# AN INFORMATION-THEORETIC TREATMENT OF HOMOPHONIC SUBSTITUTION

Hakon N. Jendal, Yves J.B. Kuhn & James L.Massey

Institute for Signal and Information Processing
Swiss Federal Institute of Technology
CH-8092 Zürich, Switzerland

## 1. INTRODUCTION

The history of cryptology shows that most secret-key cipher systems that have been broken were broken by exploiting the departure of the plaintext statistics from those of a completely random sequence. The technique of "homophonic substitution" is an old technique for converting an actual plaintext sequence into a (more) random sequence. At EUROCRYPT '88, Günther [1] introduced an important generalization of homophonic substitution, which we will call "variable-length homophonic substitution". The purpose of this paper is to give an information-theoretic treatment of Günther's type of homophonic substitution.

In Section 2, we give a rather careful discussion of Shannon's concept of a "strongly-ideal" cipher system, as this provides the motivation for any type of homophonic substitution. Section 3 gives the precise definition of variable-length homophonic substitution together with the necessary and sufficient condition for such substitution to be perfect, i.e., to create a completely-random sequence. Section 4 shows that perfect homophonic substitution can be achieved by the introduction of less than 2 bits of entropy into each source letter that is coded, and Section 5 shows that such perfect homophonic substitution can be realized using less than 4 random bits per letter coded. Section 6 indicates certain obvious generalizations of the previous results and mentions their implications for source coding (or "data compression").

The information-theoretic results used in this paper are quite basic and may be found in any good textbook on information theory, e.g., the book by Gallager [2].

J.J. Quisquater and J. Vandewalle (Eds.): Advances in Cryptology - EUROCRYPT '89, LNCS 434, pp. 382-394, 1990.

## 2. STRONGLY-IDEAL AND UNBREAKABLE CIPHER SYSTEMS

The purpose of "homophonic substitution" can be explained by considering a secret-key cipher system as diagrammed in Fig. 1. For ease of notation, let $X^n$ and $Y^n$ denote the plaintext and ciphertext sequences $[X_1, X_2,..., X_n]$ and $[Y_1, Y_2, ..., Y_n]$, respectively. As customary and as Fig. 1 suggests, we assume always that the secret key Z is statistically independent of the plaintext sequence $X^n$ for all n. We shall call the cipher non-expanding if the plaintext digits and ciphertext digits take values in the same D-ary alphabet and there is an increasing infinite sequence of positive integers $n_1, n_2, n_3, ...$ such that, when Z is known, $X^n$ and $Y^n$ uniquely determine one another for all $n \in S = \{n_1, n_2, n_3, ...\}$. We shall also call a sequence of D-ary random variables completely random if each of its digits is statistically independent of the preceding digits and is equally likely to take on any of the D possible values. The following proposition is proved in the Appendix by elementary information-theoretic arguments.

Proposition 1: If the plaintext sequence encrypted by a non-expanding secret-key cipher is completely random, then the ciphertext sequence is also completely random and is also statistically independent of the secret-key.

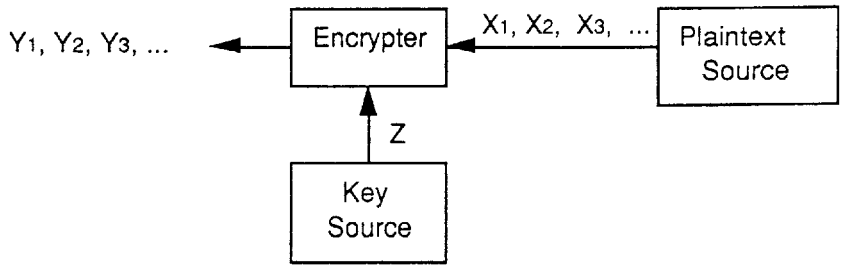


Fig. 1: A secret-key cipher system

Shannon [3] has defined the key-equivocation function f(n) of a secret-key cipher system to be the conditional entropy of the key given the first n digits of ciphertext, i.e., $f(n) = H(Z \mid Y^n)$. The key-equivocation function f(n) is thus a measure of the number of values of the secret key Z that are consistent with the first n digits of ciphertext. Because f(n) can only decrease as n increases, Shannon called a cipher system ideal if f(n) approaches a non-zero value as n tends toward infinity, and strongly ideal if f(n) is constant, i.e., if $H(Z \mid Y^n) = H(Z)$ for all n, which is equivalent to the statement that the ciphertext sequence is statistically independent of the secret key.

Corollary 1 to Proposition 1: If the plaintext sequence encrypted by a non-expanding secret-key cipher is completely random, then the cipher system is strongly ideal (regardless of the probability distribution for the secret key).

Virtually all useful non-expanding ciphers have the property, which we call "non-degeneracy", that changing the value of the secret key Z, without changing the value of the plaintext sequence $X^n$, will change the value of the ciphertext sequence for all n sufficiently large, except for a negligibly small fraction (often 0) of possible key values for any given value of $X^n$. Equivalently, a non-expanding cipher is non-degenerate if

$$H(Y^n \mid X^n) \approx H(Z)$$

holds for all sufficiently large n and all probability distributions for $X^n$ when all possible values of the secret key Z are equally likely. But, as shown in the Appendix,

$$H(Y^n \mid X^n) = H(X^n \mid Y^n)$$

holds for all n in a non-expanding cipher when the plaintext sequence $X_1$, $X_2$, ... is completely random. The following conclusion is immediate.

Corollary 2 to Proposition 1: If the plaintext sequence encrypted by a non-expanding secret-key cipher is completely random and all possible key values are equally likely, then the conditional entropy of the plaintext sequence given the ciphertext sequence satisfies

$$H(X^n \mid Y^n) \approx H(Z)$$

for all n sufficiently large.

This corollary implies in particular that, in a ciphertext-only attack, the cryptanalyst can do no better to find $X^n$ than by guessing at random from among as many possibilities as there are possible values of the secret key Z. In other words, the cipher system is unbreakable in a ciphertext-only attack when the number of possible key values is large.

The foregoing has shown that virtually any non-expanding secret key cipher can be used as the cipher in an unbreakable cipher system, provided that the plaintext source emits a completely random sequence. But it is precisely the goal of "homoponic substitution" to convert a source not of this type into such a source. When the homophonic coding is "perfect", it is then a trivial task to build unbreakable secret-key cipher systems in the form shown in Fig. 2.

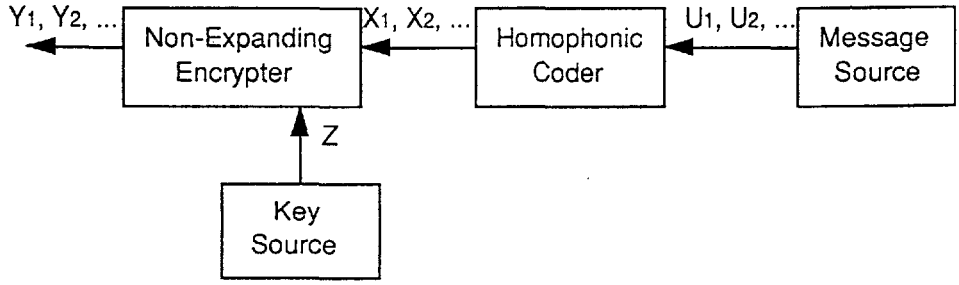


*Fig. 2: Use of homophonic substitution within a secret-key cipher system*

## 3. VARIABLE-LENGTH HOMPHONIC SUBSTITUTION

Here and hereafter, we will look upon the plaintext source of the previous section as the result of coding the actual message source, whose output sequence we denote by $U_1, U_2, U_3, \ldots$, into the D-ary sequence $X_1, X_2, X_3, \ldots$ . We assume that the random variables $U_i$ take values in an alphabet of L letters where $2 \leq L < \infty$. Until further notice, we assume that the source is memoryless and stationary or, equivalently, that $U_1, U_2, U_3, \ldots$ is a sequence of independent and identically-distributed (i.i.d.) L-ary random variables. The coding problem for the actual message source then reduces to the coding problem for the single random variable $U = U_1$. To avoid uninteresting complications, we assume hereafter that all L values of D have non-zero probability.

Note that, when $L = D^W$ for some positive integer w and when all L possible values of U are equally likely, the simple coding scheme of assigning a different one of the $D^W$ D-ary sequences of length w to each value of U makes the codeword $X_1, X_2, \ldots, X_W$ completely random. Conventional homophonic substitution attempts to achieve this same result when the values of U are not equally likely by choosing (if possible) an appropriate w with $D^W > L$, partitioning the $D^W$ D-ary sequences of length w into L subsets, placing these subsets in correspondence with the values of U in such a manner that the number of sequences in each subset is proportional to the probability of the corresponding value of U, and then choosing the codeword for a particular value u of U by an equally-likely choice from the subset of sequences corresponding to u. (Successive letters from the message source are independently coded in this manner.) When such a partitioning of the D-ary sequences of length w is possible, the codeword $X_1, X_2, \ldots, X_W$ is equally likely to be any of the D-ary sequences of length w so that the sequence $X_1, X_2, \ldots, X_W$ is completely random. The different codewords that represent the same value u of U are traditionally called the "homophones" for U, but we shall soon use this terminology in a slightly different and more fundamental sense. It is easy to see that conventional homophonic substitution for which $X_1, X_2, \ldots, X_W$ is

completely random is possible if and only if each value $u_i$ of U has probability $n_i/D^W$ for some integer $n_i$, in which case $n_i$ is the number of homophones that must be assigned to $u_i$.

Variable-length homophonic substitution, introduced by Günther [1], generalizes the conventional scheme in that the D-ary sequences used can have different lengths, and the seqences in the subset corresponding to a given value u of U can be selected with unequal probabilities as the codeword for u. The length W of the codeword $X_1, X_2, \ldots, X_W$ for U can thus be a random variable. For an arbitrary probability distribution for U, Günther [1] gave an algorithm for such variable-length homophonic substitution with D = 2 that makes the resulting binary codeword $X_1, X_2, \ldots, X_W$ completely random. He also noted that, when $L = 2^n$ so that the "natural coding" of a value of U would be a binary sequence of length n, his algorithm sometimes gave an expected codeword length E[W] less than n so that his algorithm also performed "data compression".

Fig. 3 diagrams a coding scheme of sufficient generality to include conventional homophonic substitution and variable-length homophonic substitution, as well as conventional source coding (or "data compression"). By the homophonic channel of Fig. 3, we mean a memoryless channel whose input alphabet $\{u_1,u_2,\ldots,u_L\}$ coincides with the set of possible values of U, whose output alphabet $\{v_1,v_2,v_3,\ldots\}$ is either finite or countably infinite, and whose transition probabilities $P(V=v_j \mid U=u_i)$ have the property that for each j there is exactly one i such that $P(V=v_j \mid U=u_i) \neq 0$. We shall consider those $v_j$ for which $P(V = v_j \mid U = u_i) > 0$ to be the homophones for $u_i$, rather than considering the codewords into which these $v_j$ are encoded to be the "homophones." By the D-ary prefix-free encoder of Fig. 3, we mean a device that assigns a D-ary sequence to each $v_j$ under the constraint that this codeword is neither the same as another codeword nor forms the first part (or "prefix") of a longer codeword. This provision, which is satisfied by Günther's coding scheme [1], ensures that, when $X_1, X_2, \ldots$ is a sequence of codewords, the end of each codeword can be recognized without examining any following symbols in the sequence. It is well-known in information theory (cf. [2, p.49] ) that such coding is general in the sense that for any D-ary uniquely-decodable code there is a D-ary prefix-free code with exactly the same codeword lengths.

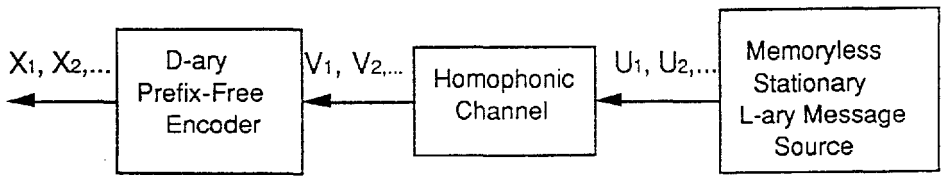


Fig. 3: A general scheme for homophonic substitution

When the homophonic channel of Fig. 2 is deterministic in the sense that all non-zero transition probabilities are 1 (so that we might as well say V = U), then Fig. 3 depicts the usual source coding (or "data compression") situation considered in information theory. When the homophonic channel is non-trivial but the binary encoding is trivially prefix-free because all codewords have the same length m (i.e., the code is a "block code"), then Fig. 3 depicts conventional homophonic substitution. In the case where both the homophonic channel is deterministic and the binary encoding is non-trivially prefix-free, then Fig. 3 depicts variable-length homophonic substitution as introduced by Günther [1].

Fig. 4 gives two examples of the general homphonic-substitution scheme illustrated in Fig. 3, both for the same binary (i.e., L=2) message source. We will soon see that both schemes in Fig. 4 are perfect. The upper system exemplifies conventional homophonic substitution into binary sequences of length w = 2. The lower system illustrates Günther's variable-length homophonic substitution. Note that the variable-length scheme has an expected codeword length of E[W] = 3/2 digits compared to E[W] = 2 for the conventional scheme. This reduction of coded symbols is an advantage offered by perfect variable-length homophonic substitution even when perfect conventional homophonic substitution is possible.

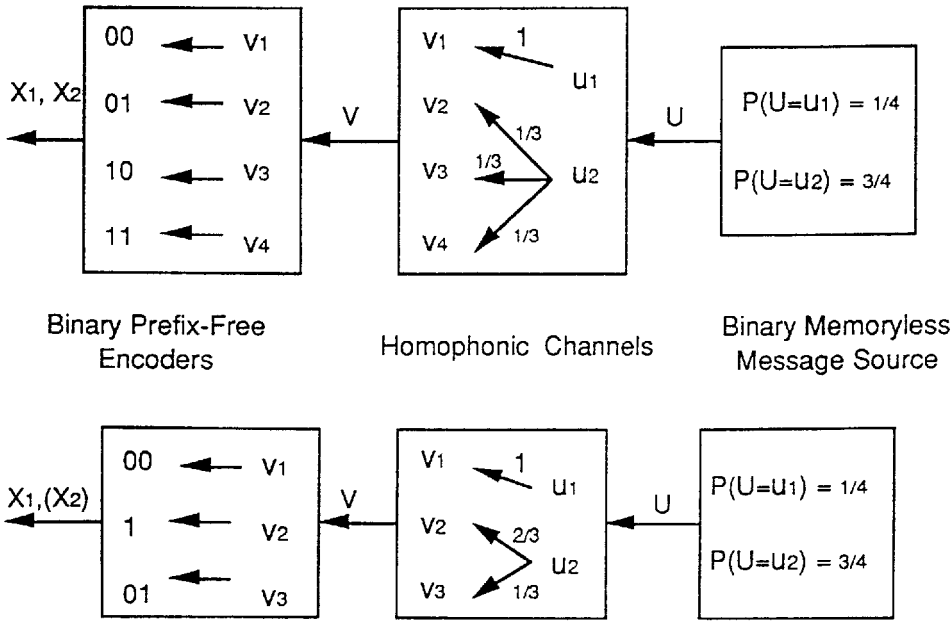


Fig. 4: Two examples of perfect homophonic substitution for the same binary memoryless message source

We will call a homophonic-substitution scheme perfect if the encoded D-ary sequence $X_1, X_2, \ldots$ is completely random. For the memoryless (source and channel) case considered in Fig. 3, this is equivalent to the condition that the codeword $X_1, X_2, \ldots, X_W$ for $V = V_1$ be completely random. Hereafter, all entropies are assumed to be in bits and all logarithms are understood to be to the base 2.

Proposition 2: For the homphonic-substitution scheme of Fig. 3,

$$H(U) \leq H(V) \leq E[W] \log D \qquad (1)$$

with equality on the left if and only if the homophonic channel is deterministic, and with equality on the right if and only if the homophonic-substitution scheme is perfect. Moreover, there exists a D-ary prefix-free coding of V such that the scheme is perfect if and only if $P(V = v)$ is a negative integer power of D for all possible values v of V. When this condition is satisfied, the scheme is perfect if and only if $P(V = v_i) = D^{-w_i}$ holds for all values $v_i$ of V where $w_i$ is the length of the D-ary codeword assigned to $v_i$.

Proof: It is well-known in information theory that $H(V) \leq E[W] \log D$ holds for every D-ary prefix-free coding of U (cf. [2, p.50]) and that equality can be achieved if and only if $P(V = v)$ is a negative integer power of D for all values v of V. Moreover, equality (when possible) is achieved by and only by a D-ary prefix-free code that assigns a codeword of length w to a value v of V with $P(V = v) = D^{-w}$. It is further well-known, (cf. [2, p.47]) that $X_1, X_2, \ldots, X_W$ is completely random for, and only for, such a code. It remains only to verify the left inequality in (1). Because the output V of the homophonic channel uniquely determines the input U, i.e., $H(U \mid V) = 0$, and because $H(U,V) = H(U) + H(V \mid U) = H(V) + H(U \mid V)$, it follows that

$$H(V) = H(U) + H(V \mid U). \qquad (2)$$

The fact that $H(V \mid U) \geq 0$ now gives the left inequality in (1). This inequality holds with equality if and only if $H(V \mid U) = 0$, i.e., if and only if the channel input U also uniquely determines the channel output V, which is equivalent to saying that the homophonic channel is deterministic.

From the facts that, in the two homophonic-substitution schemes of Fig. 3, values of V with probability 1/4 are assigned binary (D = 2) codewords of length 2 and the single value of V with probability 1/2 is assigned a binary codeword of length 1, it follows from Proposition 2 that both schemes are perfect.

When $P(V = v_i) = D^{-w_i}$ for a positive integer $w_i$ holds for all values $v_i$ of V, it is well-known (cf. [2, p.48]) that a D-ary prefix-free code in which the codeword for $v_i$ has length $w_i$ may be simply constructed as follows: Choose any distinct D-ary sequences of length 1 to be codewords for those $v_i$ (if any) with $P(V = v_i) = D^{-1}$, choose any distinct D-ary sequences of length 2 not

having any already-chosen shorter codeword as a prefix to be the codewords for those $v_i$ (if any) with $P(V = v_{ij}) = D^{-2}$, etc.

## 4. OPTIMUM HOMOPHONIC SUBSTITUTION

We will call a D-ary homophonic-substitution scheme (as in Fig. 3) for a given message source optimum if it is perfect and minimizes the expected length E[W] of the D-ary codeword assigned to the homophonic channel output V when the input is the message source output U. Let $C_D(U)$ denote the set of all homophonic channels with the property that the output letters all have probabilities that are negative integer powers of D when the input is U. Proposition 2 shows that finding an optimum homophonic-substitution scheme reduces essentially to finding a homophonic channel in $C_D(U)$ that minimizes the output entropy V and thus we shall also call such a homophonic channel optimum. We shall soon see that the optimum homophonic channel is essentially unique. For simplicity, we will take D = 2 in the remainder of this section and the next; the required generalization will be indicated in Section 6.

We begin by noting that the channels in $C_2(U)$ are characterized by the fact that, for each value u of U, the probabilities of the homophones for u form a decomposition of P(U = u) as a sum of negative integer powers of 2. For example, the upper and lower homophonic channels in Fig. 4 are both in $C_2(U)$ and decompose $P(U = u_2) = 3/4$ as $1/4 + 1/4 + 1/4$ and $1/2 + 1/4$, respectively. We next note that if the channel is optimum then the decomposition of P(U = u) for every u must consist of distinct negative powers of 2. The reason for this is that two terms equal to $2^{-n}$ would contribute $2(-2^{-n} \log 2^{-n}) = n\,2^{-n+1}$ to the entropy H(V), whereas their replacement by a single term equal to their sum $2^{-n+1}$ would contribute only $-2^{-n+1} \log 2^{-n+1} = (n-1)2^{-n+1}$, which is always smaller. Our assumptions that $L \geq 2$ and that all L possible values of U have non-zero probabilities ensures that $0 < P(U = u) < 1$ for all u. But any real number r satisfying $0 < r < 1$ either has no decomposition as a finite sum of distinct negative powers of 2, in which case its decomposition as an infinite sum of distinct negative powers of 2 is unique, or it has such a finite decomposition together with a unique decomposition as an infinite sum of distinct negative powers of 2 in which the smallest term in the former sum is replaced by an infinite sum of successive negative powers of 2. For example, 3/8 can be decomposed as $1/4 + 1/8$ or as $1/4 + 1/16 + 1/32 + 1/64 + \ldots$ . This finite decomposition (if possible) of P(U = u) always contributes less to H(V) than does the infinite one because the contribution of the successive powers of two is

$$-\sum_{n=k+1}^{\infty} 2^{-n} \log (2^{-n}) = (k + 2)2^{-k}$$

and s always greater than $-2^{-k} \log (2^{-k}) = k2^{-k}$. We have thus proved the following characterization of optimum homophonic channels.

Proposition 3: A homophonic channel in $C_2(U)$ is optimum if and only if, for every value u of U, its transition probabilities $P(V = v \mid U = u)$ for the homophones of v cause the probabilities $P(V = v) = P(V = v \mid U = u)P(U = u)$ of these homophones to equal (in some order) the terms in the unique decomposition of $P(U = u)$ as a finite sum of distinct negative powers of 2 when $P(U = u) = i/2^n$ for some positive integers i and n, and as an infinite sum of distinct negative powers of 2 otherwise.

It follows from Proposition 3 that the lower homophonic channel in Fig. 4 is optimum, and hence that $E[W] = H(V) = 3/2$ is the minimum value of E[W] for perfect homophonic-substitution for the message source of Fig.4. Proposition 3 also answers in the negative the question raised by Günther [1] as to whether his algorithm for perfect homophonic substitution is always optimum. It is easily checked that, for some message sources, in Günther's algorithm the same value u of U can result in two differenct codewords of the same length or, equivalently in our language, two of the homophones for u can have the same probability.

It remains only to find a tight upper bound on H(V) for an optimum homophonic channel. Let

$$P(U = u) = \sum_{n \in I} 2^{-n}$$

where the sum on the right is the decomposition of $P(U = u)$ created by an optimum homophonic channel.

Then

$$H(V \mid U = u) = \sum_{n \in I} (2^{-n}/P(U = u)) \log (2^{-n}/P(U = u))$$
$$< - \sum_{n \in I} 2^{-n} \log 2^{-n}$$

where the inequality is strict because the sum it bounds increases monotonically with $P(U = u)$ but $P(U = u) < 1$. Thus,

$$H(V \mid U = u) < \sum_{n \in I} n\, 2^{-n} < \sum_{n=1}^{\infty} n\, 2^{-n} = 2 \qquad (3)$$

where the second strict inequality results from the fact that I must be a proper subset of the positive integers. Multiplying by $P(U = u)$ in (3) and summing over u gives

$$H(V \mid U) < 2. \qquad (4)$$

Using (4) in (2) and making use of Proposition 2, we obtain our desired bounds on H(V).

Proposition 4: For an optimum binary homophonic-substitution scheme,

$$H(U) \leq H(V) = E[W] < H(U) + 2.$$

The somewhat remarkable conclusion from this proposition is that an optimum homophonic channel never increases the entropy of its input U by more than 2 bits, regardless of how large H(U) might be! It is easy to see that the upper bound in Propositon 4 is as tight as possible by considering

the the binary source with $P(U = u_1) = 2^{-n}$ and $P(U = u_2) = 1 - 2^{-n}$. As n increases, H(U) tends to 0 but the entropy $H(V) = 2(1-2^{-n})$ of the output of the optium homophonic channel for U tends to 2 bits.

## 5. REALIZATION OF OPTIMUM HOMOPHONIC SUBSTITUTION

The question now arises as to how one can conveniently realize the "awkward" transition probabilities that are required in an optimum homophonic channel such as the lower channel of Fig. 4. We assume that the only source of randomness available to the implementer is a binary symmetric source (BSS), i.e., a device whose output sequence $R_1, R_2, R_3, \ldots$ is a completely-random binary sequence.

The simple way to realize the transition probabilities of an optimum homophonic channel is best explained by an example. Suppose that $P(U = u) = 13/32 = 1/4 + 1/8 + 1/32$. By Proposition 3, the transition probabilities to the three homophones for u are the "awkward" numbers $(1/4)/(13/32) = 8/13$, $(1/8)/(13/32) = 4/13$ and $(1/32)/(13/32) = 1/13$. The key point is that these three probabilities are proportional to 1/4, 1/8 and 1/32 and hence also proportional to 1/2, 1/4, and 1/16. Now consider the random experiment illustrated in Fig. 5 in which a binary rooted tree is traversed, starting at the root, until a leaf is reached; the experiment halts if the leaf has been assigned to one of the three homophones $v_1$, $v_2$ or $v_3$, otherwise it returns to the root for another traversal. In any one traversal, the probabilities of reaching $v_1$, $v_2$ and $v_3$ are proportional to 1/2, 1/4 and 1/16. Thus the probabilities of the experiment halting on $v_1$, $v_2$ and $v_3$ must also be proportional to 1/2, 1/4 and 1/16 so that these probabilities can only be 8/13, 4/13 and 1/13, respectively.

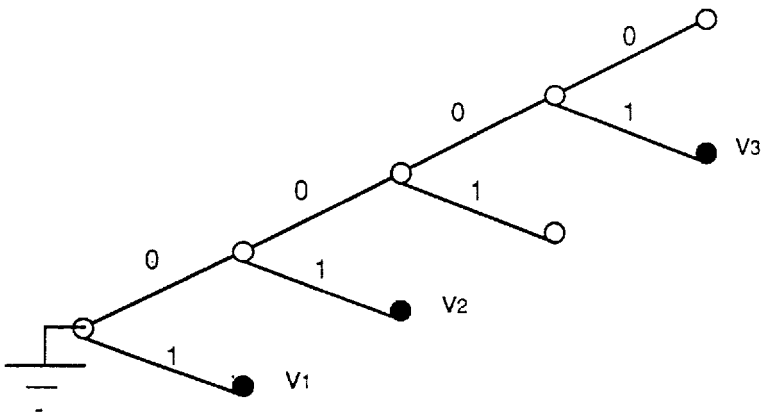


Fig. 5: A rooted tree which, if traversed continually from the root until a labelled leaf is reached, results in probabilities 8/13, 4/13 and 1/13 of terminating on the leaves $v_1$, $v_2$, and $v_3$, respectively

Consider now the average number of binary digits emitted by the BSS before the scheme just described terminates on a homophone for u. Because one node at depth 1 will always be assigned to a homophone and at least one other leaf must be assigned to a homophone, the probability exceeds 1/2 that the experiment will terminate on any one traversal of the tree. Thus, an average of less than 2 traversals will be needed. The average number of bits used in one traversal will be greatest if the tree is infinitely long, in which case this average is exactly

$$\sum_{n=1}^{\infty} n\, 2^{-n} = 2$$

because the traversal then ends with probability $2^{-n}$ on the leaf at depth n. Thus, less than 4 bits from the BSS will be required on the average to select the homophone for any value u of U. (Of course, no bits from the BSS are needed when $P(U = u) = 2^{-n}$ for some integer n.)

Proposition 5: An optimum homophonic channel for any message random variable U can be realized with a BSS as the only source of randomness in a manner such that the expected number of bits $E[B \mid U = u]$ from the BSS required to determine the homophone for u satisfies

$$E[B \mid U = u] < 4$$

for every value u of U.

## 6. GENERALIZATIONS AND REMARKS

The results of Sections 4 and 5 are easily generalized to the case of D-ary homophonic coding with $D > 2$. In Proposition 3, 2 must be replaced by D everywhere, and "distinct" must be replaced by "at most D-1 times occurring". The inequalities in Proposition 4 become

$$H(U) \le H(V) = E[W] \log D < H(U) + \frac{D}{D-1} \log D. \qquad (5)$$

The realization of optimum D-ary homophonic substitution requires that the BSS of Section 5 be replaced by the D-ary symmetric source whose output is a completely random D-ary sequence, and that the binary rooted tree traversed to obtain a homophone be replaced by a D-ary rooted tree with D-1 leaves at each depth and one node that is extended to the next depth (except in the case of a finite tree where there are D leaves at the maximum depth). The bound of Proposition 5 changes to

$$E[B \mid U = u] < \left(\frac{D}{D-1}\right)^2 \qquad (6)$$

where now of course B is the count of D-ary letters from the random source. The interested reader should have no difficulty in verifying the validity of these generalizations as no new arguments are needed.

It is also easy to generalize all the results of this paper to the case of an arbitrary L-ary message source. For the homophonic coding of $U_i$, one merely needs to replace $P(U = u)$ by $P(U_i = u \mid U_1 \ldots U_{i-1} = u_1 \ldots u_{i-1})$ where $u_1$, $u_2$, ..., $u_{i-1}$ is the sequence of message digits already coded. The homophonic channel is now a channel with memory as its transition probabilities will now depend on the past of the input sequence.

Finally, we mention the implications of Proposition 4 for source coding. Suppose that U is actually a sequence $T_1$, $T_2$, ..., $T_N$ of N digits from a memoryless and stationary source with entropy H(T). Then $H(U) = NH(T)$ so that proposition 4 shows that average number of binary encoded digits per true source letter, $E[W]/N$, satisfies

$$E[W]/N < H(T) + 2/N, \tag{7}$$

which can be made as close to H(T) as desired by choice of N. Inequality (7) differs from the traditional source-coding result of information theory (cf. [2, p. 51]) only in that the latter has 1/N in place of 2/N. The interesting fact is that the encoded digits are completely random when the optimum homophonic-substitution scheme is used to achieve the near-ideal data compression described by (7), but are only "roughly" completely random in the traditional deterministic source coding scheme. Whether this true complete randomness might be useful in source coding is a question that we leave to others to answer.

## APPENDIX

The definition in Section 2 of a non-expanding cipher is equivalent to the condition that

$$H(Y^n \mid X^n, Z) = 0 \tag{A1}$$

and

$$H(X^n \mid Y^n, Z) = 0$$

for all $n \in S$, regardless of the statistics for $X^n$ and for Z. Thus, with the aid of the identities

$$H(X^n, Y^n, Z) = H(X^n) + H(Z \mid X^n) + H(Y^n \mid X^n, Z)$$

$$= H(Y^n) + H(Z \mid Y^n) + H(X^n \mid Y^n, Z),$$

it follows that

$$H(Y^n) = H(X^n) + H(Z \mid X^n) - H(Z \mid Y^n).$$

But the independence of the key Z and plaintext sequence $X^n$ is equivalent to $H(Z \mid X^n) = H(Z)$ so that

$$H(Y^n) = H(X^n) + H(Z) - H(Z \mid Y^n)$$

holds for all $n \in S$. Thus, the inequality $H(Z \mid Y^n) \leq H(Z)$, which holds with equality if and only if $Y^n$ and Z are independent, implies

$$H(Y^n) \geq H(X^n) \qquad \text{(A2)}$$

for all $n \in S$ with equality if and only if $Y^n$ and Z are independent.

The assumption that $X^n$ is completely random gives $H(X^n) = n \log D$ bits and thus implies $H(Y^n) \geq n \log D$. On the other hand, $H(Y^n) \leq n \log D$ also holds and equality occurs if and only if $Y^n$ is completely random. Thus, if $X^n$ is completely random, equality must hold in (A2), which implies both that $Y^n$ is completely random and that $Y^n$ and Z are independent for all $n \in S$. But, the complete randomness of $Y^n$ and its independence from Z imply the complete randomness of $Y^m$ and its independence from Z for all m with $1 \leq m < n$. Because the set S contains arbitrarily large positive integers, it follows that the entire ciphertext sequence $Y_1, Y_2, Y_3, \ldots$ is completely random and independent of the key Z, which is the claim in Proposition 1.

Beginning with the identities

$$H(X^n,Y^n) = H(X^n)+H(Y^n \mid X^n)$$

$$= H(Y^n)+H(X^n \mid Y^n)$$

and recalling that if $X^n$ is totally random then so is $Y^n$ and thus $H(X^n) = H(Y^n)$, we see that

$$H(Y^n \mid X^n) = H(X^n \mid Y^n)$$

holds for all n in a non-expanding cipher when the plaintext sequence is completely random. This is the claim preceding Corollary 2 to Proposition 1.

## REFERENCES

[1] Ch.G. Günther, "A Universal Algorithm for Homophonic Coding", pp. 405-414 in Advances in Cryptology - Eurocrypt '88, Lect. Notes in Comp. Sci. No. 330. New York and Heidelberg: Springer 1988.

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

[3] C.E. Shannon, "Communication Theory of Secrecy Systems", Bell Sys. Tech. J., vol. 28, pp. 656-715, Oct. 1949.