# Methods of Cryptanalysis
# Lecture 7

Lecture Notes for a course given by Alex Biryukov
at the Weizmann Institute of Science, Israel.
Notes taken by Olga Grinchtein

May 2, 2000

# Lecture 7

# Shannon's Theory of Secrecy Systems

Claude Elwood Shannon was born in Petoskey, Michigan, on 30 April, 1916. He majored in electrical engineering and mathematics at the University of Michigan and there developed interest in communications and cryptology. He was awarded Ph.D. in mathematics at MIT in 1940. After a year at the Institute of Advanced Study in Princeton, he joined the staff of the Bell Telephone Laboratories. There he has built a maze-solving mouse, a chess-playing machine. He was also an expert tightrope walker and uncycle rider. During the World War II Bell Labs were working on secrecy systems. Shannon was working both on communication systems and was appointed as a member of several committees studying cryptanalytic techniques. Thus his work on mathematical theory of communications [6] (where he laid foundations of information-theoretic approach to computer science) and cryptology went forward concurrently since 1941 and till 1944, when most of the work was finished. Both works were published later in the Bell System Technical Journal (1948 and 1949 respectively).

Shannon's work on secrecy systems, contains two main parts: theoretical secrecy and practical secrecy. In the first part Shannon applies mathematical apparatus developed by the "theory of communications" to cryptography. He defines random, pure, perfect and ideal types of cryptosystems. He also defines the notion and gives the formula for the "unicity distance" of a random cipher. This is the amount of intercepted text required to solve the cipher in a unique way. He studies the conditional entropy of the message $H_E(M, N)$ and the key $H_E(K, N)$ (also called equivocation) given an intercepted cryptogram $E$ of length $N$. Shannon shows that perfect security in information-theoretic setting can be obtained only in the extreme cases (when the amount of the key is longer or equal to the amount of communication, or when ciphers are tailored to language redundancy types) and that most of the classical ciphers (pure) leak information that can be used for their cryptanalysis.

He then turns to the "practical secrecy" and puts forward computational complexity approach to cipher design. He shows that using the ideas of ergodic theory (a theory that studies measure preserving mixing transformations) one can construct strong ciphers. Although not unbreakable in theory (exhaustive search attack is always possible), these ciphers will require exponentially high effort from cryptanalysis. He suggested to use repeated products of non-commuting operations as good mixers. As an example one may consider a sequence of substitutions and linear transformations (for example adding adjacent letters mod 26, or bitwise permutations).

## 7.1 Secrecy Systems

A secrecy system is defined as a set of transformations $T$ of the set of possible messages $M$ into the set of possible cryptograms $E$. Each particular transformation $T_k : M \to E$ of the set $T$ corresponds to enciphering with a particular key $k$. Transformations are supposed to be $1-1$ so that unique decryption is possible when the key is known. A diagram of a general secrecy system is shown in Figure 7.1. At the transmitting end there are a message source and a key source. The key

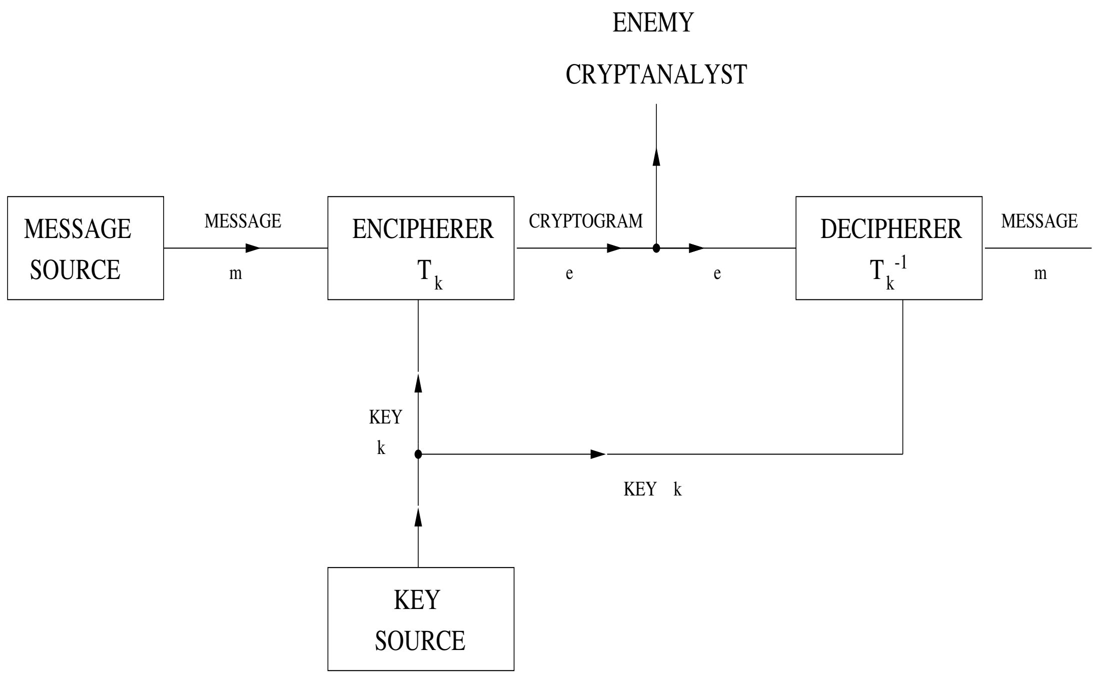


Figure 7.1: Schematic of a general secrecy system

source produces a key from among those which are possible in the system. This key is transmitted in a non-interceptable way to the receiving end. The message source produces a message which is encyphered and the resulting cryptogram is sent to the receiving end. At the receiving end the cryptogram and the key are combined in the decipherer to recover the message. If $m$ is the message, $k$ the key, and $e$ the enciphered message, or cryptogram, we may present the enciphering and deciphering operations as

$$e = f(m, k)$$
$$m = g(e, k)$$

However, it is preferable to think of this, not as a function of two variables but as a family of transformations

$$e = T_k m$$

At the receiving end it must be possible to recover $m$, knowing $e$ and $k$. Thus the transformations $T_k$ in the family must have unique inverses $T_k^{-1}$ such that $T_k T_k^{-1} = I$, the identity transformation.

Thus

$$m = T_k^{-1} e.$$

This inverse must exist uniquely for every $e$ which can be obtained from an $m$ with key $k$.

Assume that there are only a finite number of possible keys, each has an associated probability $p_i$ and there are a finite number of possible messages $m_1, m_2, \ldots, m_n$ with associated *a priori* probabilities $q_1, q_2, \ldots, q_n$. The possible messages might be the possible sequences of English letters all of length $N$, and the associated *a priori* probabilities are then the relative frequencies of occurrence of these sequences in normal English text. If the enemy intercepts the cryptogram he can calculate from it *a posteriori* probabilities of the different possible messages and keys which might have produced this cryptogram. Thus the *a posteriori* probabilities constitute knowledge of the key and message after interception.

Consider a simple substitution with random key. There are 26! transformations, corresponding to the 26! ways we can substitute for 26 different letters. These are equally likely and each therefore has an *a priori* probability 1/26!. If the enemy intercepts $N$ letters of cryptogram in this system the probabilities of different messages for him change. If $N$ is large enough there is usually a single message of *a posteriori* probability nearly unity, while all others have a total probability nearly zero. Thus there is essentially a unique solution to the cryptogram. For smaller $N$ there will usually be many messages and keys of comparable probability. In this case there are multiple solutions to the cryptogram.

## 7.2 Representation of Systems

Information theory defines the amount of information in a message as the minimum number of bits needed to encode all possible meanings of that message, assuming all messages are equally likely. The amount of information in a message $m$ is measured by the entropy of a message, denoted by $H(m)$. In general, the entropy of a message measured in bits is $\log_2 n$, where $n$ is the number of possible meanings. This assumes that each meaning is equally likely.

Let $M$ be a random variable which takes on a finite set of values $m_1 m_2 m_3 \ldots m_n$ with probabilities $P(M = m_i) = p_i$, $0 \leq p_i \leq 1$, $\sum_{i=1}^{n} p_i = 1$.

**Definition 7.2.1** *The* **Entropy** *or* **Uncertainty** *of $M$ is defined to be $H(M) = -\sum_{i=1}^{n} p_i \log p_i$, where by convention if $p_i = 0$ the corresponding term is omitted from the sum.*

Properties of this function are:

1. $0 \leq H(M) \leq \log n$
2. $H(M) = 0 \Leftrightarrow \exists i$ such that $p_i = 1$, $\forall j \neq i$, $p_j = 0$
3. $H(M) = \log n \Leftrightarrow \forall i \; p_i = 1/n$

Denote by $A$- alphabet of the language. If $M \in A^N$ is an $N$-gram of the language, the corresponding entropy is called $N$**-gram entropy** and is denoted by $H_N$. Then the $N$-gram entropy of the language is:

$$H_N = H(X_1, \ldots, X_N) = \sum_{s \in A^N} -P(\chi_N = s) \log P(\chi_N = s),$$

where $\chi_N = (X_1, \ldots, X_N) \in A^N$.

**Definition 7.2.2** *For a given language, the* **rate** *or* **entropy** *of the language is define as*

$$R = \lim_{N\to\infty} \frac{H_N}{N},$$

In his paper "Mathematical Theory of Communication" [8] Shannon described series of approximations to the natural language:

- $0^{th}$ order- symbols are independent and equiprobable

  XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.

- $1^{st}$ order- symbols independent but with frequencies of English text

  OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.

- $2^{nd}$ order- digram structure as in English.

  ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.

- $3^{rd}$ order- trigram structure as in English

  IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE.

In the case of the English language successive approximations of $H_N$ go as: $\log_2 26 \approx 4.7$, the first order approximation (English letter frequencies) gives $\approx 4.2$, digram frequencies give $\approx 3.9$. Shannon [8] suggests a value of 2.3 bits per letter for eight letter blocks. By various experiments, for large $N$ the entropy decreases into the interval $1.0 \leq H_N \leq 1.5$. By a gambling technique Cover and King [2] give an estimate of 1.3 bits of information per letter. In [1] an upper bound of 1.75 bit per letter is estimated from a large sample of ASCII encoded English text.

**Definition 7.2.3** *The* **absolute rate** *of a language $R_0$ is the maximum number of bits that can be coded in each character, assuming each character sequence is equally likely.*

If there are $L$ characters in a language, the absolute rate is $R_0 = \log_2 L$. This is the maximum entropy of the individual characters. For English, with 26 letters, the absolute rate is $\log_2 26$ or about 4.7 bits/letter. The actual rate of English is much less than the absolute rate, since natural language is highly redundant.

**Definition 7.2.4** *The* **redundancy** *of a language, denoted by $D$ is defined as $D = R_0 - R$.*

This number shows how much a text in the language can be reduced in length without losing any information. As an example, since $u$ always follows $q$ in English words, $u$ may be omitted without loss. Given that the rate of English is 1.5, the redundancy is 3.2 bits/letter. This means that each English character carries 3.2 bits of redundant information.

Define a model in which the messages are divided into two subsets (see Figure 7.2). The first contains $2^{RN}$ meaningful messages, each with the same *a priori* probability $2^{-RN}$. The remaining messages are meaningless in the language and are assigned *a priori* probabilities of 0. We also

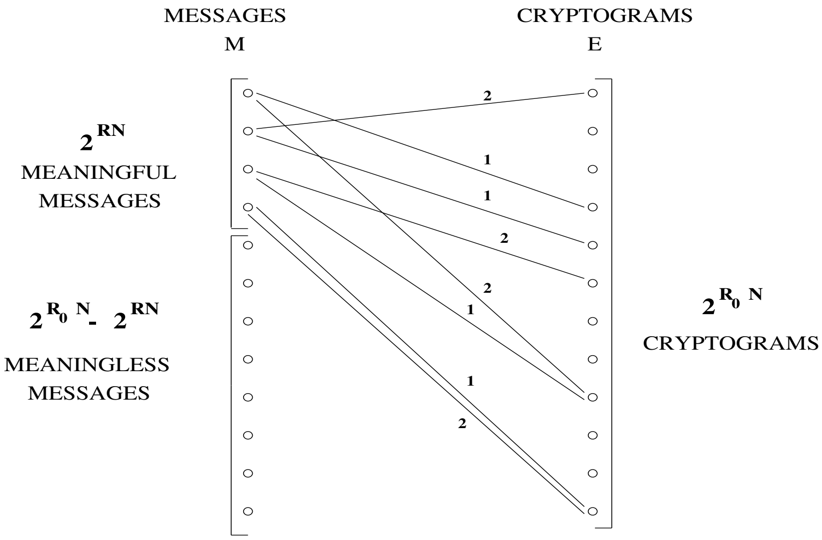


Figure 7.2: Representation of cipher

assume that there are $2^{H(k)}$ keys, all equally likely *a priori* and independent of the message. $H(k)$ is then the entropy of the key measured in bits. Referring to Figure 7.2, a line with a number $i$ in it indicates that the message at the left is encrypted into the cryptogram at the right when the $i$th key is used. If the first meaningful message in Figure 7.2 is enciphered with key #1, then the cryptogram is breakable, since the resultant cryptogram has no other meaningful messages into which it can be deciphered. However, if message #1 is enciphered with key #2, the resultant cryptogram is unbreakable since meaningful message #3 can be enciphered with key #1 to yield the same cryptogram. Such a solution will be called a false message decipherment. The number of false message decipherments is a random variable and is determined by the cipher and the cryptogram. Another possibility is shown in encryption of message #4. Seeing a cryptogram #12 *a posteriori* probability of message #4 is 1, while all the other messages are impossible. In this case however we can not find the key which was used during the encryption, since cryptogram #12 does not decrease our uncertainty about the encryption key. This situation is called false key decipherment.

## 7.3 Unicity Distance of Random Cipher

Shannon defined an important notion of a random cipher:

**Definition 7.3.1 Random cipher** *is one in which $g(e,k)$, the decipherment of $e$ under key $k$, is a uniformly distributed random variable on the set of $2^{R_0 N}$ messages, both meaningful and not.*

**Theorem 7.3.1** *Over the ensemble of random ciphers, the expected number of false key decipherments is $2^{H(k)-ND}$.*

**Proof.** There are $2^{R_0 N}$ possible messages of length $N$ and $2^{RN}$ meaningful messages. Take a key from set $K$ random and decipher the text. Then it is possible to receive a meaningful text with probability $2^{RN}/2^{R_0 N} = 2^{-ND}$. There are $|K| = 2^{H(k)}$ possible keys. Therefore the uncertainty about the key is equal to $2^{H(k)}2^{-ND} = 2^{H(k)-ND}$. □

Shannon defined the unicity distance of cipher, $U$, as the minimum amount of ciphertext required to yield a unique solution to a cryptogram. He then went on to show that ciphertexts longer than this distance are reasonably certain to have only one meaningful decryption. Ciphertexts significantly shorter than this are likely to have multiple, equally valid decryptions and therefore gain security from the opponent's difficulty in choosing the correct one.

**Definition 7.3.2** *The* **unicity distance** *is defined as the entropy of the encryption key divided by the redundancy of the language* $U = H(k)/D$.

Consider a simple substitution cipher on English,

$$\begin{aligned} H(k) &= \log_2(26!) = 88.4 \text{ bits} \\ R_0 &= \log_2(26) = 4.7 \text{ bits/character} \\ R &= 1.5 \text{ bits/characher} \\ D &= 3.2 \text{ bits/character,} \end{aligned}$$

and therefore the random cipher model would predict a unicity distance $U = 28$ characters. According to Friedman [3] simple substitution with more than 25 letters can be readily solved. The unicity distance is inversely proportional to the redundancy. As redundancy approaches zero, even a trivial cipher can be unbreakable under a ciphertext-only attack. Unicity distance is not a measure of the amount of ciphertext required for cryptanalysis. A cryptosystem may be computationally infeasible to break even if it has small unicity distance and thus is theoretically breakable with a small amount of ciphertext.

## 7.4 Pure Cipher

**Definition 7.4.1** *A system is called* **endomorphic** *if $M$ and $E$ spaces are identical.*

In the case of most classical ciphers it means that the message space and cryptogram space consist of sequences of letters. We can define a pure cipher as one whose $T_i$ form a group. This, however, would be too restrictive since such definition will work only for endomorphic systems. The proper definition is as follows:

**Definition 7.4.2** *A cipher $T$ is* **pure** *if for every $T_i$, $T_j$, $T_k$ there is a $T_s$ such that $T_iT_j^{-1}T_k = T_s$ and every key is equally likely.*

Simple substitution and transposition are pure ciphers since operations $T_iT_j^{-1}T_k$ is itself substitution or transposition. The Enigma cipher studied in previous lectures is not pure since $T_iT_k$ is not an involution.

**Theorem 7.4.1** *In a pure cipher the operations $T_i^{-1}T_j$ which transform the message space into itself form a group whose order is $m$, the number of different keys.*

**Proof.** Suppose that $T$ is a pure cipher. Since

$$T_j^{-1}T_kT_k^{-1}T_j = I,$$

it follows that each element has an inverse. The associative law is true since these are operations. Since $T$ is a pure cipher, there exist $s$ such that

$$T_i^{-1}T_j = T_s^{-1}T_k$$

Then the group property follows from

$$T_i^{-1}T_jT_k^{-1}T_l = T_s^{-1}T_kT_k^{-1}T_l = T_s^{-1}T_l$$

□

If $T$ is endomorphic and the $T_i$ are group and equally likely, then $T$ is pure, since

$$T_iT_j^{-1}T_k = T_iT_k = T_s.$$

**Theorem 7.4.2** *The product of two pure ciphers which commute is pure.*

**Proof.** Suppose $T$ and $R$ are two pure ciphers which commute. Then $\forall i, j\ \exists k, l$ such that

$$T_iR_j = R_kT_l$$

This can hold only if T and R work over the same domain, and image of T is the domain of R and vice versa, i.e. they are endomorphic. As pure ciphers which are also endomorphic they are groups by Theorem 7.4.1. As a result any combination of $T_i$, $R_j$, and their inverses is interchangeable (for ex. $\forall i, j \exists k, l$ such that $T_i^{-1}R_j = R_kT_l^{-1}$). Then

$$\begin{aligned} T_iR_j(T_kR_l)^{-1}T_mR_n &= T_iR_jR_l^{-1}T_k^{-1}T_mR_n \\ &= R_uR_v^{-1}R_wT_r^{-1}T_sT_l \\ &= R_hT_g \end{aligned}$$

□

Note that "commutation" condition is sufficient but not necessary for the product to be a pure cipher.

Consider a pure cipher shown in Figure 7.3. The messages fall into certain subsets which Shannon call residue classes, and the possible cryptograms are divided into corresponding residue classes. There is at least one line from each message in a class to each cryptogram in the corresponding class, and no line between classes which do not correspond. The number of messages in a class is a divisor of the total number of keys. The number of lines "in parallel" from a message $M$ to a cryptogram in the corresponding class is equal to the total number of keys divided by the number of messages in the class containing the message. These properties hold for pure ciphers in general.

**Theorem 7.4.3** *In a pure system the messages can be divided into a set of "residue classes" $C_1, C_2, \ldots, C_s$ and the cryptograms into corresponding set of residue classes $C'_1, C'_2, \ldots, C'_s$ with the following properties:*

1. *The message residue classes are mutually exclusive and collectively contain all possible messages. Similarly for the cryptogram residue classes.*

2. *Enciphering any message in $C_i$ with any key produces a cryptogram in $C'_i$. Deciphering any cryptogram in $C'_i$ with any key leads to a message in $C_i$.*

3. *The number of messages $\varphi_i$ in $C_i$ is equal to the number of cryptograms in $C'_i$ and is a divisor of $|K|$ the number of keys.*

4. *Each message in $C_i$ can be enciphered into each cryptogram in $C'_i$ by exactly $|K|/\varphi_i$ different keys. Similarly for decipherment.*

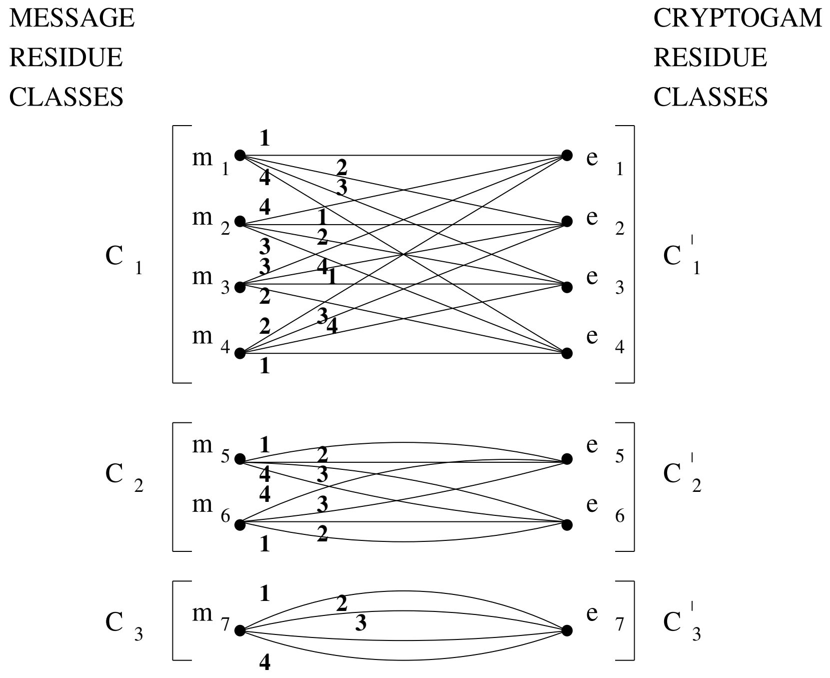


Figure 7.3: Pure system

**Proof.** Consider a class of cryptograms $C_1' = \{T_i m \ \forall i\}$ that can be obtained from some message $m$ by any encryption operation $T_i$. Let $C_1 = \{T_j^{-1} T_i m \ \forall i, j\}$ be a class of messages that can be obtained from $m$ by $T_j^{-1} T_i m$.
Pick $\overline{m} \in C_1$. It follows that $\overline{m} = T_j^{-1} T_i m$. Then

$$\forall k \quad T_k \overline{m} = T_k T_j^{-1} T_i m = T_l m \in C_1'$$

Thus we obtained the same $C_1'$ starting from any other $\overline{m} \in C_1$. Similarly the same $C_1$ would be obtained.
Picking, if it is possible, $n \notin C_1$ we construct a second class in the same way. Continuing this process we have the residue classes with properties (1) and (2).
Note that $\forall m_1, m_2 \in C_1 \ \exists i, j$ such that $m_2 = T_i^{-1} T_j m_1$. This is true since $\exists m \in M$,

$$\begin{aligned} \exists h, q \in M \quad m_1 &= T_h^{-1} T_q m \quad \textit{and} \\ \exists i, s \in M \quad m_2 &= T_i^{-1} T_s m \end{aligned}$$

and we have

$$m_2 = T_i^{-1} T_s T_q^{-1} T_h m_1 = T_i^{-1} T_j m_1.$$

Suppose that $e_1 \in C_1^{-1}$ can be reached from $m_2$ by

$$e_1 = T_\alpha m_2 = T_\beta m_2 = T_\gamma m_2 = \dots$$

Then

$$\begin{aligned} e_1 &= T_\alpha T_i^{-1} T_j m_1 = T_\beta T_i^{-1} T_j m_1 = \dots \\ &= T_a m_1 = T_b m_1 = \dots \end{aligned}$$

So $e_1$ is reachable from $m_1$ by the same number of keys as from $m_2$. Thus each $m \in C_1$ reaches $e_1$ by this number of keys. It follows that this number is a divisor of $|K|$, and it is $|K|/\varphi_i$. Thus we prove properties (3) and (4). □

**Theorem 7.4.4** *In a pure system the a posteriori probabilities of various messages $P_e(m)$ are independent of the key that is chosen. The a posteriori probabilities of the keys $P_e(k)$ are the same in value but undergo a permutation with a different key choice.*

**Proof.** Using Bayes' theorem we obtain

$$P_e(m) = P(m|e) = \frac{P(m)P(e|m)}{P(e)} = \frac{P(m)P(e|m)}{\sum_{n\in C_i} P(n)P(e|n)} = \frac{P(m)}{P(C_i)} \tag{7.1}$$

where $m \in C_i$ , $e \in C'_i$. Note that if $m, n \in C_i$, then $P(e|m) = P(e|n)$.

$P(m)$=*a priori* probability of message $m$.

$P(e|m)$=conditional probability of cryptogram $e$ if message $m$ is chosen or the sum of the probabilities of all keys which produce cryptogram $e$ from message $m$.

$P(e)$=probability of obtaining cryptogram $e$ from any cause.

$P(m|e)$=*a posteriori* probability of message $m$ if cryptogram $e$ is intercepted.

If $e$ and $m$ are not in corresponding residue classes, $P_e(m) = 0$. Similarly it can be shown that the a posteriori probabilities of the different keys are the same in value but these values are associated with different keys when a different key is used. The same set of values of $P_e(k)$ have undergone a permutation among the keys. □

**Example 1.** Consider a simple substitution with all keys equally likely. Applying the operation $T_i^{-1}$ with any key to a given cryptogram $e \in C'_n$ produces a message in $C_n$.
Thus if the cryptogram is

$$e = XAABLL$$

then

$$\begin{aligned} m_1 &= URRCDD \\ m_2 &= ZDDSTT \\ m_3 &= ACCESS \end{aligned}$$

are in the same residue class. The residue class is determined by a letter repetition pattern. However only $m_3$ is a meaningful message. Since English is a highly redundant language, most of the messages in a residue class are meaningless, so knowing the residue class information may be enough to break the cipher. For example, among $2^{64}$ 8-byte blocks only about $2^{2.3\cdot 8} \approx 2^{18}$ are meaningful English.

**Example 2.** Consider the Vigenere cipher with a random key of period $d$. In this case the message residue class consists of all sequences with the same first differences as the cryptogram, for letters separated by distance d. Let $e = c_1c_2\ldots$ be the cryptogram, $c_i \in A$ and $m = p_1p_2\ldots$ be any message in the corresponding residue class, $p_i \in A$. Then for $d = 3$ the residue class is defined by:

$$\begin{aligned} p_1 - p_4 &= c_1 - c_4 \\ p_2 - p_5 &= c_2 - c_5 \\ p_3 - p_6 &= c_3 - c_6 \\ p_4 - p_7 &= c_4 - c_7 \end{aligned}$$

**Example 3.** Consider transposition cipher of period $d$ with random key. Let $e = c_1c_2\dots$ be a cryptogram, $c_i \in A$. The residue class consists of all arrangements of the $c_i$ in which no $c_i$ is moved out of its block of length $d$, and any two $c_i$ at a distance $d$ remain at this distance. Then we can break this cipher by the following steps. The cryptogram is written in blocks of length d, one under another (d=5):

$$\begin{array}{ccccc} c_1 & c_2 & c_3 & c_4 & c_5 \\ c_6 & c_7 & c_8 & c_9 & c_{10} \\ c_{11} & c_{12} & . & . & . \\ . & . & . & . & . \end{array}$$

The columns are then cut apart and rearranged to make meaningful text. When the columns are cut apart, the only information remaining is the residue class of the cryptogram.

## 7.5 Perfect Cipher

Suppose that in formula (7.1) $\exists i$ such that $P(C_i) = 1$, then such system consists of a single large residue class (see Figure 7.4). For this system it holds $P_e(m) = P(m)$ and thus *a posteriori* probability is equal to the *a priori* probability. This condition is called **perfect secrecy**.

**Definition 7.5.1** *Cryptosystem is called* **perfect** *if* $P_e(m) = P(m)$.

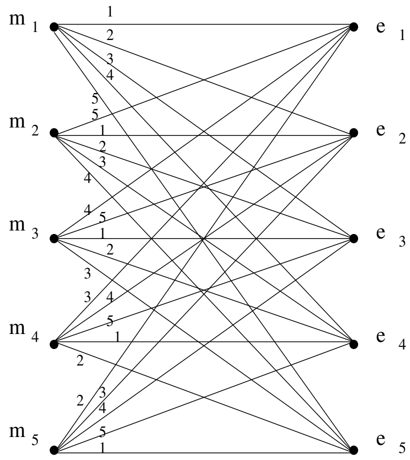


Figure 7.4: Perfect system

**Theorem 7.5.1** *A necessary and sufficient condition for perfect secrecy is that* $P_m(e) = P(e)$ *for all* $m$ *and* $e$*. That is* $P_m(e)$ *must be independent of* $m$.

**Proof.** By Bayes theorem

$$P_e(m) = \frac{P(m)P_m(e)}{P(e)}$$

In a perfect system $P_e(m)$ must be equal to $P(m)$ for all $e$ and all $m$. Hence either $P(m) = 0$, a solution that must be excluded since we demand the equality independent of the values of $P(m)$, or

$$P_m(e) = P(e)$$

for every $m$ and $e$. Conversely if $P_m(e) = P(e)$ then

$$P_e(m) = P(m)$$

and we have a perfect system. □

Perfect systems in which the number of cryptograms, the number of messages, and the number of keys are all equal are characterized by the properties that

1. each $m$ is connected to each $e$ by exactly one line
2. all keys are equally likely.

Such systems can be represented by a "Latin Square" (see Figure 7.5). The famous one-time pad

| | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ |
|---|---|---|---|---|---|
| $m_1$ | 1 | 2 | 3 | 4 | 5 |
| $m_2$ | 5 | 1 | 2 | 3 | 4 |
| $m_3$ | 4 | 5 | 1 | 2 | 3 |
| $m_4$ | 3 | 4 | 5 | 1 | 2 |
| $m_5$ | 2 | 3 | 4 | 5 | 1 |

Figure 7.5: "Latin" square

cipher is a perfect cipher. The disadvantage of perfect systems for large communication systems is that $|K| \geq |M|$. Thus prior to communication of length $|M|$ sides have to exchange the key of at least the same size. This is impractical in most of the cases.

## 7.6 Ideal Cipher

*This section is currently "under construction".*

If we compress data, redundancy of the language tends to 0. Then unicity distance tends to infinity. Such system is an ideal. Thus the data compression increases the security of a cryptosystem. (see Figure 7.6).

The ideal cipher can sometimes be obtained by matching the structure of the cipher to that of the language.

**Example 1.** Consider an artificial language in which the 26 letters occur with equal frequency. Letter choices are independent but once a letter is chosen, it is repeated tree times. Thus a typical sentence in this language is "$FFFXXXLLLAAAXXXRRR''$. Apply a simple substitution cipher with $2^{H(k)}$ keys chosen at random from among the 26! possible keys. The above sentence would be enciphered into a cryptogram of the form "$SSSBBBMMMYYYBBBFFF''$, assuming $F$ is

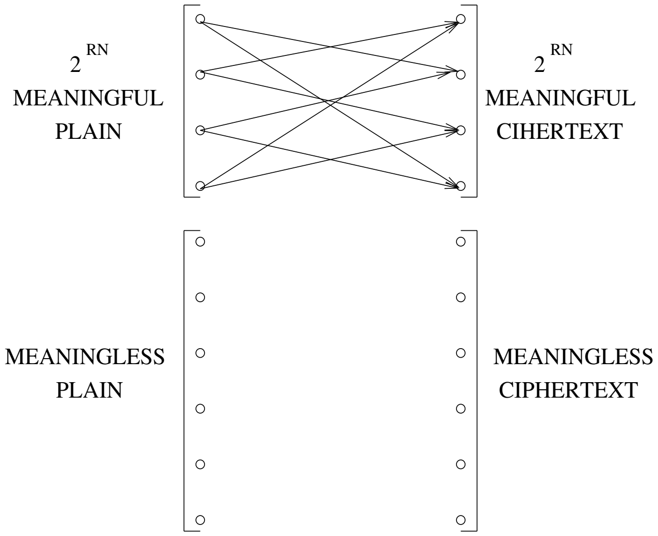


Figure 7.6: Ideal cipher

enciphered as $S$, $X$ as $B$ and so on. A cryptanalyst who is given such a cryptogram has no way of choosing among the $2^{H(k)}$ possible keys. Each key yields a typical sentence as its decipherment and the cryptogram itself looks like a meaningful message.
On the other hand transposition cipher with the same number of keys $2^{H(K)}$ chosen random from $N!$ possible transpositions can break using the redundancy of the language. In this case a typical cryptogram looks like "$XLFFXARARXLXXRAXFL''$.

**Example 2.** Consider a second artificial language in which letters are independent random variables, each having the same probability distribution as in English. Then "$RESNFEALTMISEQ''$ is a typical sentence in this language. Using a simple substitution cipher yields a cryptogram such as "$FXOBRXYMGUNOXC''$. Then cryptanalyst can break this cipher using frequencies analysis. However if transposition is applied to the same language we obtain a transposed message looks as meaningful as the original message.

# Bibliography


[1] P. F. Brown, V. J. Della Pietra, R. L. Mercer, S. A. Della Pietra, *An Estimate of an Upper Bound for the Entropy of English*, Computational Linguistics, Vol. 18, N. 1, pp.31–40, 1992.

[2] T. M. Cover, R. King, *A Convergent Gambling Estimate of the Entropy of English*, IEEE Transactions on Information Theory, Vol. 24, N. 4, pp.413–421, 1978.

[3] W. F. Friedman *"Cryptology"*, in Encyclopedia Brittannica, p.848, 1973.

[4] Martin Hellman, *An Extension of the Shannon Theory Approach to Cryptography*, IEEE Transactions on information theory, Vol. IT-23, No.3, May 1977.

[5] Bruce Schneier, *Applied Cryptography*, John Wiley & Sons, 1996.

[6] C. Shannon, *A Mathematical Theory of Communication*, Bell Systems Technical Journal, Vol. 27, Part 1, pp.479-523, Part 2, pp.623-656, 1948.

[7] C. Shannon, *The Communication Theory of Secrecy Systems*, Bell System Technical Journal, Vol.28, pp. 656-715, Oct. 1949

[8] C. Shannon, *Prediction and Entropy in Printed English*, Bell Systems Technical Journal, Vol. 30, N. 1, pp.50-64, 1951.