THE POSSIBILITY OF SECURE NON-SECRET DIGITAL ENCRYPTION

J. H. Ellis, January 1970

Introduction

1. It is generally regarded as self-evident, that, in order to prevent an interceptor from understanding a message which is intelligible to the authorised recipient, it is necessary to have some initial information known to the sender and to the recipient but kept secret from the interceptor. This information can take many forms, such as the method of encipherment itself, the construction of a cipher machine, a key setting or a one-time tape. All these methods require that there is a route by which this secret information can be sent without fear of interception. Only then can the cipher text be sent safely in a non-secret manner, and large quantities of cipher text of high security thus tend to need the parallel transmission of smaller, but still substantial quantities of secret information.

2. This report demonstrates that this secret information is not theoretically necessary and that, in principle, secure messages can be sent even though the method of encipherment and all transmissions between the authorised communicators are known to the interceptor. This is what is meant by "non-secret encryption". It must be emphasised however that this demonstration has only the status of an existence theorem. It shows only that such a system is theoretically possible, and not that a practical form exists. The demonstration consists of showing that a particular, but unfortunately as yet highly impractical, system has the desired properties. This is followed by an heuristic discussion which attempts to establish the necessary properties of a system and indicate the likely form of a practical solution.

3. As the title indicates we are concerned here with digital transmission. Analogue systems have essentially different properties and possibilities with regard to non-secret encipherment, and they will be discussed only as one illustrative example in the following section.

Possibility of Non-Secret Encryption

4. In what follows the originator of the message will be referred to as the "sender", the authorised recipient as the "recipient", and the unauthorised interceptor as the "interceptor".

5. It seems apparent that the recipient must be in a special position with respect to the interceptor to enable him to understand the message, and as they are both assumed to know the cipher text equally this would seem to imply that the recipient has some knowledge denied to the interceptor, also that this knowledge must be shared with the sender (otherwise how could it be put to use?). At first sight the idea that the recipient might be in the necessary special position just because he was the authorised recipient seems impossible. However consider the following case.

6. An ingenious scheme intended for the encipherment of speech over short metallic connections was proposed by Bell Telephone Laboratories (Ref. 1) in which the recipient adds noise to the line over which he receives the signal. If this noise is sufficiently large compared with the message it can effectively disguise it. The recipient however can subtract the noise from the signal he receives and so obtain the original message. This method has obvious disadvantages and limitations which are irrelevant to this discussion, but it has an important property; if the interceptor were provided with a receiver identical with that of the recipient and connected to the same point on the line, then the two terminals would be identical for all practical purposes and could be interchanged without altering the situation. Nevertheless the interceptor would not be able to read the message as he would not know the noise which had to be subtracted from the line signal. Thus this system fulfils the condition which was discussed above, that the recipient is able to decipher the message because he is the authorised recipient and not because of any special physical position or prior secret knowledge. Clearly, should the interceptor try to usurp this authority by adding his own noise to the line, all that he will succeed in doing will be to block the message from the genuine recipient, revealing his own presence without acquiring any information,

7. Because such a system exists we know that the required property is not an impossibility and we can now try to devise a useful system which possesses it. In the "added noise" scheme the receiver has information not known to the interceptor, namely the exact form of the noise which has been added, but this has been generated by the recipient and is not known to the sender or anyone else. The reason that the recipient is able to use this knowledge is that he takes an active part in the encipherment process, and his special position is due to this fact. In the nature of things the interceptor cannot take an active part in any communication system without actually posing as the recipient and also ensuring that the recipient is not himself participating.

8. The above system is essentially analogue, and the question of whether or not it can be developed into a useful technique is outside the scope of this paper. The problem with which we are now concerned is that of trying to find a digital system which is non-secret. It is implied by the foregoing discussion that it is necessary for the recipient to take an active part in the encipherment process; and, as the only connection we can assume between the sender and the recipient is the ability to pass digita information, this participation by the recipient must consist of sending digits to the sender. We shall now describe a theoretical model of a system which has these properties.

A Theoreticacal Model

9. First we shall generalise the idea of a cipher machine. Essentially acipher machine (or other encipherment process) can be regarded as a device which takes an input (key, text, setting etc.) and produces an output from it. If there is no random element in the machine then the output is defined by the input. Thus we can regard our machine as a look-up table containing one value of output for each possible input value. Here an 'input value' means a complete set of data necessary to define the value of the output; it may be in the form of a setting and serial text or of any other form. Often, but not necessarily, the output defines the input. As the input may consist of data of more than one type we can consider the machine as a one-dimensional or multidimensional table as is most convenient. For instance, if a machine uses 100 bits of message and 100 bits of key to obtain 100 bits of cipher text this could equally well be regarded as a table with 2^{200} possible input values or as a two- dimensional table with two sets of inputs, each of 2^{100} different possibilities. Clearly here is no essential difference in these two views, which correspond merely to envisaging the table arranged as a square oras a column.

10. In regarding cipher machines as look-up tables there is, of course, no suggestion that they should have a physical form of this kind. The table is merely a general method of defining the operation of a machine without having to consider its internal working, and it is one which can be applied to any machine or system. Clearly a table can be formed for any given machine by enumeration and, conversely, a machine could, in principle, be constructed to reproduce the results of any given table. Indeed the table itself, if actually reproduced, would constitute a cipher machine. Thus, if a table can be shown to be possible with certain properties, then a machine with these properties is possible.

11. For brevity we shall call one-dimensional tables "linear", and two-dimensional tables "rectangular" (or "square" if the sides are equal). The output value of a machine M whose inputs have values x, y, z, ... will be written M(x,y,z,...). The terms "output" and "input" will be used to include "output value" and "input value" where it is felt that no confusion will arise.

12. We can now describe the proposed system, and a block diagram is shown in Figure 1. The sender has a message p which he wishes to transmit securely to the recipient. The recipient generates a random key k which he enciphers by means of the machine Ml, forming the enciphered key x. The sender uses x and M2 to encipher p to form the enciphered message z. Finally the recipient uses k and M3 to decipher z and so obtain p.

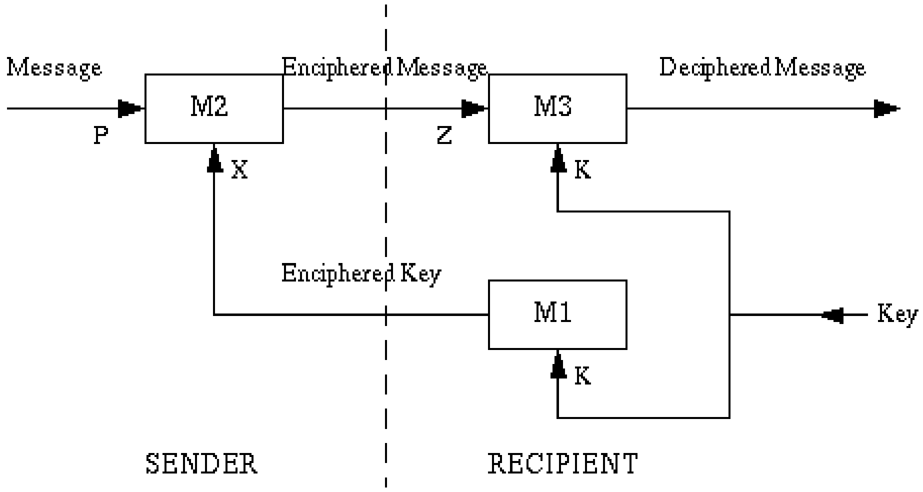

Fig. 1

13. The following properties are clearly essential. It must be impossible for the interceptor to obtain p from z without knowing k even though he knows x. Also, since a knowledge of k would enable him to decipher z, he must be unable to obtain k from x. Finally M3 must have the property of being able to decipher z. To obtain these properties we specify the look-up tables corresponding to MI, M2 and M3 in the following way: -

- **a.** Let k have n different possible values and p have m different possible values, for simplicity take them to be the integers 1 to n and 1 to m respectively. Let x have the same range of values as k, and z have the same range as p.

- **b.** MI can be defined as a linear look-up table of n entries whose contents are the numbers 1 to n in a random order, where "random" implies that the output is sufficiently uncorrelated with the input so that the position of a particular entry in the table cannot be found in a simpler way than by searching through the table.

- **c.** M2 corresponds to an n by m rectangular table in which the entries for a fixed value of x consists of the numbers 1 to m in random order, and where the columns for the various values of x are suitable uncorrelated with one another.

- **d.** M3 is an n by m table in which each entry is the value of p corresponding to the values of k and z which locate it. In other words the kth column of M3 is obtained from the xth column of M2 [x = M1(k)] by making the zth entry in the M3 column the address of the number z in the M2 column. This is shown in Figure 2.

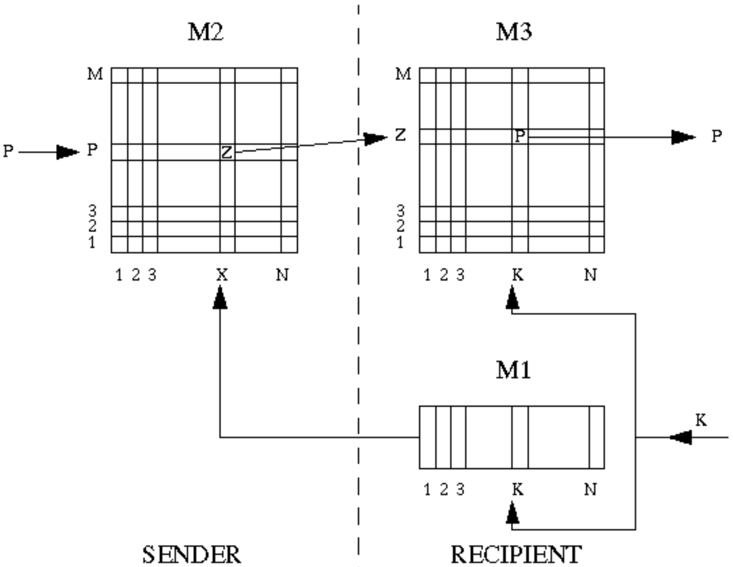

Fig. 2

14. It is evident that this process gives the recipient the message. Now consider the position of the interceptor. Firstly he has the values of x and z and the details of M2. This means that he has enough information to define p, for he has only to search column x of M2 to find the entry z in order to establish p. However this column has m entries in random order, and so he would need to make a search of these m values since we have defined "random" to imply just this. m is thus the work factor in this form of attack, and as m is 2^{100} for a 100 bit message (about 20 letters) this factor can easily be made adequately large. Similarly the work factor in obtaining k from x is n, which can also be made large. It is clear from the random nature of the tables that there is no simpler attack; thus we have shown that the system can be guaranteed to have a work factor which is the smaller of m and n, both of which can easily be made amply large. This is in spite of the fact that details of all machines are known to the interceptor, who may be supposed to have copies of them, and that no prior secret information is held in common by the sender and the recipient. Thus the system satisfies our requirements.

15. This may seem a specious argument on the grounds that the effort of making, say, 2^{100} trials is regarded as prohibitive for the interceptor while the making of a device containing 2^{200} entries is accepted for the authorised communicator. However there is no suggestion that this is a proof that such a system can be made, certainly not that a practical machine would be constructed in the form of a look-up table. The argument is that the operation of a machine is defined by its equivalent look-up table, and that if a system with a look-up table defined in a certain non-contradictory way has certain properties then there is a hope that a real system can be devised having these properties. All that has been proved is that a theoretical system having such properties exists; and clearly there is an enormous number of functions which would satisfy such a system.

Possibility of a Practical Model

16. To help us assess the probability of being able to devise a practical model we shall examine two questions. Firstly "Would a practical model be similar to the theoretical one?" and secondly, "Can machines having the essential properties of Ml, M2 and M3 be devised?".

17. To answer the first question let us return to the basic problem, which is that we have a situation in which the sender and the recipient have some form of cipher machine whose details are known to the interceptor, who also has full knowledge of all traffic which passes between them, but where the sender needs to send a message to the recipient without the interceptor being able to read it. It is clear that the recipient must introduce some random element into his transmissions otherwise they would be entirely predictable and serve no useful purpose. The sender may or may not produce some random contribution, but there is no indication of the need for this. The interchange could consist of a continuous dialogue or simply a block of data sent by the recipient and replied to by the sender.

18. Some properties of the system can be deduced.
An exhaustion attack is always possible.

19. Since the message must be uniquely decipherable, only one message can correspond to the signals transmitted. Therefore any message (including any random element which may be included in the sender's process) which, when combined with the recipient's signals, gives the sender's signals must be the correct one. Hence an exhaustion attack by trying all possible messages will always succeed. In this case "Messages" includes any random element although the random element need not, of course, be decipherable. (In what follows the word "message" will always have this meaning, as a random element could always be added to the message, and although the fact that the random element does not need to be decipherable could be of importance in a practical case it does not effect arguments of the present paper. Therefore we shall neglect the possibility of the sender using a random element).

20. In a similar way an exhaustion attack is always possible with the random element used by the recipient, which we shall call the key. From this it clearly follows that:
The maximum possible work factor of the system is the smaller of the number of different possible messages and the number of different possible keys.

21. These facts establish that it is important that the message and keys are each used in large blocks to obtain security and it does not seem practicable to perform the interchange between sender and recipient in dialogue of less than block length as the initial exchanges would depend on less message and key and would thus provide a weakness. This latter point is clearly not strictly true, but, as the encipherment must be based essentially on long blocks, the dialogue idea will not be further considered. It is also impossible to improve the security by making the message of the form of a random number such as a key-setting. This is unfortunate as one very profitable use for such a system would be remote key setting. However the work factors obtainable from moderate lengths of message can be very large. It is also evident that stylised messages must be padded with random material since complete cribs can be readily tested.

22. In operating this system both sender and recipient will produce cipher text using either message or key in conjunction with data which has been received from the other and is thus known to the interceptor. Clearly it is essential that the interceptor cannot work backwards and find the message or key, and so in this sense we have:

The encipherment must be irreversible.

23. These results indicate that a system will be essentially of the form where the sender has a message of a standard length which is long enough to defeat an exhaustion attack (say upwards of 100 bits), and the recipient generates a similar length of random key. The recipient enciphers the key by some process and sends the result, which we will call "enciphered key", to the sender, who uses it in turn to encipher the message to form an "enciphered message" which he transmits to the recipient. Finally the recipient must use his original key to decipher the message. This is the system illustrated in Figure 1, and thus we imply an affirmative answer to the first question of Paragraph 16.

24. To answer the second question let us look rather more closely at the general properties of cipher machines.

25. In the case of a linear machine there is one output for each input, and if these outputs are all different from one another then the input is defined by the output and we will say that in this case the machine is "defined". The ess correlation there is between input and output the better the encipherment will be. Intuitively one would say that, ideally, the output should depend on the input in a random manner; but, as the input-output relationship is fixed, and any relationship is possible in a random choice, we will say that the machine is "random" if the input cannot be correlated with the output in any way which is simpler than by enumeration. We shall not attempt to define "simple". There will be degrees of randomness which will depend on the ingenuity of the cipher machine construction in a practical case, in the same way that the stream of key from a good key generator is apparently random although produced according to a relatively simple law. Also in practice it will be necessary for the output to be produced from the input by some means other than enumeration. We shall therefore replace the concept of randomness by that of "Irreversibility". A linear machine will be regarded as perfectly irreversible if no way of obtaining the input from a given output could be found which is simpler than enumeration, and satisfactorily irreversible if no way can be found which has a work factor smaller than that which is regarded as acceptable. It follows that a random machine is also irreversible. A machine which is both irreversible and defined we shall call "ideal".

26. We may extend these definitions to include higher dimensions as follows:-

a. A machine is irreversible with respect to an input if, for any given set of values of the other inputs, there exists no way of deriving the input value from a given output value which is simpler than by enumeration of the values of the input.

b. A machine is defined with respect to a given input if, for any fixed set of values of the other inputs, there is a different value of the output for each different value of the input.

c. A machine is ideal with respect to an input if it is both irreversible and defined with respect to that input.

27. If, when the output of a certain linear machine is used as the input of another linear machine, the output of the second machine is always the same as the input to the first we shall say that the second machine is the inverse of the first. Clearly if the inputs and outputs have the same range of values the first machine is also the inverse of the second. A pair of rectangular machines, each of which has a fixed value applied to one input could be inverse in this way (e.g. M2 and M3). It is clear that for any defined machine there exists a possible inverse, although the inverse of an ideal machine would have to be produced by enumeration.

28. We can now say that MI is an ideal linear machine, that M2 is a rectangular machine ideal with respect to the p input, and that M3 is a rectangular machine which is the inverse of M2 when M3 is fed with k and M2 with x=M1(k).

29. There is no problem in devising an ideal linear machine. Any good key generator with the settings regarded as the input and the key as output is a good enough approximation to this. Similarly there is no problem in making an ideal M2; but when we add the requirement of the existence of a practical M3 we meet difficulties. Firstly there appears to be a contradiction in the requirement for M2 to be ideal and to have an inverse (M3). This would be a genuine impossibility for a linear machine, for our definition of irreversibility requires that there shall be "no way of obtaining the input from a given output which is simpler than enumeration"; a suitable work factor would, by definition, preclude enumeration, and so the existence of an inverse machine would establish that the first was reversible. In the case of a rectangular machine however it is possible for it to be irreversible with respect to one input and yet still have a machine which is its inverse for that input under certain conditions. This is because it is irreversible given the output and the other input, not for a fixedvalue of the other input. Thus in the case of M2 it could be reversible with respect to p for each fixed value of x, in the sense that, for every value of x there exists a linear machine which is the inverse of M2 for that particular x input; but if the search for the particular machine corresponding to a given x input involves enumeration comparable to that of enumerating p then M2 is still irreversible with respect to p. In other words a multidimensional machine can be irreversible with respect to an input even though the linear machines formed from it by fixing the values of the other inputs are reversible.

30. Resolving this apparent contradiction has demonstrated the real problem of designing M2, which is that of producing a rectangular machine which is ideal with respect to p but where the linear machines formed by fixing the x input are reversible.

31. Two possible methods of solving this problem seem plausible. The first is to make M2 some sort of reflex process involving a linear irreversible machine in such a way that the inverse process involves the input to this machine, but where the encipherment can be done using only the output; in other words the input is never used directly in the encipherment process, only via the machine. The machine could then become MI and the required properties of M2 follow. A solution of this form would be very useful as it would enable any convenient type of key-generator to be used for MI. However such a solution may prove to be essentially impossible. The other approach is to find a process in which finding the inverse depends on knowing the value of an inverse function of some value used in the process. The function in question then becomes MI. This would depend on special properties of particular functions.

32. Attempts to find solutions in this way have so far been unsuccessful but have seemed tantalisingly near to success at times, so that one feels that a solution probably does exist. There is certainly no apparent reason to believe it impossible.

Conclusions

33. In assessing the implications of the above arguments it is necessary to distinguish carefully between fact and opinion, i.e. between that which has actually been proved and that which seems likely. It is particularly difficult to do this in this case because we have established something, which, to most people, seems inherently impossible. Our tacit assumptions and inferences have therefore to be watched with particular care.

34. What has been shown rigorously, is that a digital system in which no secret information is shared by the sender and recipient is theoretically plausible. Of this the disproof of the assumption that the sender and the recipient must share secret information for any secure communication to be possible is the major step, and this is achieved by considera- tion of the system in which the recipient adds noise to the line. The demonstration of the digital system adds little definite to our knowledge, since it does not show that it can be done, only that it is theoretically possible in the digital case. However there are some indirect advantages which may be obtained from consideration of the digital system. It seems likely that many people would assume, in the absence of contrary evidence, that the properties of the analogue system were dependent entirely on the physical conditions prevailing in it, and so would not consider a study of a system using orthodox digital transmission worthwhile. Moreover it is possible that the heuristic discussion contained in the previous section may indicate the necessary form of practical fulfilment, although assumptions made on this basis should be scrutinised carefully as the arguments are by no means rigorous.

35. Another potential advantage to be obtained from the theoretical system is that it could provide a counter-example of a tentative hypothesis. The practical realisation of a non-secret digital scheme depends of course on there not being some basic theorem which forbids it. Any postulated theorem which forbade the theoretical existence of the system presented would thus be untenable and not worth pursuing. This may be more worthwhile than at first seems likely as the only apparent disadvantage of the system is the enormous size and cost, and if counter theorems are to be limited to this aspect then the ultimate possibility is raised substantially above that which would prevail if any sort of theoretical limitation were possible.

36. A consideration which encourages belief in the existence of a practical solution is that the number of different functions which satisfy the conditions for the look-up tables of M1, and M3 is clearly enormous and only one set capable of practical generation is needed. This is not to say, of course, that a solution of the problem seems at all imminent. Even if there is no fundamental difficulty, we know that the gap between showing something to be possible (which we have not yet done, only that it may be possible) and actually doing it can be immense.

37. It may be that the best way to continue this study is to assume properties for the functions MI, M2 and M3 (such as, for example, association or permutation) and see what can be proved as a consequence. This may serve to narrow the vast field of possible functions which exists at present and makes a random search impracticable. In any case it is hoped that this report will stimulate those proficient in these matters to find a practical solution. The potential advantages are too obvious to need specification.

Reference: (1) "Final report on project C43." Bell Telephone Laboratory, October, 1944, p.23.