An Extension of the Shannon Theory Approach to Cryptography

MARTIN E. HELLMAN, MEMBER, IEEE

Abstract—**Shannon's information-theoretic approach to cryptography is reviewed and extended. It is shown that Shannon's random cipher model is conservative in that a randomly chosen cipher is essentially the worst possible. This is in contrast with error-correcting codes where a randomly chosen code is essentially the best possible. The concepts of matching a cipher to a language and of the trade-off between local and global uncertainty are also developed.**

I. Introduction

Interest in commercial applications of cryptography has recently experienced a dramatic upsurge, largely motivated by the growing use of time-shared remote access computers and the large monetary losses possible through their misuse. Privacy legislation has also had an important effect.

This and companion papers [1], [2] emphasize increasing the theoretical basis on which commercial and other nonsecret cryptographic systems can be built. This paper is concerned with the "classical" information-theoretic, or Shannon theory, approach to cryptography in which unlimited cryptanalytic computational abilities are assumed. This approach was pioneered by Shannon [3]. The companion papers are concerned with another theoretical approach, more closely related to the theory of computational complexity, in which the cryptanalyst has a large but finite computational ability. As argued in [2], the current emphasis of computational complexity on worst case computation time is inadequte for use in cryptography and ensemble arguments are needed. We therefore believe that both aspects of the theory can benefit from attention by information theorists. It is our opinion that the finite computational theory will bear more directly on the practice of cryptography and that the classical theory will be useful mostly in providing insights into design principles.

Fig. 1 depicts the flow of information in a cryptographic, or cipher, system. A message (plaintext) M is to be communicated over an insecure communication channel such as radio or a tapped telephone line. To prevent unauthorized personnel from learning the contents of the message, it is enciphered prior to transmission to produce a cryptogram (ciphertext)

$$C = T_K(M). \tag{1}$$

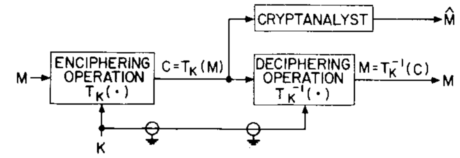
CRYPTANALYST
$\hat{M}$
M
ENCIPHERING OPERATION $T_K(\cdot)$
$C = T_K(M)$
DECIPHERING OPERATION $T_K^{-1}(\cdot)$
$M = T_K^{-1}(C)$
M
K

Fig. 1. Flow of information in cryptographic system.

The transformation T_K is invertible and depends on a key K, known to the legitimate transmitter and receiver, but unknown to the cryptanalyst who seeks to gain unauthorized access to the message through his knowledge of the cryptogram, the nature of the message (e.g., the language in which it was written), and the general cipher system being used (i.e., the set of transformations $\{T_k(\cdot)\}$). The only information available to the legitimate receiver, but not to the cryptanalyst, is the identity of the key being used. A secure channel, indicated by the shielded cable, is used for transmission of the key. Typically, this secure channel is a courier and is too slow for transmission of the message.

We may represent the enciphering, deciphering, and cryptanalytic operations as

$$C = f(M,K) \triangleq T_K(M) \tag{2}$$

$$M = g(C,K) \triangleq T_K^{-1}(C) \tag{3}$$

$$\hat{M} = h(C), \tag{4}$$

and our goal is to preserve the security of the message (e.g., by keeping Pr $(\hat{M} = M)$ close to 0 for even the best cryptanalytic function h).

As discussed in [2], the cryptanalyst often has side information available to him, and this can greatly ease his task. This side information usually takes the form of partial knowledge of the message. Such knowledge can be viewed as increasing the redundancy of the message and can easily be taken into account in what follows. The chosen plaintext attack, described in [2], is associated with a cryptanalytic attack of finite computational abilities and is therefore not of interest in this paper.

If each message or cryptogram consists of N characters from a finite alphabet of L symbols, we can represent an arbitrary cipher in the form of Fig. 2. Letting $R_o = \log_2 L$ denote the absolute rate of the language, there are 2^{R_oN} possible messages and an equal number of cryptograms. Unless otherwise stated, all ciphers are tacitly noncompressive and nonexpansive and are therefore of this form.

For purposes of this paper, we define a model in which the messages are divided into two subsets. The first contains 2^{RN} meaningful messages, each with the same *a priori* probability 2^{-RN}. R is, of course, the rate of lan-

Manuscript received October 31, 1975; revised August 3, 1976. This work was supported by the National Science Foundation under Grants GK-33250 and ENG-10173. Paper was previously presented at the IEEE International Symposium on Information Theory, Notre Dame, IN, October 27–31, 1974.

The author is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305.

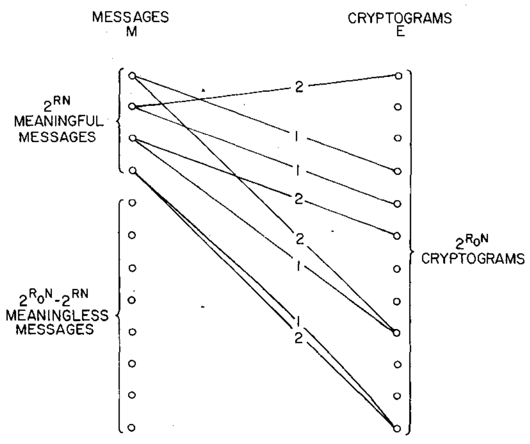

Fig. 2. Representation of cipher.

guage. The remaining messages are meaningless in the language and are assigned *a priori* probabilities of 0.

This model is adopted in preference to the more usual ergodic source model for several reasons. A natural language matches neither model exactly, and adoption of our model greatly simplifies the proofs of several theorems. Also, from a cryptographic point of view, there appears to be little difference between the results of one model or the other. In particular, perfect noiseless source coding is impossible to achieve in practice under either model's approximation to a natural language. This is attributable in large part to the fact that there is no way of devising a simple, implementable scheme that maps the meaningful messages into indices 1 through 2^{RN} and the meaningless ones into the remaining indices. The importance of this will become clear in the next section. The word "meaningful" almost can be replaced by "typical" in an ergodic source model.

We also 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 Fig. 2, a line with a number i in it indicates that the message at the left is encrypted as the cryptogram at the right when the ith key is used. The figure shows only twelve messages and two keys for the sake of clarity. Since meaningless messages are never enciphered (at least in our simple model), their encryptions are not shown.

If the first meaningful message in Fig. 2 is enciphered with key #1, then the cryptogram is breakable. This is because the resultant cryptogram (the fourth) has no other meaningful messages into which it can be deciphered. If, however, message #1 is enciphered with key #2, the resultant cryptogram (the ninth) is unbreakable since another meaningful message (#3) can be enciphered with another key (#1) to yield this same cryptogram. Such a false solution will be called a *spurious message decipherment.* The number of spurious message decipherments n_m is a random variable and is determined by the cipher and C. If n_m takes on large values with probability close to one, then the system will be secure even if the cryptanalyst is allowed unlimited computation. The best he can do is to make a list of all meaningful solutions and to choose $\hat{M}$ randomly from this list.

A different situation exists when the fourth message is encrypted with key #1. The resultant cryptogram (#12) is deciphered as message #4 with both key #1 and key #2. Thus the cryptanalyst knows which message was sent, and $n_m = 0$, but he does not know whether key #1 or key #2 was used. This will be termed a *spurious key decipherment,* and here the number of spurious key decipherments n_k is one. Clearly, $n_k \geq n_m$ in general. A cipher which makes good use of its key will have $n_k \doteq n_m$.

While n_m is of greater interest, it is much simpler to evaluate n_k. In particular, if $l(C)$ denotes the number of "lines" ending on cryptogram C, then $n_k(C)$, the number of spurious key decipherments when C is intercepted, is

$$n_k(C) = \max\{[l(C) - 1], 0\}. \tag{5}$$

Here, "line" means a line connecting a meaningful message to a cryptogram as in Fig. 2. All of our theorems therefore deal with n_k, and toward the end of the paper we relate these results to n_m.

Even n_m does not totally summarize the security level of a cipher, since it is a measure of uncertainty on a block basis rather than on a per digit basis. Finding tractable approaches to per digit uncertainty is an important problem, and Lu [9] has made a start at treating it. In cryptography, as elsewhere, there is a trade-off between assuming a model which is tractable and one which closely resembles reality. The n_k approach used in this paper is extremely tractable and, as later discussed, appears to model reality closely in all but pathological systems.

II. Random Ciphers

As originally defined by Shannon [3], a random cipher is one in which $g(C,K)$, the decipherment of C under key K, is a uniformly distributed random variable on the set of 2^{R_oN} messages, both meaningful and not. Further, $g(C_o,K_o)$ is independent of the rest of the $\{g(C,K)\}$. The motivation behind this definition is easily understood, since in most cryptographic systems use of an incorrect key in decipherment produces a garbled random-looking message. Indeed, a cryptomachine can be used as a good source of pseudorandom numbers. It appears that this intuitive reasoning, developed by Shannon during World War II, led to his less intuitively understandable, but more famous, random coding arguments for error-correcting codes.

The following definition is slightly different and ensures that the cipher chosen is both uniquely encipherable and uniquely decipherable, as are most real ciphers.

Definition: A random cipher is one in which, for each (C_o,K_o) pair, $g(C_o,K_o)$ has a uniform marginal distribution on all 2^{R_oN} messages. However, the $\{g(C,K)\}$ are dependent in that, for any set S of cryptograms not including C_o, the distribution of $g(C_o,K_o)$ given $\{g(C,K_o)\}_{C\in S}$ is uniform over all messages not in $\{g(C,K_o)\}_{C\in S}$. There is no dependence between $g(C,K_o)$ and $g(C',K_1)$, for $K_o \neq K_1$.

As usual, we are really defining an ensemble of ciphers to allow the calculation of ensemble averages. Use of the term "random cipher" is a slight abuse that should produce no more confusion that that associated with referring to a random code ensemble as a "random code."

The following theorem and corollary constitute a simplified version of one of Shannon's main results [3].

Theorem 1: Over the ensemble of random ciphers, the expected number of spurious key decipherments $\overline{n}_k$ is

$$\overline{n}_k = (2^{H(K)} - 1)\, 2^{-ND} \doteq 2^{H(K)-ND}, \tag{6}$$

where

$$D = R_o - R \tag{7}$$

is the redundancy of the language in bits/character. The approximation (6) is very accurate, provided $H(K)$ is of reasonable size. For example, if $H(K) \geq 10$ bits, it is accurate to at least three significant decimal digits.

Proof: Because of the symmetry induced by the ensemble, $\overline{n}_k$ is independent of C_o, the particular cryptogram intercepted. There is one correct meaningful solution, and there are $2^{H(K)} - 1$ remaining keys, each of which has the same probability p of yielding a spurious key decipherment. Therefore,

$$\overline{n}_k = (2^{H(K)} - 1)p. \tag{8}$$

From the definition of a random cipher, we see that the dependencies do not come into play since we are dealing with $2^{H(K)}$ different keys. Therefore,

$$p = 2^{RN}/2^{R_oN} = 2^{-ND}, \tag{9}$$

and

$$\overline{n}_k = (2^{H(K)} - 1)2^{-ND}, \tag{10}$$

completing the proof.

Corollary 1, stated below, follows directly from this reasoning.

Corollary 1: Over the ensemble of random ciphers, the random variable n_k has a binomial distribution with parameters $n = (2^{H(K)} - 1)$ and $p = 2^{-ND}$. Therefore, when the product np is large (equivalently, $\overline{n}_k$ is large), the probability that the system is secure (i.e., Pr (n_k is large)) is close to one. Conversely, when the product np is small, the system is insecure with probability close to one. Because of the rapid variation of the exponential with increasing N, we may take

$$N_o = H(K)/D \tag{11}$$

as a good measure of the minimum amount of text necessary for there to be a unique solution to a random cipher. That is, for $N < N_o$, the system is considered secure, while for $N > N_o$, it is considered insecure.

Shannon calls the minimum amount of ciphertext required to yield a unique solution to a cryptogram the *unicity distance* of the cipher. We see that N_o in (11) is a good measure of the unicity distance for a random cipher.

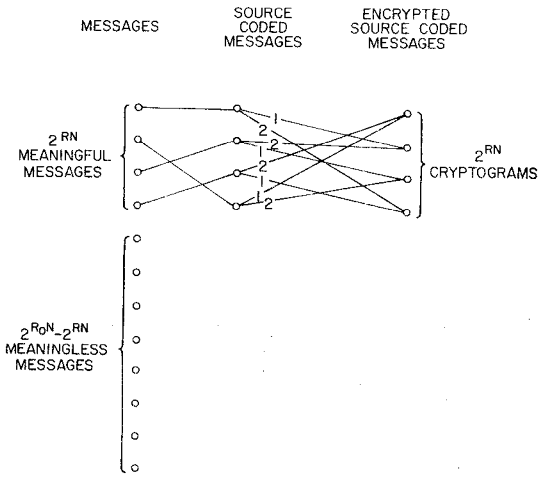

Fig. 3. Perfect source coding followed by encryption produces unbreakable cipher.

Taking a simple substitution cipher on English as an example,

$$H(K) = \log_2 (26!) = 88.4 \text{ bits} \tag{12}$$

$$R_o = \log_2 (26) = 4.7 \text{ bits/character} \tag{13}$$

$$R \doteq 1.5 \text{ bits/character} \tag{14}$$

$$D = 3.2 \text{ bits/character,} \tag{15}$$

and therefore the random cipher model would predict a unicity distance

$$N_o \doteq 28 \text{ characters.} \tag{16}$$

When $N = 40$, $\overline{n}_k = 1 \times 10^{-12}$, and when $N = 20$, $\overline{n}_k = 2 \times 10^7$, indicating the rapid variation of $\overline{n}_k$ in the vicinity of N_o.

According to Friedman [5], "practically every example of 25 or more characters representing the monoalphabetic encipherment of a 'sensible' message in English can be readily solved." Since the random cipher model is only a model, this close agreement is rather remarkable. The material that follows helps to explain this agreement. If a cipher is used which produces a good mixing of messages as the key is varied, the random cipher model should produce a reasonable estimate of unicity distance. We will also see that by avoiding this mixing, it is possible to design a cipher whose performance is much superior to that of the random cipher model, and that random ciphers have essentially the poorest performance possible.

In particular, (11) indicates that if perfect noiseless source coding were to precede encipherment, N_o would be infinite since the output of a perfect source coder is a "language" with no redundancy. Hence, $D = 0$ for the language acted on by the cipher. Fig. 3 is a cipher diagram illustrating this situation. It is seen that data compression has the perhaps unexpected benefit of increasing the security of a cryptosystem. Of course, it is impossible to implement a perfect source code because of cost, but even partial compression is useful in increasing N_o. This observation is due to Shannon [3].

Note that the source code is not secret and is assumed known to the cryptanalyst. The source coding operation would provide no security by itself, and only adds to the security of the system in an indirect, although important, way.

If it is necessary that the system be capable of transmitting even meaningless messages, a cipher of the form shown in Fig. 4 can be used to achieve an unbounded unicity distance. The important point is that the meaningful and meaningless messages are not intermixed as the key is varied. There must be a set of 2^{RN} "meaningful" cryptograms into which all (meaningful message, key) pairs are enciphered. Although Fig. 4 indicates an intermediate source coding stage, in the next section we will see that certain ciphers can be "matched" to a language to produce the same effect without any explicit source coding.

The fact that random ciphers perform so much more poorly than carefully designed ciphers (e.g., those of Figs. 3 and 4) is at first surprising since random error-correcting codes are virtually optimal [6]. In fact, random ciphers are essentially the worst possible, in the sense of the following two theorems.

Theorem 2: If each meaningful message has *a priori* probability 2^{-RN} and each key has *a priori* probability $2^{-H(K)}$ independent of the message, then

$$\bar{n}_k \geq 2^{H(K)-ND} - 1, \tag{17}$$

for any uniquely encipherable, uniquely decipherable cipher. The expectation is now only over C since the cipher is fixed.

Remark: This theorem says that whenever the random cipher model predicts security through a large expected number of spurious decipherments, any cipher whatsoever must also possess a large expected number and therefore be secure. However, when the random cipher model predicts few spurious decipherments, the theorem does not preclude other ciphers from still having many.

Proof: Using the assumed *a priori* probabilities and (5), we obtain

$$\begin{aligned} \bar{n}_k &= \sum_C \Pr(C) n_k(C) \\ &\geq \sum_C [l(C)/2^{H(K)+RN}][l(C) - 1] \qquad (18) \\ &= \left[2^{-H(K)-RN} \sum_C l^2(C) \right] - 1. \qquad (19) \end{aligned}$$

Using the convexity of the function x^2, we see that the last term in brackets is minimized when each of the 2^{R_oN} values of $l(C)$ equals its average value $2^{RN+H(K)}/2^{R_oN}$, although there might not exist any cipher that achieves this minimum. Therefore,

$$\begin{aligned} \bar{n}_k &\geq \{2^{-H(K)-RN} 2^{R_oN} [2^{RN+H(K)-R_oN}]^2\} - 1 \qquad (20) \\ &= 2^{H(K)-ND} - 1, \qquad (21) \end{aligned}$$

completing the proof.

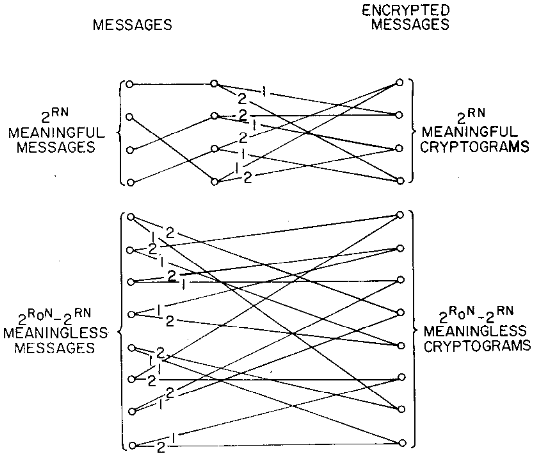
MESSAGES
ENCRYPTED MESSAGES
2^{RN} MEANINGFUL MESSAGES
2^{RN} MEANINGFUL CRYPTOGRAMS
$2^{R_oN}-2^{RN}$ MEANINGLESS MESSAGES
$2^{R_oN}-2^{RN}$ MEANINGLESS CRYPTOGRAMS
1
2

Fig. 4. Unbreakable cipher without source coding.

The above theorem still leaves something to be desired in that it does not guarantee that the actual number of spurious key decipherments n_k will be large with high probability when its expected value $\bar{n}_k$ is large. For example, if $\bar{n}_k = 10^{10}$, it may be possible for n_k to take on the value zero with probability $1 - 10^{-10}$ and to take on the value 10^{20} with probability 10^{-10}. Such a cipher would be highly insecure in spite of the large value of $\bar{n}_k$. The following theorem removes this weakness.

Theorem 3: If each (meaningful message, key) pair has equal *a priori* probability, then any uniquely encipherable and decipherable cipher whatsoever must satisfy

$$\Pr(n_k \leq m - 1) \leq m/2^{H(K)-ND}. \tag{22}$$

Remark: This theorem says that if $2^{H(K)-ND}$ (which is $\bar{n}_k$ for a random cipher) is large, then the probability that the actual number of spurious key decipherments n_k is also large must be close to one for all ciphers. For example, if $\bar{n}_k = 10^{10}$ for a random cipher, then $\Pr(n_k < 10^5) \leq 10^{-5}$ for any cipher; if $\bar{n}_k = 10^{20}$ for a random cipher, then $\Pr(n_k < 10^{10}) \leq 10^{-10}$ for any cipher; etc.

Proof:

$$\begin{aligned} \Pr(n_k \leq m - 1) &= \sum_{\{C:l(C)\leq m\}} \Pr(C) \\ &= \sum_{\{C:l(C)\leq m\}} l(C)/2^{RN+H(K)} \\ &\leq 2^{-RN-H(K)} \sum_{\{C:l(C)\leq m\}} m \\ &\leq m 2^{-RN-H(K)} 2^{R_oN} \\ &= m/2^{H(K)-ND}. \qquad (23) \end{aligned}$$

Q.E.D.

These theorems show that random ciphers are essentially the worst possible. The value of N_o given by (11) is thus conservative.

As has been noted, the number of spurious message decipherments satisfies $n_m \leq n_k$ so that dealing with n_k can make a cipher look more secure than it really is, which somewhat weakens the statement that random ciphers are essentially the worst possible. It is possible to design a cipher which is less secure than a random cipher by making many keys equivalent. For example, if $T_K(M) = M$ for all keys K, then, even if $H(K)$ is huge, there are no spuriously deciphered messages, although there are many spuriously diciphered keys. It would appear to be a simple matter to ensure that the key is well-used, thereby revalidating the conservative nature of the random cipher model.

For random ciphers, the difference between n_k and n_m is unimportant as far as N_o is concerned. When N is near the unicity distance, very few of the meaningful messages are possible decipherments, and the chance that two of these will be the same is infinitesimal since they are chosen uniformly from a set of size 2^{RN}.

III. Matching Ciphers to Languages and Locally Informative Ciphers

We have seen that there are ciphers, such as those depicted in Figs. 3 and 4, which are much superior to random ciphers. Basically, what is required is a clustering of meaningful messages into a set of meaningful cryptograms. It might appear that such clustering would require a two-stage encipherment; the first stage being a source coding operation. However, we shall show that the same effect can sometimes be obtained by matching the structure of the cipher to that of the language. A trade-off between local and global security will also be developed.

As a first step, consider an artificial language in which the 26 letters occur with equal frequency and successive letter choices are independent except that once a letter is chosen, it is repeated three times. Thus a typical sentence in this language is *"FFFXXXLLLAAAXXXRRR."* Now consider applying 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 enciphered as S, X as B, etc. A cryptanalyst who is given such a cryptogram has no way of choosing among the $2^{H(K)}$ possible keys since a simple substitution operation is measure preserving. Each key yields a typical sentence as its decipherment, and even the cryptogram itself looks like a meaningful message! A simple substitution cipher is thus secure when applied to this simple artificial language. The effect is the same as if we had first done perfect source coding and then encrypted. Contrast this with the use of a transposition cipher with the same number of keys $2^{H(K)}$ chosen at random from among the N! possible transpositions. A typical cryptogram looks like *"XLFFXARARXLXXRAXFL."* The redundancy of the language is now quite useful in cryptanalysis, and (11) is a good measure of the unicity distance.

Next, consider first-order English as a second artificial language. That is, successive letters are independent random variables, each having the same probability distribution as in English (i.e., $P(E) = 0.123$, $P(T) = 0.096, \cdots, P(Z) = 0.001$). A typical sentence in this language looks like *"RESNFEALTMISEQ."* Using a simple substitution cipher yields a cryptogram such as *"FXOBRXYMGUNOXC."* This type of cipher is no longer secure, and frequency analysis is an obvious cryptanalytic technique. For example, X is seen to occur three times in 14 letters. It is likely to represent E or another frequently occurring letter. It is almost definitely not Z or a similar, infrequently occurring letter. However, now a transposition cipher is secure since none of the language's redundancy has to do with the relative positions of letters. Equivalently, transposition is a measure preserving transformation, and a transposed message looks as meaningful as the original message.

Thus in some sense, a simple substitution cipher is perfectly matched to the first artificial language, and a transposition cipher is perfectly matched to the second. However, note that we have not done any source coding. Even meaningless messages can be unambiguously enciphered and deciphered. Furthermore, we did not have to know the exact structure of the language. For example, in the second case, if the letters had occurred with the same relative frequencies as in French, the cipher would still be secure. There is an obvious connection to recent work on universal codes [7].

On a natural language such as English, neither cipher is perfectly secure. Yet, the second artificial language more closely resembles English than does the first. And, as noted by Shannon [3], a transposition cipher on English has a larger unicity distance than that given by (11). This is because a typical English message will have about 12.3 percent E's, 9.6 percent T's, etc. When a transposition is applied, the resultant cryptogram will also have about 12.3 percent E's, 9.6 percent T's, etc. Therefore, a typical message of length N cannot be encrypted into an arbitrary cryptogram out of the $2^{R_0 N}$ possible, but rather, it must be encrypted into one of the $2^{R_1 N}$ cryptograms which have 12.3 percent E's, 9.6 percent T's, etc. R_1 is the rate of a first order approximation to English [8] and is approximately 4.17 bits/character. Going through arguments similar to those for random ciphers, we thus find that for a transposition cipher,

$$\bar{n}_k \geqq 2^{H(K)-N(R_1-R)} - 1. \tag{24}$$

Similarly, all theorems of the preceding section still apply if D is replaced by $R_1 - R$. The effective redundancy is thus reduced from 4.7–1.5 = 3.2 bits/character to 4.17–1.5 = 2.67 bits/character. The unicity distance is increased by a factor of 3.2/2.67 = 1.2. And, from a computational point of view, first-order frequency characteristics are of no use in cryptanalysis, thereby removing a frequently used entry into a cipher.

By extension, we could use a transposition on pairs of letters, or on words, and remove even more of the effective redundancy. Using frequency characteristics of common English words, we estimate that the rate of a first-order word approximation to English (a language in which successive words are independent random variables, and in which words occur with the same relative frequencies as in English) is $R_{w1} = 2.22$ bits/character. The effective value of D becomes $2.22–1.5 = 0.72$ bits/character, and the bound on unicity distance increases by a factor of $3.2/0.72 = 4.4$ over that guaranteed by a random cipher. Also, the redundancy that is left is more difficult to describe, making cryptanalysis computationally more difficult.

We could go further and transpose sentences, paragraphs, whole reports, etc., but an obvious problem appears as we consider such extensions. While it is true that a cryptanalyst has more difficulty breaking these larger transpositions, it is also true that even without cryptanalysis, one can glean important local information. For example, in a word transposition cipher, knowledge that certain words have occurred can be damaging in and of itself. Even though global security is preserved, the local information which seeps through can cause damage.

The above arguments indicate that a word transposition, followed by a simple cipher to hide local information, might be good in practical use. A similar idea, known as bisection, has found use in classical cryptography to remove stylized beginnings (e.g., "Dear Sir:") and endings of messages as entry points for cryptanalysis.

Another example concerns numerical data. This is often viewed as possessing little redundancy, and (11) would therefore predict a large unicity distance. If, however, the data is stored in *ASCII* or *EBCDIC,* a great deal of redundancy is available to a cryptanalyst. By designing a cipher which always replaces numerals by other numerals, this redundancy is made useless for cryptanalysis. Of course, then the cryptanalyst gains local information as to where text ends and data begins, but, in most cases, this seems worth trading in return for global security. Lu [9] has recently studied a problem which can be partially viewed as a trade-off between local and global uncertainty.

IV. Discussion

As mentioned in the introduction, it is our belief that the classical, or Shannon theory, approach taken in this paper is not, in general, directly applicable to designing practical cryptographic systems. Rather, it appears to be useful mostly for gaining qualitative insights into the design of practical systems. For example, we have seen the value of data compression in increasing security and that ciphers can be matched to the structure of a data source.

Another example of qualitative insight which can be gained from the Shannon theory approach concerns the wiretap channel introduced by Wyner [10]. Through insightful techniques, Wyner developed necessary and sufficient conditions on an achievable region of information rates. Carleial and Hellman [11] later derived a modified version of Wyner's result which appears useful in quantifying the concept of a discrete mixing transformation. This concept was introduced by Shannon [3] in describing a technique for the design of computationally secure ciphers. There are undoubtedly other connections between the classical information-theoretic approach to cryptography and the more directly useful computational theory approach. The need to study ensemble behavior in the computational theory is one area which deserves attention, and we hope that others will come to light as work progresses.

One last comment concerns a type of duality between ciphers and error-correcting codes. Since random error-correcting codes are essentially the best possible and random ciphers are essentially the worst possible, and since redundancy is needed for error-correction but weakens a cipher, we are tempted to label these problems as duals or opposites. The reason for this duality stems from the very opposite goals of error-correcting codes and ciphers. An error-correcting code is designed to be uncertainty (i.e., noise) immune, whereas a cipher is designed to uncertainty (i.e., key) sensitive. The cryptanalyst is trying to remove the "noise", and we want his job to be as difficult as possible. This duality should not be taken too far though, especially when the cryptanalyst has a finite computational ability. Concatenated codes [12] possess nice error-correcting properties, and concatenated or product ciphers also appear to be useful for computationally secure ciphers [13]–[15].

References

[1] W. Diffie and M. Hellman, "Multiuser cryptography," presented at 1975 National Computer Conference, New York, June 1976.

[2] ——, "New directions in cryptography," *IEEE Trans. Inform. Theory,* vol. IT-22, pp 644–654, Nov. 1976.

[3] C. Shannon, "Communication theory of secrecy systems," *Bell Sys. Tech. J.*, vol. 28, pp. 656–715, Oct. 1949.

[4] ——, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, Part I, pp. 479–523, Part II, pp. 623–656, 1948.

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

[6] R. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[7] L. Davisson, "Universal noiseless coding," *IEEE Trans. Inform. Theory,* vol. IT-19, pp. 783–795, Nov. 1973.

[8] C. Shannon, "Prediction and entropy of printed English," *Bell Sys. Tech. J.*, vol. 30, pp. 50–64, Jan. 1951.

[9] S. C. Lu, "Sphere packing bounds for additive block ciphers and memoryless binary sources," University of Hawaii Report 75-EE-CAN-2, July 1975.

[10] A. Wyner, "The wiretap channel," *Bell Sys. Tech. J.*, vol. 54, pp. 1355–1387, Oct. 1975.

[11] A. Carleial and M. Hellman, "A note on Wyner's wiretap channel," *IEEE Trans. Inform. Theory,* this issue, pp. 387–390.

[12] G. D. Forney, Jr., *Concatenated Codes.* Cambridge, MA: MIT Press, 1966.

[13] H. Feistel, "Cryptography and computer privacy," *Sci. Amer.*, vol. 228, pp. 15–23, May 1973.

[14] E. Grossman, "Group theoretic remarks on cryptographic systems based on two types of addition," IBM Report RC-4742, Yorktown Heights, NY, Feb. 1974.

[15] D. Coppersmith and E. Grossman, "Generators for certain alternating groups with applications to cryptography," *SIAM J. Appl. Math,* vol. 29, pp. 624–627, Dec. 1975.