CHAPTER TWELVE

Measures of Secrecy for Cryptographic Systems

We agree with the statement that "cryptography is currently an engineering subject in which there are more facts and rules of thumb than theorems or systematic developments" [1]. The science of cryptography has evolved only recently as a result of attempts to explain or define in mathematical terms the facts and rules of thumb that have evolved from the practiced art of cryptography. This chapter reflects this quality and of necessity combines a wide variety of material ranging from the very simple to the very complex.

Cryptographic protection (secrecy) is attainable if plaintext can be recovered from ciphertext only by those authorized. There are in fact two types of secrecy that can be achieved with a cryptographic algorithm: *theoretical secrecy* and *practical secrecy*.

Theoretical secrecy is based on a single axiom: that the information available to or intercepted by an opponent is insufficient for the derivation of a unique cipher solution. In other words, there is always a measure of uncertainty, regardless of what method of analysis is used, as to which candidate among a set of possible values (keys or messages) is correct. For example, cryptographic protection may be based on the assumption that an opponent has only ciphertext available (called a ciphertext-only attack), and that the amount of ciphertext intercepted by an opponent would be insufficient to allow the plaintext or key to be recovered.[1] Today, with the vast amounts of data being transmitted in communication networks, such an assumption cannot be justified. In fact, the designers of a cryptographic system should assume that an opponent can obtain plaintext and matching ciphertext in sufficient quantities to determine the key uniquely (see Cryptographic Algorithms, Chapter 2).Therefore, by and large, theoretical secrecy is unattainable in today's data processing systems and networks.

On the other hand, practical secrecy assumes sufficient information is available to break the cipher, and is measured by the work (work factor)

[1] The following are examples of ciphertext-only attacks in which there are multiple solutions. With transposition on English, cryptogram "nde" would have at least two solutions: "den" and "end." With a simple substitution on English, cryptogram "nde" would have several solutions: "the," "and," "but," and so forth.

required to find the solution to a given cryptanalytical problem (see Cryptographic Algorithms, Chapter 2). This type of secrecy is achieved (e.g., in the DES) by designing the cryptographic algorithm so that it is computationally infeasible to solve for a message or key, even if the analyst has specific knowledge of the cryptographic algorithm and large amounts of chosen ciphertext/plaintext and corresponding plaintext/ciphertext.

Experience has shown that it is difficult to devise a cryptographically strong algorithm. (Note for example the successful cryptanalysis of the German Enigma Cipher and the Japanese PURPLE Cipher used during World War II). To understand why this is so, it is helpful to investigate the mathematical foundations of cryptography. (Theoretical secrecy, which is primarily of value to a general study of cryptography, has been the subject of extensive mathematical analyses. Significant results have been obtained. Practical secrecy, which is significant to the specialized study of cryptographic algorithms, has also been widely investigated; but few substantive results have been obtained.)

Shannon [2,3] invented a particularly useful theoretical model called a *random cipher*. Using information theory, he described the relationship between the amount of intercepted ciphertext and the likelihood of a successful attack. With the model, he was able to determine the *unicity distance* (ud) of a cipher, which he described as follows: with more than ud characters of ciphertext there is only one solution to the cipher, with less than that amount there are several so-called solutions.

More accurate results can be obtained if other probabilistic measures (not based on information theory) are used. The amount of (intercepted) ciphertext defines a specific probability that the ciphertext has a unique solution (only one meaningful decipherment is obtained using the set of all possible keys). By calculating the probability that a correct key or correct plaintext can be obtained for a given amount of ciphertext, more precise statements can be made concerning a cipher's vulnerability.

ELEMENTS OF MATHEMATICAL CRYPTOGRAPHY

The analysis that follows assumes a conventional cryptographic algorithm (an algorithm in which the enciphering and deciphering keys are equal). However, the results can be adapted to public-key algorithms as well. (For a definition of conventional and public-key algorithms, see Cryptographic Algorithms, Chapter 2.)

Information Flow in a Conventional Cryptographic System

Figure 12-1 illustrates the information flow in a conventional cryptographic system. A message (plaintext x) is generated by the sender. An enciphering algorithm E, which depends on a secret key k, is used to encipher x into a cryptogram (ciphertext y):

$$E_k(x) = y$$

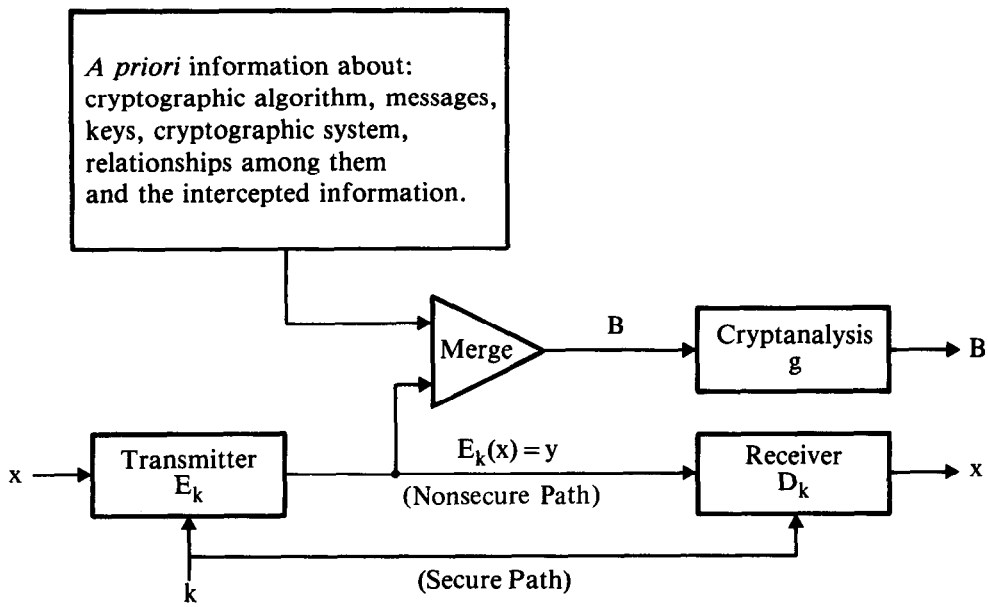


Figure 12-1. Information Flow in a Conventional Cryptographic System

Cryptogram y is then transmitted to the receiver where the deciphering algorithm D, which also depends on secret key k, is used to recover x:

$$D_k(y) = x$$

It is assumed that an opponent does not possess k, and hence cannot recover x from y using D. (Note that algorithms D and E may or may not be kept secret.) In order for k to remain secret, a secure communication path is needed between the sender and receiver.

The opponent's initial information is a variable that can be as little as only ciphertext or as much as complete knowledge of the system (except for the key). If B represents the information available to an opponent and g represents the process of cryptanalysis, then the deduced information, B′, can be expressed as

$$B' = g(B)$$

Practical secrecy assumes that the computational resources and time available for analysis must be within practical bounds. By making these bounds high enough, a sufficiently high work factor is achieved. Theoretical secrecy, on the other hand, assumes that the analyst has unlimited computational resources.

A Cipher with Message and Key Probabilities

A mathematical analysis of ciphers is made possible by assigning probabilities to messages and keys and by making certain simplifying assumptions about

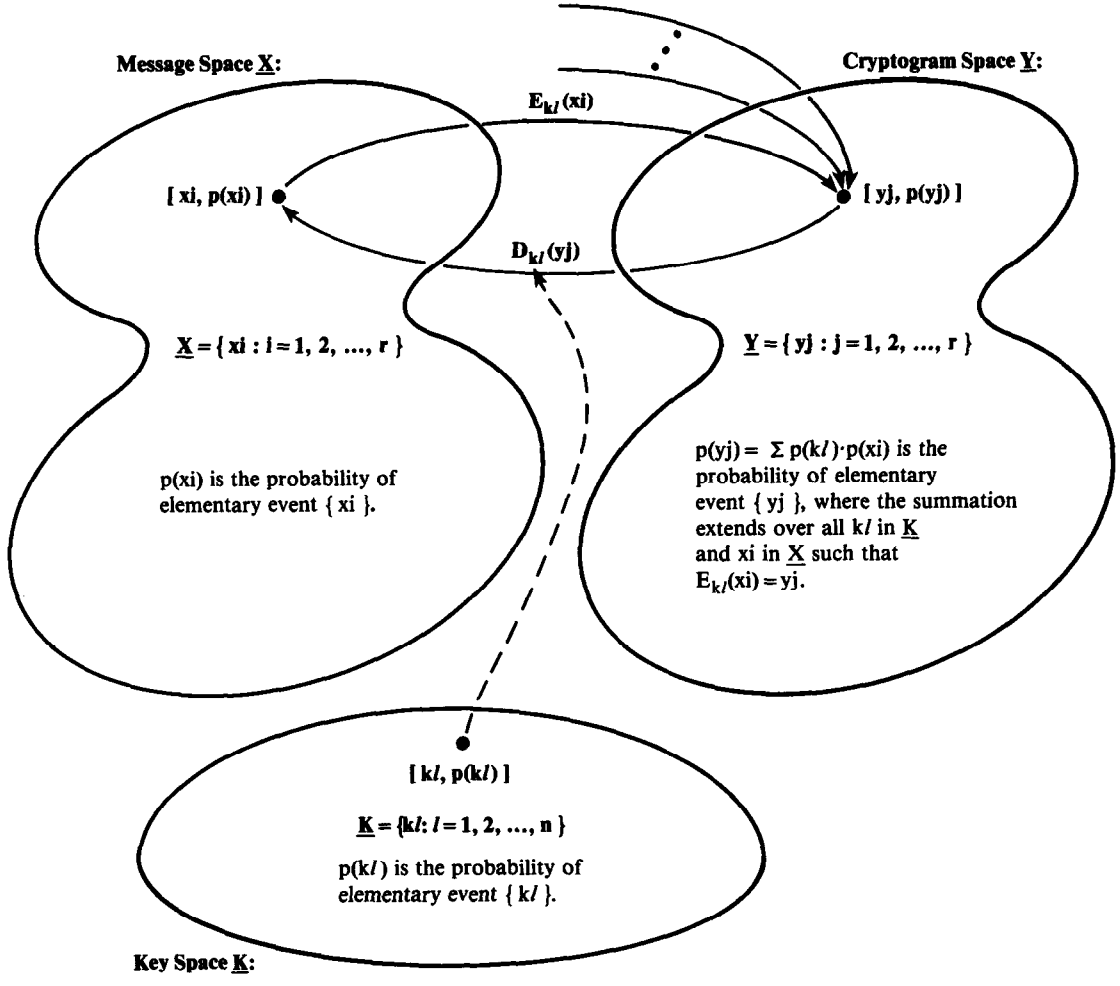


Figure 12-2. A Cipher with Message and Key Probabilities

the enciphering and deciphering transformations, messages, cryptograms, and keys (Figure 12-2). For given enciphering (E) and deciphering (D) algorithms, they are:

1. $\underline{X} = \{xi:i = 1, 2, \ldots, r\}$ is a finite set of r unique messages having associated probabilities of occurrence, p(x1), p(x2), . . . , p(xr).
2. $\underline{K} = \{kl:l = 1, 2, \ldots, n\}$ is a finite set of n unique keys having associated probabilities of occurrence, p(k1), p(k2), . . . , p(kn).
3. $\underline{E} = \{E_k:k = k1, k2, \ldots, kn\}$ is a finite set of one-to-one enciphering functions from the message space ($\underline{X}$) to the cryptogram space ($\underline{Y}$).
4. For every key (k*l* in $\underline{K}$) and message (xi in $\underline{X}$) there is a cryptogram (yj in $\underline{Y}$) such that $E_{kl}(xi) = yj$. The probability of elementary event {yj} is given by

$$p(yj) = \Sigma\, p(kl)p(xi)$$

   where the summation extends over all k*l* in $\underline{K}$ and xi in $\underline{X}$ such that $E_{kl}(xi) = yj$. It is assumed that messages and keys are independently chosen, that is, p(k*l*, xi) = p(k*l*)p(xi).
5. The condition that E_{kl} is a one-to-one function means that the number of elements in $\underline{Y}$, denoted $|\underline{Y}|$, must be equal to or greater than the number of elements in $\underline{X}$. For the special case where $|\underline{X}| = |\underline{Y}|$, each enciphering function in $\underline{E}$ is not only one-to-one but also onto (see Cryptographic Algorithms, Chapter 2), and D_k is the inverse function of E_k. To simplify the analysis, assume that $|\underline{X}| = |\underline{Y}|$. Therefore,

$$\underline{Y} = \{yj:j = 1, 2, \ldots, r\}$$

   is a finite set of r unique cryptograms, and

$$\underline{D} = \{D_k:k = k1, k2, \ldots, kn\}$$

   is a finite set of one-to-one deciphering functions from the cryptogram space ($\underline{Y}$) to the message space ($\underline{X}$).

When no ambiguity exists, the indices (i, j, and *l*) associated with x, y, and k will be omitted from the discussion.

The probabilities assigned to messages and keys represent the analyst's prior knowledge (or assumptions) about the messages and keys selected for encipherment. For example, if there were a known bias in the key selection process, the analyst would assign highest probability to those keys with the greatest chance of being selected. In effect, this would reduce the average number of keys needed to be searched before finding the correct key. However, if keys are randomly selected, or if the selection process is unknown, the analyst would assign equal probability to each key:

$$p(k) = 1/n \qquad \text{for each k in } \underline{K}$$

In the analysis that follows, keys are assumed to be equally probable. Probabilities are assigned to messages using a method suggested by Shannon [2,3]. The message space ($\underline{X}$) is divided into two sets: (1) a set of s meaningfully distinct, or *meaningful messages*, denoted by $\underline{X}'$, and (2) a set of r–s *meaningless messages* denoted by $\underline{X}''$. By assuming that almost all enciphered messages will be meaningful, it follows that the sum of the probabilities of the messages in $\underline{X}'$ is approximately equal to one, and the sum of the probabilities of the messages in $\underline{X}''$ is approximately equal to zero. In the analysis below, assume that the sums of the possibilities of the messages in $\underline{X}'$ and $\underline{X}''$ are one and zero, respectively.[2]

For mathematical simplicity, assume that the analyst has no prior knowledge of the messages' contents and that each message in $\underline{X}'$ is assigned equal probability:

$$p(x) = 1/s \qquad \text{for each x in } \underline{X}'$$

The cryptogram space ($\underline{Y}$) can therefore be divided into a set of *possible cryptograms* (those that can be generated from at least one meaningful message), denoted by $\underline{Y}'$,

$$\underline{Y}' = \{E_k(x) : k \text{ in } \underline{K} \text{ and } x \text{ in } \underline{X}'\}$$

and a set of *impossible cryptograms* (those that can be generated only from meaningless messages), denoted by $\underline{Y}''$,

$$\underline{Y}'' = \underline{Y} - \underline{Y}'$$

where $\underline{Y} - \underline{Y}'$ is the *difference* of $\underline{Y}$ and $\underline{Y}'$, defined as the elements in $\underline{Y}$ that are not in $\underline{Y}'$. (The probability, p(y), of each cryptogram (y) in $\underline{Y}'$ is determined by the probabilities of the various messages and keys, as shown above.)

In the definitions given below, y stands for an intercepted cryptogram, that is, y is an element of the set $\underline{Y}'$.

1. M is the random variable defined as the number of keys that will decipher a given intercepted cryptogram (yj) into a meaningful message.
2. M′ is the random variable defined as the number of keys, except for the key originally used to produce the given cryptogram, that will decipher the intercepted cryptogram into a meaningful message (M′ = M − 1).
3. U is the random variable defined as the number of different meaningful messages that are produced when a given intercepted cryptogram (yj) is deciphered with all possible keys.

[2] Of course, if the sender purposely enciphers random data, then r = s.

4. U′ is the random variable defined as the number of different meaningful messages, except for the message originally used to produce the given cryptogram, that are produced when the intercepted cryptogram is deciphered with all possible keys (U′ = U − 1).

The relationship between M and U is illustrated by the following example in which x1 and k1 are a message and key originally used to produce cryptogram y1; x2 and x3 are incorrect meaningful decipherments; and k2, k3, and k4 are incorrect keys leading to meaningful decipherments (Figure 12-3).

Since 4 keys produce only 3 different meaningful decipherments, it follows that M = 4 and U = 3. It can be seen from the example that the following relations hold in general:

$$u \leqslant m$$
$$m' = m - 1$$
$$u' = u - 1$$

(m and u denote specific values of the random variables M and U, respectively.)

The *probability distribution* of M (the probabilities associated with the occurrence of all possible values of M) is of primary interest to the analyst since the probability of solving successfully for the correct key, denoted by p(SK) (where S stands for success), can be evaluated from M.

Assume that m different keys will decipher an intercepted cryptogram into meaningful messages. Since any of these m keys could be the correct key (the key originally used to produce the given cryptogram), the proba-

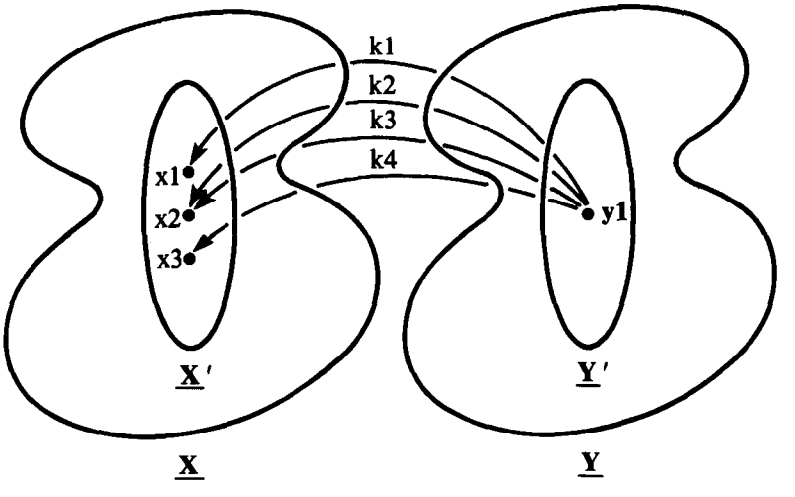


Figure 12-3. Example in which Cryptogram y1 has Four Meaningful Decipherments

bility of guessing or randomly selecting the correct key from among this set is thus

$$p(SK|M = m) = 1/m$$

Since m can range from 1 to n (note that n is the total number of keys, see Figure 12-2), the probability of obtaining the correct key is given by:

$$p(SK) = \sum_{m=1}^{n} (1/m)p(M = m) \tag{12-1}$$

where $p(M = m)$ is the probability that $M = m$. Similarly, the distribution of U allows the probability of successfully solving for the correct message, denoted by p(SM), to be calculated:

$$p(SM) = \sum_{u=1}^{s} (1/u)(p(U = u) \tag{12-2}$$

(s is the total number of meaningful messages.) In general, the mathematical relationships that define a cipher—the complex structural relationships dictated by the enciphering and deciphering functions as applied to the set of messages—render it impossible to determine the distributions of M and U. At best, only approximations to these distributions can be found.

Shannon [2] overcame this problem by defining a special theoretical cipher, called a *random cipher*, with the property that the distribution of M is easily determined. He demonstrated that the results obtained with a random cipher are consistent with those obtained with some actual ciphers.

The Random Cipher

Let Q denote the set of all possible one-to-one and onto functions from the set of messages $\underline{X}$ to the set of cryptograms $\underline{Y}$. Select at random, *with replacement,*[3] n enciphering functions, f1, f2, . . . , fn from Q and denote this set by $\underline{E}$:

$$\underline{E} = \{fi : i = 1, 2, \ldots, n\}$$

Corresponding to each enciphering function fi in Q, there is an inverse (deciphering) function, fi^{-1} also in Q, such that

$$fi^{-1}(fi(x)) = x$$

for all x in $\underline{X}$. Let it be assumed that for each enciphering function fi selected

[3]The term "with replacement" means that each selected element is replaced before the next element is selected.

from Q, the corresponding inverse (deciphering) function fi^{-1} is also selected from Q. Thus the set of selected enciphering functions $\underline{E}$ defines a set of associated deciphering functions $\underline{D}$:

$$\underline{D} = \{fi^{-1} : i = 1, 2, \ldots, n\}$$

If f is replaced by E, fi^{-1} is replaced by D, and subscripts are redefined as keys, then $\underline{E}$ and $\underline{D}$ can be rewritten as

$$\underline{E} = \{E_k : k = k1, k2, \ldots, kn\}$$
$$\underline{D} = \{D_k : k = k1, k2, \ldots, kn\}$$

A random cipher can thus be described in terms of the notation used in prior chapters. There is also a natural one-to-one correspondence among ki, E_{ki}, and D_{ki}; selecting any one of the elements is the same as selecting all three elements. The definition of a random cipher given here is different from that given by Shannon [2]. Shannon's random cipher is not a true cipher, since it is possible for the same key to decipher two different cryptograms into the same message. The random cipher defined here eliminates this problem—it is a true cipher.

However, in eliminating one problem, another is created. Because functions are selected from Q with replacement, it may happen that $\underline{E}$ contains two enciphering functions, fi and fj (selected at the ith and jth trials, respectively, such that fi = fj and $i \neq j$ (in other words, $\underline{E}$ may contain an E_{ki} and E_{kj} such that $E_{ki} = E_{kj}$ and $ki \neq kj$). That is, *equivalent* enciphering functions (keys) can occur. But the occurrence of equivalent keys in the definition of a random cipher is not a problem if the total number of such keys remains very small. (Note that in some ciphers it is nearly impossible to prove that equivalent keys do or do not exist.)

Consider a random cipher used to model the DES algorithm. Such a cipher could be constructed by randomly selecting (with replacement) 2^{56} functions (or keys) from Q, where Q contains $(2^{64})!$ different one-to-one and onto functions from the set of all 64-bit messages to the set of all 64-bit cryptograms. Since 2^{56} is very small in comparison to $(2^{64})!$, the probability of selecting the same function (key) twice would be extremely small.

Before proceeding with a mathematical analysis of the random cipher, a computational procedure used to estimate the number of meaningful messages in $\underline{X}$ is discussed. This will allow the results obtained with the random cipher to be applied to several examples of actual ciphers.

Number of Meaningful Messages in a Redundant Language[4]

If an opponent has only ciphertext, but enough is available for analysis, then a necessary and generally sufficient condition for breaking a cipher via

[4] A detailed analysis of the number or meaningful messages is given in Appendix F. Some of the important results derived there are summarized in this section.

the brute-force methods discussed below is that the underlying language from which the messages are selected possess the property known as *redundancy*. A language has redundancy if for any N it can be shown that the possible sequences of N letters are not all equally probable. All natural languages possess redundancy. That English is redundant is demonstrated by a table of the number of times each letter appears in a sample of text. The results of such an experiment using 4 million letters of English text are shown in Table 12-1.

Exactly how redundancy facilitates the process of cryptanalysis can be stated in the following way:

> Redundancy is essentially a series of conditions on the letters of a message, which insure that it be statistically reasonable. These consistency conditions produce corresponding consistency conditions in the cryptogram. The key gives a certain amount of freedom to the cryptogram but, as more and more letters are intercepted, the consistency conditions use up the freedom allowed by the key. Eventually, there is only one [combination of] message and key which satisfies all the conditions and we have a unique solution. [2]

In effect, the reason that a cryptogram eventually has a unique solution, if enough text is available for analysis, is that in a redundant language the messages of N letters can be divided into two sets, those which are intelligible or meaningful and those which are not, and as N is increased the ratio of meaningful to meaningless messages approaches zero. In the English language, for

a	Freq (a)	p(a)	a	Freq (a)	p(a)
A	321712	.0804	N	283561	.0709
B	61472	.0154	O	303844	.0760
C	122403	.0306	P	79845	.0200
D	159726	.0399	Q	4226	.0011
E	500334	.1251	R	244867	.0612
F	92100	.0230	S	261470	.0654
G	78434	.0196	T	370072	.0925
H	219481	.0549	U	108516	.0271
I	290559	.0726	V	39504	.0099
J	6424	.0016	W	76673	.0192
K	26972	.0067	X	7779	.0019
L	165559	.0414	Y	69334	.0173
M	101339	.0253	Z	3794	.0009

p(a) = Freq(a)/4,000,000.
Based on a sample of 8000 excerpts of 500 letters taken from the Brown University Corpus of Present-Day American English. [3]

Table 12-1. Individual Letter Frequencies in 4 Million Characters of English Text

example, the meaningful sequences are just those that are encountered in normal text.

An approximation of s, the number of meaningful messages in $\underline{X}$, can be obtained using a zero-order approximation of message probability (Equation F-2 in Appendix F). When message length N is very large, each message contains about Np1 occurrences of the first letter, Np2 occurrences of the second letter, and so on, where pi is the probability of occurrence of letter i. Hence, for N very large, most messages have a probability p approximately equal to ($\simeq$)

$$p \simeq p1^{Np1} p2^{Np2} \ldots pn^{Npn}$$

where n is the number of different characters in the language. Ignoring statistical variations in p between messages, assume as a first order approximation that all s meaningful sequences have the same probability p. Since the probabilities of all meaningful sequences add up to 1 (as assumed before), it follows that $sp \simeq 1$ and thus

$$s \simeq 1/p$$

Hence,

$$\log_2 s \simeq -\log_2 p$$

$$\simeq -N \sum_{i=1}^{n} pi \log_2 pi$$

If

$$G_1 = -\sum_{i=1}^{n} pi \log_2 pi$$

is defined as the *entropy per character* (measured in bits per character)[5] for the message source, it follows that

$$s \simeq 2^{NG_1}$$

(See Appendix F.) Using the values for p(a) in Table 12-1, a value of 4.17 bits per character is obtained for G_1. Thus s can be expressed as

$$s \simeq 2^{N4.17}$$

Taking into account probabilities of pairs of letters (digrams), triplets (trigrams), and so on, it is possible to obtain a correspondingly higher-

[5] Since NG_1 denotes the number of bits required to represent s messages and N is the number of characters in each message, G_1 is expressed in bits per character.

order approximation for s (see Appendix F). Groups of n (n = 1, 2, ...) contiguous letters are also referred to as n-grams.

As N approaches infinity (all statistical information about the language is known), one obtains

$$\lim_{N \to \infty} (\log_2 s)/N = R$$

and therefore

$$s = 2^{NR}$$

where R, called the *rate of the language*, is a constant determined by the particular language. For English (26 letters), R is about 1.2 bits per character [4].

PROBABILISTIC MEASURES OF SECRECY USING A RANDOM CIPHER

Probability of Obtaining the Key When Only Ciphertext Is Available for Analysis

Given a random cipher, let y be an intercepted cryptogram which has been enciphered from an unknown meaningful message x. The opponent, who has intercepted y, knows only that x is an element of $\underline{X}'$ and y is an element of $\underline{Y}'$. Assume that kj is the key originally used to encipher x into y. Hence the probability that kj deciphers y into a meaningful message is 1:

$$p[D_{kj}(y) \text{ is meaningful}] = 1$$

In the construction of a random cipher, the enciphering functions in $\underline{E}$ are independently selected from the set Q. Therefore, the process of deciphering y with each of the (n − 1) incorrect keys

$$k1, k2, \ldots, kj - 1, kj + 1, \ldots, kn$$

can be thought of as (n − 1) Bernoulli trials,[6] where s/r represents the probability that a key will successfully decipher y into a meaningful message:

$$p[D_{ki}(y) \text{ is meaningful}] = s/r$$

and 1 − (s/r) is the probability that a key will fail to decipher y into a meaningful message:

$$p[D_{ki}(y) \text{ is meaningless}] = 1 - (s/r)$$

[6] Many problems in probability theory involve independent, repeated trials of an experiment whose outcomes can be classified into two categories called successes and failures. An experiment which has only two possible outcomes is called a *Bernoulli trial* [5].

for all values of i not equal to j. Hence it follows that M′ has a *binomial distribution* [5]:

$$p(M' = m') = \binom{n-1}{m'}(s/r)^{m'}(1-(s/r))^{n-1-m'} \quad \text{for } m' = 0, 1, \ldots, n-1 \tag{12-3a}$$

The expected value (E) and variance (Var) of M′ are, respectively,

$$E(M') = (n-1)(s/r) = \lambda' = \lambda(n-1)/n$$
$$\text{Var}(M') = (n-1)(s/r)(1-(s/r)) = \lambda'(1-(s/r))$$

where parameter λ equals ns/r. For s/r much less than 1 (written $s/r \ll 1$), the binomial distribution (Equation 12-3a) can be approximated by the *Poisson distribution* [5], so that

$$p(M' = m') \simeq e^{-\lambda'}(\lambda')^{m'}/m'! \qquad \text{for } s/r \ll 1 \tag{12-3b}$$

Since m′ equals m − 1, it follows that

$$p(M = m) = \binom{n-1}{m-1}(s/r)^{m-1}(1-(s/r))^{n-m} \quad \text{for } m = 1, 2, \ldots, n \tag{12-4a}$$

The expected value and variance of M are, respectively,

$$E(M) = E(M' + 1) = \lambda' + 1 = (\lambda(n-1)/n) + 1 \tag{12-4b}$$

$$\text{Var}(M) = \text{Var}(M' + 1) = \lambda'(1-(s/r)) \tag{12-4c}$$

The Poisson approximation for p(M = m) is given by

$$p(M = m) \simeq e^{-\lambda'}(\lambda')^{m-1}/(m-1)! \qquad \text{for } s/r \ll 1$$

Using information theory and the above mathematical relationships, Shannon defined the *unicity distance* of a random cipher (the point where there is no uncertainty over which key was used for enciphering) as the value of N(N = cryptogram length in characters) for which λ(λ = ns/r) becomes equal to one [2]. (See also An Expansion of Shannon's Approach Using Information Theory.)

Essentially, when language redundancy is present, the ratio s/r gets smaller as message length (or cryptogram length) gets larger. At some point, s/r is small enough so that ns/r equals 1. However, if data are composed of random, independently-selected characters, in which case there is no language redundancy, s equals r and unicity distance equals infinity.

Unicity distance is often given the following interpretation. *Below the uni-*

city distance ($N < ud$), an attack on the key will not succeed; above the unicity distance ($N \geqslant ud$), an attack on the key will succeed.

However, the interpretation is not strictly correct; there is no abrupt change between the point where the key is ($N \geqslant ud$) and is not ($N < ud$) obtainable. A more precise statement would be that for every cryptogram of N characters, there is an associated probability, p(SK), of obtaining the key used to produce that given (known) cryptogram from the selected (unknown) message.

If information measures are used to determine unicity distance, one concludes that a cipher is vulnerable to attack when λ is close to one. But how vulnerable the cipher is cannot be said. If a probabilistic approach is used instead, more precise statements can be made about the cipher's vulnerability. A value for the probability of successfully obtaining the correct key, p(SK), is derived by combining Equations 12-1 and 12-4a:

$$p(SK) = \sum_{m=1}^{n} (1/m)p(M = m) = (1/\lambda)(1 - (1 - (\lambda/n))^n) \qquad (12\text{-}5)$$

Using the Taylor [6] series expansion for $\ln(1 - (\lambda/n))^n$, it follows that

$$p(SK) \simeq (1/\lambda)(1 - e^{-\lambda}) \qquad \text{for } \lambda/2 \ll 1 \qquad (12\text{-}6)$$

(ln is the natural logarithm to the base e = 2.7182818. . . .) The accurate result for p(SK) (Equation 12-5) depends on n as well as λ, whereas the approximation for p(SK) (Equation 12-6) depends only on λ.

Equations 12-5 and 12-6 show that p(SK) equals one if either the number of keys in $\underline{K}$ is equal to one (n = 1), or $\lambda = 0$ (e.g., if the number of characters in the intercepted cryptogram approaches infinity, $N \to \infty$). Except for the trivial case where $\underline{K}$ contains only one key, the result implies that a random cipher can be broken with certainty only when an infinite amount of ciphertext is available for analysis.

Table 12-2 contains computed values of p(SK) for different values of $\log_2\lambda$ and $\log_2 n$. The values of n for n much greater than 1 ($n \gg 1$) are computed using the approximation for p(SK) (Equation 12-6), while the remainder of the table entries are computed using the accurate expression for p(SK) (Equation 12-5). It can be seen from Table 12-2 that the approximation for p(SK) can be used in all situations where $\log_2 n$ is greater than 10, without much loss of accuracy. Furthermore, it can be seen that the values of interest are all located in a narrow band on either side of the point where $\log_2\lambda$ equals 0.

Shannon [2] defined unicity distance (ud) as the value of N for which λ equals 1. Note that the condition $\{\lambda = 1\}$ is equivalent to the condition $\{\log_2\lambda = 0\}$. Thus when N = ud, it follows from Equation 12-6 that

$$p(SK) = (e - 1)/e = 0.6321 \qquad \text{for } n \gg 1$$

However, if p(SK) is plotted against N, one observes that the transition be-

$\log_2\lambda$ (Bits)	$\log_2 n$ (Bits)				
	0	1	5	10	n>>1
−14	1.0000	0.9999	0.9999	0.9999	0.9999
−12	1.0000	0.9999	0.9999	0.9999	0.9999
−10	1.0000	0.9998	0.9995	0.9995	0.9995
− 9	1.0000	0.9995	0.9990	0.9990	0.9990
− 8	1.0000	0.9990	0.9981	0.9981	0.9980
− 7	1.0000	0.9980	0.9962	0.9961	0.9961
− 6	1.0000	0.9961	0.9925	0.9922	0.9922
− 5	1.0000	0.9922	0.9850	0.9845	0.9845
− 4	1.0000	0.9844	0.9703	0.9694	0.9693
− 3	1.0000	0.9688	0.9418	0.9401	0.9400
− 2	1.0000	0.9375	0.8879	0.8849	0.8848
− 1	1.0000	0.8750	0.7917	0.7871	0.7869
0	1.0000	0.7500	0.6379	0.6323	0.6321
1	0.0	0.5000	0.4366	0.4325	0.4323
2		0.0	0.2465	0.2455	0.2454
3			0.1250	0.1250	0.1250
4			0.0625	0.0625	0.0625
5			0.0313	0.0313	0.0313
6			0.0	0.0156	0.0156
7				0.0078	0.0078
8				0.0039	0.0039
9				0.0020	0.0020
10				0.0010	0.0010
12				0.0	0.0002
14					0.0

Values in the column denoted "n >> 1" were computed using the equation $p(SK) \simeq (1/\lambda)(1 - e^{-\lambda})$, which holds when $\lambda/2n << 1$. Values in all other columns were computed using the equation $p(SK) = (1/\lambda)(1 - (1 - (\lambda/n))^n)$.

Table 12-2. p(SK) Values for a Random Cipher

tween $p(SK) \simeq 0$ and $p(SK) \simeq 1$ is indeed very sharp. This is illustrated below in an example of simple substitution on English. Hence the loose interpretation of unicity distance resulting from information theory is quite good.

An Example of Simple Substitution on English (Ciphertext Only)

An example is given below in which a random cipher is used to model simple substitution on English. It is shown that the unicity distance is about 22 characters, which agrees quite well with reported values for simple substitution ciphers.

In simple substitution on English, there are $n = 26!$ ways in which a 26-letter plain alphabet can be transformed into a 26-letter cipher alphabet (i.e., the maximum number of possible keys is 26!). However, for small and

moderate values of N, the number of different letters in the message is usually less than 26. Therefore, the *effective* number of keys is less than 26!.

The average numbers of different letters that occur in messages of N characters, for values of N from 5 to 1500 characters, are shown in Table 12-3. Messages of 25 characters contain about 14 different letters. Therefore, the effective number of keys the analyst must cope with is about

$$\begin{aligned} n &= (26)(25)(24) \ldots (13) \\ &= \frac{26!}{12!} \\ &= 8.4 \times 10^{17} \end{aligned}$$

instead of

$$26! = 4.0 \times 10^{27}$$

Message Length N (Characters)	Average Number of Different Letters per Message
5	4.5
10	7.8
15	10.2
20	12.0
25	13.4
30	14.5
40	16.1
50	17.3
75	19.2
100	20.4
200	22.4
300	23.0
400	23.4
500	23.7
700	24.2
1000	24.6
1500	25.2

The samples were taken from the Brown University Corpus of Present-Day English. [3]
Number of sampled messages = 1000.
On the average, 13.4 different letters occur in a sample of 1000 messages of 25 characters.

Table 12-3. Average Number of Different Letters in N Letters of English Text

Since the unicity distance for simple substitution on English is shown to be about 22 characters, no more than 22-gram statistics should be used to approximate s. In the present analysis, 15-gram statistics are used.

Recall that when J-gram statistics are used ($J \geqslant 1$), the number of meaningful messages ($s_{N,J-1}$) can be approximated by

$$s_{N,J-1} \simeq 2^{NF_J}$$

(See Equation F-14 in Appendix F) where N ($\geqslant$J) is the number of characters in the sample messages, J − 1 is the order of the Markov approximation to message probability, and F_J (see Equation F-9 in Appendix F) is a measure of the conditional entropy of the message source.

Using the value $F_{15} = 2.02$ bits per character,[7] $s_{N,14}$ evaluates to

$$s_{N,14} \simeq 2^{N2.02}$$

(To simplify the notation in the discussion below, let s be used in place of $s_{N,14}$.) The total number of messages r is equal to 26^N, which can also be written as

$$r = 2^{(\log_2 26)N} = 2^{4.70N}$$

Therefore, it follows that

$$\begin{aligned} \log_2 \lambda &= \log_2 (ns/r) \\ &= \log_2 n + \log_2 s - \log_2 r \\ &\simeq 59.5 + 2.02N - 4.70N \end{aligned}$$

which can be used to show that $\log_2 \lambda$ is close to 0, or equivalently, that λ is close to 1, when

$$N = 22.2 \text{ characters}$$

This result is interpreted to mean that the unicity distance for simple substitution on English is 22.2 characters when the values of n, s, and r are taken as 26!/12!, $2^{2.02N}$, and $2^{4.70N}$, respectively. That is, ud = 22.2 characters provided that the message contains about 14 different letters and the cryptanalysis makes use of 15-gram statistics.

Referring now to Table 12-2, one finds that

$$p(SK) = 0.63$$

for a random cipher in which n = 26!, $s = 2^{2.02N}$, $r = 2^{4.70N}$, and N = ud.

Just how rapidly p(SK) approaches 1 for values of N above 22.2 charac-

[7] F_{15} is computed, using Equation F-18 and values from Table F-3, as follows: $F_{15} = (5.5/4.5)((2.1 - 1.2)/2) = 2.02$.

N (Characters)	18.9	20.0	21.1	22.2	23.3	24.4	25.6
p(SK)	.0020	.0156	.1250	.6321	.9400	.9922	.9990

Values of p(SK) were obtained from Table 12-2 using $\log_2 \lambda$; values of $\log_2 \lambda$ were calculated from the expression $\{\log_2 \lambda = 59.5 + 2.02N - 4.70N\}$ using N.

Table 12-4. Values of p(SK) for N Near ud Given that a Random Cipher is used to Model Simple Substitution

ters, and 0 for values of N below 22.2 characters, can be seen from Table 12-4.

The p(SK) values obtained with a random cipher agree with empirical observations for simple substitution on English. Friedman indicates that

> Practically every example of 25 or more characters representing monoalphabetic encipherment of a 'sensible' message in English can be readily solved [7].

According to Shannon,

> The unicity point . . . can be shown experimentally to lie between the limits 20 and 30. With 30 letters there is nearly always a unique solution to a cryptogram of this type and with 20 it is usually easy to find a number of solutions [2].

Such a close agreement between theoretical and empirical results indicates that the underlying assumptions of a random cipher are good. This same general agreement holds for other ciphers as well (e.g., Caesar, transposition, and Vigenere) [2]. A more detailed treatment of unicity distance computations can be found in Appendix G.

Probability of Obtaining the Key When Plaintext and Corresponding Ciphertext Are Available for Analysis

Consider now a cryptanalysis involving plaintext and corresponding ciphertext. Again, a random cipher is assumed. Let y be the cryptogram produced when a known message x is enciphered with an unknown key kj. In the analysis that follows, x may be a meaningful or meaningless message (i.e., x may be any of the r messages in $\underline{X}$). As before, the probability that kj deciphers y into x is 1:

$$p[D_{kj}(y) \text{ equals } x] = 1$$

Likewise, the process of deciphering y with each of the $(n - 1)$ incorrect keys, k1, k2, . . . , $kj - 1$, $kj + 1$, . . . , kn, can be thought of as $(n - 1)$ Bernoulli trials, except now the probability that a key will successfully decipher y into x is 1/r:

$$p[D_{ki}(y) \text{ equals } x] = 1/r$$

and the probability that a key will fail to decipher y into x is $1 - 1/r$:

$$p[D_{ki}(y) \text{ not equal } x] = 1 - 1/r$$

for all values of i not equal to j.

Clearly, cryptanalysis involving plaintext and corresponding ciphertext (where the number of meaningful messages s equals 1) is a special case of the previous analysis. It follows therefore that p(SK) can be calculated from Equations 12-5 and 12-6, except that λ equals n/r instead of ns/r. Since λ is reduced by a factor of s, it is not surprising that the cipher is more vulnerable to attack.

Probability of Obtaining the Plaintext

The emphasis here is on analyzing a random cipher from the viewpoint of obtaining the correct plaintext for a given intercepted cryptogram, without regard to whether one obtains the correct key. Recall that U is defined as the number of meaningful messages that can be recovered from the intercepted cryptogram.

In Appendix H, accurate expressions are derived for the distribution of U (see Equations H-1a and H-3). Using these equations, accurate values were computed for the expected value of U, denoted E(U), and the probability of successfully obtaining the correct plaintext, denoted by p(SM) (see Table 12-5). In particular, values of E(U) and p(SM) are given for the case where the number of keys n equals 32 ($\log_2 n = 5$), and for different values of the number of meaningful messages (s = 1, 4, 8, 16, and 32) and λ($\lambda = 2^{-6}, 2^{-3}, 2^{-2}, 2^{-1}, 2^{0}, 2^{1}, 2^{2}$, and 2^{3}). Recall that $\lambda = ns/r$, that is, λ equals the number of keys n multiplied by the number of meaningful messages s divided by the number of cryptograms r.

For purposes of comparison, values for p(SK) from Table 12-2 and values for E(M) computed from Equation 12-4b are also given in Table 12-5. This shows that the difference between p(SM) and p(SK), and the difference between E(U) and E(M), are not too great even for small values of s and n. In an actual cipher, s and n would be much larger than 32; thus no distinction needs to be made between p(SK) and p(SM).

Approximations for p(U = u), E(U), and p(SM) are as follows (see also Equations H-7 thru H-9 in Appendix H):

$$\begin{aligned} p(U = u) &\simeq p(M' = u - 1)[1 + \lambda'/s + \\ &\qquad (u - 1)(\lambda' - 1)/2s - (u - 1)^2/2s] \\ &\qquad \text{for } u \geqslant 1 \\ E(U) &\simeq E(M)[1 - (\lambda'/2s)(\lambda' + (2/\lambda') + 1)] \\ p(SM) &\simeq p(SK)(1 + \lambda'/2s) \end{aligned}$$

where $p(M' = u - 1)$ can be deduced from Equation 12-4a, since $p(M' = u - 1)$ equals $p(M = u)$. The values obtained with these approximations differ by less than 5 percent from the corresponding more accurate values given in Table 12-5.

$\log_2 n = 5$

$\log_2 \lambda$ (Bits)	p(SM) $\log_2 s$ (Bits)					p(SK)
	1	2	3	4	5	
−6	0.9962	0.9943	0.9934	0.9929	0.9927	0.9925
−3	0.9706	0.9561	0.9490	0.9454	0.9436	0.9418
−2	0.9430	0.9153	0.9016	0.8949	0.8915	0.8879
−1	0.8925	0.8416	0.8168	0.8047	0.7986	0.7917
0	0.8080	0.7211	0.6800	0.6600	0.6502	0.6379
1	0.6869	0.5533	0.4928	0.4642	0.4503	0.4366
2	0.5676	0.3891	0.3130	0.2785	0.2622	0.2465
3	0.5080	0.2882	0.1974	0.1587	0.1412	0.1250
			E(U)			E(M)
−6	1.008	1.0113	1.0132	1.0142	1.0147	1.0151
−3	1.060	1.0895	1.1052	1.1131	1.1171	1.1211
−2	1.1140	1.1763	1.2087	1.2253	1.2337	1.2422
−1	1.2151	1.3421	1.4113	1.4473	1.4657	1.4844
0	1.3839	1.6453	1.7983	1.8813	1.9244	1.9688
1	1.6263	2.1588	2.5109	2.7139	2.8230	2.9375
2	1.8648	2.8788	3.7040	4.2376	4.5420	4.875
3	1.9841	3.5943	5.3839	6.7940	7.6910	8.75

Note that $0 \leqslant s/r \leqslant 1$, $0 \leqslant \lambda = ns/r \leqslant n$, and $p(SK) = 0$ for $\lambda > n$.

Table 12-5. Values of p(SM) and E(U) where the Number of Keys is Fixed ($n = 32$)

AN EXPANSION OF SHANNON'S APPROACH USING INFORMATION THEORY

In this section, the unicity distance of a cipher with message and key probabilities is discussed in terms of *information theory* [8]. (No assumption is made about the distribution of M, that is, the discussion is not limited to random ciphers but pertains to ciphers in general.)

Consider a message to be a string of symbols, where the symbols belong to a source alphabet. Information theory applies a numerical measure of information to a message, whose value is frequently given in "bits".[8] Based upon this measure is the notion of *entropy*, whose value is frequently given in terms of "bits per symbol". The value of the entropy depends on the statistical or probabilistic properties of the set of messages composed from the source alphabet, rather than the semantics of the particular message. Let $\underline{X} = \{x1, x2, \ldots, xr\}$ denote r different messages with probabilities p1, p2, . . . , pr. The information measure associated with the selection of one member xi from $\underline{X}$, is "$-\log_2 pi$" bits of information. When each message is equally likely, the probability of each message is "1/r" bits, and each message has information value "$\log_2 r$". This is often written as "$-\log_2 (1/r)$"; which is the negative of the $\log_2$ of the probability. For set $\underline{X}$, the average information per message is defined to be the entropy of $\underline{X}$, denoted $H(\underline{X})$. Entropy is defined by the expression:

$$H(\underline{X}) = \sum_{i=1}^{r} -(pi)\log_2 (pi)$$

If the messages are equally likely, then $H(\underline{X})$ assumes its maximum value of $\log_2 r$. In that case, $\log_2 r$ bits are needed to encode or represent each message and the message bears all the information that is received (i.e., the receiver has no information about which message is selected and sent). For example, if $\underline{X} = \{x1, x2, x3, x4\}$ and $p1 = p2 = p3 = p4 = 1/4$, $H(\underline{X})$ equals 2. Thus two bits are needed to represent each message.

However, if the messages are unequally likely, one has in advance something that any gambler, speculator, or forecaster would instantly recognize as information. The additional information contributed by the received message is lessened by that amount. For example, if the 4 messages above are assigned probabilities $p1 = 1/2$, $p2 = 1/4$, and $p3 = p4 = 1/8$, then $H(\underline{X})$ equals 1.75 bits. On the other hand, if $p1 = 1$ and $p2 = p3 = p4 = 0$, then $H(\underline{X})$ equals 0. That is, the received message, which is predictable, provides no additional information.

Alternatively, $H(\underline{X})$, the entropy function of p1, p2, . . . , pr, can be interpreted as a measure of the *uncertainty* over which message the sender will select and transmit to the receiver. (Recall that $H(\underline{X})$ assumes values in the interval 0 to $\log_2 r$.) When $H(\underline{X}) = \log_2 r$, there is maximum uncertainty (i.e., the receiver has no information about the message that will be transmitted).

[8] In the pure binary system, a bit is either 0 and 1.

When $H(\underline{X}) = 0$, there is no uncertainty (i.e., the receiver knows in advance which message will be transmitted).

Information theory relies heavily on the mathematical science of probability. For this reason, information theory has been applied to other probabilistic studies in communication theory, cryptanalysis, and the like. In the study of cryptanalysis, $H(\underline{X})$ and $H(\underline{K})$ represent the analyst's prior information over which message and key are selected for encipherment.

Information measures provide an alternative approach for discussing unicity distance. However, because of certain required approximations, the results obtained with this approach are different from those obtained using other probabilistic measures (not based on information theory). In the former case, the relationship between the probability of obtaining the key (or data) and cryptogram length is a step function: the probability is zero when the cryptogram's length is less than the unicity distance, and one when its length is greater than the unicity distance. However, because the transition region—defined by the values of N for which the probability of obtaining the key (or data) is neither close to zero nor close to one—is very small, either approach provides useful results.

Information Measures[9]

The following is a list of common information measures useful to a discussion of theoretical secrecy.

1. Entropy of U:
$$H(U) = -\sum_{u} p(u)\log_2 p(u) \qquad (12\text{-}7a)$$

2. Conditional entropy of U given element v:
$$H(U|v) = -\sum_{u} p(u|v)\log_2 p(u|v) \qquad (12\text{-}7b)$$

3. Equivocation of U given V:
$$H(U|V) = \sum_{v} p(v)H(U|v) \qquad (12\text{-}7c)$$

4. Entropy of U and V:
$$H(U, V) = -\sum_{u,v} p(u, v)\log_2 p(u, v) \qquad (12\text{-}7d)$$

5. Equivocation of U given V and W:
$$H(U|V, W) = -\sum_{u,v,w} p(u, v, w)\log_2 p(u|v, w) \qquad (12\text{-}7e)$$

6. Equivocation of U and V given W:
$$H(U, V|W) = -\sum_{u,v,w} p(u, v, w)\log_2 p(u, v|w) \qquad (12\text{-}7f)$$

[9] Information measures are discussed in greater detail in Appendix F. See also reference 9.

7. Entropy of U, V, and W:

$$H(U, V, W) = -\sum_{u,v,w} p(u, v, w)\log_2 p(u, v, w) \qquad (12\text{-}7g)$$

The following is a list of information identities and relations the proofs of which are left to the reader.

1. $H(U, V) = H(U|V) + H(V)$ (12-7h)
2. $H(U, V, W) = H(U|V, W) + H(V, W) = H(U, V|W) + H(W)$ (12-7i)
3. $H(U) = H(U|V)$; if U and V are independent, i. e., $p(u, v) = p(u)p(v)$ (12-7j)
4. $H(U) > H(U|V)$; if U and V are dependent, i. e., $p(u, v) \neq p(u)p(v)$ (12-7k)
5. $H(U|V, W) + H(V|W) = H(V|U, W) + H(U|W)$ (12-7l)

In the expressions above, U, V, and W are finite sets whose elements have been assigned probabilities such that

$$\sum_u p(u) = \sum_v p(v) = \sum_w p(w) = 1$$

Unicity Distance for a Cipher When Only Ciphertext Is Available for Analysis

Let it be shown first that $H(\underline{K}, \underline{Y})$ equals $H(\underline{K}, \underline{X})$. From the general relation

$$H(U, V, W) = H(U|V, W) + H(V, W)$$

(see Equation 12-7i) it follows, with an appropriate change of variables, that

$$H(\underline{X}, \underline{K}, \underline{Y}) = H(\underline{X}|\underline{K}, \underline{Y}) + H(\underline{K}, \underline{Y})$$

and

$$H(\underline{Y}, \underline{K}, \underline{X}) = H(\underline{Y}|\underline{K}, \underline{X}) + H(\underline{K}, \underline{X})$$

Hence it follows that

$$H(\underline{K}, \underline{Y}) - H(\underline{K}, \underline{X}) = H(\underline{Y}|\underline{K}, \underline{X}) - H(\underline{X}|\underline{K}, \underline{Y})$$

But since a cipher satisfies $y = E_k(x)$ and $x = D_k(y)$, a knowledge of k and y permits x to be derived, and a knowledge of k and x permits y to be derived.[10]

[10] An exception to this rule is a homophonic substitution cipher (see Appendix G) where $E_k(x)$ defines a set of cryptograms. Encipherment includes the additional step of selecting (usually randomly) one of the cryptograms from this set.

Therefore,

$$H(\underline{X}|\underline{K}, \underline{Y}) = H(\underline{Y}|\underline{K}, \underline{X}) = 0$$

and consequently

$$H(\underline{K}, \underline{Y}) = H(\underline{K}, \underline{X})$$

But, by Equation 12-7h, $H(\underline{K}, \underline{Y})$ can be rewritten as

$$H(\underline{K}, \underline{Y}) = H(\underline{K}|\underline{Y}) + H(\underline{Y})$$

Moreover, since messages and keys are selected independently, that is, $p(x, k) = p(x)p(k)$ for all x in $\underline{X}$ and k in $\underline{K}$, it follows that

$$H(\underline{K}, \underline{X}) = H(\underline{K}) + H(\underline{X})$$

A general equation for $H(\underline{K}|\underline{Y})$ is thus obtained:

$$H(\underline{K}|\underline{Y}) = H(\underline{K}) - H(\underline{Y}) + H(\underline{X})$$

This relationship can now be used to derive the unicity distance of a cipher. Since

$$H(\underline{K}|\underline{Y}) = \sum_{y} p(y)H(\underline{K}|y)$$

$H(\underline{K}|\underline{Y})$ measures the average uncertainty over which key was used to encipher the selected (unknown) message into the given (known) cryptogram. The condition $\{H(\underline{K}|\underline{Y} = 0\}$ implies that there is no uncertainty over which key was used for enciphering (the produced cryptogram is assumed available for analysis). The following definition for unicity distance can now be given.

The *unicity distance* (ud) of a cipher in which only ciphertext is available for analysis is the value of N for which

$$H(\underline{K}) - H(\underline{Y}) + H(\underline{X}) = 0 \qquad (12\text{-}8)$$

provided that such an N exists.

Assume that the analyst has no prior information concerning which message(s) and key(s) are selected for encipherment. In that case, the analyst considers keys to be equally likely, and therefore assigns equal probability to each key in $\underline{K}$. Hence,

$$H(\underline{K}) = \log_2 n$$

(n denotes the number of keys in $\underline{K}$). From the concept of meaningful and meaningless messages, it follows that

$$H(\underline{X}) \simeq \log_2 s$$

(s denotes the number of meaningful messages). Assuming that cryptograms are nearly equally probable, it follows that

$$H(\underline{Y}) \simeq \log_2 r$$

(r denotes the number of cryptograms in $\underline{Y}$). An approximation for $H(\underline{K}|\underline{Y})$ is therefore obtained as follows:

$$\begin{aligned} H(\underline{K}|\underline{Y}) &= H(\underline{K}) - H(\underline{Y}) + H(\underline{X}) \\ &\simeq \log_2 n - \log_2 r + \log_2 s \\ &\simeq \log_2 (ns/r) \\ &\simeq \log_2 \lambda \end{aligned} \tag{12-9}$$

Equation 12-9 shows that $H(\underline{K}|\underline{Y})$ is near zero when λ equals 1, and therefore that no uncertainty should remain regarding which key was used to encipher the selected (unknown) message into the given (known) cryptogram. However, the results obtained with a random cipher (Table 12-2) are different. When λ equals 1, the probability of obtaining the correct key, p(SK), is 0.6321. The reason for the discrepancy is as follows. In a random cipher, $H(\underline{Y})$ is about equal to $\log_2 r$ only when λ ($\lambda = ns/r$) is much greater than 1. When λ is near 1 the approximation $\{H(\underline{Y}) \simeq \log_2 r\}$ is no longer valid. In that case, the value of $H(\underline{Y})$ is strictly less than $\log_2 r$. This means that the value of N for which the expression $H(\underline{K}|\underline{Y}) - H(\underline{Y}) + H(\underline{X})$ equals zero is greater than the value of N for which λ equals 1. (Since the number of keys n is constant, if an N exists for which $H(\underline{K}|\underline{Y})$ equals zero, then as N becomes very large the ratio s/r will approach 1/n and λ will approach 1.) Equation 12-9 therefore permits only a rough approximation of unicity distance.

Unicity Distance for a Cipher When Plaintext and Corresponding Ciphertext Are Available for Analysis

From the general relationship

$$H(U|V, W) + H(V|W) = H(V|U, W) + H(U|W)$$

(see Equation 12-71) and an appropriate change of variables, it follows that

$$H(\underline{K}|\underline{Y}, \underline{X}) + H(\underline{Y}|\underline{X}) = H(\underline{Y}|\underline{K}, \underline{X}) + H(K|\underline{X})$$

But since a knowledge of k in $\underline{K}$ and x in $\underline{X}$ implies a knowledge of $y = E_k(x)$ in $\underline{Y}$,

$$H(\underline{Y}|\underline{K}, \underline{X}) = 0$$

and keys and messages are selected independently,

$$H(\underline{K}|\underline{X}) = H(\underline{K})$$

it follows that

$$H(\underline{K}|\underline{Y}, \underline{X}) = H(\underline{K}) - H(\underline{Y}|\underline{X})$$

The condition $\{H(\underline{K}|\underline{Y}, \underline{X}) = \text{zero}\}$ means that there is no uncertainty regarding which key was used to encipher the given (known) plaintext into the given (known) ciphertext. Therefore, the *unicity distance* (ud) of a cipher in which plaintext and corresponding ciphertext are available is the value of N for which

$$H(\underline{K}) - H(\underline{Y}|\underline{X}) = 0 \tag{12-10}$$

provided that such an N exists. Recall that Shannon defined the ud of a random cipher as the value of N for which ns/r ($ns/r = \lambda$) equals one (see Probability of Obtaining the Key When Only Ciphertext is Available for Analysis).

When plaintext and corresponding ciphertext are available for analysis, the set of meaningful messages can be thought of as containing only a single element (the given plaintext):

$$s = 1$$

The remaining $r - 1$ messages are therefore treated as meaningless. Thus the ud of a random cipher in which plaintext and corresponding ciphertext are available for analysis is the value of N for which n/r equals 1. By taking the logarithm (base 2) of each side of the equation and replacing $\log_2 n$ with $H(\underline{K})$, ud becomes the value of N for which

$$H(\underline{K}) - \log_2 r = 0$$

Comparing this with Equation 12-10, one can see that the derived expression provides only a rough approximation to ud. The condition $\{H(\underline{Y}|\underline{X}) \simeq \log_2 r\}$ does not hold when n/r ($n/r = \lambda$) is near 1.

Relationships Among $H(\underline{X}|\underline{Y})$, $H(\underline{K}|\underline{Y})$, and $H(\underline{K}|\underline{X}, \underline{Y})$

The information measures $H(\underline{X}|\underline{Y})$, $H(\underline{K}|\underline{Y})$, and $H(\underline{K}|\underline{X}, \underline{Y})$ are of particular interest in cryptanalysis. In each case, the value of N for which the respective measure is equal to zero can be used to define the ud of the cipher. The measure $H(\underline{X}|\underline{Y})$ corresponds to the case where the analyst solves for the plaintext instead of the key, under the assumption that only ciphertext is available for analysis. The measures $H(\underline{K}|\underline{Y})$ and $H(\underline{K}|\underline{X}, \underline{Y})$ have already been discussed.

From Equation 12-71 and an appropriate change of variables, it follows that

$$H(\underline{K}|\underline{X}, \underline{Y}) + H(\underline{X}|\underline{Y}) = H(\underline{X}|\underline{K}, \underline{Y}) + H(\underline{K}|\underline{Y})$$

But a knowledge of k in $\underline{K}$ and y in $\underline{Y}$ implies a knowledge of $x = D_k(y)$ in $\underline{X}$:

$$H(\underline{X}|\underline{K}, \underline{Y}) = 0$$

Therefore, it follows that

$$H(\underline{X}|\underline{Y}) = H(\underline{K}|\underline{Y}) - H(\underline{K}|\underline{X}, \underline{Y}) \qquad (12\text{-}11)$$

But

$$H(\underline{X}|\underline{Y}) \geqslant 0$$

implies that

$$H(\underline{K}|\underline{Y}) \geqslant H(\underline{K}|\underline{X}, \underline{Y})$$

Thus in a cipher where $H(\underline{X}|\underline{Y})$ and $H(\underline{K}|\underline{Y})$ approach zero as N becomes large and the number of keys in $\underline{K}$ remains constant, it follows that the information measures $H(\underline{X}|\underline{Y})$, $H(\underline{K}|\underline{Y})$, and $H(\underline{K}|\underline{X}, \underline{Y})$ can be plotted as depicted in Figure 12-4. This conclusion can be reached via the following:

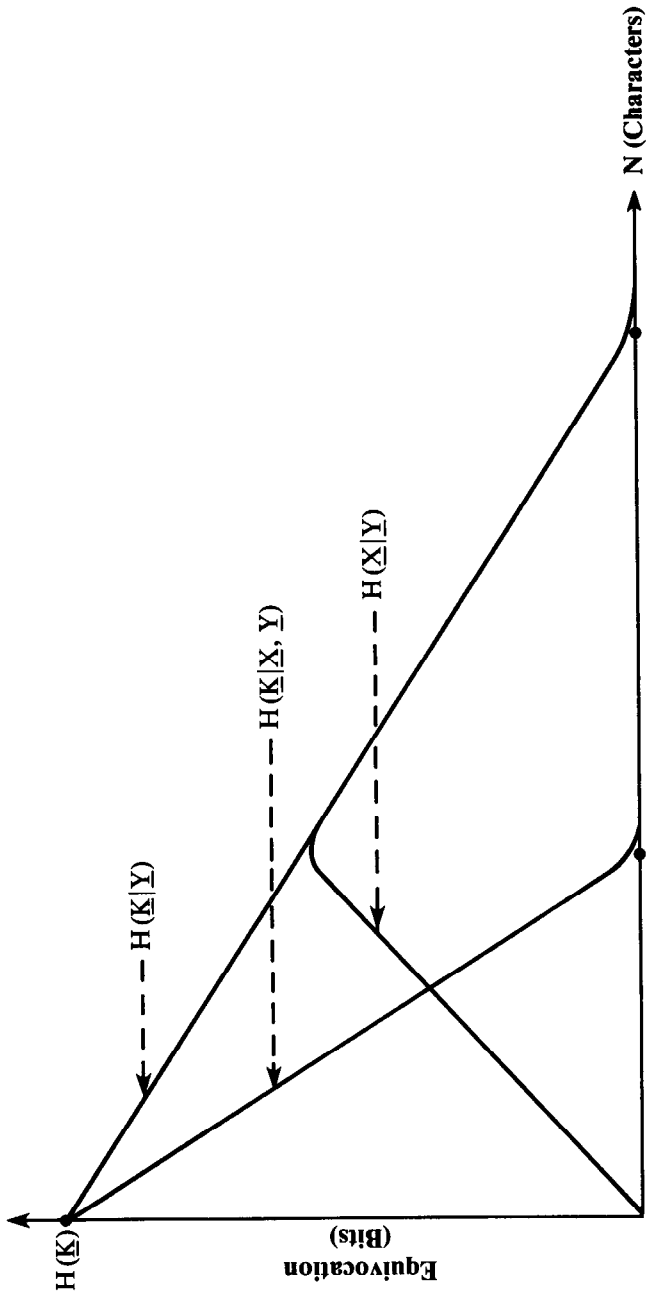
1. When N equals zero, it is assumed that $\underline{X}$ and $\underline{Y}$ each contain one element (i.e., $\underline{X}$ contains a null message x_0, and $\underline{Y}$ contains a null cryptogram y_0). Each of the n keys in $\underline{K}$ map x_0 to y_0. Hence $H(\underline{K}|\underline{Y})$ and $H(\underline{K}|\underline{X}, \underline{Y})$ are both equal to $H(\underline{K})$, and $H(\underline{X}|\underline{Y})$ is equal to zero.
2. Generally, when plaintext and corresponding ciphertext are available for analysis, one can solve for the key more easily than when only ciphertext is available for analysis. Thus when N is greater than zero, the value $H(\underline{K}|\underline{X}, \underline{Y})$ is strictly less than the value $H(\underline{K}|\underline{Y})$:

   $$H(\underline{K}|\underline{Y}) > H(\underline{K}|\underline{X}, \underline{Y})$$

   for $N > 0$. This means that $H(\underline{K}|\underline{X}, \underline{Y})$ will approach zero more rapidly than will $H(\underline{K}|\underline{Y})$.
3. When $H(\underline{K}|\underline{X}, \underline{Y})$ is near zero, Equation 12-10 indicates that $H(\underline{X}|\underline{Y})$ is approximately equal to $H(\underline{K}|\underline{Y})$:

   $$H(\underline{X}|\underline{Y}) \simeq H(\underline{K}|\underline{Y})$$

   This in turn says that $H(\underline{X}|\underline{Y})$ and $H(\underline{K}|\underline{Y})$ will nearly coincide and approach zero together. That $H(\underline{X}|\underline{Y})$ is about equal to $H(\underline{K}|\underline{Y})$ at the ud where $H(\underline{K}|\underline{Y})$ equals zero agrees with the previous result for a random cipher (Table 12-5) indicating that p(SK) is about equal to p(SM) at the ud (when λ equals 1).



Note: $H(X|Y) = H(K|Y) - H(K|X, Y)$

Figure 12-4. Plot of $H(X|Y)$, $H(K|Y)$ and $H(K|X, Y)$ Against N for a Cipher in which $H(X|Y)$ and $H(K|Y)$ Approach Zero as N Becomes Large

Unicity Distance for the Data Encryption Standard

If plaintext and corresponding ciphertext are available for analysis, the unicity distance of the DES algorithm can be approximated using Equation 12-10. That is, ud is the value of N for which

$$H(\underline{K}) - H(\underline{Y}|\underline{X}) = 0$$

Since there are 2^{56} possible keys in DES, it follows that

$$H(\underline{K}) = 56$$

and therefore, ud is the value of N (in 8-bit characters) for which $H(\underline{Y}|\underline{X}) = 56$. From Equations 12-7b and 12-7c,

$$H(\underline{Y}|\underline{X}) = -\sum_{y,x} p(y, x)\log_2 p(y|x)$$

Thus, $H(\underline{Y}|\underline{X}) = 56$ when $p(y|x) = 1/2^{56}$ for each x and each y produced from x, and $p(y|x) = 0$ for each x and each y not produced form x (i.e., when no two keys map x to the same cryptogram).

If the 2^{56} cryptograms produced by mapping x under each of the 2^{56} keys are considered to be selected at random using replacement from the set of 2^{8N} possible cryptograms, then almost all of the cryptograms will be unique for values of N greater than 8. Thus, a ud of about 8 characters is obtained for DES (i.e., one block of ciphertext is ordinarily enough to determine the key). The precise calculations are omitted.

If only ciphertext is available for analysis, then ud is the value of N for which

$$H(\underline{K}) - H(\underline{Y}) + H(\underline{X}) = 0$$

(see Equation 12-8). If messages in $\underline{X}$ consist of English text (26 letter alphabet with no spaces) and each letter is represented by an 8-bit character, then 26 of the 256 possible 8-bit characters have probabilities corresponding to normal English text, and the other 230 characters have zero probability.

When individual letter probabilities are taken into consideration, the number of meaningful N-character messages, s, can be approximated as follows:

$$s \simeq 2^{4.17N}$$

(See Number of Meaningful Messages in a Redundant Language.) Therefore, it follows that

$$H(\underline{X}) \simeq 4.17N$$

As a first order approximation, assume that each cryptogram in $\underline{Y}$ is equally probable, and hence that

$$H(\underline{Y}) \simeq \log_2 r$$
$$\simeq 8N$$

Substituting the values $H(\underline{K}) = 56$, $H(\underline{Y}) = 8N$, and $H(\underline{X}) = 4.17N$ into equation $\{H(\underline{K}) - H(\underline{Y}) + H(\underline{X}) = 0\}$, one obtains:

$$N = 56/(8 - 4.17)$$
$$= 14.6 \text{ characters}$$

Thus 15 characters or 2 blocks (after rounding up to the nearest block) of ciphertext are enough in theory to solve for the key.

The fact that the ud of DES is only a few characters clearly demonstrates that a cipher which has good practical secrecy does not necessarily have good theoretical secrecy. The strength of DES is based entirely on the prohibitive amount of time and resources required to break it. (See Work Factor as a Measure of Secrecy in this chapter, and Cryptographic Algorithms, Chapter 2.)

Other examples of unicity distance computations can be found in Appendix G.

WORK FACTOR AS A MEASURE OF SECRECY

To show the relationship among the work factor (the time and resources it takes to break a cipher), the sophistication of the attack, and the amount of information available to the cryptanalyst, a simple substitution cipher is analyzed. Although the example is that of a weak cipher, the analysis provides an insight into the approach to be taken with stronger ciphers.

The Cost and Time to Break a Cipher

No matter what method of cryptanalysis is used, the analyst must always expend some amount of time and resources (defined as work factor) to reach his goal. Usually, there is a cost associated with each of the resources used, permitting the overall cost of recovering the key or message to be determined. By increasing available resources, such as computing power, storage, human efforts, and the like, the time required to attack the cipher successfully can often be reduced. Consequently, there is a relationship between cost and time for any given cryptanalytic attack against a cipher.

The information obtained by breaking a cipher also has a value (expressed in financial terms) based on what it is worth to the opponent. Usually, the information decreases in value over its lifetime, which permits a relationship to be established between value and time similar to that between cost and time. The relationships between cost and time and between value and time can then be used in determining the practical secrecy of the cipher.

The cost and time to break a cipher are functions of how it is attacked. Since there may be many different ways to cryptanalyze a cipher, many cost-time relationships are possible. Usually, the cost of breaking a cipher is estimated on the basis of the best known method of attack, even though it may not be the best method altogether.

Cryptanalysis involves high-speed computers and complex, sophisticated computer programs. This includes the following:

1. Computer processors, including special-purpose hardware used to execute the logical and arithmetic operations needed to obtain the solution.
2. Computer storage for the analysis programs and data.
3. Human resources to devise and write analysis programs, gather data, and oversee the analysis.

Simple Substitution on English—Some Preliminaries

An example of simple substitution on English (in which only ciphertext is available for analysis) shows the relationships existing among the cost and time for analysis, the language statistics used for analysis, and the amount of available ciphertext. Results are obtained empirically.

Two different approaches are considered: single-letter frequency analysis, and digram-frequency analysis. To evaluate both approaches, a plaintext is enciphered with a randomly chosen key. The resulting ciphertext is then analyzed to determine how many characters of the key and how many characters of the plaintext are correctly obtained. An important factor is knowing how much better the obtained solution is than a result obtained by pure guessing (random selection).

Let t be the number of characters in the key and p(w) be the probability that w characters of the key are properly obtained by random choice, $0 \leqslant w \leqslant t$. It can be shown [5] that

$$p(w) = (1/w!) \sum_{i=0}^{t-w} (-1)^i (1/i!) \tag{12-12}$$

For a large t, the finite series above can be replaced by an infinite series whose sum is given by 1/e:

$$\lim_{t \to \infty} p(w) = (1/w!)(1/e) = (1/w!)0.368 \tag{12-13}$$

which represents the Poisson distribution with mean equal to 1.

A comparison with the Poisson distribution shows that for $t \geqslant 10$, the values obtained with Equations 12-12 and 12-13 agree to 4 decimal places. Thus when t = 26 (26 letters) or t = 27 (26 letters and space), the Poisson distribution is an excellent approximation to Equation 12-12. Let

$$p(\text{number of correctly guessed key symbols} > c) = a \tag{12-14a}$$

$$p(\text{number of correctly guessed key symbols} \leqslant c) = 1 - a \qquad (12\text{-}14b)$$

where c can be any value from 0 to t (t = the number of characters in the key). With the aid of a table of Poisson probabilities, the values of a and $1 - a$ can be evaluated for different values of c (Table 12-6).

If the number of correctly obtained key characters is greater than 5, as one might anticipate when cryptanalysis is performed, then the hypothesis that keys were obtained by random guessing can be rejected at a level of confidence of 99.94% (Table 12-6).

c	6	5	4	3	2	1
a	.0001	.0006	.0037	.0190	.0803	.2692
1 – a	.9999	.9994	.9963	.9810	.9197	.7358

Table 12-6. Values of "a" and "1 – a" for Different Values of c

In one set of tests, using a single-letter frequency analysis on simple substitution on English (26 letters and space), it was determined that about 6 key characters are recovered from a plaintext containing 250 characters. This result is not too useful by itself, since text with only 6 correct (21 incorrect) characters looks more like a cryptogram than an intelligent message. However, a single-letter frequency analysis is helpful if it is used to obtain an initial key for a more powerful digram-frequency analysis. This initial key is usually better than could be obtained using random selection.

In a digram-frequency analysis an initial key (obtained via a single-letter frequency analysis), is used to decipher the cryptogram. The new digram statistics associated with the trial decipherment are then evaluated and used as a basis for adjusting the initial key. This process is repeated several times, so that the final key is likely to contain more correct characters than the starting key. During these iterations, a certain element of randomness is purposely introduced into the algorithm. This has the effect that repeated analysis of the same cryptogram does not (except with low probability) produce the same path to a solution. In that case, repeated analyses with the algorithm can be considered as statistically independent events, and therefore the probability of success (breaking the cipher) can be increased by increasing the number of trials. It is assumed that the probability of success at each trial is the same and that the number of trials are selected in advance. With the assumption of statistical independence, the distribution of the number of successful trials is therefore given by the binomial distribution.

Based on the observation that a text which is 90% recovered can still be read, the analysis is considered a success if at least 90% of the plaintext characters are successfully recovered. A partial printout of a message which is 91.7% correct (21 correct key characters) is shown below.

NATIONAL BUREAU OW STANFARFS CRYPTODRAPHIC ALDORITHMS WOR PROTECTION OW COMPUTER FATA FURIND TRANSMISSION ANF FORMANT

STORADE SOLICITATION OW PROPOSALS THE NATIONAL BUREAU OW STANFARFS UNFER FEPARTMENT OW COMMERCE AUTHORITIES ANF RESPONSIBILITIES WOR WOSTERIND PROMOTIND ANF FEVELOPIND US TRAFE AND COMMERCE ANF BASEF ON THE NATIONAL BUREAU OW STANFARFS RESPONSIBILITY WOR THE CUSTOFY MAINTENANCE ANF FEVELOPMENT OW THE NATIONAL STANFARFS OW MEASUREMENT ANF PROVISION OW MEANS ANF METHOFS WOR MAKIND MEASUREMENTS CONSISTENT GITH THOSE STANFARFS SOLICITS PROPOSALS WOR THE ENCRYPTION OW COMPUTER FATA

Let

$$p(SM) = \text{the probability that at least 90\% of the plaintext is recovered as the result of cryptanalysis} \tag{12-15}$$

Using a method of confidence limits [9], it can be shown that

$$p(pmin \leqslant p(SM) \leqslant pmax) = \gamma \tag{12-16}$$

where

$$pmin = x/(x + (n - x + 1)F\gamma)$$

$$pmax = (x + 1)F\gamma/((n - x) + (x + 1)F\gamma)$$

$$x = \text{the number of successful attacks in n trials}$$

$F\gamma$ is the *F distribution* with $[2(n - x + 1), 2x]$ degrees of freedom in pmin, and $[2(x + 1), 2(n - x)]$ degrees of freedom in pmax. Thus x/n can be used as an estimate for p(SM).

When a binomial distribution can be approximated by a normal distribution (whenever $Var(x) > 3$), the following mathematically more convenient approach can be used [10].

$$pmin = (1/(n + z^2))[x - 0.5 + (z^2/2) - z[(x - 0.5)((n - x + 0.5)/n) + (z^2/4)]^{1/2}] \tag{12-17a}$$

$$pmax = (1/(n + z^2))[x + 0.5 + (z^2/2) + z[(x + 0.5)((n - x - 0.5)/n) + (z^2/4)]^{1/2}] \tag{12-17b}$$

The value of z is determined by the chosen level of confidence γ (Equation 12-16) and the normal distribution function Φ (whose mean is 0 and variance is 1) as follows

$$\Phi(z) - \Phi(-z) = \gamma$$

If $\gamma = 0.95$, then $z = 1.96$. Since the intent here is only to demonstrate the basic approach, the approximations given by Equations 12-17a and 12-17b are used in the computations.

Now, let

$$p(SM, m) = \text{the probability that at least 90\% of the plaintext is recovered in at least one out of m repeated trials of the analysis} \quad (12\text{-}18)$$

By evaluating p(SM), one is able to approximate p(SM, m). From earlier remarks, it follows that

$$p(SM, 1) = p(SM) \quad (12\text{-}19a)$$

$$p(SM, m) = 1 - (1 - p(SM, 1))^m \quad (12\text{-}19b)$$

In practical situations, Equation 12-19b will be useful for moderate values of p(SM). If p(SM) is very small, it means that there is not enough ciphertext available for analysis. Hence allowing large values of m will not result in a significant improvement. Furthermore, as m becomes large, it also becomes impractical for a person to scan all the recovered plaintext solutions. (Remember that the figure of 90% is based on a *person's* ability to enlarge upon a solution known to be incomplete.)

Empirical Results for Simple Substitution on English Using a Digram-Frequency Analysis

The first part of the analysis provides a statistical estimate for p(SM) (defined in Equation 12-15). The following procedure is used. A plaintext and random key are selected and used to produce the ciphertext to be analyzed. Prior to each cryptanalysis of the ciphertext, a starting key is produced using a single-letter frequency analysis. The success of the attack varies according to the search characteristics (determined by a random process). The procedure is executed n times as n independent trials of an experiment. Thus an estimate for p(SM) can be obtained using a sample size of n.

The basic idea (attributed to D. Coppersmith, IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y.) is to make repeated pairwise changes to the starting key, eventually producing a final key close or equal to the actual key. The method used is to interchange the plain characters assigned to two randomly selected cipher characters. If the new digram matrix based on these changes is closer to a standard digram matrix, the key is changed. The measure of closeness is based on whether the dot product of the vectors, defined by the affected rows and columns of the two digram matrices, increases or not.

Analysis is carried out with ciphertext of length N = 250, 275, 300, 350, 400, 500, 600, 700, 800, 900, and 1000 characters, respectively, and a value of n = 120. By rearranging the 120 observed values into 40 groups of 3 each, one obtains an estimate for the probability of success of a multiple digram-frequency analysis with 3 repetitions. If at least 90% of the plaintext is recovered in at least one of the 3 trials, the attack is considered a success.

Plaintext* Length N	Sample Size, n = 120			Sample Size, n = 40		
	p(SM)	Confidence Limits $\gamma = 95\%$		p(SM, 3)	Confidence Limits $\gamma = 95\%$	
250	.008	.000	.052	.025	.001	.147
275	.025	.006	.077	.075	.020	.215
300	.200	.135	.285	.525	.363	.682
350	.433	.344	.527	.850	.695	.938
400	.508	.416	.600	.850	.695	.938
500	.567	.473	.656	.925	.785	.980
600	.817	.733	.879	1.000	.891	1.000
700	.942	.879	.974	1.000	.891	1.000
800	.958	.901	.985	1.000	.891	1.000
900	.967	.912	.989	1.000	.891	1.000
1000	.900	.828	.945	1.000	.891	1.000

*Alphabet consists of 26 letters and space.

Table 12-7. Statistical Estimates for Probability of Successful Message Attack for Simple Substitution on English Using a Digram Frequency Analysis

Thus with 40 groups of 3 trials each, the number of successes can range from 0 to 40. In this way, the value for p(SM, 3) can be estimated for each value of N. The point estimates for p(SM) and p(SM, 3) are given in Table 12-7. The confidence intervals are computed using equations 12-17a and 12-17b, at a 95% level of confidence ($z = 1.96$).[11]

In addition to p(SM) and p(SM, 3), the mean and standard deviation for the number of correctly recovered plaintext characters, the number of correctly recovered key characters and the computation time to perform the analysis are also evaluated. Assuming a normal distribution for the underlying population, which may not be strictly justified, the confidence limits for each of these parameters are obtained via [9]

$$\bar{x} - (t_\alpha/2)(s/n^{1/2}) < u < \bar{x} + (t_\alpha/2)(s/n^{1/2}) \qquad (12\text{-}20)$$

where

u = parameter whose confidence limits are determined

$\bar{x}$ = sample mean $= (1/n) \sum_{i=1}^{n} xi$

[11] It was shown before that the distribution of successes, which is a binomial distribution, led to Equation 12-16 and that Equation 12-17 is an approximation of Equation 12-16.

$$s = \text{sample standard deviation} = [(1/(n-1)) \sum_{i=1}^{n} (\bar{x} - xi)^2]^{1/2}$$

n = sample size

$\{x1, x2, \ldots, xn\}$ = the observed values

t is the *student's t distribution* with n degrees of freedom, and $t_\alpha/2$ is related to the level of confidence γ. For $\gamma = 0.95$, which implies $\alpha = 1 - \gamma = 0.05$, one obtains a value of $t_\alpha/2 = 1.98$ when $n = 120$.

The results are shown in Table 12-8. More elaborate statistical tests could certainly be devised, but the emphasis here is to illustrate only the principles involved.

Empirical Results for Simple Substitution on English Using Single-Letter Frequency Analysis

A single-letter frequency analysis is quite elementary. It is discussed here so that the reader can contrast these results with those obtained for the digram-frequency analysis. The following procedure is used in conjunction with plaintext consisting of 26 letters (no space). First the letters are rearranged according to their relative frequency, from highest to lowest:

E T A O I N S R H L D C U M F P G W Y B V K X J Q Z

For each cryptogram under analysis, this vector is used as the basis for assigning plaintext equivalents to each character of the cryptogram (i.e., the most frequently occurring character in the cryptogram is assigned letter E, the next most frequently occurring character in the cryptogram is assigned letter T, etc.) The recovered plaintext is then compared to the original so that its correctness can be evaluated. The results of this experiment are given in Table 12-9.

Comparison of Results

Figures 12-5 and 12-6 illustrate the superiority of the digram-frequency analysis over the single-letter frequency analysis. They also confirm that the unicity distance is a function of the language statistics used to attack the cipher, and that ud becomes lower as more language statistics are effectively incorporated into the analysis. From Figure 12-6, it can be deduced that analysis with 3-grams, 4-grams, and so on, would give rise to a series of similar curves to the left of that obtained with digrams (the digram curve). In the limit, this series of curves would approach a curve that corresponds to a cryptanalysis performed by a human (90% recovery or more).

Each analysis (1-gram, 2-gram, etc.) has a certain cost associated with it. The single-letter frequency analysis took less than 1.5 CPU seconds on the IBM System 370, Model 168, not counting the input of the ciphertext itself, and required 500 bytes of storage. The digram-frequency analysis, on the

Plaintext* Length N	Sample Mean and Standard Deviation			Confidence Limits for the Mean at Level of Confidence $\gamma = 95\%$		
	Characters of Key Correctly Recovered	% of Plaintext Correctly Recovered	CPU Time (sec)	Characters of Key Correctly Recovered	% of Plaintext Correctly Recovered	CPU Time (sec)
250	10.3 4.8	56.7 23.6	27.6 6.1	9.5 11.2	52.5 61.0	26.5 28.7
275	11.1 4.5	60.2 22.1	28.7 7.4	10.3 12.0	56.2 69.2	27.4 30.0
300	13.8 5.4	69.3 24.4	27.5 8.1	12.9 14.8	64.9 73.7	26.1 29.0
350	20.1 3.8	86.3 14.3	28.8 6.7	19.4 20.8	83.7 88.9	27.6 30.0
400	19.5 4.3	84.6 16.4	30.2 6.7	18.8 20.3	81.7 87.6	29.0 31.4
500	23.2 3.8	90.8 12.8	28.0 6.3	22.5 23.9	88.5 93.1	26.9 29.2
600	24.8 3.3	95.1 11.4	31.0 6.3	24.2 25.4	93.0 97.2	29.9 32.2
700	25.4 2.5	97.6 9.9	31.0 7.2	24.9 25.8	95.8 99.4	29.7 32.3
800	25.9 2.5	97.1 8.2	31.3 6.3	25.4 26.3	95.6 98.6	30.2 32.5
900	25.9 1.8	97.2 6.0	29.6 6.0	25.6 26.2	96.1 98.3	28.5 30.7
1000	25.3 3.2	94.4 12.6	28.9 5.8	24.7 25.8	92.2 96.7	27.8 30.0

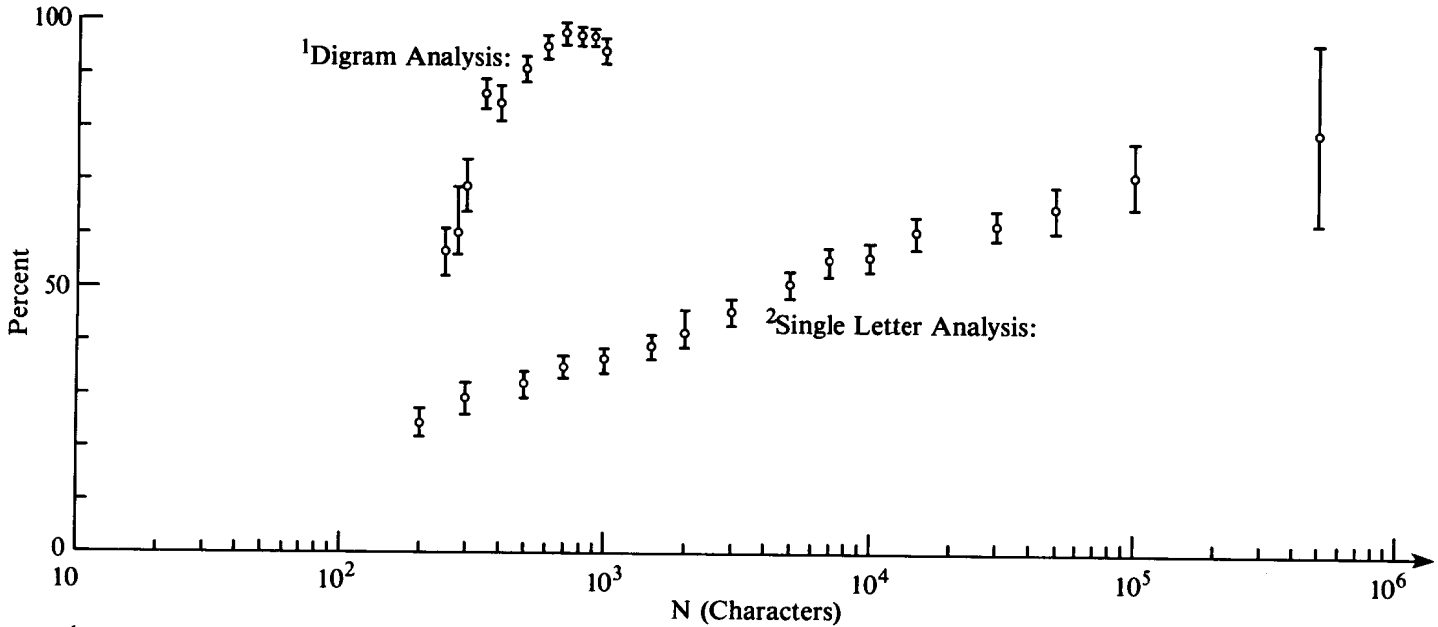
Sample Size = 120.
Analysis was performed on an IBM System/370, model 168.
*Alphabet consists of 26 letters and space.

Table 12-8. Statistical Estimates for Key Recovery, Message Recovery, and Processing Time Using a Digram Frequency Analysis

Plaintext Length* N	Sample Size	p(SM)	Confidence Limits $\gamma = 95\%$ (Eqs. 12-17a and 12-17b, $z = 1.96$)	Sample Mean and Standard Deviation		Confidence Limits for Mean at Level of Confidence $\gamma = 95\%$ (Eq. 12-20)	
				Characters of Key Correctly Recovered	% of Plaintext Correctly Recovered	Characters of Key Correctly Recovered	% of Plaintext Correctly Recovered
200	120			5.1 2.1	24.5 12.8	4.7 5.5	22.1 26.8
300	120			6.2 2.5	29.3 14.5	5.7 6.7	26.6 31.9
500	120			7.0 2.1	31.7 12.0	6.5 7.3	29.5 33.9
700	120			7.7 2.6	35.1 11.7	7.2 8.2	33.0 37.2
1000	120			7.9 2.9	36.4 12.4	7.3 8.4	34.1 38.6
1500	120			8.2 2.9	38.8 12.4	7.6 8.7	36.6 41.1
2000	120			9.1 2.9	41.6 12.7	8.5 9.6	39.3 43.9
3000	120			10.0 3.0	45.4 13.1	9.4 10.6	43.0 47.8
5000	120			11.4 3.2	50.5 13.7	10.7 12.0	48.0 53.0
7000	120			12.6 3.2	55.0 13.3	11.9 13.1	52.6 57.4
10000	120	.017	.003 .065	12.9 3.7	55.7 14.7	12.2 13.6	53.0 58.4
15000	120	.025	.006 .077	14.1 3.3	60.5 15.3	13.5 14.8	57.7 63.3
30000	120	.025	.006 .077	14.4 3.8	61.6 14.7	13.6 15.1	58.9 64.3
50000	80	.100	.047 .193	14.6 5.3	64.8 18.0	13.4 15.8	60.7 68.8
100000	40	.150	.062 .306	16.4 6.0	70.7 18.8	14.4 18.3	64.6 76.7
500000	8	.375	.102 .741	19.5 6.8	79.2 20.1	13.8 25.2	62.4 96.0

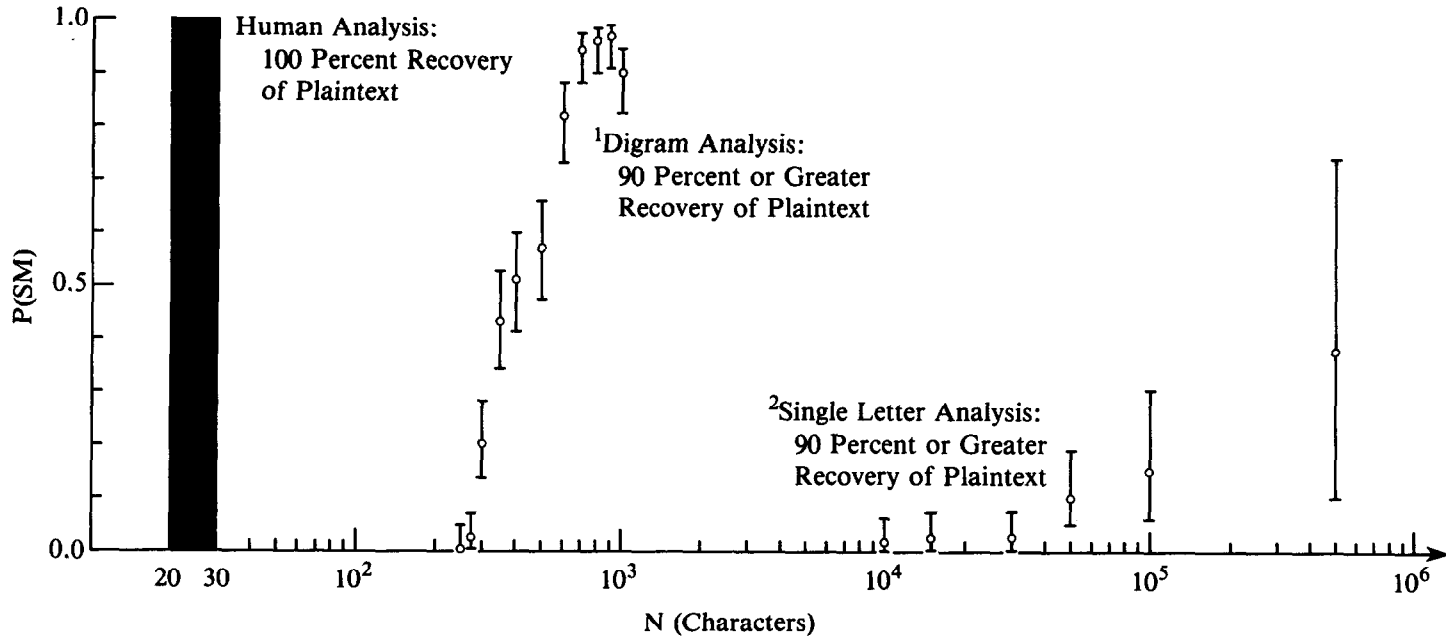
*26 Letters (No Space)

Table 12-9. Statistical Estimates for Probability of Successful Message Attack, Key Recovery, and Message Recovery for Simple Substitution on English Using a Single-Letter Frequency Analysis



[1]Digram Analysis Using 26 Letters and Space
[2]Single Letter Frequency Analysis Using 26 Letters (No Space)

Figure 12-5. Percent of Plaintext Recovered as a Function of Ciphertext Length Using a Single-Letter Frequency Analysis and a Digram Frequency Analysis for Simple Substitution on English



[1]Digram Analysis Using 26 Letters and Space
[2]Single Letter Frequency Analysis Using 26 Letters (No Space)

Figure 12-6. Comparison of p(SM) as a Function of Ciphertext Length Using a Single-Letter Frequency Analysis and a Digram Frequency Analysis for Simple Substitution on English

other hand, required about 30 CPU seconds on the same machine, and required 3000 bytes of storage. Both CPU time and storage can easily be converted to a monetary value. Hence it generally follows that the more powerful the attack (when higher order language statistics are used), the greater the associated cost. It follows also that the opponent may have some degree of freedom in selecting a method of analysis which will both be successful and keep his cost to a minimum. For example, with simple substitution on English, 1000 characters of ciphertext are not enough to allow a solution using only 1-grams. On the other hand, there may be no advantage in using trigrams in the analysis when digrams will do the job.

The analysis using digrams presented here shows that about 500 characters of ciphertext are needed for a successful attack. A more recent result obtained by Bahl [11] indicates that only about 300 characters are required.

REFERENCES

1. Diffie, W. and Hellman, M. E., "Privacy and Authentication: An Introduction to Cryptography," *Proceedings of the IEEE*, **67**, No. 3, 397–427 (1979).
2. Shannon, C. E., "Communication Theory of Secrecy Systems," *Bell System Technical Journal*, **28**, 656–715 (1949).
3. Francis, W., *A Standard Sample of Present-Day Edited American English for Use with Digital Computers*, Linguistics Department, Brown University, Providence, RI, 1964.
4. Shannon, C. E., "Predictions of entropy in printed English," *Bell System Technical Journal*, **30**, 50–64 (1951).
5. Parzen, E., *Modern Probability Theory and Its Applications*, Wiley, New York, 1960.
6. Hildebrand, F. B., *Advanced Calculus for Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1962.
7. Friedman, W. F., "Cryptology," *Encyclopedia Britannica*, p. 848 (1973).
8. Raisbeck, G., *Information Theory, An Introduction for Scientists and Engineers*, M.I.T. Press, Cambridge, 1964.
9. Gallagher, R., *Information Theory and Reliable Communication*, Wiley, New York, 1968.
10. Browlee, K. A., *Statistical Theory and Methodology in Science and Engineering*, Wiley, New York, 1961.
11. Bahl, L. R., *An Algorithm For Solving Simple Substitution Cryptograms*, International Symposium on Information Theory, Ithaca, NY, October 10–14, 1977.

Other Publications of Interest

12. Peleg, S. and Rosenfeld, A., "Breaking Substitution Ciphers Using a Relaxation Algorithm," *Communications of the ACM*, **22**, No. 11, 598–605 (1979).
13. Hellman, M. E., "An Extension of the Shannon Theory Approach to Cryptography," *IEEE Transactions on Information Theory*, **IT-23**, No. 3, 289–294 (1977).
14. Blom, R. J., "Bounds on Key Equivocation for Simple Substitution Ciphers," *IEEE Transactions on Information Theory*, **IT-25**, No. 1, 8–18 (1979).
15. Lu, S. C., "The Existence of Good Cryptosystems for Key Rates Greater than the Message Redundancy," *IEEE Transactions on Information Theory*, **IT-25**, No. 4, 475–480 (1979).

16. Lu, S. C., "Random Ciphering Bounds on a Class of Secrecy Systems and Discrete Message Sources," *IEEE Transactions on Information Theory*, **IT-25**, No. 4, 405–414 (1979).

17. Kullback, S., *Statistical Methods in Cryptanalysis*, Aegean Park Press, Laguna Hills, CA, 1976.

18. Dunham, J. G., "On Message Equivocation for Simple Substitution Ciphers," *IEEE Transactions on Information Theory*, **IT-26**, No. 5, 522–527 (1980).