A CHANNEL WITH INFINITE MEMORY

J. WOLFOWITZ
CORNELL UNIVERSITY

Let $w(j|i)$, where $i, j = 1, \cdots, a$, be such that

$$\min_{i,j} w(j|i) \geqq \alpha > 0 \tag{1}$$

and

$$\sum_{j=1}^{a} w(j|i) = 1 \qquad i = 1, \cdots, a. \tag{2}$$

To avoid the trivial we assume that the positive integer a is greater than one. Call any element in the set $\{1, \cdots, a\}$ a "letter," and any sequence of n letters an "n-sequence" or a "word." Let $g_1, g_2, \cdots$ be an infinite sequence of nonnegative numbers such that

$$\sum_{i=1}^{\infty} g_i = 1, \qquad \sum_{i=1}^{\infty} i g_i < \infty. \tag{3}$$

We will consider a stationary, nonanticipating channel with an infinite past R (say); the latter may be a mathematical idealization of a physical situation. For reasons which will appear in a moment it will at any time completely describe the past history to give the (infinite) sequence of letters which have been transmitted over the channel. Thus, if

$$r = (\cdots, r_{-2}, r_{-1}, r_0) \tag{4}$$

is a sequence of letters, the statement $R = r$ means that the last letter transmitted was r_0, the one before that was r_{-1}, and so forth. Suppose the letter i, where $i = 1, \cdots, a$, is sent or transmitted over the channel when the past history is r, that is, $R = r$. The chance letter $v(i)$ received (by the receiver) has the distribution

$$P\{v(i) = j|R = r\} = w(j|i)g_1 + \sum_{k=0}^{\infty} w(j|r_{-k})g_{k+2} \qquad j = 1, \cdots, a. \tag{5}$$

A code (n, N, λ) for this channel is a set

$$\{(u_1, A_1), \cdots, (u_N, A_N)\} \tag{6}$$

where $u_1, \cdots, u_N$ are n-sequences, $A_1, \cdots, A_N$ are disjoint sets of n-sequences, and

$$P\{v(u_i) \in A_i|R = r\} \geqq 1 - \lambda, \tag{7}$$

This research was supported by the U. S. Air Force under contract AF 18(600)-685, monitored by the Office of Scientific Research.

for $i = 1, \cdots, N$ and every sequence r. Here $v(u_i)$ is the chance n-sequence (word) received when the n-sequence (word) u_i is sent. The number n is the length of each word, N is the length of the code, and λ is the probability of error (more fully, an upper bound on the probability of error for each word). The practical application of the code is as follows: When the sender wishes to send the ith word ($i = 1, \cdots, N$) he sends the n-sequence u_i. When the receiver receives an n-sequence which lies in A_j he concludes that the jth word (u_j) has been sent. If the received sequence does not lie in $A_1 \cup \cdots \cup A_N$ the receiver may draw any conclusion he wishes about the word which has been sent. The probability that any word sent will be incorrectly decoded (understood) is thus at most λ, no matter what the past history of the channel may be.

In this note we shall sketch briefly a method by which the capacity of this channel can be determined (theorem 1 below) and a coding theorem and strong converse (theorem 2) proved. More details will appear elsewhere. The same channel as above, except that the number of letters in the alphabet of letters sent is different from the number of letters in the alphabet of letters received, is only trivially different from the present channel. The channel where the distribution of each letter received depends, in a manner similar to (5), on all the letters previously sent and received will be treated elsewhere.

We shall base ourselves on the ideas of [1] and [2] and use the method of generated sequences employed there. In [2], where m was the duration of memory, one "ignored" all the received letters whose serial numbers were congruent to $1, \cdots, m$, modulo $(m + \chi)$, where χ was a suitably chosen number. Of course now there is no m because the duration of memory is not finite. However, let d and χ be positive integers, and let us ignore all received letters whose serial numbers are congruent to $1, \cdots, d$, modulo $(d + \chi)$.

Write

$$t_i = g_i + g_{i+1} + g_{i+2} + \cdots \tag{8}$$

$$t'_i = g_i + 2g_{i+1} + 3g_{i+2} + \cdots . \tag{9}$$

Let

$$r = (\cdots, r_{-2}, r_{-1}, r_0) \tag{10}$$

be any sequence of letters, and

$$s = (\cdots, s_{-2}, s_{-1}, s_0) \tag{11}$$

be a fixed sequence. Consider the probability, under the condition $R = r$ and under the condition $R = s$, of the subsequence (of the n-sequence received) which consists of the letters not ignored, that is, of those letters whose serial numbers are not congruent to $1, \cdots, d$, modulo $(d + \chi)$. The ratio of these probabilities (under the two conditions) and its reciprocal can be shown to be both bounded above by $\beta = \exp_e \{t'_{(d+2)}/\alpha\}$. Suppose then one demonstrates the existence of a system (6) in which the assignment of an n-sequence to each of $A_1, \cdots, A_N$ is based solely on the subsequence of its nonignored letters (that is, two n-sequences with the same subsequence of nonignored letters belong to the same one of $A_1, \cdots, A_N$), and which, in place of (7), satisfies

$$P\{v(u_i) \in A_i | R = s\} \geqq 1 - \frac{\lambda}{\beta}, \qquad i = 1, \cdots, N. \tag{12}$$

Then clearly such a system is a code (n, N, λ) for our channel, that is, it satisfies (7) for all r.

Now suppose we obtain an upper bound on the length N of all systems (6) such that the assignment of an n sequence to each of $A_1, \cdots, A_N$ is based solely on the subsequence of its nonignored letters, and such that

$$P\{v(u_i) \in A_i | R = s\} \geqq 1 - \lambda, \qquad i = 1, \cdots, N, \tag{13}$$

is satisfied. Then, as in [2], one can show that $a^{dn/(d+\chi)}$ times this upper bound is an upper bound on the length of *all* systems (6) which satisfy (13), and is therefore a fortiori an upper bound on the length of all systems (6) which satisfy (7) (for all r), that is, an upper bound on the length of all codes (n, N, λ) for our channel.

We now consider the discrete memoryless channel $T(r)$ whose input alphabet consists of all $(d + \chi)$-sequences, whose output alphabet consists of all χ-sequences, and whose channel probability function is given by the probability distribution of the χ-sequence of letters not ignored when, in the channel of the present note, we set $n = (d + \chi)$ and $R = r$. Let $C(d, \chi)$ be the capacity of the channel $T(s)$, and let $C_r(d, \chi)$ be the capacity of the channel $T(r)$. Then it can be shown without much difficulty that, for any sequence r,

$$\left| \frac{C(d, \chi)}{(d + \chi)} - \frac{C_r(d, \chi)}{(d + \chi)} \right| < \Delta(d, \chi), \tag{14}$$

where $\Delta(d, \chi)$ will be given as follows: First define

$$\phi(d, \chi) = \prod_{i=d+1}^{d+\chi} \frac{\alpha + t_{i+1}}{\alpha} \tag{15}$$

$$h_1(d, \chi) = \phi(d, \chi) \log_2 \phi(d, \chi) \tag{16}$$

$$h_2(d, \chi) = \phi(d, \chi) - 1. \tag{17}$$

Then

$$\Delta(d, \chi) = 2h_2(d, \chi) \log_2 a + \frac{2h_1(d, \chi)}{d + \chi}. \tag{18}$$

We now use the method of generated sequences, used in [1] and [2], with modifications as follows: Let $n = n'(d + \chi)$, with n' large. Compute all probabilities as if $R = s$; thus, when the second subsequence of length $(d + \chi)$ is sent, the past history consists of the first $(d + \chi)$-sequence sent, and the sequence s before that, and so forth. A received n-sequence is generated by a transmitted n-sequence if the number of pairs (of letters of the input and output alphabets, as in [1]), differs from the expected number by less than a suitable multiple of the standard deviation. Taking into account the argument which involved (12) and (13) above one can obtain, after an argument which utilizes the ideas of [1] and [2], the following

LEMMA. *Let $\epsilon > 0$ and λ, with $0 < \lambda < 1$, be arbitrary. Let d and χ be any integers. For all n sufficiently large there exists, for the channel of this paper, a code (n, N, λ), with*

$$N > \exp_2 \left\{ n \left[\frac{C(d, \chi)}{(d + \chi)} - \Delta(d, \chi) - \epsilon \right] \right\}. \tag{19}$$

Any code (n, N, λ) for the channel satisfies (when n is sufficiently large)

$$N < \exp_2 \left\{ n \left[\frac{C(d, \chi)}{(d + \chi)} + \Delta(d, \chi) + \epsilon + \frac{d \log_2 a}{(d + \chi)} \right] \right\}. \tag{20}$$

Now apply (19) to the arbitrarily chosen pair (d_1, χ_1) and (20) to the arbitrarily chosen pair (d_2, χ_2). Since ϵ was arbitrary we obtain that

$$\frac{C(d_1, \chi_1)}{(d_1 + \chi_1)} - \frac{C(d_2, \chi_2)}{(d_2 + \chi_2)} \leqq \Delta(d_1, \chi_1) + \Delta(d_2, \chi_2) + \frac{d_2 \log_2 a}{(d_2 + \chi_2)}. \tag{21}$$

Since (d_1, χ_1) and (d_2, χ_2) were arbitrary we can reverse their roles in (21). From this and (21) we obtain

$$\left| \frac{C(d_1, \chi_1)}{(d_1 + \chi_1)} - \frac{C(d_2, \chi_2)}{(d_2 + \chi_2)} \right| \leqq \Delta(d_1, \chi_1) + \Delta(d_2, \chi_2) + \frac{d_1 \log_2 a}{(d_1 + \chi_1)} + \frac{d_2 \log_2 a}{(d_2 + \chi_2)}. \tag{22}$$

From (22) we obtain at once

THEOREM 1. *We have*

$$\lim \frac{C(d, \chi)}{(d + \chi)} = C, \tag{23}$$

say, as $d \to \infty$, $d/\chi \to 0$. For any (d, χ) we have

$$\left| C - \frac{C(d, \chi)}{(d + \chi)} \right| \leqq \Delta(d, \chi) + \frac{d \log_2 a}{(d + \chi)}. \tag{24}$$

C is the capacity of the channel, because from theorem 1 and the lemma we obtain at once

THEOREM 2. *Let $\epsilon > 0$ and λ, with $0 < \lambda < 1$, be arbitrary. For all n sufficiently large there exists a code (n, N, λ), with*

$$N > \exp_2 \{n(C - \epsilon)\}. \tag{25}$$

Any code (n, N, λ) for the channel satisfies (when n is sufficiently large)

$$N < \exp_2 \{n(C + \epsilon)\}. \tag{26}$$

The relation (24) has as a consequence that, at least in principle, the capacity C can be computed to within any specified error bound. The relation (26) is the strong converse, so called because it holds for any λ such that $0 < \lambda < 1$, and not only for small enough λ.

REFERENCES

[1] J. WOLFOWITZ, "The coding of messages subject to chance errors," *Illinois J. Math.*, Vol. 1 (1957), pp. 591–606.

[2] ———, "Strong converse of the coding theorem for the general discrete finite-memory channel," *Information and Control*, Vol. 3 (1960), pp. 89–93.