SOME NEW APPROACHES TO RANDOM-ACCESS COMMUNICATIONS

(3)

Ret
- Massey
- FF

James L. MASSEY

Institute for Signal and Information Processing
Swiss Federal Institute of Technology
8092 Zurich, Switzerland

Random-accessing is defined as any technique to accomplish unscheduled seizure of a many-user communications channel; its purpose is to reduce transmission delay below what can be achieved by scheduled-accessing or by channel division. Some general principles regarding channel division, channel seizure, and the effect of feedback are formulated. The "classical" approach to random-accessing, i.e., ALOHA-like techniques, is seen to be subject to instability. A newer approach, collision-resolution algorithms (CRA's), is shown to avoid this problem. The analysis of CRA's has led to bounds on the performance of any random-access system that are briefly discussed. Two new approaches to random-accessing without feedback information are described, viz., protocol sequences for the M-user collision channel and coding for the M-active-out-of-T-user collision channel. Examples are generously used throughout the paper, and some speculations on the practicality of the new approaches are offered.

1. INTRODUCTION

Before describing "new approaches to random-access communications", we should make clear what we mean by "random-accessing" and what we see as its main purpose. To do this, we must first say a few words about "multiple-accessing" in general.

A multiple-access technique is any technique that permits two or more senders to operate on a single communications channel. Time-division multiple-accessing (TDMA), frequency-division multiple-accessing (FDMA) and code-division multiple-accessing(CDMA) are well-known multiple-access schemes of the channel-division type; i.e., they divide the single channel into many "smaller" channels, one for each sender. This division may be fixed, or it may be adjusted from time to time to correspond to the changing needs of the senders as in so-called "demand assignment" schemes. A second class of multiple-access schemes is that of what we shall call the channel-seizure type. In this type of multiple-accessing, a single sender can use the full (time and frequency) resources of the channel for himself alone on some sort of temporary basis. An example of a channel-seizure scheme is a token-ring in which, when the "token" arrives at a sender's station on the ring, that sender can remove the token, send his own message as if he were the only sender on the ring, and then reinsert the token.

A random-access technique can be defined as a multiple-accessing scheme of the channel-seizure type (i) in which it can happen that two or more senders may simultaneously attempt to seize the channel, and (ii) which provides in some way for the recovery from such "access conflicts". In a random-access system, a sender generally "takes a chance" when he attempts to seize the channel, and he relies on the access protocol to repair the damage when he encounters "bad luck".

In some communication scenarios (as we shall see later), access conflicts cannot be avoided. More often, however, it is a matter of choice whether or not to allow access conflicts and hence whether or not to use random-accessing. The obvious question is: why should anyone choose to allow such an obviously bad thing as access conflicts? The answer can be put as a second question: why should anyone demand that a sender always wait for a guarantee of exclusive access before he attempts to seize the channel? When traffic on the channel is

Reprinted with permission from *Performance '87*, pp. 551–569, 1988.

light, the bold sender will be almost sure to succeed in his gamble for access and can thus avoid the delay that a timid sender would incur. The primary purpose of random-accessing is to reduce the delay between the time that a sender obtains an information input and the time that he transmits this information successfully over the channel. Random-accessing is a gamble, but one in which the odds can be on the side of the player rather than on the side of the "house".

In Section 2 of this paper, we show why channel seizure is generally preferable to channel division for multiple-accessing, and we examine the role of channel feedback information. Section 3 describes the ALOHA approach to random-accessing and points out its virtues and defects. In Section 4, we describe one new approach to random-accessing, viz. collision resolution, and we contrast it with the ALOHA approach. Section 5 considers certain general bounds on the throughput of random-access schemes. Section 6 describes two new approaches to random-accessing without feedback. Some concluding remarks are given in Section 7.

2. SOME GENERAL MULTIPLE-ACCESS PRINCIPLES

The simplest multiple-access channel is surely the two-sender binary adder channel (2SBAC) shown in Fig. 1. Each time instant, each sender sends a binary digit (0 or 1) and the received digit is the sum (0, 1 or 2) of these two numbers, i.e.,

$$Y_n = X_{1n} + X_{2n}$$

where X_{1n} and X_{2n} are the binary digits sent by senders 1 and 2, respectively, at time n and Y_n is the received digit. The "wall" shown between the two senders in Fig. 1 signifies that the user on one side is not privy to the information to be sent on the other side, although the two users are allowed in advance to have formulated a common strategy for sending this information.

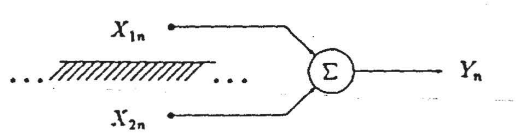

Fig.1 : The two-sender binary adder channel (2SBAC)
$X_{1n} \in \{0,1\}, X_{2n} \in \{0,1\}, Y_n \in \{0,1,2\}$

Fig. 2 shows the pentagonal "capacity region" of the 2SBAC, i.e., the region of rate pairs (R_1, R_2) such that Sender 1 can send data at the rate R_1 (bits per channel use) and Sender 2 can send at the rate R_2, both with arbitrarily small error probability.

It is easy to see how the point $(R_1,R_2) = (1,0)$ on the capacity-region boundary can be achieved. Sender 2 simply always sends 0's (and thus $R_2 = 0$) so that $Y_n = X_{1n}$, and hence Sender 1 can directly send his "raw" information bits over the channel with no need for coding ($R_1 = 1$). The point $(R_1,R_2) = (0,1)$ can be similary achieved. By agreeing to alternate between these two schemes for appropriate periods, Senders 1 and 2 can achieve any point (R_1,R_2) such that $R_1 + R_2 = 1$, i.e., at any point on the "time-sharing line" shown in Fig. 2.

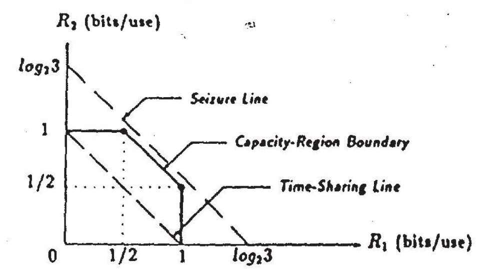

Fig.2 : Capacity Region of the 2SBAC of Fig.1

It is almost as easy to see how the point $(R_1, R_2) = (1, 1/2)$ on the capacity-region boundary can be approached. Sender 1 transmits his raw information bits ($R_1 = 1$). This causes the channel seen by Sender 2 to be that shown in Fig. 3, because, for instance, if Sender 2 should send a 1 then with probability 1/2 Sender 1 will also send a 1 and 2 will be received, while with probability 1/2 Sender 1 will send a 0 and 1 will be received. But the channel of Fig. 3 is the familiar binary erasure channel (in which a received 1 is the "erasure symbol") with erasure probability $\delta = 1/2$ and capacity $C = 1-\delta = 1/2$. Thus, Shannon's noisy coding theorem ensures the existence of a coding scheme that will allow Sender 2 to send information at a rate R_2 arbitrarily close to 1/2 with arbitrarily small error probability. After the receiver has decoded Sender 2's codeword, he can subtract it from the received sequence to obtain the uncoded sequence that was transmitted by Sender 1. The price of making R_2 closer to the capacity 1/2 is an increasingly longer codeword length or, equivalently, a longer delay in recovering the information at the receiver. The point $(R_1, R_2) = (1/2, 1)$ can, of course, be similarly approached. By appropriately alternating between coding schemes, any point (R_1, R_2) on the capacity-region boundary line $R_1 + R_2 = 3/2$ between the points (1, 1/2) and (1/2, 1) can be approached.

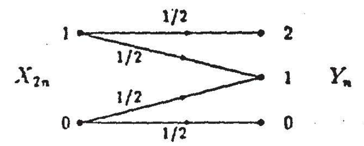

Fig.3 : The binary erasure channel seen by Sender 2 when Sender 1 sends random binary digits over the 2SBAC of Fig.1.

Perhaps the best interpretation of the "wall" shown in Fig. 1 is as a prohibition against seizure of the channel by a single sender. If a single sender is allowed to control both X_{1n} and X_{2n}, then he can by choosing (X_{1n}, X_{2n}) to be (0,0),(0,1) or (1,1) cause Y_n to be 0, 1 or 2, respectively, i.e., he can create a noiseless ternary channel with capacity $\log_2 3$ (bits per use). By alternating appropriately between such seizures, two senders could achieve any point on the "seizure line" shown in Fig. 2 that lies strictly outside the (seizure-prohibited) capacity region.

Suppose now that there is a feedback channel from the receiver to the two senders in Fig. 1 so that each sender learns the value of Y_n immediately after X_{1n} and X_{2n} have been sent. The point $(R_1,R_2) = (1, 1/2)$ can now be achieved with the greatest of ease. Sender 1 still sends his raw information bits ($R_1 = 1$) so that Sender 2 still sees the binary erasure channel of Fig. 3. Sender 2, however, can now (because of the feedback of Y_n) simply send each of his information bits repeatedly until it is received "unerased", i.e., until $Y_n = 0$ when this information bit is a 0 or until $Y_n = 2$ when this information bit is a 1. Because the erasure probability δ is 1/2, Sender 2 will be sending information at the rate $R_2 = 1-\delta = 1/2$ bits/use. Moreover, the average delay between first transmission and successful transmission is only 2-1 = 1 time instant. Something even more remarkable, however, results from the availability of feedback (as was first shown by Gaarder and Wolf [1]): points outside the capacity region of Fig. 2 can be achieved! This was quite surprising when first discovered because it had long been known that feedback could not increase the capacity of a single-sender memoryless channel. The actual capacity region of the 2SBAC with feedback was only recently determined by Willems [2]; it differs from the capacity region without feedback, shown in Fig. 2, in that the boundary line between the points (1, 1/2) and (1/2, 1) is bowed slightly outward (but still well away from the "seizure line").

The simple 2SBAC of Fig. 2 is a rich source of lessons about multiple-accessing. With its help, we have been able to illustrate all of the following <u>general principles of multiple-access communications</u>:

(1) Channel seizure, when possible, is the most effective way to utilize a multiple-access channel.

(2) When channel seizure is prohibited, time-sharing (or other types of channel division) generally is still sub-optimum in the sense that it cannot be used to achieve all points in the capacity region.

(3) Feedback, when available, can be exploited to reduce the coding delay and complexity required to achieve a given transmission rate.

(4) When channel seizure is prohibited, feedback can also enlarge the capacity region.

The first of these principles supports the way that computer communications is carried out today. Virtually all newer local area networks (LAN's) operate on a channel seizure basis, sometimes with deterministic access (as in a token ring) and sometimes with random access (as in Ethernet). The third principle suggests that feedback will play an especially crucial role in random-accessing, because some kind of "coding" is absolutely necessary to overcome the losses due to access conflicts.

3. THE ALOHA APPROACH TO RANDOM-ACCESSING

The ALOHA system, devised by Abramson [3] and his colleagues at the University of Hawaii, was the first random-access system; its approach underlies most present-day random-access systems, e.g. Ethernet. To illustrate the ALOHA approach, we now describe the ALOHA system, including the modification of "time slotting" that was introduced by Roberts [4].

Suppose that all data to be sent is in the form of "packets", all of which have the same length (measured in transmission time on the seized channel) that we take to be the unit of time. We define the time interval $(n-1) \leq t < n$ to be the n-th channel "slot". "Time-slotting" means that senders can transmit packets *only by beginning transmission at a slot boundary. Thus,* transmitted packets from two senders will either overlap completely at the receiver or not at all.

The channel model postulated by Abramson was that, when 2 or more transmitted packets overlap at the receiver, then they mutually destroy one another, but otherwise packets are received error-free. Moreover, there is feedback from the receiver at the end of each slot so that all users learn whether or not a collision occurred (collision/no-collision binary feedback).

The information-generation model postulated by Abramson was that of a very large number (essentially infinite) of identical sources, each with an associated sender, such that the number of new information packets generated during any slot is a Poisson random variable with mean λ (packets/slot), independent of previously generated packets. The essentially infinite number of senders means that access conflicts cannot be entirely avoided, i.e., random-accessing becomes a necessity. [In fact, the original operational ALOHA system had a very small number of transmitters so that random-accessing was a matter of choice, made by Abramson and his colleagues for the express purpose of reducing access delay.]

The random-access protocol devised by Abramson was ingeniously simple. A new packet must be transmitted in the slot immediately following that in which it was generated. When a collision occurs, each "colliding" sender must retransmit in a randomly-selected later slot. Each such sender, of course, independently makes this random selection of retransmission delay.

Abramson's analysis of the ALOHA system was equally ingenious, if not rigorous. He postulated that the retransmission policy could be shaped in such a way that the number of retransmitted packets in any slot would also be a Poisson random variable, independent from slot to slot and independent of the new-packet generation process, with a mean of λ_r (packets/slot). Because the sum of independent Poisson random variables is again Poisson, this implies that the total number of packets transmitted in any slot is also a Poisson random variable with mean $\lambda_t = \lambda + \lambda_r$. Because the throughput τ of successful packets at the receiver is the fraction of slots in which exactly one packet is transmitted, it follows that τ is just the probability that a Poisson random variable with mean λ_t takes on the value 1, i.e.,

$$\tau = \lambda_t e^{-\lambda_t}. \qquad (1)$$

Equation (1), which is the so-called throughput equation for slotted-ALOHA, is shown graphically in Fig. 4. It is easy to check from (1) that τ is maximized when $\lambda_t = 1$ (packet/slot), which seems quite natural, and that this maximum is

$$\tau_{max} = e^{-1} \approx .368 \text{ (packets/slot)},$$

which seems quite fundamental. It is common to say that e^{-1} is the "capacity of the slotted-ALOHA channel", but, as we shall see, this description is misleading.

The reader may (and should) be disturbed by the fact that the new-packet arrival rate λ appears nowhere in the throughput equation (1). To bring λ into the picture, one must invoke the equilibrium hypothesis which states that packets are entering and leaving the system at the same rate, i.e.,

$$\tau = \lambda.$$

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE