Perfect Secrecy

We shall show in these notes that a probabilistic form of "perfect secrecy" can be achieved, if circumstances permit the use of a sufficiently large key space. The setting we shall work with is that of a *random cryptographic system.*

As in our previous handouts, the ingredients are:

a) A "MESSAGE SPACE"
$$M = \{m_1, m_2, \ldots, m_n\},$$

b) A "CIPHER SPACE"
$$C = \{c_1, c_2, \ldots, c_r\},$$

c) A "KEY SPACE"
$$K = \{k_1, k_2, \ldots, k_s\}$$

d) A set of one-to-one maps of M into C
$$E_k(m) : M \to C$$

e) Two sets of probabilities
$$\{p_1, p_2, \ldots, p_n\} \quad \text{and} \quad \{q_1, q_2, \ldots, q_s\}$$

This given, a cryptographic transaction in such a system takes place as follows:

1) The sender produces a message M which is a random variable with
$$P[M = m_i] = p_i$$

2) The sender selects a key K by an independent mechanism with
$$P[K = k_i] = q_i$$

3) The sender encrypts M into $C = E_K(M)$ and sends it to the receiver.

Thus our cryptographic transaction here may be viewed as an *experiment* producing the three random variables
$$M, K \quad \text{and} \quad C = E_K(M)$$
This given, our aim here is to find out under what circumstances the opponent, even if completely aware of the mechanism involved, upon intercepting C, can draw no conclusions whatsoever about the original message M . When this happens we shall say that our system achieves "*Perfect Secrecy*".

Probabilistically, C yielding no information about M can only mean one thing: that M and C are **independent** random variables. This means that we have "*Perfect Secrecy*" if and only if, for all choices of m in M and c, in C we have
$$P[M = m, C = c] = P[M = m]P[C = c] \tag{1}$$

From this definition, we easily see that

Theorem 1 *Perfect secrecy is achieved when*

1. *All keys are equally likely*
2. *For each pair (m_i, c_j) there is a unique key, k_s, such that*

$$E_{k_s}(m_i) = c_j$$

Proof.

$$P(C = c_j) = \sum_{i=1}^{N} P(M = m_i) \sum_{E_{k_s}(m_i)=c_j} P(K = k_s)$$

But if there is only one key k_s yielding $E_{k_s}(mi) = c_j$ then the inner sum reduces to a single term, and if all keys are equally likely then the inner sum reduces to $1/S$ and

$$P(C = c_j) = \sum_{i=1}^{N} P(M = m_i)\frac{1}{S} = \frac{1}{S}$$

On the other hand

$$\begin{aligned} P(M = m_i, C = c_j) &= \sum_{E_{k_s}(m_i)=c_j} P(M = m_i)P(K = k_s) \\ &= P(M = m_i)\frac{1}{S} \\ &= P(M = m_i)P(C = c_j) \end{aligned}$$

QED

Now it develops that this definition places some severe restrictions on our cryptographic system and that we can in fact give a complete description of all such systems. To this end note first that under our hypotheses, (more particularly the independence of M and K) , we have

$$P[M = m, C = c] = \sum_{k \in K \& E_k(m)=c} P[K = k]P[M = m]$$

The reason for the presence of summation is that we do not exclude that there may be more than one key that sends m into c. Now, substituting this in (1) and cancelling the common factor $P[M = m]$, condition (1) can be rewritten as follows:

$$P[C = c] = \sum_{k \in K \& E_k(m)=c} P[K = k] \tag{2}$$

For a better understanding of what is going on here, let us construct a bipartite graph $\mathcal{G}$ with opposing vertex sets M and C[1] where, between a message m and a cipher c we place an edge *labelled* k if and only if $E_k(m) = c$.

Now we can see a bit better the implications of (2). We assume that each cipher may occur with non zero probability, this is to say that for each c, $P[C = c] > 0$. Now the left hand side of

[1]This simply means that the edges of $\mathcal{G}$ join elements of M to elements of C.

(2) being different from zero implies that at least one term in the right hand side must not vanish. Therefore at least for one k we must have $E_k(m) = c$. Since this must also be true for each message, we deduce that, in our graph $\mathcal{G}$, each pair of vertices (m, c) must be joined by at least one edge.

Now fix a message m and consider all edges coming out of it. Clearly there will be an edge for each $k \in K$. Since there are a total of s different keys, there will necessarily be exactly s edges coming out of m. On the other hand by our previous observation m is connected to each $c \in C$, and since there are r different ciphers, there must also be at least r edges coming out of m. This gives $s \geq r$ that is

$$\#K \geq \#C \tag{3}$$

Moreover, since each encrypting transformation $E_k(m)$ is an injection of M into C, we must have at least as many ciphers as messages. This gives $r \geq n$ that is

$$\#C > \#M \tag{4}$$

In summary, we are led to the following remarkable conclusion

Theorem 2 *In a system with perfect secrecy there must be at least as many keys as messages.*

Proof

The conclusion follows by combining (3) and (4).

However, we can deduce further consequence from (2). Let us see how we can characterize at least the simplest such system. In an effort to keep the key space as small as possible we shall require that

$$\#K = \#C \tag{5}$$

Let us now consider the consequences of this assumption. First of all, we see that this forces the summation in (2) to be a single summand. For if there was more than one key sending m to a given c, then there would have to be at least one more key than there are ciphers and (5) could not possibly hold true. In other words, for given m and c, the equation

$$E_k(m) = c$$

has only one solution in k. Call the solution

$$k_m(c).$$

This given, we immediately deduce that, for any fixed m, the map

$$c \to k = k_m(c) \tag{6}$$

must be a bijection of C onto K. Using the same symbols to denote keys and ciphers, the map in (6) may be viewed as a permutation of $1, 2, \ldots, s$. Moreover, we must have altogether n such permutations: namely one for each message.

We can easily see that the graph $\mathcal{G}$ itself as well as the maps $E_k(m)$ can be recovered from the knowledge of the permutations defined by (6). Indeed, we simply construct $\mathcal{G}$ by joining a given m to a given c by an edge labelled k whenever

$$k_m(c) = k.$$

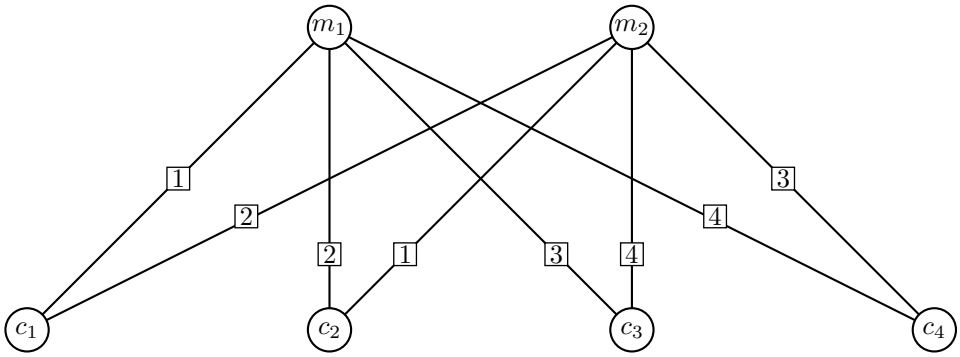
For example in case $n = 2$, $r = s = 4$ with

$$M = \{m_1, m_2\} \quad \text{and} \quad C = \{c_1, c_2, c_3, c_4\}$$

we could take

$$k_{m_1} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix} \quad \text{and} \quad k_{m_2} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix}$$

This gives the graph



In fact, from what we have shown, it follows that the graph $\mathcal{G}$ under the condition (5) is obtained by simply joining each m to each c . That is what is usually referred to as a "*complete bipartite graph*". Once we have drawn such a graph to obtain a perfect secrecy system we must assign the labels to the edges. This amounts to defining the permutations $k \to k_m(c)$. Finally we must give the keys probabilities in such a manner that (2) holds true. Now the latter can easily be done. Indeed, as we have observed the summation in (2) reduces to one term, namely (2) reduces to

$$P[k = k_m(c)] = P[c]$$

This implies that all the keys leading to the same c, must have the same probability. To make matters as simple as possible, we give all keys the same probability $1/s$. That forces all the ciphers to have the same probability as well.

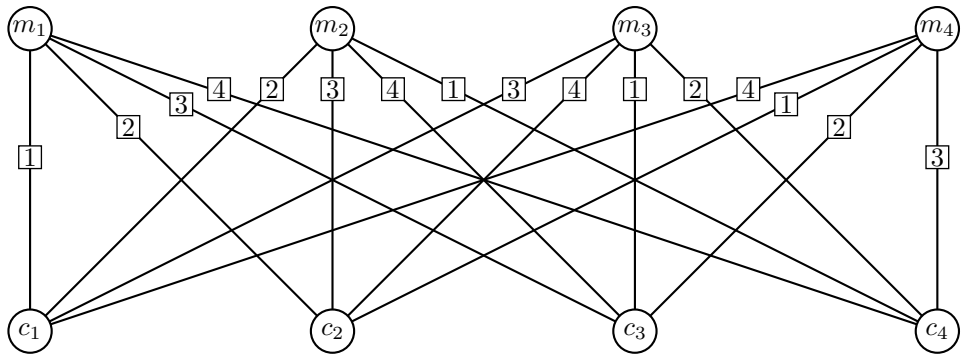
The simplest secrecy systems are those for which the cipher space, the message space and the key space have all the same number of elements. For simplicity we identify messages and ciphers with the integers $1, 2, \ldots, n$. Now, there is a very nice description of such a system. Namely, we shall have one such system for each $n \times n$ board whose squares are filled with the integers 1,2,...,n in such a manner that in each row and each column appears a permutation of the numbers $1, 2, \ldots, n$. These configurations are called *latin squares*. An example of one such square for n=4 is given below:

$$M = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{bmatrix} \tag{7}$$

Given such a configuration the encrypting transformations are easily constructed. For instance to find the image of message m_3 under key 4 , we look at column 3 and see in which row 4 lies, in this case 4 is in the 2nd row, that gives

$$E_4(3) = 2.$$

Thus the graph G corresponding to this cryptographic system is as given below:



We may look at the configuration in (7) as a matrix and decompose it as a linear combination of 4 permutation matrices. Namely,

$$M = 1 \times \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} + 2 \times \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} + 3 \times \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + 4 \times \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

We should point out that the simplest way to construct a latin square for any given n is to take the matrix of the n-cycle

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & \dots & n-1 & n \\ 2 & 3 & 4 & 5 & 6 & \dots & n & 1 \end{pmatrix}$$

For instance, for $n = 5$, that is

$$C = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Then the successive powers of this matrix yield a latin square by the formula

$$M = C + 2 \times C^2 + 3 \times C^3 + ... + n \times C^n \tag{8}$$

The reader may experiment with the case $n = 5$ and perhaps find a proof of this assertion for the general case.

Finally we should point out that the *first available matching* algorithm may be used to construct all possible such latin squares. For in essence we have to decompose a matrix with all entries equal to one, into a sum of permutation matrices. Each of these permutations may be considered as a *match* of *boys* (= messages) to *girls* (=ciphers). After a match is found, we remove from our graph the edges corresponding to this match, then start all over again to find another match in the remaining graph. The process is repeated n times until all edges have been removed. It can be shown that the process can stop only when there are no more edges. Furthermore, it can also be shown that this procedure will produce all possible latin squares.

Exercises:

1. Construct a 5×5 latin square and draw the graph of the corresponding perfect secrecy system.

2. Verify the relation (8) for n=5.

3. Suppose a random cryptographic system is obtained as follows: The message consists of two random variables (X_1, X_2) obtained by spinning twice a roulette with arcs "A" , "B" of lengths 1/4 and 3/4 respectively. Let there be two equally probable keys "1" and "2" and let

$$\begin{aligned} E_1(X_1, X_2) &= (X_1, X_2) \\ E_2(X_1, X_2) &= (X_2, X_1) \end{aligned}$$

   in other words E_1 leaves the message alone and E_2 interchanges the two components.

   (a) Is this a perfect secrecy system?

   (b) Calculate, $H(K)$, $H(K|C)$, $H(M)$, $H(K|M, C)$.

4. Let $X_1, X_2, X_3, \ldots$ be generated by successive spins of the roulette of problem 3. Suppose the message M consists of $2N$ letters. Let the key be as in problem 3. with E_1 leaving the message alone and E_2 interchanging X's in odd position with X's in even position.

   (a) Show that $H(M|C) = H(K) - H(K|M, C)$

   (b) Show that $H(M|C) \to H(K)$ as $N \to \infty$.

5. Use the some of the basic theorems of information theory to show that a system achieves perfect secrecy if and only if

$$H(K|C) = H(M) + H(K|M, C)$$