Information of Partitions with Applications to Random Access Communications

BRUCE E. HAJEK, MEMBER, IEEE

Abstract—**The minimum amount of information and the asymptotic minimum amount of entropy of a random partition which separates the points of a Poisson point process are found. Related information theoretic bounds are applied to yield an upper bound to the throughput of a random access broadcast channel. It is shown that more information is needed to separate points by partitions consisting of intervals than by general partitions. This suggests that single-interval conflict resolution algorithms may not achieve maximum efficiency.**

I. Introduction

Let $U = (U^{(1)}, U^{(2)}, \cdots, U^{(N)})$ represent a random number N of random points in an interval $[0, T]$, listed in increasing order. That is, $U^{(1)}$, $U^{(2)}$, $\cdots$ are a sequence of random variables (i.e., functions of ω for ω in some underlying probability space) with values in the interval $[0, T]$, such that with probability one $U^{(1)}(\omega) < U^{(2)}(\omega) < \cdots$, and N is a nonnegative, integer-valued random variable. Consider also a random partition $A(\omega) = (A_1(\omega), A_2(\omega), \cdots, A_{N(\omega)}(\omega))$ of the interval $[0, T]$. Thus, for each ω, $A_1(\omega), A_2(\omega), \cdots, A_{N(\omega)}(\omega)$ are disjoint subsets of $[0, T]$. The random partition A will be said to separate U for a given ω if each set $A_i(\omega)$ for $1 \le i \le N(\omega)$ contains one of the points $U_j(\omega)$ with $1 \le j \le N(\omega)$. This concept is illustrated in Fig. 1. If A separates U for all ω in a set of ω's having probability one, then A is said to separate U.

Usually in order for A to separate U there must be some statistical dependency between the random objects A and U. A measure of the dependence is the mutual information $I(A; U)$. A main result of this paper, given in Section II, is the finding of the minimum (infimum to be precise) of the mutual information $I(A; U)$ over all random partitions A which separate U when U is either a Poisson point process or when U consists of the order statistics of n independent random variables uniformly distributed on $[0, T]$.

Note that the partition sets A_i need not be random intervals. However, in Section IV we restrict our attention to partitions A for which the random sets A_i are random intervals. It is shown that separating partitions which consist of intervals must have a greater mutual information with U than the minimum needed for general separating partitions.

Manuscript received October 10, 1980; revised October 26, 1981. This work was supported by JSEP Grant N00014-79-C-0424. The material in this paper was presented at the IEEE International Symposium on Information Theory, Santa Monica, CA, February 1981.

The author is with the Coordinated Science Laboratory and Department of Electrical Engineering, University of Illinois, Urbana, IL 61801.

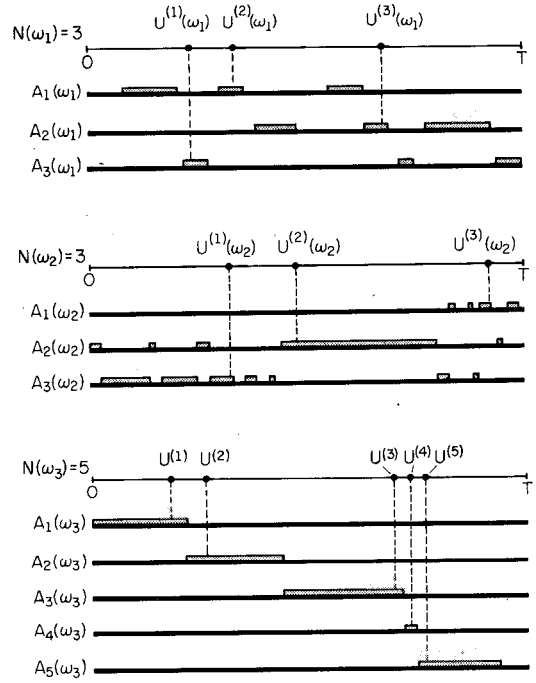

Fig. 1. A random partition which separates for three values of ω.

The problems considered in this paper are motivated by the theory of conflict resolution algorithms—in particular the tree algorithm first considered by Capetanakis [4] and Tsybakov and Mikhailov [17] (which is related to an algorithm derived earlier by Hayes [7]). Roughly speaking, the goal of a conflict resolution algorithm is to partition a population of transmitters in such a way that only one active transmitter is in each set of the partition, thus reducing conflicts among active transmitters. Furthermore, the algorithm must operate in a decentralized way with only small amounts of information available. Thus, lower bounds on the necessary amount of information needed to separate points can be translated into upper bounds on the possible efficiency of conflict resolution algorithms. This idea was first implicitly exploited by Pippenger [16]. In Section III Pippenger's bound is improved by tightening up his method. Recently, tighter bounds have been developed by noninformation theoretic methods—see the remark following Theorem 3.

As noted in Section IV, the result of that section is also motivated by the still open problem of determining the maximum throughput for certain conflict resolution algorithms.

0018-9448/82/0900-0691$00.75

II. The Amount of Information Needed to Separate Random Points

A random partition of an interval $[0, T]$ is an ordered collection $A = (A_1, A_2, \cdots)$ of disjoint random subsets of $[0, T]$[1]. Note that we have abused the term "partition" since we do not require that $\cup_i A_i = [0, T]$. Suppose that $U = (U^{(1)}, \cdots, U^{(N)})$ is a random vector such that each coordinate $U^{(i)} \in [0, T]$ and such that the number of coordinates N is possibly random. A random partition $A = (A_1, \cdots, A_N)$ is said to separate (or "encapsulate") U if with probability one $A_i \cap \{U^{(1)}, \cdots, U^{(N)}\}$ contains exactly one element for each i. Clearly the partition A and the random vector U must be highly dependent. A measure of the dependence of U and A is the mutual information between U and A, and the main results of this section are to find for two different distributions on U just how large the mutual information between U and A must be if A separates U.

We are mainly interested in the case where $U = (U^{(1)} < \cdots < U^{(N)})$ is a homogeneous Poisson point process with fixed rate $\lambda > 0$ restricted to the fixed interval $[0, T]$. Thus, if $I_1, \cdots, I_k$ are disjoint (nonrandom) subintervals of $[0, T]$, then the number of the points of U in these intervals yield independent Poisson random variables with the expected number of points in I_j equal to $\lambda \times$ (length of I_j). An equivalent characterization of the distribution of U is that N is a Poisson random variable with mean λT and given $N = n$ the conditional distribution of U has probability density

$$f_n(u) = \begin{cases} n!/T^n, & \text{if } 0 < u_1 < \cdots < u_n < T, \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

The other distribution that we consider for U is the distribution with density (2.1) where n is some fixed positive integer (Theorem 2). The density (2.1) also arises when n independent, uniformly distributed random variables on $[0, T]$ are arranged to be in increasing order. The main results of this section will now be stated. Their proofs follow some remarks.

Theorem 1: (Information of separating a Poisson point process) Let $U = (U^{(1)} < \cdots < U^{(N)})$ be the points of a homogeneous Poisson point process on $[0, T]$ with intensity $\lambda > 0$. Denote by $\tilde{\mathcal{S}}$ the set of random partitions $(A_1, \cdots, A_N)$ of $[0, 1]$ which separate U. Let $\mu(A_i)$ denote the Lebesgue measure of A_i. Then for each $A \in \tilde{\mathcal{S}}$

$$I(A; U) \geq -E\left[\sum_{i=1}^{N} \log\left(\frac{\mu(A_i)}{T}\right)\right] - \lambda T \log(\lambda T/e), \tag{2.2}$$

$$\inf_{A \in \tilde{\mathcal{S}}} I(A; U) = E[N \log N] - \lambda T \log\left(\frac{\lambda T}{e}\right) \tag{2.3}$$

and

$$\lim_{T \to \infty} \frac{1}{T} \inf_{A \in \tilde{\mathcal{S}}} I(A; U) = \lambda \log e. \tag{2.4}$$

Theorem 2: (Information of separating n independent uniformly distributed points) Let $U = (U^{(1)} < \cdots < U^{(n)})$ be a random vector with probability density given by (2.1). Denote by $\mathcal{S}$ the set of random partitions $A = (A_1, \cdots, A_n)$ of $[0, T]$ which separate U. Then for each $A \in \mathcal{S}$,

$$I(A; U) \geq -E\left[\sum_{i=1}^{n} \log\left(\frac{\mu(A_i)}{T}\right)\right] - \log n!, \tag{2.5}$$

$$\inf_{A \in \mathcal{S}} I(A; U) = \log \frac{n^n}{n!}, \tag{2.6}$$

and

$$\lim_{n \to \infty} \frac{1}{n} \inf_{A \in \mathcal{S}} I(A; U) = \log e. \tag{2.7}$$

Remarks:

1) Theorem 1 can be considered as a corollary of Theorem 2 and the fact noted above that given $N = n, U$ has conditional density (2.1).
2) The inequality (2.5) and hence (2.2) are implicit in Pippenger [16].
3) Note that if $\mu(A_i)$ is small, this gives a large contribution to the right sides of (2.2) and (2.5). This justifies the intuitive reasoning that if $\mu(A_i)$ is small, then the knowledge of A_i narrows down the location of a point of U to a set of small measure, thus giving a relatively large contribution to $I(A; U)$.
4) An intuitive justification of (2.5) is as follows. U is uniformly distributed over a set with (n-dimensional) volume $T^n/n!$. If an observer were told what partition A is, then he would know that U is contained in the set

$$\{u: u_1 < \cdots < u_n\} \cap \left(\bigcup_{\pi} A_{\pi(1)} \times \cdots \times A_{\pi(n)}\right),$$

where $\pi = (\pi(1), \cdots, \pi(n))$ ranges over all $n!$ permutations of $1, \cdots, n$. This set has volume $\prod_{i=1}^{n} \mu(A_i)$. Thus, the volume of the region in which U could have been located is reduced by a factor of $(T^n/n!)/(\prod_{i=1}^{n} \mu(A_i))$ when A is learned. The logarithm of this ratio represents the minimum (random) amount of information learned by observing A. The right side of (2.5) is simply the average of this bound.

[1] Throughout this paper $(\Omega, \mathcal{F}, P)$ will be a fixed, complete probability space. It will be assumed that $(\Omega, \mathcal{F}, P)$ is sufficiently large to accommodate the construction of various ensembles of random variables with consistent specified distributions. A random subinterval of $\mathbb{R}$ is an interval $(A, B) = \{x: A(\omega) < x < B(\omega)\}$ (or $[A, B) = \{x: A \leq x < B\}$, $(A, B] = \{x: A < x \leq B\}$ or $[A, B] = \{x: A \leq x \leq B\}$) for each ω, where A and B are possibly dependent real random variables defined on $(\Omega, \mathcal{F}, P)$. To avoid measurability technicalities, each random set in this paper will be assumed to be a finite or countably infinite union of random intervals. Thus, for example, the Lebesgue measure of a random set is a random variable.

5) Our main motivation in discovering (2.6) and its analog (2.3) was to see how tight the bound (2.5) is. As seen in the proof, equality holds in (2.5) if and only if conditioned on the partition A, the points of U are *independently* distributed *uniformly* over the corresponding partition sets of A. Thus, if for some reason such a uniform conditional distribution is not always possible, maybe the bound (2.5) on $I(A;\ U)$ in terms of the distributions of the lengths $\mu(A_i)$ could be improved. Perhaps the main import of (2.6) is that this is not the case, at least when the partition lengths $\mu(A_i)$ are constrained to be equal. That is, in proving (2.6) we show that when A is restricted so that $\mu(A_1) = \cdots = \mu(A_n)$, (this minimizes the right side of (2.5)), then the inequality (2.5) cannot be improved, except to perhaps a strict inequality.
6) On the basis of certain measure theoretic considerations, I conjecture that the infimum in (2.6) and (2.3) is not actually a minimum.

Proof of Theorem 1: We shall prove Theorem 1 using Theorem 2, which is proved below. Let $A \in \tilde{\mathcal{S}}$. Then N, the number of points of U, is completely determined by either A or U. Moreover, given $N = n$, A is a random partition which separates $(U^{(1)}, \cdots, U^{(N)})$, and the conditional distribution of $(U^{(1)}, \cdots, U^{(N)})$ given $N = n$ is the same as the distribution of U in Theorem 2. Thus (2.5) of Theorem 2 immediately implies a lower bound on $I(A;\ U \mid N = n)$. We thus have

$$\begin{aligned} I(A;\ U) &= I(A, N;\ U) \\ &= I(N;\ U) + I(A;\ U \mid N) \\ &= H(N) + I(A;\ U \mid N) \\ &\geq H(N) - E\left[\sum_{i=1}^{N} \log\left(\frac{\mu_i(A_i)}{T}\right)\right] \\ &\quad - E[\log(N!)]. \end{aligned} \tag{2.8}$$

Since N is a Poisson random variable with mean λT

$$H(N) = -\lambda T \log(\lambda T/e) + E[\log(N!)]. \tag{2.9}$$

Substituting (2.9) into (2.8) yields (2.2). By the same reasoning, but now using (2.6) instead of (2.5),

$$\begin{aligned} \inf_{A \in \tilde{\mathcal{S}}} I(A;\ U) &= H(N) + \inf_{A \in \tilde{\mathcal{S}}} I(A;\ U \mid N) \\ &= H(N) + E\left[\log\left(\frac{N^N}{N!}\right)\right] \end{aligned}$$

which, by (2.9), is equivalent to (2.3). Because $\lim_{T \to \infty} E[((N/\lambda T) - 1)^2] = 0$ it follows that $\lim_{T \to \infty} E[(N/T)\log(N/\lambda T)] = 0$. This fact and (2.3) imply the final assertion (2.4). □

Proof of Theorem 2: We shall prove (2.5) first. A quick proof of (2.5) is suggested in the fourth remark above. The following alternative proof keys on the conditional distribution of a single point of U given the set A_i containing it. A simple rescaling argument shows that is suffices to prove Theorem 2 for $T = 1$. Thus, we assume $T = 1$.

A random permutation $\pi = (\pi(1), \cdots, \pi(n))$ is called uniformly distributed if it takes on each of its $n!$ possible values with equal probability. Choose a uniformly distributed permutation $\pi = (\pi(1), \cdots, \pi(n))$ which is independent of (A, U), and define $\tilde{U} = (\tilde{U}_1, \cdots, \tilde{U}_n)$ by $\tilde{U}_i = U^{(\pi(i))}$. Note that $\tilde{U}$ is uniformly distributed over $[0, T]^n$. Let $\gamma = (\gamma(1), \cdots, \gamma(n))$ be the random permutation of $\{1, \cdots, n\}$ such that $\tilde{U}_i \in A_{\gamma(i)}$ for $i = 1, \cdots, n$. Note that γ is also uniformly distributed. By the conditional independence of $\tilde{U}$ and A given U and the fact that U is a function of $\tilde{U}$,

$$\begin{aligned} I(A;\ U) &= I(A;\ \tilde{U}) \\ &= I(A, \gamma;\ \tilde{U}) - I(\gamma;\ \tilde{U} \mid A) \\ &\geq I(A, \gamma;\ \tilde{U}) - H(\gamma) \\ &= H_n(\tilde{U}) - H_n(\tilde{U} \mid A, \gamma) - \log n!, \end{aligned} \tag{2.10}$$

where H_n denotes differential entropy (see [3, p. 86], [15]) relative to Lebesgue measure on $[0, 1]^n$. Since $\tilde{U}$ is uniformly distributed over $[0, 1]^n$, $H_n(\tilde{U}) = 0$. Now

$$\begin{aligned} H_n(\tilde{U} \mid A, \gamma) &\leq \sum_{i=1}^{n} H_1(\tilde{U}_i \mid A, \gamma) \\ &\leq \sum_{i=1}^{n} H_1(\tilde{U}_i \mid A_{\gamma(i)}) \\ &\leq \sum_{i=1}^{n} E[\log \mu(A_{\gamma(i)})] \\ &= \sum_{i=1}^{n} E[\log \mu(A_i)]. \end{aligned} \tag{2.11}$$

The second inequality results from the fact that the random set $A_{\gamma(i)}$ is a function of (A, γ). The third inequality is a consequence of the fact that $P(\tilde{U}_i \in A_{\gamma(i)}) = 1$ and thus $H_i(\tilde{U}_i \mid A_{\gamma(i)})$ is maximized when, conditioned on $A_{\gamma(i)}$, $\tilde{U}_i$ is uniformly distributed over $A_{\gamma(i)}$. Combining (2.10), (2.11), and the fact $H_n(\tilde{U}) = 0$ yields (2.5).

Since (2.7) is a consequence of (2.6) and Sterling's formula, it remains to prove (2.6). One half of (2.6) ((2.13) below) follows from (2.5). Indeed, for any $A \in \mathcal{S}$,

$$\begin{aligned} -E\left[\sum_{i=1}^{n} \log \mu(A_i)\right] &= -nE\left[\frac{1}{n}\sum_{i=1}^{n} \log \mu(A_i)\right] \\ &\geq -n \log E\left[\frac{1}{n}\sum_{i=1}^{n} \mu(A_i)\right] \\ &= \log n^n - n \log E\left[\sum_{i=1}^{n} \mu(A_i)\right] \\ &\geq \log n^n \end{aligned} \tag{2.12}$$

by Jensen's inequality and the convexity of $-\log(x)$. Substitution of (2.12) into (2.5) yields

$$\inf_{A \in \mathcal{S}_n} I(A;\ U) \geq \log \frac{n^n}{n!}. \tag{2.13}$$

Equation (2.5) and hence Theorem 2 will now be established once we prove (2.13) with the inequality reversed.

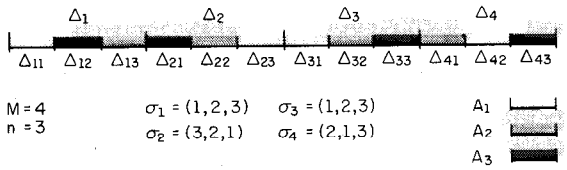

Fig. 2. A realization of the random partition A.

Let us first outline the proof.

Examination of the proof of (2.13) shows that $I(A; U) = \log(n^n/n!)$ if and only if $P(\mu(A_i) = (1/n)) = 1$ for all i and, given (A, γ), then $\tilde{U}_1, \cdots, \tilde{U}_n$ are conditionally independent and $\tilde{U}_i$ is uniformly distributed over $A_{\gamma(i)}$. Although we conjecture that such a separating partition does not exist, we can construct separating partitions which exhibit approximately such behavior. (See Remark 5.) Since we want to construct a partition separating points with a specified conditional distribution of points given the partition, it is useful to work backwards and construct the partition first and then construct the random points. The thing to check at the end is that the unconditional distribution of the points has density (2.1). Unfortunately we cannot quite arrange this so a final step is required to patch up the argument.

Fix a large integer M and subdivide the interval $[0, 1]$ into M equal-length disjoint subintervals $\{\Delta_j\colon 1 \le j \le M\}$. Further subdivide each Δ_j into n equal-length disjoint sub-subintervals $\{\Delta_{jk}\colon 1 \le k \le n\}$. Choose independent uniformly distributed random permutations $\sigma_1, \cdots, \sigma_M$ of $\{1, \cdots, n\}$. Define a random partition $A = (A_1, \cdots, A_n)$ of $[0, 1]$ by $A_i = \cup_{j=1}^{M} \Delta_{j, \sigma_j(i)}$. Note that $\mu(A_i) = (1/n)$ for all i, ω. See Fig. 2.

Next construct a random vector $\tilde{V} = (\tilde{V}_1, \cdots, \tilde{V}_n)$ such that, given A $\tilde{V}_1, \cdots, \tilde{V}_n$ are conditionally independent and $\tilde{V}_i$ is uniformly distributed over $A_i \subset [0, 1]$. Clearly A is a separating partition for $\tilde{V}$. $\tilde{V}$ is almost uniformly distributed over $[0, 1]^n$. Indeed, define $F \subset [0, 1]^n$ by $F = \cup(\Delta_{j_1} \times \cdots \times \Delta_{j_n})$, where the union is over all n-tuples of *distinct* $j_1, \cdots, j_n \in \{1, \cdots, M\}$. Then $\tilde{V}$ has a probability density function $f_{\tilde{V}}$ with respect to Lebesgue measure μ_n on $[0, 1]^n$ satisfying $f_{\tilde{V}}(v) = 1$ for $v \in F$ and $f_{\tilde{V}}(v) \le n!$ for all $v \in [0, 1]^n$. Also, $P(\tilde{V} \in F) = \mu_n(F) = (M!/M^n(M-n)!) \triangleq \delta_M \to 1$ as $M \to +\infty$. It follows by the dominated convergence theorem that

$$H_n(\tilde{V}) = -\int_{[0,1]^n} f_V(v) \log f_V(v)\, dv \underset{M \to \infty}{\to} 0. \tag{2.14}$$

Note also that

$$H_n(\tilde{V} \mid A) = \sum_{i=1}^{n} H_n(\tilde{V}_i \mid A_i) = -n \log(n) = -\log n^n. \tag{2.15}$$

Combining (2.14) and (2.15) yields

$$I(\tilde{V}; A) = H_n(\tilde{V}) - H_n(\tilde{V} \mid A) \underset{M \to \infty}{\to} \log n^n. \tag{2.16}$$

Define $V = (V^{(1)}, \cdots, V^{(n)})$ by requiring $V^{(1)} < \cdots < V^{(n)}$ and $\{V^{(1)}, \cdots, V^{(n)}\} = \{\tilde{V}_1, \cdots, \tilde{V}_n\}$. We then have $V^{(i)} = \tilde{V}_{\sigma(i)}$, where $\sigma = \{\sigma(1), \cdots, \sigma(n)\}$ is a random permutation which, by symmetry, is uniformly distributed and independent of V. Note, however, that σ is a function of (V, A) and note that $\tilde{V}$ and (V, σ) are functions of each other. By these facts

$$\begin{aligned} I(\tilde{V}; A) &= I(V, \sigma; A) \\ &= I(V; A) + I(\sigma; A \mid V) \\ &= I(V; A) + \log n!. \end{aligned} \tag{2.17}$$

Hence, in view of (2.16),

$$I(V; A) \underset{M \to \infty}{\to} \log \frac{n^n}{n!}. \tag{2.18}$$

V is nearly uniformly distributed over $\{v \in [0, 1]^n\colon v_1 < \cdots < v_n\}$. Indeed, by the properties of $\tilde{V}$ discussed above,

$$P(V \in F) = P(\tilde{V} \in F) = \delta_m \underset{M \to \infty}{\to} 1$$

and V has a bounded (independently of M) density f_V satisfying $f_V(v) = n!$ on $\{v \in [0, 1]^n\colon v_1 < \cdots < v_n\} \cap F$. In light of (2.18), if V were exactly (rather than nearly) uniformly distributed over $\{v \in [0, 1]^n\colon v_1 < \cdots < v_n\}$ then the proof of (2.6) would be complete.

Let $U = (U^{(1)}, \cdots, U^{(n)})$ be uniformly distributed over $\{u \in [0, 1]^n\colon u_1 < \cdots < u_n\}$. Let $B = (B_1, \cdots, B_n)$ be a random partition separating U such that B is a function of U and $H(B) < +\infty$. Such a partition is provided, for example, by the Capetanakis tree algorithm [4]. Modify B by defining $B^0 = B$ if $U \in F^c$ and $B^0 = \{[0, 1], \varnothing, \cdots, \varnothing\}$ if $U \notin F^c$. Construct another random partition A^0 by defining $A^0 = \{[0, 1], \varnothing, \cdots, \varnothing\}$ if $U \in F^c$ and requiring (A^0, U) to have the same conditional distribution given $U \in F$ as (A, V) given $V \in F$. This, of course, is possible since U and V have the same probability distribution on the subset F. Finally, let $C^0 = A^0$ if $U \in F$ and let $C^0 = B^0$ if $U \ne F$. Let $Z = 1$ if $U \in F$ and $Z = 0$ if $U \notin F$. C is a random partition which separates U, and by Fano's inequality [11]

$$\begin{aligned} I(C; U) &\le I(A^0; U \mid Z) + I(B^0; U \mid Z) + H(Z) \\ &\le I(A; V) + H(B^0) + H(Z). \end{aligned}$$

As $M \to +\infty$, $I(A; V) \to \log(n^n/n!)$, $H(Z) = -\delta_n \log \delta_n - (1 - \delta_n)\log(1 - \delta_n) \to 0$, and $H(B^0) \to 0$ since $P(U \in F) = \delta_n \to 1$. This completes the proof of (2.6) and hence, the theorem. □

III. An Upper Bound on the Efficiency of Certain Random Access Algorithms

In this section we find an upper bound on the efficiency of any conflict resolution algorithm (CRA) for a Poisson model of active users and 0-, 1-, e-feedback. To begin, we will describe the particular model used and describe how two well-known CRA's fit into the model.

Suppose that a large number of communication stations (called "users") must share a multiaccess broadcast channel such as a satellite relay channel, and suppose that some

finite subset of the users (called "active users") actually have a data packet to transmit. Time is divided into slots. It is assumed that at most one packet can be transmitted in any one slot. If two or more transmissions take place in one slot, they "collide" and the transmitting users remain active. If exactly one transmission takes place in a slot, it will be successful and the transmitting user becomes inactive.

We assume that each user (or at least each active user) is assigned a point in a fixed interval $[0, T]$. We will mention three possibilities for this. First, the interval $[0, T]$ may be coordinates of space, so that each user's point corresponds to his location. Second, $[0, T]$ might represent a time interval and an active user's point might correspond to the time at which the user received a packet to transmit. A third possibility is that each user simply generates a number at random in the interval $[0, T]$. For example, the basic Capetanakis–Tsybakov–Mikhailov CRA [10], [4] requires users to "flip a fair coin" to introduce randomization. For our purposes, we could pretend that each user has already flipped a fair coin an infinite number of times, and stored away the results ("predestination"). Then a user who generated the list $R_1, R_2, \cdots$ ($R_i = 0$ or 1) in this fashion could be assigned the number $.R_1R_2 \cdots \in [0, 1]$ expressed in binary notation. Note that in each case each user knows the number assigned to him.

Since the total number of users is large and each has a small probability of being active it is reasonable to suppose that the points $U = (U^{(1)} < \cdots < U^{(N)})$ in $[0, T]$ corresponding to active users form a Poisson point process with some intensity $\lambda > 0$.

A characterizing feature of the conflict resolution scenario we are considering is that the users are not in direct communication (at least prior to the conflict resolution) so that each user has a limited amount of information about the status of the other users. We will use an information model which we call the 0-, 1-, e-feedback model—it has been used in [9], [4] and elsewhere. For this model it is assumed at the end of a given slot that all users are correctly informed of one of the three outcomes:

0) no users transmitted in slot;
1) one user transmitted (successfully) in slot;
e) two or more users transmitted (unsuccessfully) in slot.

Note that when e) occurs (a collision) the users are not informed of how many users were involved.

The action of a CRA can now be described. For each slot i, the algorithm determines a subset B_i of $[0, T]$ such that all active users in B_i transmit their packets in slot i. Then, by the 0-, 1-, e-feedback assumption, at the end of the slot each user learns X_i, where $X_i = 0$, $X_i = 1$, or $X_i = e$ if B_i contains zero, one, or more than one active user, respectively. Now at the beginning of slot i, each active user must know whether or not to transmit in the slot. That is, each user must know whether or not it is in B_i. Since the only information available to the users at the beginning of the slot is the past feedback information

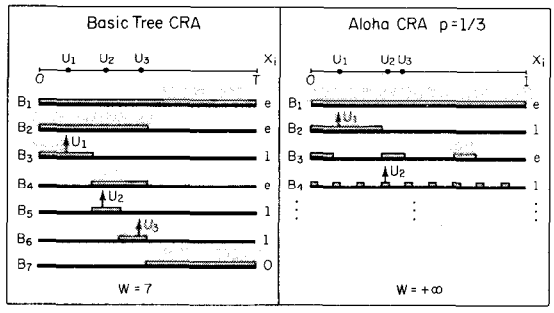

Fig. 3. Two examples of a conflict resolution algorithm.

$(X_1, \cdots, X_{i-1})$, we must require that the set B_i be a function of $(X_1, \cdots, X_{i-1})$[2].

Two particular conflict resolution algorithms are illustrated in Fig. 3. The first is the basic tree algorithm of Capetanakis–Tsybakov–Mikhailov (see [10]). In our terminology, it is described as follows. In the first slot, all active users transmit (i.e., $B_1 = [0, T]$). Then for $i \geq 1$, if there is a conflict in slot i (so $X_i = e$) then B_{i+1} consists of the first half of B_i. If there is not a conflict in slot i (so $X_i = 0$ or $X_i = 1$) then B_{i+1} is the second half of the interval B_j where j is the largest index smaller than i such that all originally active users in the first half of B_j have already transmitted. Note that since the distribution of a typical active user is uniform over the interval $[0, T]$, given that the user was involved in a conflict in slot i, the probability that the user will be enabled in slot $i + 1$ is one half, independently of which slots the user transmitted in during previous slots.

The second algorithm illustrated in Fig. 3 is the basic ALOHA algorithm with retransmission probability one third. For simplicity, set $T = 1$. The set B_i is then given by the set of x in $[0, T]$ with a zero in the $i - 1$th place to the right of the decimal point in their base three representation. Thus, since the distribution of a typical active user is uniform over the interval $[0, 1]$, the probability that an active user transmits in slot i is $1/3$, independently of which of the previous slots the user has previously transmitted in. For this example the sequence of sets (B_i) does not depend on the feedback sequence (X_i).

[2]Note that it is not appropriate to let B_i be an arbitrary randomized function of $(X_1, \cdots, X_{i-1})$ because the lack of communication among the users would prevent them from agreeing on an outcome to the randomization. However, some realistic strategies do allow users to independently generate numbers to aid in their decision making. Although we require B_i to be a (nonrandomized) function of $(X_1, \cdots, X_{i-1})$, our framework can take into account such randomized strategies. Indeed, suppose that each user generates a number at random uniformly distributed in $[0, 1]$. The digits in the binary expansion of such numbers would provide an infinite sequence of coin toss variables by which the users could make randomized decisions. As before, each user would also have a point in $[0, T]$ (see the introduction). If the distribution of points in $[0, T]$ corresponding to active users is a Poisson point process then the pairs of points (original plus randomly generated) associated with active users would form a Poisson point process in the rectangle $[0, T] \times [0, 1]$. Hence a CRA specifying subsets of $[0, T] \times [0, 1]$ by a nonrandomized function of $(X_1, \cdots, X_{i-1})$ can be used to model a CRA with user randomization. We need only note that the results of Sections II and III carry over with no essential modification when the interval $[0, T]$ is replaced by the rectangle $[0, T] \times [0, 1]$.

A CRA is defined to be completely executed at the end of W slots if there are no more active users left after W slots and if, moreover, the feedback variables $X_1, \cdots, X_W$ indicate that no more active users are left. (Note, for example, it may happen that the initial number of active users N may be zero—however the CRA will not have completed execution until the feedback sequence $X_1 = 0$, $X_2 = 0, \cdots$ is long enough to *imply* that $N = 0$). For the first example in Fig. 3, the CRA completed execution after seven slots. On the other hand, unless the ALOHA algorithm is modified, it will never complete execution whenever $X_1 = e$ because $B_2 \cup B_2 \cup \cdots \cup B_n$ does not cover the interval [0, 1] for any finite n.

If a CRA is completely executed in W slots, the efficiency of the algorithm is defined to be $\eta = EN/EW = T/EW$.

Consider those sets B_i corresponding to slots i with successful transmissions. These sets form a random partition of $[0, T]$ which separate the active users. Thus, the bound of Theorem 1 can be applied to establish the following theorem.

Theorem 3: For any CRA as described above, $\eta \leq \eta^* \approx 0.711386$, where η^* is defined in Lemma 1 below.

Remark: Our proof of Theorem 3 refines arguments of Pippenger [16] which he used to establish an upper bound of 0.744. Molle [12], [13] has recently obtained an improved upper bound of 0.6731. This was subsequently improved to 0.6125 by Cruz and Hajek [5]. At present the best upper bound is 0.587, given by Tsybakov and Mikhailov [18]. Bounds for different feedback models have begun to appear [2]. The arguments of these papers are quite different (at least on the surface) from the information theoretic ones given in Pippenger [16] and here. It is felt that the maximum efficiency is considerably smaller than 0.587, and hopefully the techniques of these papers can be combined to obtain a tighter upper bound. The maximum asymptotic efficiency of known CRA's for 0-, 1-, e-feedback and Poisson source is $0.488 \cdots$ [14].

Proof: The idea of the proof is as follows. Consider some conflict resolution algorithm and then define the random variables

$$l_i = \mu(B_i),$$
$$p_{0,i} = P[X_i = 0 \mid X_1, \cdots, X_{i-1}],$$
$$p_{1,i} = P[X_i = 1 \mid X_1, \cdots, X_{i-1}],$$
$$p_{e,i} = P[X_i = e \mid X_1, \cdots, X_{i-1}].$$

These variables characterize the conditional probability distribution of the outcome of step i, conditioned on the outcome of the previous steps. Following Pippenger [16] we shall derive three inequalities involving these variables (see (3.1), (3.3), and (3.6)). We then note that the inequalities depend only on the distribution of $(p_{0,i}, p_{1,i}, p_{e,i}, l)$ after averaging over slots i. The resulting probability distribution ν, defined precisely below, can also be described as follows. For a set A, $\nu(A)$ is proportional to the expected number of slots i with $1 \leq i \leq W$ such that $(p_{0,i}, p_{1,i}, p_{e,i}, l) \in A$. ν is then normalized to be a probability distribution. The final step of the proof, given by Lemma 1, is to perform a maximization over certain probability distributions ν. This maximization is achieved by showing that the maximizing measure is concentrated on a single point.

Without loss of generality, we will assume that $B_i \cap B_j = \varnothing$ whenever $i > j$ and $X_j \neq e$. If this were not already true, the algorithm could be modified to give a more efficient algorithm. For convenience, define $B_i = \varnothing$ and $X_i = 0$ for all $i > W$.

First, since exactly N of $X_1, X_2, \cdots$ are equal to one,

$$E[N] = \sum_{i=1}^{\infty} P[X_i = 1] = \sum_{i=1}^{\infty} E[p_{1,i}] = E\left[\sum_{i=1}^{W} p_{1,i}\right]. \tag{3.1}$$

Secondly, by the assumption that $B_i \cap B_j = \varnothing$ if $i > j$ and $X_j \neq e$,

$$\sum_{i=1}^{\infty} l_i I_{\{X_i \neq e\}} \leq \mu([0, T]) = T. \tag{3.2}$$

Using the fact that l_i is a function of $X_1, \cdots, X_{i-1}$,

$$\begin{aligned} E\left[\sum_{i=1}^{\infty} l_i I_{\{X_i \neq e\}}\right] &= \sum_{i=1}^{\infty} E\left[l_i I_{\{X_i \neq e\}}\right] \\ &= \sum_{i=1}^{\infty} E\left[E\left[l_i I_{\{X_i \neq e\}} \mid X_1, \cdots, X_{i-1}\right]\right] \\ &= \sum_{i=1}^{\infty} E[l_i(p_{0,i} + p_{1,i})] \\ &= E\left[\sum_{i=1}^{W} l_i(p_{0,i} + p_{1,i})\right]. \end{aligned}$$

Equation (3.2) thus implies that[3]

$$E\left[\sum_{i=1}^{W} l_i(p_{0,i} + p_{1,i})\right] \leq T. \tag{3.3}$$

Since $(X_1, X_2, \cdots)$ is a function of $(X_1, \cdots, X_W)$, the entropy of $(X_1, \cdots, X_W)$ can be expressed as

$$\begin{aligned} H(X_1, \cdots, X_W) &= \sum_{i=1}^{\infty} H(X_i \mid X_1, \cdots, X_{i-1}) \\ &= \sum_{i=1}^{\infty} E[h(p_{0,i}, p_{1,i})] \\ &= E\left[\sum_{i=1}^{W} h(p_{0,i}, p_{1,i})\right], \end{aligned} \tag{3.4}$$

where h is the ternary entropy function defined by

$$h(\alpha, \beta) \triangleq -\alpha \log \alpha - \beta \log \beta - (1 - \alpha - \beta)\log(1 - \alpha - \beta).$$

[3]Our bound on η is smaller than Pippenger's bound [16] since his bound is based on (3.3) with $p_{0,i} + p_{1,i}$ replaced by $p_{1,i}$ only, which is weaker than (3.3). The fact that $p_{0,i}$ should be incorporated was suggested by Berger [1].

The partition $A = (A_1, \cdots, A_n)$ is a function of $(X_1, \cdots, X_W)$ which in turn is a function of U. Using this fact and then applying (2.2) of Theorem 1 yields that

$$\begin{aligned} H(X_1, \cdots, X_W) &= I(U; X_1, \cdots, X_W) \\ &\geq I(U; A) \\ &\geq E\left[\sum_{i=1}^{N} -\log \frac{\mu(A_i)}{T}\right] - T\log(T/e) \\ &= E\left[\sum_{i=1}^{N} -\log \mu(A_i)\right] + T\log e \\ &= E\left[\sum_{i=1}^{W} -p_{1,i}\log l_i\right] + T\log e. \end{aligned} \tag{3.5}$$

Comparing (3.4) and (3.5) yields that

$$E\left[\sum_{i=1}^{W} h(p_{0,i}, p_{1,i})\right] \geq E\left[\sum_{i=1}^{W} -p_{1,i}\log l_i\right] + T\log e. \tag{3.6}$$

Our bound on the efficiency of the algorithm is based completely on the relationships (3.1), (3.3), and (3.6).

The algorithm generates a sequence $(p_{0,i}, p_{1,i}, l_i)$, $i \geq 1$ of random vectors with values in the set

$$\Sigma = \{(p_0, p_1, l): 0 \leq p_0, p_1 \leq 1, p_0 + p_1 \leq 1, l \geq 0\}.$$

Denote the set of probability measures on Σ by Σ^*. The expectation relative to any $\nu \in \Sigma^*$ is denoted

$$E_\nu[\Phi] \triangleq \int_\Sigma \Phi(p_0, p_1, l)\nu(dp_0, dp_1, dl)$$

for any bounded, Borel measurable function Φ on Σ. The algorithm induces a probability measure $\nu \in \Sigma^*$ which is conveniently defined by requiring that

$$E_\nu[\Phi] = \frac{E\left[\sum_{i=1}^{W} \Phi(p_{0,i}, p_{1,i}, l_i)\right]}{E[W]}$$

for any bounded Borel measurable function Φ on Σ. With this choice of ν, (3.1), (3.3), and (3.6) may be respectively rewritten (divide through by $E[W]$ and use $\eta = E[N]/E[W] = T/E[W]$) as

$$\begin{aligned} E_\nu[p_1] &= \eta, \\ E_\nu[(p_0 + p_1)l] &\leq \eta, \\ E_\nu[h(p_0, p_1)] &\geq E_\nu[-p_1\log l] + \eta\log e. \end{aligned}$$

Thus, an upper bound on η is

$$\eta^* \triangleq \max_{\nu \in \Sigma^*} E_\nu[p_1], \tag{3.7}$$

subject to the two constraints

$$E_\nu[(p_0 + p_1)l - p_1] \leq 0, \tag{3.8}$$

$$E_\nu[-h(p_0, p_1) - p_1\log l/e] \leq 0. \tag{3.9}$$

The proof of Theorem 3 is completed by the following lemma, which is proved in the Appendix.

Lemma 1: The maximum defining η^* in (3.7)–(3.9) is achieved uniquely by a distribution ν concentrated at one point. Thus

$$\eta = \max_{(p_0, p_1, l) \in \Sigma} p_1 \tag{3.10}$$

subject to

$$(p_0 + p_1)l - p_1 \leq 0, \tag{3.11}$$

$$-h(p_0, p_1) - p_1\log(l/e_n) \leq 0. \tag{3.12}$$

Furthermore, the point $(p_0^*, p_1^*, l^*) \in \Sigma$ which achieves the maximum in (3.10)–(3.12) satisfies (3.11) and (3.12) with equality. We obtain $\eta^* = 0.711386 \cdots$ which is achieved by $(p_0^*, p_1^*, l^*) = (0.083739, 0.711386, 0.894684)$. □

IV. The Amount of Information Required to Separate Points by Intervals

Let $U = (U_1 < \cdots < U_n)$ be a random vector in $[0, T]^n$ with the probability density (2.1). In Section II the minimum mutual information with U of a random partition which separates U was found. In this section we shall impose the additional constraint that each of the n sets of the separating partition consist of a subinterval of $[0, T]$. It will be shown that the minimum achievable mutual information with U of random partitions which separate U is larger when the partition sets are required to be intervals. The argument is based on the following simple observation: consider a partition consisting of intervals which separates $U_1 < \cdots < U_n$ and fix an i with $1 < i < n$. Since the partition interval which contains U_i must be contained in the interval $[U_{i-1}, U_{i+1}]$, the partition interval containing U_i must have length at most $U_{i+1} - U_{i-1}$. Thus, when it happens that $U_{i+1} - U_{i-1}$ is relatively small, the value of U_i can be deduced with high precision from the knowledge of the partition interval containing U_i. (For example, in Fig. 1, $A_4(\omega_3)$ is forced to be small.) This implies a lower bound on the mutual information of the partition with U.

Proposition 1: Let $U = (U_1 < \cdots < U_n)$ be a random vector in $[0, T]^n$ with the probability density (2.1). Denote by $\mathcal{S}_I$ the set of random partitions $A = (A_1, \cdots, A_n)$ of $[0, T]$ which separate U, and for which each of the random sets $A_1, \cdots, A_n$ is a random interval. Then

$$\lim_{n \to \infty} \frac{1}{n} \inf_{A \in \mathcal{S}_I} I(A; U) \geq (1.07)\log e, \tag{4.1}$$

and for the special case $n = 2$,

$$\inf_{A \in \mathcal{S}_I} I(A; U) \geq \tfrac{1}{4}\log e + \log 2 \simeq 0.943\log e. \tag{4.2}$$

Remarks:

1) Inequality (4.1) should be compared with the result $\lim_{n\to\infty}(1/n)\inf_{A\in\mathcal{S}} I(A; U) = \log e$ from Theorem 2. The inequality (4.2) should be compared with (2.6) which, for $n = 2$, becomes $\inf_{A\in\mathcal{S}} I(A; U) = \log 2 = 0.693\log e$. An interpretation of (4.1) is that it takes at least 7 percent more information to separate a large number of points by partitions consisting of intervals than by more general partitions.

2) Proposition 1 has implications for conflict resolution algorithms. A CRA is called a single-interval algorithm if

during each time slot the collection of users enabled to transmit consists of a subinterval of $[0, T]$ (as opposed to more general subsets of $[0, T]$). It is shown in [14] that the algorithm achieving throughput 0.488 has the maximum asymptotic efficiency among all CRA's which are first-come, first-serve, single-interval algorithms. Thus, if it turned out that these additional constraints were not effective, then $0.488\cdots$ could be the maximum asymptotic efficiency of any CRA. However, since by Proposition 2 we see that more information is required to separate the users by partitions consisting of intervals rather than by arbitrary partitions, we are led to conjecture that the maximum achievable efficiency for unrestricted CRA's is greater than the maximum achievable by single-interval algorithms. This, in fact, is our main motivation for presenting Proposition 1.

Proof: Suppose that $U = (U_1 < \cdots < U_n)$ is a random vector in $[0, T]^n$ with the probability density (2.1), and suppose that $A = (A_1, \cdots, A_n) \in \mathbb{S}_I$. Without loss of generality, assume that $P(U_i \in A_i) = 1$ for each i—if this is not already true, simply reorder the intervals $A_1(\omega), \cdots, A_n(\omega)$ for each ω so that $x \in A_i(\omega)$, $y \in A_j(\omega)$ for $i < j$ implies that $x < y$. The resulting partition still separates U and has no more entropy than the original partition since it is a function of the original partition.

It is clear that if Proposition 1 is true for one value of T then it is true for any $T > 0$. For technical reasons, we will set $T = n$.

Define $U_0 = 0$ and $U_{n+1} = 1$ and let l_i denote the length of A_i. Since the interval A_i is contained in the interval $[U_{i-1}, U_{i+1}]$, we have that $l_i \le X_i$, where $X_i = U_{i+1} - U_{i-1}$ for $1 \le i \le n$. Elementary calculations yield that $X_1, \cdots, X_n$ are identically distributed with the beta density

$$f_{X,n}(x) = \left(\frac{n-1}{n}\right) x \left(1 - \frac{x}{n}\right)^{n-2}, \qquad x \in [0, n]. \tag{4.3}$$

Now, (2.5) with $T = n$ can be expressed as

$$\frac{1}{n} I(A; U) \ge \frac{1}{n} \sum_{i=1}^{n} E[-\log(l_i)] - \frac{1}{n} \log\left(\frac{n!}{n^n}\right), \tag{4.4}$$

and we have the constraints

$$l_i \le X_i, \qquad 1 \le i \le n, \tag{4.5}$$

$$\frac{1}{n} \sum_{i=1}^{n} E[l_i] \le 1. \tag{4.6}$$

Lemma 2:

$$\frac{1}{n} \sum_{i=1}^{n} E[-\log(l_i)] \ge \delta_n \triangleq E[-\log(X_1 \wedge c)], \tag{4.7}$$

where c is the unique positive constant such that $E[X_1 \wedge c] = 1$, and $x_1 \wedge c$ denotes the minimum of X_1 and c.

Proof of Lemma: Let c be determined so that $E[X_1 \wedge c] = 1$ and define $F^*(x) = P(X_1 \wedge c \le x)$, the distribution function of $X_1 \wedge c$. Clearly $F^*(x) = 1$ for $x \ge c$ and

$$1 = E[X_1 \wedge c] = \int_0^\infty x F^*(dx) = \int_0^\infty (1 - F^*(x))\, dx. \tag{4.8}$$

Define $F_l(x) = (1/n)\Sigma_{i=1}^n F_i(x)$ where F_i is the distribution function of l_i for $1 \le i \le n$. Note that

$$\frac{1}{n} \sum_{i=1}^{n} E[-\log(l_i)] = \int_0^\infty -\log(x) F_l(dx) \tag{4.9}$$

and, by (4.6),

$$1 \ge \frac{1}{n} \sum_{i=1}^{n} E[l_i] = \int_0^\infty x F_l(dx) = \int_0^\infty (1 - F_l(x))\, dx. \tag{4.10}$$

(Since $l_i \le n$ for each i, the quantities on each side of (4.9) are well defined with values in the extended interval $[-\log n, +\infty) \cup \{+\infty\}$.) The constraint (4.5) implies that $F_i(x) \ge F^*(x)$ for each $x < c$, and hence also $F_l(x) \ge F^*(x)$ for $x < c$. In addition, the fact $P(l_i \le n) = 1$ for $1 \le i \le n$ implies that $F_l(x) = 1$ for $x \ge n$. For $x \ge 0$, define $G(x) = F_l(x) - F^*(x)$. The above properties of F_l and F^* imply that $G(0) = G(x) = 0$ if $x \ge n$, that $G(x) \ge 0$ if $x < c$, and that $G(x) \le 0$ if $x > c$. Also, by (4.8) and (4.10),

$$\int_0^\infty G(x)\, dx = \int_0^\infty (1 - F^*(x))\, dx - \int_0^\infty (1 - F_l(x))\, dx \ge 0.$$

Now

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} E[-\log l_i] - \delta_n \\
&= \int_0^\infty -\log(x) G(dx) \\
&= \lim_{\epsilon \downarrow 0} \left\{ \int_0^\infty -\log(x+\epsilon) G(dx) \right\} \\
&= \lim_{\epsilon \downarrow 0} \left\{ \int_0^\infty \frac{1}{x+\epsilon} G(x)\, dx - G(x) \log(x+\epsilon) \Big|_{x=0}^{x=+\infty} \right\} \\
&= \int_0^\infty \frac{1}{x} G(x)\, dx \\
&\ge \int_0^c \frac{1}{x} G(x)\, dx + \frac{1}{c} \int_c^\infty G(x)\, dx \\
&\ge \int_0^c \left(\frac{1}{x} - \frac{1}{c}\right) G(x)\, dx \ge 0.
\end{aligned}$$

This completes the proof of Lemma 2. □

We claim that

$$\lim_{n \to \infty} \delta_n = E[-\log(X \wedge \alpha)] \triangleq \delta, \tag{4.11}$$

where X has the gamma density of order two $f_X(x) = xe^{-x}$ and α is determined by the equation $E[X \wedge \alpha] = (2 +$

$\alpha)e^{-\alpha} = 1$. Indeed, using (4.3) establishes the estimate

$$|f_{X,n}(x) - f_X(x)| = x\left[\frac{n-1}{n}\left(1 - \frac{x}{n}\right)^{n-1} - e^{-x}\right] \le x\epsilon_n, \quad 0 \le x \le L, \quad (4.12)$$

where $\varepsilon_n \to 0$ as $n \to \infty$ for any fixed L. This estimate implies that the constant c defined in Lemma 2 converges to α as $n \to \infty$ and then $\lim_{n\to\infty}\delta_n = \delta$ is established using (4.12) and the dominated convergence theorem.

Solving $(2 + \alpha)e^{-\alpha} = 1$ for α yields that $\alpha \approx 1.146$ and then a numerical integration yields that $\delta = 1.07$. Combining (4.4), (4.7), (4.8), and Sterling's formula $\lim_{n\to\infty} \log(n^n/n!) = \log e$ proves (4.1).

For the remainder of this proof $n = 2$. Denote the right endpoint of the first partition interval A_1 by Y. Then $l_1 \le Y$ and $l_2 \le 2 - Y$ so $l_1 l_2 \le Y(2 - Y)$, and since $U_1 \le Y \le U_2$ we have $Y(2 - Y) \le Z$ where

$$Z = \begin{cases} U_2(2 - U_2), & \text{if } U_2 \le \frac{1}{2}, \\ 1, & \text{if } U_1 \le \frac{1}{2} \le U_2, \\ U_1(2 - U_1), & \text{if } \frac{1}{2} \le U_1. \end{cases}$$

Then by (4.4) for $n = 2$,

$$I(A; U) \ge E[-\log(l_1 l_2)] + \log 2$$
$$\ge E[-\log(Z)] + \log 2 = \tfrac{1}{4}\log e + \log 2,$$

proving (4.2).

V. The Minimum Entropy Rate of Partitions Which Separate a Poisson Process

In Section II the minimum amount of information needed to separate a Poisson point process $U^{(1)} < \cdots < U^{(N)}$ was found. A natural related problem is to find the minimum entropy of the separating partitions. In this section a source coding argument is applied to show that the minimum entropy of separating partitions is asymptotically the same as the minimum information found earlier.

Theorem 4: Let $U^{(1)} < \cdots < U^{(N)}$ be the points of a homogeneous Poisson point process on a fixed interval $[0, T]$ with unit intensity. (Thus $E[N] = T$.) Denote by $\tilde{\mathcal{S}}$ the set of random partitions of $[0, T]$ which separate U. Then

$$\lim_{T\to\infty} \frac{1}{T}\left\{\inf_{A\in\tilde{\mathcal{S}}} H(A)\right\} = \log e. \quad (5.1)$$

Remark: The proof is based on Shannon's source coding theorem and the following idea. Mapping a point process into a partition which separates it can be achieved by a zero-distortion encoding of the point process into partitions. The appropriate measure of distortion between the point process and a partition is equal to zero if the points are separated by the partition and is equal to one if the partition fails to separate the points.

Proof of Theorem 4: Since $H(A) \ge I(A; U)$ for each $A \in \tilde{\mathcal{S}}$, the inequality

$$\lim_{T\to\infty} \frac{1}{T}\left\{\inf_{A\in\tilde{\mathcal{S}}} H(A)\right\} \ge \log e$$

follows immediately from (2.4). The reverse inequality must be proved by the construction of separating partitions with sufficiently small entropy.

For any $T > 0$ and integer $n > 0$, a Poisson point process $U^{(1)} < \cdots < U^{(M)}$ on $[0, T]$ with unit intensity can be decomposed into n independent Poisson point processes, each with total intensity $\lambda = T/n$, by letting the ith subprocess consist of the points in $(\lambda(i-1), \lambda i] \cap \{U^{(1)}, \cdots, U^{(M)}\}$. If the ith subprocess is separated by a random partition of $(\lambda(i-1), \lambda i]$ for each i, then the union of the n partitions forms a partition which separates $(U^{(1)}, \cdots, U^{(M)})$. Hence, to prove the theorem it suffices to show that, given any $\epsilon > 0$, n and λ are sufficiently large, and an independent Poisson point process $U_i = (U_i^{(1)} < \cdots < U_i^{(N_i)})$, $i = 1, \cdots, n$, each with expected total count $E[N_i] = \lambda$, then there exists a collection $(A_1, \cdots, A_n)$ of random partitions such that A_i separates U_i and

$$\frac{1}{n\lambda} H(A_1, \cdots, A_n) \le (1 + \epsilon)\log e. \quad (5.2)$$

Fix $\epsilon > 0$. Suppose that $U = (U^{(1)} < \cdots < U^{(N)})$ is a Poisson point process on the interval $[0, \lambda]$ with expected total count $E[N] = \lambda$. U is distributed in S, where S is the disjoint union

$$S = \bigcup_{k=0}^{\infty} \{u \in \mathbb{R}^k : 0 < u_1 < \cdots < u_k < \lambda\}.$$

By (2.4) of Theorem 1, if λ is sufficiently large then there is a random partition A of $[0, \lambda]$ which separates U and such that $(1/\lambda)I(A; U) \le (1 + \epsilon/3)\log e$. In addition, by our construction, A is discretely distributed—i.e., A takes values in a countable set $\mathcal{P}$ of partitions of $[0, \lambda]$. Define a distortion measure on $S \times \mathcal{P}$ by $d(u, A) = 0$ if A separates u and $d(u, A) = 1$ otherwise. For $n \ge 1$ define the extension of d to $S^n \times \mathcal{P}^n$ by

$$d_n((u_1, \cdots, u_n), (A_1, \cdots, A_n)) = \frac{1}{n}\sum_{i=1}^{n} d(u_i, A_i),$$

whenever $u_i \in S$ and $A_i \in \mathcal{P}$ for each i. We have

$$I(A; U) \le \lambda(1 + \epsilon/3)\log e,$$
$$E[d(A; U)] = 0.$$

Hence, since the distortion function d is bounded, the abstract alphabet, memoryless source coding theorem ([3, p. 281]) implies that for any $\epsilon' > 0$, there is an n sufficiently large and a mapping e_n: $S^n \to \mathcal{P}^n$ such that the range of e_n contains at most $e^{n\lambda(1+2\epsilon/3)}$ n-sequences of partitions and such that the following is true: If $U_1, \cdots, U_n$ are independent Poisson point processes, each with the same distribution as U, then

$$E\left[d_n\left((U_1, \cdots, U_n), (A_1^0, \cdots, A_n^0)\right)\right] \le \epsilon', \quad (5.3)$$

where $(A_1^0, \cdots, A_n^0) = e_n(U_1, \cdots, U_n)$.

Define random variables $Z_1, \cdots, Z_n$ by $Z_i = d(U_i, A_i^0)$. Let $B_1, \cdots, B_n$ each be random partitions of $[0, \lambda]$ such that for each i, B_i is obtained by applying the Capetanakis tree algorithm [4] to the ith point process U_i. Then define

$B_1^0, \cdots, B_n^0$ by

$$B_i^0 = \begin{cases} \{\varnothing\}, & \text{if } Z_i = 0, \\ B_i, & \text{if } Z_i = 1. \end{cases}$$

Hence, B_i^0 is a function of U_i and since $H(B_i) < +\infty$, there is a bounded continuous function $\delta = (\delta(s): 0 \le s \le 1)$, depending only on λ, such that $\delta(0) = 0$ and $H(B_i^0) \le \delta(P(Z_i = 1))$.

Now define a modification of $A_1^0, \cdots, A_n^0$ by

$$A_i = \begin{cases} A_i^0, & \text{if } Z_i = 0, \\ B_i^0, & \text{if } Z_i = 1. \end{cases}$$

Then A_i separates U_i for $i = 1, \cdots, n$ and by Fano's inequality

$$\begin{aligned} &\frac{1}{n} H(A_1, \cdots, A_n) \\ &\le \frac{1}{n} H(A_1^0, \cdots, A_n^0) + \frac{1}{n} H(B_1^0, \cdots, B_n^0) \\ &\quad + \frac{1}{n} H(Z_1, \cdots, Z_n) \\ &\le \lambda(1 + \tfrac{2}{3}\epsilon) \log e \\ &\quad + \frac{1}{n} \sum_{i=1}^{n} (\delta(P[Z_i = 1]) + h(P[Z_i = 1])), \end{aligned} \tag{5.4}$$

where h is the binary entropy function. By (5.3), $(1/n)\Sigma_{i=1}^n P(Z_1 = 1) < \epsilon'$ and ϵ' can be made arbitrarily small by choosing n large enough. Hence, since δ and h are bounded, continuous and equal to zero at zero, the term involving the summation in (5.4) can be made less than $(\lambda\epsilon \log e)/3$ by choosing n large enough. The inequality (5.2) is then obtained, completing the proof. □

Appendix
Proof of Lemma 2

The following proof of Lemma 2 consists of three steps. The first step of the proof is to find a point (p_0^*, p_1^*, l^*) in the interior of Σ for which p_1^* is a local extremum subject to the constraints (3.11) and (3.12) which, for now, are required to hold with equality. By the Lagrange method this is equivalent to finding the stationary point(s) of

$$\lambda_1\{-h(p_0, p_1) - p_1 \log(l/e)\} + \lambda_2\{(p_0 + p_1)l - p_1\} - p_1, \tag{A.1}$$

where λ_1 and λ_2 are adjusted to satisfy the constraints

$$(p_0 + p_1)l - p_1 = 0, \tag{A.2}$$

$$-p_1 \log l - h(p_0, p_1) + p_1 \log e_n = 0. \tag{A.3}$$

Throughout this proof, logarithms will be taken to base e. Setting the derivatives of (A.1) with respect to l, p_0, p_1 equal to zero yields, respectively, that

$$\lambda_1\left(-\frac{p_1}{l}\right) + \lambda_2(p_0 + p_1) = 0, \tag{A.4}$$

$$\lambda_1\{\ln p_0 - \ln(1 - p_0 - p_1)\} + \lambda_2 l = 0, \tag{A.5}$$

$$\lambda_1\{\ln p_1 - \ln(1 - p_0 - p_1) - \ln(l/e)\} + \lambda_2(l - 1) - 1 = 0. \tag{A.6}$$

Equations (A.2) and (A.4) combine to yield $\lambda_1 = \lambda_2$, so set $\lambda_1 = \lambda_2 = \lambda$. Taking the difference of the left sides of (A.5) and (A.6) yields, after some manipulation, that

$$l = 1 - e^{-1/\lambda}. \tag{A.7}$$

Finally, using (A.2) to eliminate p_0, (A.5) can be manipulated to yield

$$p_1 = l/(1 + (1 - l)e^l). \tag{A.8}$$

Now, (A.2), (A.7), and (A.8) allow us to express the left side of (A.6) in terms of λ (or l) alone. The resulting equation is readily solved using Newton's method. We find that (A.2)–(A.6) has a unique solution in Σ given approximately by

$$(p_0^*, p_1^*, l^*, \lambda^*) = (0.0837395, 0.711386, 0.8946839, 0.444288).$$

The second step of the proof is to show that

$$\lambda^*\{-h(p_0, p_1) - p_1 \ln(l/e_n)\} + \lambda^*\{(p_0 + p_1)l - p_1\} - p_1 + p_1^* \ge 0 \tag{A.9}$$

for all $(p_0, p_1, l) \in \Sigma$, with equality if and only if $(p_0, p_1, l) = (p_0^*, p_1^*, l^*)$. Equality indeed holds in (A.9) if $(p_0, p_1, l) = (p_0^*, p_1^*, l^*)$ since both of the bracketed quantities vanish. Furthermore, elementary arguments show that strict inequality holds in (A.9) if (p_0, p_1, l) is contained in the boundary of Σ. Finally, the equations found in the first part of this proof imply that (p_0^*, p_1^*, l^*) is the unique level extrema contained in the interior of Σ of the left side of (A.9). Thus, (A.9) is indeed true for all $(p_0, p_1, l) \in \Sigma$ with equality if and only if $(p_0, p_1, l) = (p_0^*, p_1^*, l^*)$.

The proof of Lemma 1 is now easily completed. Suppose $\nu \in \Sigma^*$ satisfies the constraints (3.8) and (3.9). Taking the expectation of each side of (A.9) relative to ν and using (3.8) and (3.9) implies that $E_\nu(p_1) \le p_1^*$. (It is crucial that $\lambda^* > 0$.) Moreover, equality holds if and only if ν is concentrated on the point $(p_0^*, p_1^*, l^*) \in \Sigma$. □

References

[1] T. Berger, "The Poisson multiple access conflict resolution problem," in *Multi-User Communications*, G. Longo, Ed., CISM Courses and Lecture Series, no. 265. New York: Springer-Verlag, 1981.

[2] T. Berger and N. Mehravari, "Conflict resolution protocols for random multiple-access-channels with binary feedback," in *Proc. Nineteenth Ann. Allerton Conf. Commun., Contr., Comput.*, Urbana, IL, 1981.

[3] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice Hall, 1971.

[4] J. I. Capetanakis, "Tree algorithms for packet broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 505–515, Oct. 1979.

[5] R. Cruz and B. Hajek, "A new upper bound to the throughput of a multi-access broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 402–405, May 1982.

[6] R. G. Gallager, "Conflict resolution in random access broadcast networks," in *Proc. AFOSR Workshop in Commun. Theory and Applications*, MA, pp. 74–76, Sept. 17–20, 1978.

[7] J. Hayes, "An adaptive technique for local distribution," *IEEE Trans. Commun.*, vol. COM-26, pp. 1178–1186, Aug. 1978.

[8] P. A. Humblet, "Bounds on the utilization of ALOHA-like multiple-access broadcast channels," Rep. LIDS-P-1000, M.I.T., June 1980.

[9] S. S. Lam and L. Kleinrock, "Packet switching in a multi-access broadcast channel: Dynamic control procedures," *IEEE Trans. Commun.*, vol. COM-23, pp. 891–904, Sept. 1975.

[10] J. L. Massey, "Collision-resolution algorithms and random-access communications," in *Multi-User Communications*, G. Longo, Ed., Springer-Verlag CISM Courses and Lectures Series, no. 265. New

York: Springer-Verlag, pp. 73–137, 1981.
[11] R. J. McEliece, *The Theory of Information and Coding: A Mathematical Framework for Communication*. Reading, MA: Addison-Wesley, 1977.
[12] M. L. Molle, "On the capacity of infinite population multiple access protocols," vol. IT-28, pp. 396-401, May 1982.
[13] ——, "Bounds on the capacity of infinite population multiple access protocols," in *Abstracts* 1981 *IEEE Int. Symp. Inform. Theory*, Santa Monica, CA, pp. 120–121.
[14] J. Mosely, "An efficient contention resolution algorithm for multiple access channels," M.S. thesis, Tech. Rep. LIDS-TH-918, M.I.T.
[15] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. Izd. Akad. Nauk. SSSR, Moscow, 1964. (Trans. San Francisco: Holden-Day, 1964.)
[16] N. Pippenger, "Bounds on the performance of protocols for a multiple-access broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 145–151, Mar. 1981.
[17] B. S. Tsybakov and V. A. Mikhailov, "Free Synchronous packet access in a broadcast channel with feedback," *Problems of Information Transmission*, vol. 14, no. 4, pp. 259–280, April 1979 (translated from Russian original in *Problemy Peredachi Informatsii*, Oct.–Dec. 1978).
[18] ——, "An upper bound to capacity of random multiple access systems," presented at the 1981 IEEE Inform. Theory Symp., Santa Monica, CA, Feb. 1981. *Problemy Peredachi Informatisii*, vol. 17, no. 1, pp. 90–95, Jan.–Mar., 1981.

Causal Source Codes

DAVID L. NEUHOFF, MEMBER, IEEE, AND R. KENT GILBERT, MEMBER, IEEE

Abstract—**Causal source codes are defined. These include quantizers, delta modulators, differential pulse code modulators, and adaptive versions of these. Several types of causal codes are identified. For memoryless sources it is shown that the optimum performance attainable by causal codes can be achieved by memoryless codes or by time-sharing memoryless codes. This optimal performance can be evaluated straightforwardly.**

I. Introduction

THE TASK of a source code, which consists of an encoder and decoder, is to encode the source output X into a compressed representation Z (we shall assume it is binary) and, subsequently, to decode Z into a reproduction $\hat{X}$ of X. Roughly speaking, we will say that a source code is *causal* if the reproduction created by the code of the present source output depends on present and past outputs but not on future ones. It is not required that the compressed representation be produced causally. Quantizers, delta modulators, differential pulse code modulators, and adaptive versions of these are all causal in the above sense. Indeed, these codes provided the initial motivation for this study. A precise definition of causality will be given in the next section.

Manuscript received June 6, 1979; revised January 6, 1982. This work was supported by NSF Grants ENG76-82531 and ECS79-21075. Portions of this work were presented at The Sixteenth Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, October 1978; and at the IEEE International Symposium on Information Theory, Grignano, Italy, June 1979.

D. L. Neuhoff is with the Department of Electrical and Computer Engineering, University of Michigan, Ann Arbor, MI 48109.

R. K. Gilbert was with the Computer, Information, and Control Engineering Program, University of Michigan, Ann Arbor, MI 48109; he is now with Bell Northern Research, Inc., Ann Arbor, MI 48106.

Source coding theorems for block, sliding-block, tree, and trellis source codes have shown that these classes contain codes that achieve the rate-distortion function $R(D)$, which is the optimum performance theoretically attainable. Although these classes contain some codes that are causal, it is widely believed that it is the noncausal codes, and only the noncausal codes, that achieve $R(D)$. Nevertheless, there are causal codes of great practical interest. Indeed, for some sources (e.g., Gaussian memoryless) there are causal codes (e.g., quantizers with entropy coding) whose performance comes quite close to $R(D)$. In other cases there are causal codes that are believed to perform quite well but have not yielded to rigorous analysis.

In this paper we define causality for source codes, describe several specific kinds of causal codes, and characterize the optimum performance theoretically attainable by them for memoryless sources. The goal is to discover how much is lost relative to $R(D)$ by the restriction to causal codes, or equivalently, how much can be gained by noncausal codes. The principal result is that for memoryless sources the optimum performance by causal source codes can be achieved by memoryless codes or by time-sharing memoryless codes. This best performance can be evaluated straightforwardly.

Lloyd [1] was the first to consider causal source codes (in the sense described herein). He determined how well sliding-block causal codes can perform for the binary symmetric memoryless source by finding a lower bound to the best performance, which was obviously achievable. Piret [2] shows that causal sliding-block codes with feedback could also achieve Lloyd's bound. Our approach is similar to Lloyd's.

0018-9448/82/0900-0701$00.75