the unused squares of the Foss sheet used for the count triangle. If the square YD of the triangle contains 15 letters with an uneven distribution in which P occurs three times and T occurs four times, then in the DY square of the Foss sheet we make the entry 15 - P^3 - T^4.

Time factor :-

Until we can mechanise the analysis the labour required to produce the count triangle is a serious obstacle. Even the production of the alphabets takes more time on the X machines than can easily be spared. It seems reasonable then that, whenever there is a good chance of ruling out places altogether at the first stage the counts should be made from the alphabets. It may be practicable to do the first stage without even having the alphabets done, though this will obviously be tedious.

W. G. W.