REFERENCES

[1] R. de Buda, *"Coherent Demodulation of Frequency-Shift Keying with Low Deviation Ratio", IEEE Trans. Commun.* (concise paper), Vol. COM-20, pp. 429-435, June 1972.
[2] M. L. Doelz et al, *"Minimum-Shift Data Communication System"*, U.S. Patent Number 2 977 417, March 1961.
[3] R. de Buda, *"The Fast FSK–A New Modulation System"*, Can. Gen. Elec. Co. Ltd. Tech Report RQ71EE2, February 1971.
[4] W. A. Sullivan, *"High Capacity Microwave System for Digital Data Transmission"*, Conf. Rec. Int. Conf. Communications, Montreal, Quebec, Canada, June 1971, pp. 23.4-23.8.
[5] W. M. Hutchinson et al, *"Data Demodulator Apparatus"*, U.S. Patent Number 3 743 755, July 1973.
[6] R. de Buda, *"About Optimal Properties of Fast Frequency-Shift Keying", IEEE Trans. Commun.*, Vol. COM-22, October 1974.
[7] R. de Buda, *"Fast FSK Signals and their Demodulation"*, Can. Elec. Eng. J. Vol. 1, Number 1, 1976.
[8] F. Amoroso, *"Pulse and Spectrum Manipulation in the Minimum (Frequency) Shift Keying (MSK) Format", IEEE Trans. Commun.*, Vol. COM-24, March 1976.
[9] S. A. Gronemeyer and A. L. McBride, *"MSK and Offset QPSK Modulation", IEEE Trans. Commun.*, VOL. COM-24, No. 8, August 1976.
[10] S. A. Rhodes, *"Effect of Noisy Phase Reference on Coherent Detection of Offset-QPSK Signals", IEEE Trans. Commun.*, Vol. COM-22, No. 8, August 1974.
[11] R. K. Kwan, *"The Effects of Filtering and Limiting a Double-Binary PSK Signal", IEEE Trans. Aerosp. Electron. Syst.* Vol. AES-5, July 1969.
[12] C. J. Wolejsza, A. M. Walker, and A. M. Werth, *"PSK Modems for Satellite Communications"*, in *Proc. First Intelsat Int. Conf. Digital Satellite Commun.*, London, England, November 1969.
[13] W. C. Lindsey, *"Synchronization Systems"*, Prentice-Hall, Englewood Cliffs, NJ, 1972, pp. 92-95.
[14] A. J. Viterbi, *"Principles of Coherent Communication"*, McGraw-Hill, New York, 1966.
[15] C. R. Wylie, *"Advanced Engineering Mathematics"*, McGraw-Hill, New York, 1969.
[16] J. N. Wozencraft and I. M. Jacobs, *"Principles of Communication Engineering"*, Wiley, New York, 1967.

★

Robert Matyas (S'70–M'73) was born in Montreal, Quebec, Canada on December 10, 1949. He received the B.Eng. degree in electrical engineering from McGill University, Montreal in 1971 and the M.Sc. degree from Queen's University, Kingston, Ontario, Canada in 1973 for his work in decision-aided carrier tracking loops.

From 1973 to 1975 he was with the Earth Station Engineering Group, Telesat Canada, Ottawa, Ontario, primarily involved in the implementation of the first commercial time division multiple access (TDMA) system for use in the Canadian domestic satellite network. In 1975 he joined the Small Terminal Satellite Communications program at the Communications Research Centre of the Department of Communications, Ottawa. Since 1975, he has been involved in the design of transportable terminals for use in satellite and land mobile systems. He is presently Project Leader for the Low Rate Digital Systems project.

The Effect of Acknowledgment Traffic on the Capacity of Packet-Switched Radio Channels

FOUAD A. TOBAGI, MEMBER, IEEE, AND LEONARD KLEINROCK, FELLOW, IEEE

Abstract—**We consider a population of terminals communicating with each other or with a central station over a packet-switched multiple access radio channel. To ensure the integrity of the transmitted data over the multi-access channel, we consider a reliable method using an error detecting block code in conjunction with a positive acknowledgment of each correct message. In this paper, we study the effect on channel capacity of the overhead created by the error-control traffic for both slotted ALOHA [1] and carrier sense multiple access (CSMA) [2]. For this we consider several implementation schemes for the two channel configurations: the common-channel configuration (a single channel for both information traffic and error-control traffic); and the split-channel configuration. The packet delay analysis will be treated in a forthcoming companion paper.**

Paper approved by the Editor for Computer Communication of the IEEE Communications Society for publication without oral presentation. Manuscript received December 20, 1976; revised December 22, 1977. This research was supported by the Advanced Research Projects Agency, Department of Defense, under Contract DAHC 15-73-C-0368.

F. A. Tobagi was with the Department of Computer Science, University of California, Los Angeles, CA 90024. He is now with the Digital Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305.

L. Kleinrock is with the Department of Computer Science, University of California, Los Angeles, CA 90024.

I. INTRODUCTION

CONSIDER an environment consisting of a large population of small (i.e. low-traffic) users communicating with each other or with a central station over a shared packet-switched multiple-access radio channel of limited bandwidth, using a random access scheme (such as slotted ALOHA [1, 3] or carrier sense multiple access (CSMA) [2, 4]). Basically, errors in multi-access radio channels are due to two major causes:

(1) random noise on the radio channel
(2) multi-use interference in the form of overlapping packets.

Indeed, it is well understood that signals for the single carrier frequency which overlap in time result in information destruction (unless a spread spectrum method such as code division multiple access (CDMA) is used which extracts a price in the form of bandwidth expansion).

To ensure the integrity of the transmitted data, a very reliable method is the use of an error detecting (e.g. cyclic) block code in conjunction with a positive acknowledgment of each correctly received message. Each packet contains a field for the cyclic checksum in its header. The receiver at the central station or the receiving terminal merely listens for a complete packet which has a correct checksum. In such an event, the receiving device transmits an acknowledgment packet back to the terminal. This acknowledgment contains (among other things) the unique identification of the originating terminal along with a checksum to ensure the integrity of the acknowledgment packet itself. The terminal constantly checks all transmissions to detect those with its ID and a correct checksum. It is only when the terminal receives a correct acknowledgment packet within an appropriate time-out interval that it considers its transmission successful. Such a scheme is referred to as a *positive acknowledgment* scheme and is used by the systems in consideration throughout this paper. If two terminals' transmissions overlap, the checksum will be wrong. The terminals become aware of their failure when they receive no acknowledgment within the time-out period, and then they retransmit their packets. To avoid continuously repeated conflicts, a *random retransmission* delay is introduced, spreading the conflicting packets over time.

In previously published analytical studies [1, 2, 3, 4], it was assumed that acknowledgment traffic was carried by a channel (assumed to be available) separate from the random access channel being examined; with that assumption, acknowledgment packets always arrive reliably and at no cost. That is, it was assumed that sufficient bandwidth is provided to the acknowledgment channel so that overlaps between acknowledgment packets are avoided; this is possible since a positive acknowledgment is created only when a packet is correctly received, and there will be at most one such packet at any given time in the environment in question. In this case, the time to receive an acknowledgment packet is simply its transmission time plus the propagation delay involved (provided that we further assume that the processing time needed to perform the error-check and to generate the acknowledgment packet is negligible). It is all too evident that in ground radio systems, one does not always enjoy the existence of this extra bandwidth, and acknowledgments will use part of the total available bandwidth (our limited resource).

The amount of overhead introduced, as well as the degradation in delay incurred, varies with the mode of operation. When the available bandwidth is provided as a single channel to be shared by both information and acknowledgment packets, then the channel performance will further suffer from interference between information packets and acknowledgment packets unless some kind of priority scheme is provided.

In this paper, we study the degradation in channel capacity due to the overhead created by the error control traffic. For this, we first present, in Section II, the throughput analysis for slotted ALOHA, nonpersistent CSMA and 1-persistent CSMA in the common-channel configuration (CC) whereby a single channel is provided for both information and error-control traffic. In Section III, we consider the split-channel configuration (SC) for which we derive the system capacity; we then compare, in terms of the channel capacity, the two channel configurations, CC and SC. The analysis of packet delay is deferred to a forthcoming paper.

Before we proceed with the analysis, let us briefly summarize the system assumptions and notation to be used. It is assumed that a terminal may either be transmitting or receiving (but not both simultaneously); however, the delay incurred to switch from one mode to the other is negligible. All packets are of constant length and are transmitted over an assumed *noiseless* channel (i.e., the errors in packet reception caused by random noise are not considered to be a serious problem and are neglected in comparison with errors caused by overlap interference). The system assumes non-capture (i.e., the overlap of any fraction of two packets results in destructive interference of both and they each must be retransmitted). We further simplify the problem by assuming the propagation delay (small compared to the packet transmission time) to be identical for all source-destination pairs*.

We now briefly review the basic notation used in this paper, and also widely used in References [2, 5, 6]. We let W denote the total bandwidth available; we denote by b_m and b_a the number of bits in a message packet and an acknowledgment packet, respectively; τ denotes the propagation delay between source and destination, and a is the ratio of propagation time to packet transmission time, $a = \tau W/b_m$**; ω is the ratio of acknowledgment length to message packet length, $\omega = b_a/b_m$. We assume that our population of terminals collectively form an independent Poisson source with an aggregate mean (message) packet generation rate of λ packets/second. Under steady state conditions, λ is also the channel throughput. The maximum packet generation rate that the total bandwidth W can ever handle is W/b_m. The normalized throughput (average number of packets per transmission time of a packet on the entire bandwidth) is denoted by S and is expressed as

$$S = \frac{\lambda b_m}{W}.$$

Here again, the maximum achievable throughput is called the capacity and is denoted by C. The message traffic offered to the channel from our collection of users consists not only of the transmission of new packets, but also of the retransmission of previously collided packets. We denote by G the mean offered traffic rate; note that $G \geqslant S$. As in [2], the offered traffic is assumed to be Poisson***.

II. COMMON-CHANNEL CONFIGURATION (CC)—THROUGHPUT ANALYSIS

A. Slotted ALOHA

Slotted ALOHA is the scheme whereby time is slotted into segments whose duration is exactly equal to the transmission

* By considering this constant propagation delay equal to the largest possible, one gets lower (i.e., pessimistic) bounds on performance.

** The bandwidth is assumed to be modulated at 1 bit/Hz·s.

*** The independence assumption in channel traffic, proven to be reasonable in the absence of acknowledgments, still holds true here despite the introduction of the acknowledgment traffic. Indeed, for analysis of the (S, G) relationships and of the channel capacity, the retransmission delay is taken to be arbitrarily large (and at channel capacity this presents no problem since packet delays are then infinite); therefore acknowledgment traffic will cause no further correlation in the channel information traffic.

time of a single packet and where users are allowed to transmit any time they desire by synchronizing the start of transmission with the beginning of a slot. In the absence of acknowledgment traffic, it was shown under the Poisson traffic assumption that [3]

$$S = Ge^{-G}. \tag{1}$$

A packet will need no further (re)transmission if and only if it is correctly received by the receiver and its acknowledgment packet is correctly received back. Two slotted schemes are considered. The first assumes no cooperation between the transmitters; information packets can interfere with acknowledgment packets. The second assumes that all transmitters, with a packet ready for transmission, perform the error-check on all packets that are transmitted: if the transmission is successful, then terminals are prohibited from transmitting their packet in the following slot (anticipating that an acknowledgment will be transmitted). In this second scheme, acknowledgment packets have priority and are always successful.

Slotted ALOHA: Common-Channel with Non-Priority Acknowledgment Traffic (CCNPA)

The channel throughput for slotted ALOHA with non-priority acknowledgment traffic is given by:

$$S = \frac{Ge^{-2G}}{1 + Ge^{-G}}. \tag{2}$$

Proof:

In this slotted (synchronized) scheme, an acknowledgment packet will use an entire slot, even if $\omega < 1$. Here we use the cycle analysis widely used in the previous papers [2, 4, 5]. Define a busy period as any complete collection of consecutive slots occupied by message packets. Because of collisions between information packets and acknowledgment packets, some among these slots will also contain colliding acknowledgment packets. An idle period is any period of time separating two consecutive busy periods. An idle period may very well contain the transmission of an acknowledgment packet if the last information packet of the preceding busy period is successful. A busy period plus the following idle period constitutes a cycle. (See Figure 1.) Let $\bar{B}$ be the expected duration of the busy period; $\bar{I}$, the expected duration of the idle period; and $\bar{B} + \bar{I}$, the expected length of a cycle. Let U denote the time during a cycle that the channel is used to transmit successful message packets. Using renewal theory arguments, the average channel utilization is simply given by

$$S = \frac{\bar{U}}{\bar{B} + \bar{I}}. \tag{3}$$

Since we require the packet and its acknowledgment to be received correctly, only one packet in a busy period (the last one!) may contribute to the throughput since it is the only packet which (when successful) has an acknowledgment packet which does not incur any interference.

Let $\{q_k\}$ be the distribution of the number of information packets arriving during a slot:

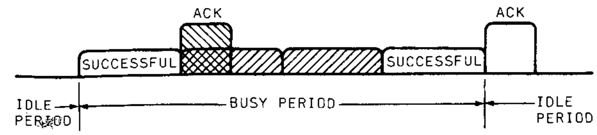


Figure 1. Slotted ALOHA: Common channel configuration with non-priority acknowledgment traffic.

$$q_k = \frac{G^k}{k!} e^{-G}. \tag{4}$$

A busy period will contain m slots occupied by information packets with probability $q_0(1 - q_0)^{m-1}$. Let P_s^i denote the probability of success of the i^{th} information packet in a busy period. This success occurs only if the previous slot is unsuccessful (otherwise an acknowledgment packet will interfere) and knowing that arrivals did occur, there is only one packet arrival. We have

$$P_s^1 = \frac{q_1}{1 - q_0} \tag{5}$$

and

$$P_s^i = (1 - P_s^{i-1}) \frac{q_1}{1 - q_0}. \tag{6}$$

Let

$$c \triangleq \frac{q_1}{1 - q_0}. \tag{7}$$

Eq. (6) can be written as

$$P_s^i = c - cP_s^{i-1} \tag{8}$$

and its solution is

$$P_s^i \triangleq \Pr\{\text{success at the } i^{\text{th}} \text{ slot of the busy period}\}$$

$$= \frac{c[1 - (-c)^i]}{1 + c}. \tag{9}$$

Since only the last packet in a busy period may contribute to the throughput, we get:

$$\bar{U} = \Pr\{\text{success during the last slot of a busy period}\}$$

$$= \sum_{m=1}^{\infty} P_s^m q_0 (1 - q_0)^{m-1}$$

$$= \frac{c}{1 + c}\left[1 + \frac{cq_0}{1 + c(1 - q_0)}\right]. \tag{10}$$

Since

$$\bar{B} = \frac{1}{q_0} \tag{11}$$

and

$$\bar{I} = \frac{1}{1 - q_0} \tag{12}$$

using Eq. (3) and substituting for all the known quantities, the throughput can be expressed in terms of q_0 and q_1; using Eq. (4), we finally get Eq. (2).

Q.E.D.

Slotted ALOHA: Common-Channel with Priority Acknowledgment Traffic (CCPA)

In this scheme, the idea is to give priority to the acknowledgment packets. To do so, the devices are considered to be able to perform the following operations:

- They have the capability of performing error-checks on any slot.†
- They can distinguish between an information packet and an acknowledgment packet.

A terminal, involved in a conversation, performs error-checks in each slot. If the generation (or rescheduling) of its packet takes place in a slot during which one of the following is true,

(i) No packets were transmitted
(ii) An acknowledgment packet was transmitted
(iii) The information packet transmitted does not have a correct checksum.

Then the terminal transmits its packet over the next slot (as in normal operation of slotted ALOHA).

If the generation (or rescheduling) of the packet takes place in a slot during which an information packet with a correct checksum is received, then the terminal delays the transmission of its packet by one slot (to avoid conflict with the acknowledgment packet which will surely follow).

Let us treat the case where $\omega = 1$. The channel throughput for slotted ALOHA with priority acknowledgment traffic is given by:

$$S = \frac{Ge^{-G}}{1 + 2Ge^{-G}(1 - e^{-G})}. \tag{13}$$

Proof:

We define a *non-zero slot* (NZS) as a slot at the beginning of which a non-zero number of packets is present for transmission. These packets may either have arrived during the immediately preceding slot or may have arrived over the slot before last and have been delayed by one (acknowledgment) slot due to the presence of an acknowledgment packet. Similarly, we define a *zero slot* (ZS). We now define a *busy period* as the time lapse from a non-zero slot immediately following a zero slot up to but excluding the first zero slot encountered (see Fig. 2). With this definition, a busy period consists of m slots occupied by *information* packets, numbered $1, 2, \cdots, m$, with probability $q_0(1 - q_0)^{m-1}$, where q_k is given by Eq. (4). Let $P_s{}^i$ again denote the probability of success of the i^{th} information packet in a busy period. This success occurs if either (i) the $(i - 1)$st information slot is unsuccessful and there is only one packet arrival, or (ii) the $(i - 1)$st information slot is successful and there is only one packet arrival during it (thus causing the busy period to continue) and no packet arrives during the acknowledgment slot. Thus, we have

† This is only possible under the assumption that all terminals are within range and in line-of-sight of each other.

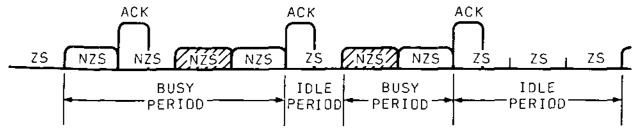


Figure 2. Slotted ALOHA: Busy and idle periods with priority acknowledgment traffic.

$$P_s{}^1 = \frac{q_1}{1 - q_0} \tag{14}$$

$$P_s{}^i = (1 - P_s{}^{i-1})\frac{q_1}{1 - q_0} + P_s{}^{i-1}\frac{q_1}{1 - q_0}q_0$$
$$= \frac{q_1}{1 - q_0} - q_1 P_s{}^{i-1} \qquad i > 1. \tag{15}$$

Let

$$c' = -q_1 \tag{16}$$

and let c be as defined in Eq. (7). We have

$$P_s{}^i = \frac{c}{1 - c'}[1 - (c')^i]. \tag{17}$$

A busy period contains information and acknowledgment packets. To obtain its length we need to compute $\bar{A}$, the expected time during a busy period occupied by acknowledgment traffic. This we do by first conditioning on m the number of information packets in a busy period and then removing the condition. Let $\bar{A}_m$ be the average time, during a busy period of m *information slots*, occupied by acknowledgment packets; we have

$$\bar{A}_m = \sum_{i=1}^{m-1} P_s{}^i = \frac{c}{1 - c'}\left[m - 1 - \frac{c'[1 - (c')^{m-1}]}{1 - c'}\right]. \tag{18}$$

(The summation is taken up to $m - 1$ since, according to our definition of the busy period, the acknowledgment corresponding to the last packet in the busy period is part of the idle period.) Still conditioned on m information slots during the busy period, we have

$$\bar{U}_m = \sum_{i=1}^{m} P_s{}^i = \frac{c}{1 - c'}\left[m - \frac{c'[1 - (c')^m]}{1 - c'}\right]. \tag{19}$$

Removing the condition on m, we get:

$$\bar{A} = \sum_{m=2}^{\infty} \bar{A}_m q_0 (1 - q_0)^{m-1}$$

$$= \frac{c}{1-c'} \left[\frac{1}{q_0} - 1 - \frac{c'}{1-c'} \left(1 - \frac{q_0}{1 - c'(1-q_0)} \right) \right]$$

$$= \frac{G}{1 + Ge^{-G} - Ge^{2G}} \tag{20}$$

and

$$\bar{U} = \sum_{m=1}^{\infty} \bar{U}_m q_0 (1 - q_0)^{m-1}$$

$$= \frac{c}{1-c'} \left[\frac{1}{q_0} - \frac{c'}{1-c'} \left(1 - \frac{q_0 c'}{1 - c'(1-q_0)} \right) \right]$$

$$= \frac{G}{(1 - e^{-G})(1 + Ge^{-G} - Ge^{-2G})}. \tag{21}$$

The throughput can then be written as:

$$S = \frac{\bar{U}}{\frac{1}{q_0} + \bar{A} + \frac{1}{1-q_0}} = \frac{Ge^{-G}}{1 + 2Ge^{-G}(1 - e^{-G})}.$$

Q.E.D.

In the case $\omega < 1$, it is possible to require that the terminals readjust the start of the slot following an acknowledgment; no channel time is then wasted. The analysis is similar to the one presented below for the 1-persistent CSMA mode; it is briefly sketched in Appendix A.

In Fig. 3 we plot the throughput S versus the channel traffic G for these systems. We show that the channel capacity drops down from $1/e = 0.368$ in the case of slotted ALOHA without acknowledgments, to 0.26 for slotted ALOHA with priority acknowledgment traffic ($\omega = 1$), and to only 0.14 for slotted ALOHA with non-priority acknowledgment traffic.

B. Carrier Sense Multiple Access with Priority Acknowledgment Traffic

In the carrier sense multiple access modes [2, 4], one attempts to avoid collisions by listening to (i.e., "sensing") the carrier due to other users' transmissions. Based on this information about the state of the channel, one may think of various actions to be taken by the terminal. Two protocols, introduced and analyzed in [2] and [4], will be considered in this study: the *nonpersistent* protocol and the *1-persistent* protocol.

In the nonpersistent CSMA protocol, a terminal with a packet ready for transmission, senses the channel and operates as follows. If the channel is sensed idle, it transmits the packet. If the channel is sensed busy, then the terminal schedules the retransmission of the packet to some later time according to the retransmission delay distribution. At this new point in time, it senses the channel and repeats the algorithm described.

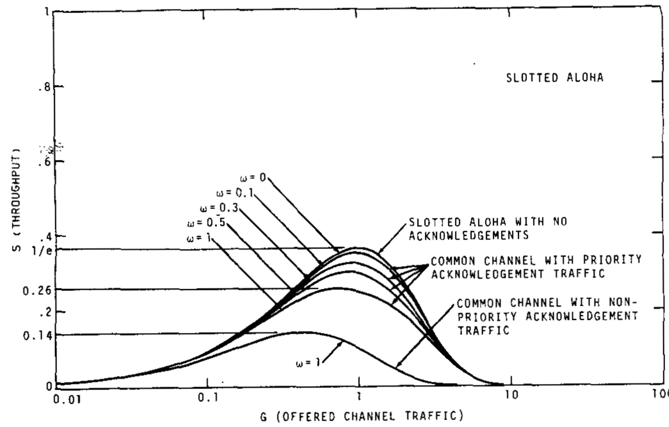


Figure 3. Slotted ALOHA: Throughput versus channel traffic.

The (S, G) relationship in the absence of acknowledgments has been shown to be [2, 4]

$$S = \frac{Ge^{-aG}}{G(1 + 2a) + e^{-aG}}. \tag{22}$$

A slotted version can be considered in which the time axis is slotted and the slot size is τ seconds (the propagation delay). All terminals are synchronized and are forced to start transmission only at the beginning of a slot. The (S, G) relationship is given by [2, 4]

$$S = \frac{aGe^{-aG}}{1 + a - e^{-aG}}. \tag{23}$$

The 1-persistent CSMA protocol differs from the nonpersistent by the fact that, if the channel is sensed busy, then the terminal waits until the channel goes idle (i.e., persisting on listening) and then transmits the packet. Here again, a slotted version can be considered in much the same way as for the previous protocol. In the absence of acknowledgments, the (S, G) relationship for the 1-persistent CSMA is shown to be [2, 4]

$$S = \frac{G[1 + G + aG(1 + G + aG/2)]e^{-G(1+2a)}}{G(1 + 2a) - (1 - e^{-aG}) + (1 + aG)e^{-G(1+a)}} \tag{24}$$

and for the slotted version to be [2, 4]

$$S = \frac{G \exp\{-G(1 + a)\}[1 + a - \exp\{-aG\}]}{(1 + a)(1 - \exp\{-aG\} + a \exp\{-G(1 + a)\}}. \tag{25}$$

In addition to its good performance, a great advantage in CSMA lies in a much simpler implementation of priority traffic. To give priority to acknowledgment traffic, the system should operate as follows:

(i) If a terminal, with a packet ready for transmission, senses the channel idle, then the terminal transmits its packet τ seconds later if and only if the channel is still sensed idle.

(ii) If such a terminal senses the channel busy, then it follows the protocol in question (nonpersistent, 1-persistent), repeating step (i) whenever the channel is sensed idle.

(iii) All acknowledgment packets are transmitted immediately, without incurring the τ seconds delay.

We shall consider only the above priority scheme for the two CSMA protocols.

Nonpersistent CSMA: Common-Channel with Priority Acknowledgment (CCPA)

The (S, G) relationship for this scheme is given in terms of a and ω by

$$S = \frac{Ge^{-aG}}{G(1 + 3a) + [1 + G(\omega + a)]e^{-aG}}. \tag{26}$$

Proof:

Using the same approach and notation introduced in [2], we define here a transmission period (TP) as the period required for transmission and reception of a packet and its (possible) overlapping packets (see Fig. 4). A transmission period is successful if only one packet is transmitted; in this case its length is equal to $1 + 2a$.†† If a transmission period is unsuccessful due to a packet overlap, its length is $1 + 2a + Y$, where Y is the time separating the arrivals of the first and last packets in the transmission period. The distribution function for Y is given by [2, 4]

$$\begin{aligned} F_Y(y) \triangleq \Pr\{Y \leqslant y\} &= \Pr\{\text{no arrival occurs in an interval of length } a - y\} \\ &= \exp\{-G(a-y)\}, \quad (y \leqslant a) \end{aligned} \tag{27}$$

and its average is given by

$$\bar{Y} = a - \frac{1}{G}(1 - e^{-aG}). \tag{28}$$

An acknowledgment period is simply the time required for transmission and reception of an acknowledgment packet; its length is equal to $\omega + a$. An idle period is denoted by I and its average is given by $\bar{I} = 1/G$. A cycle is defined as the period of time separating the starting points of two consecutive transmission periods. Again let U denote the time during a cycle that the channel is used without conflicts. We have

$$\bar{U} = e^{-aG}. \tag{29}$$

A cycle is successful if no packets arrive during its first a seconds (i.e., with probability e^{-aG}); in this case its length is equal to $1 + 2a + \omega + a + I$. A cycle is unsuccessful with probability $1 - e^{-aG}$, and its length is equal to $1 + 2a + Y + I$. Thus, the average cycle time is equal to

$$\overline{\text{Cycle}} = 1 + 2a + \bar{Y} + (\omega + a)e^{-aG} + 1/G. \tag{30}$$

†† The term $2a$ accounts for (i) the initial τ seconds delay introduced by the priority scheme and (ii) the propagation delay incurred by the packet (included in the TP, since it is only τ seconds after completion of the transmission that the channel is sensed idle).

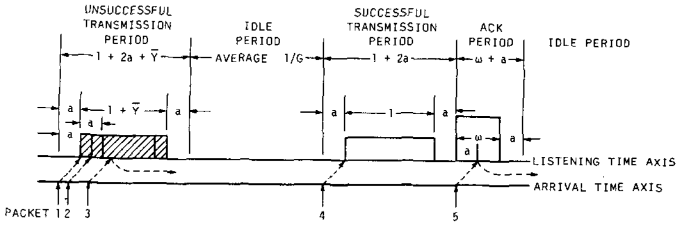


Figure 4. Nonpersistent CSMA: Common channel configuration with priority acknowledgment traffic.

Applying Eq. (3), which states that the channel utilization is simply the ratio of $\bar{U}$ to the average cycle length, we get Eq. (26).

Q.E.D.

Using similar techniques it is easy to prove that the throughput equation for the *slotted* nonpersistent CSMA with priority acknowledgment is

$$S = \frac{aGe^{-aG}}{(1 + 2a)(1 - e^{-aG}) + [(\omega + a)G + 1]ae^{-aG}}. \tag{31}$$

Indeed, the transmission period is always equal to $1 + 2a$. The idle period is geometrically distributed (its length equals k slots with probability $(1 - e^{-aG})e^{-kaG}$); its average length is given by

$$\bar{I} = \frac{ae^{-aG}}{1 - e^{-aG}}. \tag{32}$$

Similarly to Eq. (30), the average cycle length is given by

$$\overline{\text{Cycle}} = 1 + 2a + (\omega + a)\frac{aGe^{-aG}}{1 - e^{-aG}} + \frac{ae^{-aG}}{1 - e^{-aG}} \tag{33}$$

and $\bar{U}$ is given by

$$\bar{U} = \frac{aGe^{-aG}}{1 - e^{-aG}}. \tag{34}$$

Eq. (31) is obtained upon substituting $\bar{U}$ and $\overline{\text{Cycle}}$ into $S = \bar{U}/\overline{\text{Cycle}}$.

With $a = \tau W/b_m = 0.01$, we plot, in Fig. 5, S versus G for the nonpersistent CSMA without acknowledgment traffic (Eq. (22)) and for the nonpersistent CSMA with priority acknowledgment traffic (Eq. (26)) and various values of ω. We obtain similar curves for the slotted version as well as other values of a. In Fig. 6, we plot the capacity versus ω for various values of a. Clearly, the capacity should be and is a decreasing function of ω. The effect ω has on capacity is more noticeable with smaller values of a (which constitutes the more interesting range in ground radio), and the rate of degradation is more important in the range of smaller values of ω; this can simply be explained by the fact that the smaller is a, the higher is the achievable throughput, and therefore the higher is the rate of transmission of acknowledgment packets, which in turn is translated to a larger fraction of the channel bandwidth occu-

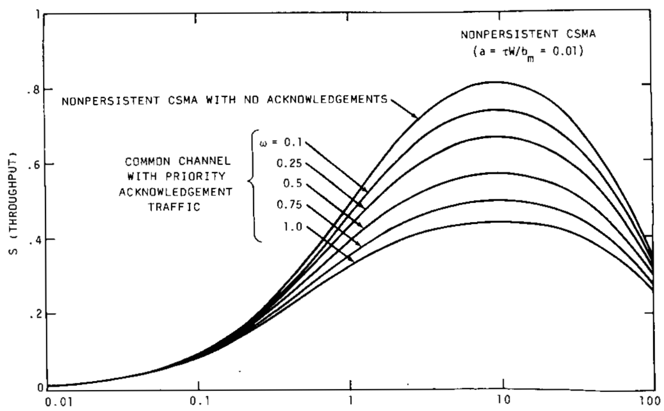


Figure 5. Nonpersistent CSMA: Throughput versus offered channel traffic.

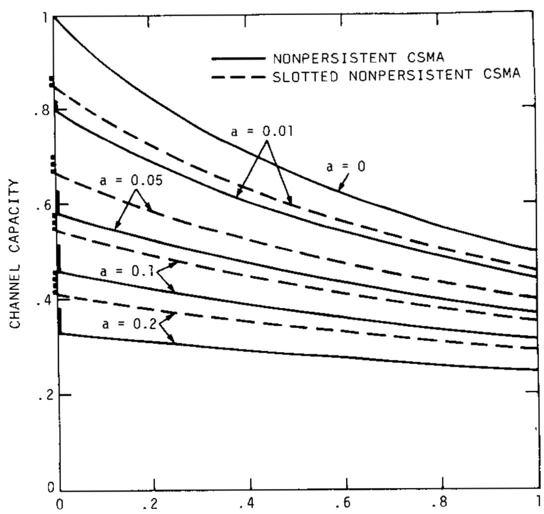


Figure 6. Nonpersistent CSMA: Capacity versus ω in common channel configuration.

pied by acknowledgment traffic. A change in ω is reflected by an important change in channel capacity; this effect is attenuated as ω increases since the channel capacity decreases with increasing values of ω as well. Moreover, the implementation of priority traffic is not achieved at zero cost; on the $\omega = 0$ vertical axis, we show this with heavy lines whose heights represent the loss in capacity due to that effect. As in [2], it is no surprise that the slotted version is consistently superior to the unslotted version.

1-Persistent CSMA: Common-Channel with Priority Acknowledgment Traffic (CCPA)

The throughput equation is given by Eq. (3) in which $\bar{U}, \bar{B}$, and $\bar{I}$ are defined in the following proof.

Proof:

In the present section, for the sake of simplicity, the analysis assumes a constant Y, denoted by $\tilde{Y}$. Two cases can be considered:

(1) $\tilde{Y} = a$ (yields a lower bound on throughput)

(2) $\tilde{Y} = \bar{Y} = a - (1/G)(1 - e^{-aG})$ (expected value of Y).

Furthermore, one can use an upper bound on the above average using

$$\tilde{Y} = \frac{a^2 G}{2}. \tag{35}$$

The upper bound is used for the numerical results. (In the slotted version Y is simply equal to a.)

Fig. 7 illustrates the behavior of the channel under the operating mode of this protocol. Also from the same figure, one can easily determine the boundaries which define a transmission period (TP) and an acknowledgment period (AP).

Under the approximations stated above, we have:

$$|\,\mathrm{TP}\,| = 1 + 2a + \tilde{Y} \tag{36}$$

$$|\,\mathrm{AP}\,| = \omega + a. \tag{37}$$

Every successful TP (denoted in the figure by S-TP) is followed by an AP. Thus, the channel looks like a sequence of TP's, AP's, and eventually some "quiet" periods (QP) as depicted in Fig. 8. At the end of each TP, some number of packets accumulate. The distribution of this number is Poisson with mean $G(1 + a + \tilde{Y})$.

Let

$$q_0 = e^{-G(1+a+\tilde{Y})} \tag{38}$$

$$q_1 = G(1 + a + \tilde{Y})e^{-G(1+a+\tilde{Y})}. \tag{39}$$

A TP at the end of which no packets accumulated is called a zero-TP (Z-TP). By the same token, we define a non-zero-TP (NZ-TP). Define here a *busy period* as the time lapse starting with the first TP following a Z-TP up to and including the first Z-TP encountered. The time separating two consecutive busy periods is called an *idle* period. Fig. 9 illustrates the various situations encountered for the idle periods. The distribution of the number of TP in a busy period is geometric with parameter $1 - q_0$, i.e.,

$$\Pr\{m \text{ TP's in a busy period}\} = q_0(1 - q_0)^{m-1}. \tag{40}$$

Let us first compute the probability of success over a TP. The existence of several situations encountered for an idle period renders the probability of success over the first TP in a busy period, denoted again by P_s^1, dependent on the way that busy period started. Let P_s^1 be, for the time being, an unknown. With arguments similar to those made in analyzing slotted ALOHA, we have for $i > 1$

$$P_s^i = (1 - P_s^{i-1}) \frac{q_1}{1 - q_0} e^{-aG} + P_s^{i-1} \frac{q_1}{1 - q_0} e^{-G(\omega+a)} e^{-aG}. \tag{41}$$

Let

$$c \triangleq \frac{q_1}{1 - q_0} e^{-aG} \tag{42}$$

$$c' = \frac{q_1}{1 - q_0} e^{-aG}(e^{-G(\omega+a)} - 1). \tag{43}$$

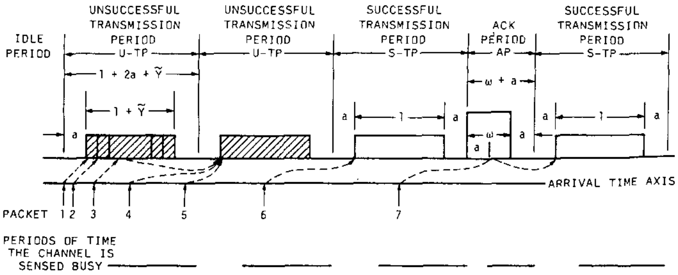


Figure 7. 1-Persistent CSMA: Common channel configuration with priority acknowledgment traffic.

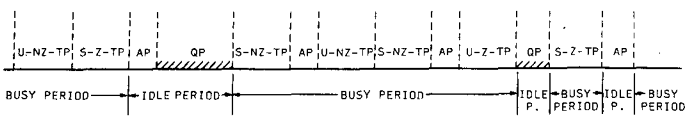


Figure 8. 1-Persistent CSMA: Busy and idle periods in common channel configuration with priority acknowledgment traffic.

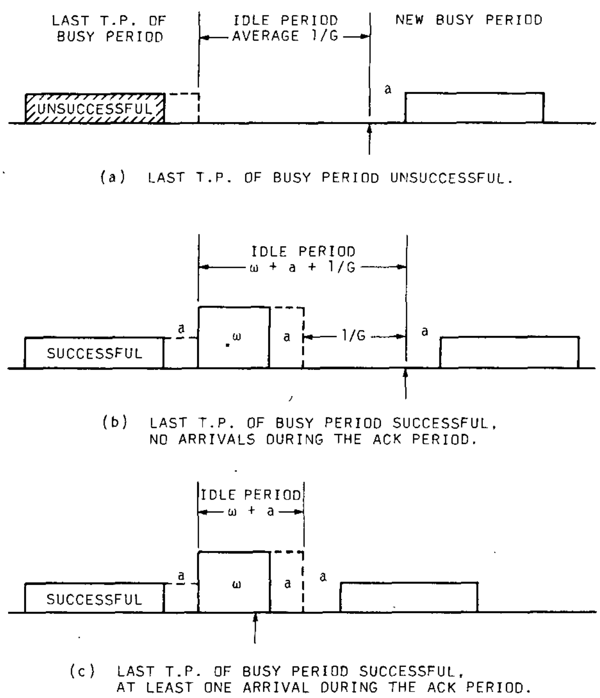


Figure 9. Various situations of idle periods in 1-persistent CSMA with priority acknowledgment traffic.

Eq. (41) reduces to

$$P_s^i = c + c'P_s^{i-1} \qquad i > 1 \tag{44}$$

whose solution is given in terms of P_s^1 by:

$$P_s^i = c\left[\frac{1-(c')^{i-1}}{1-c'}\right] + (c')^{i-1}P_s^1. \tag{45}$$

Let P_s^{last} denote the probability of success over the last TP in a busy period. We have

$$\begin{aligned} P_s^{\text{last}} &= \sum_{m=1}^{\infty} P_s^m q_0(1-q_0)^{m-1} \\ &= \frac{c}{1-c'}\left[1-\frac{q_0}{1-c'(1-q_0)}\right] + \frac{q_0}{1-c'(1-q_0)}P_s^1 \\ &= \mathrm{A} + \mathrm{B}P_s^1 \end{aligned} \tag{46}$$

where

$$\mathrm{B} = \frac{q_0}{1-c'(1-q_0)} \tag{47}$$

$$\mathrm{A} = \frac{c}{1-c'}[1-\mathrm{B}]. \tag{48}$$

By examining the three situations illustrated in Fig. 9, we can write another relation between P_s^1 and P_s^{last} as follows:

$$\begin{aligned} P_s^1 &= (1-P_s^{\text{last}})e^{-aG} + P_s^{\text{last}}[e^{-G(\omega+a)}e^{-aG} \\ &\quad + G(\omega+a)e^{-G(\omega+a)}e^{-aG}] \\ &= \mathrm{C} + \mathrm{D}P_s^{\text{last}} \end{aligned} \tag{49}$$

with

$$\mathrm{C} = e^{-aG} \tag{50}$$

$$\mathrm{D} = e^{-aG}[(1+aG+\omega G)e^{-G(\omega+a)} - 1]. \tag{51}$$

From Eq. (46) and Eq. (49), we solve for P_s^1

$$P_s^1 = \frac{\mathrm{C}+\mathrm{AD}}{1-\mathrm{DB}}. \tag{52}$$

Conditioning on m transmission periods in a busy period we now compute $\bar{U}$ and $\bar{A}_m$ as follows. For $m = 1$ we have

$$\bar{U}_1 = P_s^1 \tag{53}$$

$$\bar{A}_1 = 0. \tag{54}$$

For $m > 1$ we have

$$\bar{U}_m = P_s^1 + \sum_{i=2}^{m} P_s^i \tag{55}$$

$$\frac{\bar{A}_m}{(\omega+a)} = P_s^1 + \sum_{i=2}^{m-1} P_s^i. \tag{56}$$

Removing the condition on m and after a few derivations we get

$$\begin{aligned} \bar{U} &= \sum_{m=1}^{\infty} \bar{U}_m q_0(1-q_0)^{m-1} \\ &= P_s^1 + \frac{c}{1-c'}\left[\frac{(1-q_0)}{q_0} - \frac{c'}{1-c'}\left(1-\frac{q_0}{1-c'(1-q_0)}\right)\right] \\ &\quad + \frac{c'}{1-c'}P_s^1\left(1-\frac{q_0}{1-c'(1-q_0)}\right) \end{aligned} \tag{57}$$

$$\frac{\bar{A}}{\omega+a} = \sum_{m=2}^{\infty} \bar{A}_m q_0(1-q_0)^{m-1} = \bar{U}(1-q_0). \tag{58}$$

Since the average number of transmission periods is $1/q_0$, the average busy period is simply given by

$$\bar{B} = \frac{1}{q_0}(1 + 2a + \tilde{Y}) + \bar{A}. \tag{59}$$

The average idle period is obtained by examining Fig. 9 which leads to

$$\bar{I} = (1 - P_s^{\text{last}}) \frac{1}{G} + P_s^{\text{last}} \left[e^{-G(\omega+a)} \left(\omega + a + \frac{1}{G} \right) + (1 - e^{-G(\omega+a)})(\omega + a) \right] \tag{60}$$

where P_s^{last} is given by:

$$P_s^{\text{last}} = \frac{\text{A} + \text{BC}}{1 - \text{BD}}. \tag{61}$$

Under steady state conditions, the throughput is finally given by the ratio of $\bar{U}$ to the average cycle length $\bar{B} + \bar{I}$.

We plot S versus G in Fig. 10 and the capacity versus ω in Fig. 11. Note the similarity to the nonpersistent protocol; however, due to the lower channel capacity obtained with the 1-persistent CSMA, the sensitivity of the channel capacity to ω is not as important as with the nonpersistent CSMA, even for small a.

So far, we have derived the (S, G) relationship and the channel capacity for the common-channel configurations. In the following section we consider the split-channel configuration case.

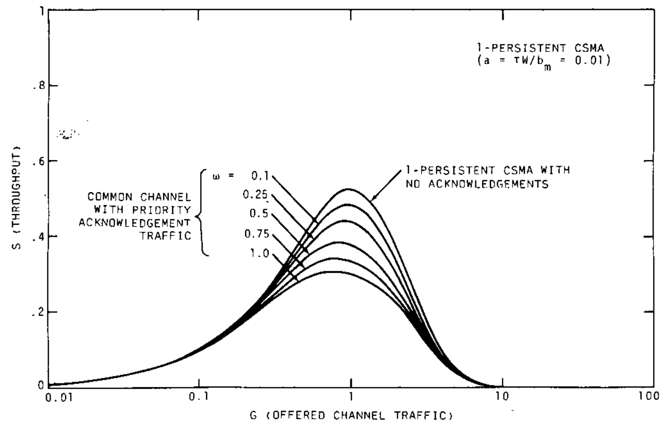


Figure 10. 1-Persistent CSMA: Throughput versus offered channel traffic.

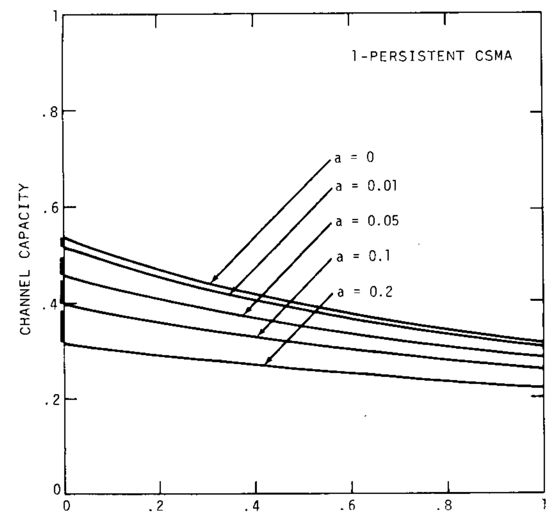


Figure 11. 1-Persistent CSMA: Capacity versus ω in common channel configuration.

III. SPLIT-CHANNEL CONFIGURATIONS (SC)–CHANNEL CAPACITY

In the split-channel configuration, the bandwidth is divided into two separate channels, a message channel and an acknowledgment channel. In one mode we consider that the receiver transmits the acknowledgment packet on the acknowledgment channel without delay. Under the assumption that the processing time needed to perform the error-check and to generate the acknowledgment is negligible, the transmission of the acknowledgment packet immediately follows a successful transmission. To guarantee success to acknowledgments, just sufficient bandwidth is provided to the acknowledgment channel so that overlap between any two consecutive acknowledgment packets is avoided. This mode will be referred to as the *split-channel real time* mode (SCRT). Another (perhaps hypothetical) mode of operation assumes the existence of a master control (the station, in a centralized environment) whose role is to perform the error checks on the transmitted packets and to issue the acknowledgment packets itself†††. Such a "station" can queue the acknowledgments, does not interfere with itself and requires less bandwidth for the acknowledgment channel than in the previous mode. It will be referred to as the *split-channel with queueing* mode (SCWQ) and is interesting mainly for comparative purposes. SCRT needs more bandwidth since it wastes some capacity due to idle periods on the acknowledgment channel.

††† Recall that the environment we are focusing on is a large population of terminals and other devices all within range. Furthermore, errors due to random noise are considered negligible compared to errors due to multiaccess interference. Moreover, this mode of operation is not so unrealistic if terminals are communicating only with the station.

Let W_m and W_a denote the bandwidth assigned to the message channel and acknowledgment channel respectively. We have

$$W = W_m + W_a. \tag{62}$$

Let $\theta \triangleq W_m/W$. Let a_m be the ratio of propagation delay to packet transmission time on the message channel. We have

$$a_m = \tau W_m/b_m = \theta a. \tag{63}$$

Let S_m be the normalized throughput on the random access message channel. We have

$$S_m = \lambda b_m/W_m = S/\theta. \tag{64}$$

We are interested in the normalized channel capacity, defined as the maximum attainable throughput, normalized to b_m/W.

It is well understood that at channel capacity, packet delays are infinite.

A. Slotted ALOHA

For SCRT, we require

$$W_a = \omega W_m. \tag{65}$$

This is achieved by taking

$$\theta = 1/(1+\omega). \tag{66}$$

The capacity of this system is

$$C_{\text{S-ALOHA-SCRT}} = \frac{C_{\text{S-ALOHA}}}{1+\omega} = \frac{1/e}{1+\omega}. \tag{67}$$

For SCWQ, we require in this capacity analysis that the bandwidth assigned to the acknowledgment channel be just sufficient to carry the acknowledgments. Therefore, W_a is related to W_m through

$$W_a = C_{\text{S-ALOHA}} \cdot \frac{W_m}{b_m} \cdot b_a = \omega W_m C_{\text{S-ALOHA}} \tag{68}$$

This is achieved by taking

$$\theta = 1\Big/\left(1+\frac{\omega}{e}\right). \tag{69}$$

The system capacity is then given by

$$C_{\text{S-ALOHA-SCWQ}} = \frac{1}{e+\omega}. \tag{70}$$

We plot in Fig. 12 the capacity of these systems versus ω, and compare it to the capacity of the CC configurations studied in Section II.A. We note that capacity-wise, SCWQ performs the best and that CCNPA performs the worst. When terminals readjust slot boundaries when an acknowledgment packet is transmitted, the capacity of CCPA is certainly not outperformed by SCWQ.

B. Carrier Sense Multiple Access

In slotted ALOHA, the normalized channel capacity for fixed b_m is independent of the bandwidth assignment θ; in CSMA, on the contrary, the capacity is a decreasing function of a_m [2, 4]. We denote this function by $C_{\text{CSMA}}(a_m)$. It is obtained by maximizing S with respect to G in Eqs. (22), (23), (24) and (25), depending on the protocol considered. As noted previously, $a_m = \theta a$. For SCRT we still require $\theta = 1/(1+\omega)$, so that the system capacity is simply given by

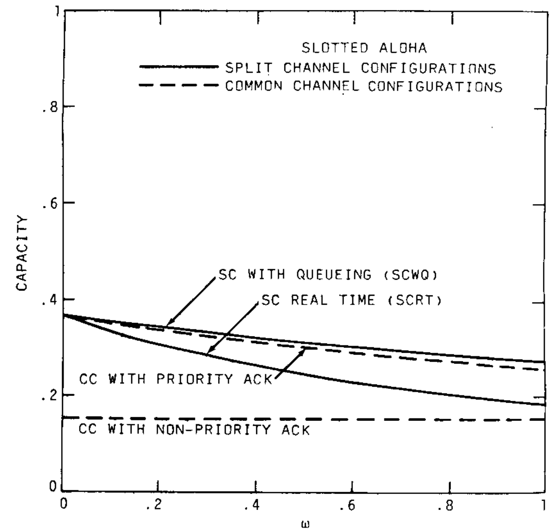


Figure 12. Slotted ALOHA: Capacity versus ω for common and split channel configurations.

$$C_{\text{CSMA-SCRT}}(a) = \frac{C_{\text{CSMA}}[a/(1+\omega)]}{1+\omega}. \tag{71}$$

For SCWQ, θ is solved for iteratively in the following equation:

$$\theta = 1/[1+\omega C_{\text{CSMA}}(\theta a)] \tag{72}$$

and the system capacity is then given by

$$C_{\text{CSMA-SCWQ}}(a) = \frac{C_{\text{CSMA}}(\theta a)}{1+\omega C_{\text{CSMA}}(\theta a)}. \tag{73}$$

For the slotted nonpersistent CSMA we plot, in Fig. 13, the capacity versus ω for CCPA, SCRT and SCWQ. It is interesting to note that for small a ($a = 0.01$) the three systems are, capacity-wise, basically equivalent. For larger a, the discrepancy among the various schemes is more important to the disadvantage of CCPA which uses the "a" slot overhead in implementing the priority scheme.

IV. CONCLUSION

In this paper we studied the effect on channel performance of the overhead created by the error-control traffic (i.e., acknowledgments), for both slotted ALOHA and CSMA. Common-channel and split-channel configurations were analyzed and compared with respect to channel capacity. Our major conclusions are as follows. In the common-channel configurations, priority acknowledgment schemes (CCPA) perform significantly better than non-priority schemes (CCNPA), as indicated in Fig. 13 for slotted ALOHA. Moreover, the implementation of priority traffic in CSMA is very simple. The hypothetical split-channel configuration with queueing (SCWQ) is certainly best capacity-wise since it wastes no channel bandwidth. However, for slotted ALOHA, CCPA with adjustable slot boundaries (accommodating the shorter acknowledgment packets) is certainly, for all practical purposes, as good as

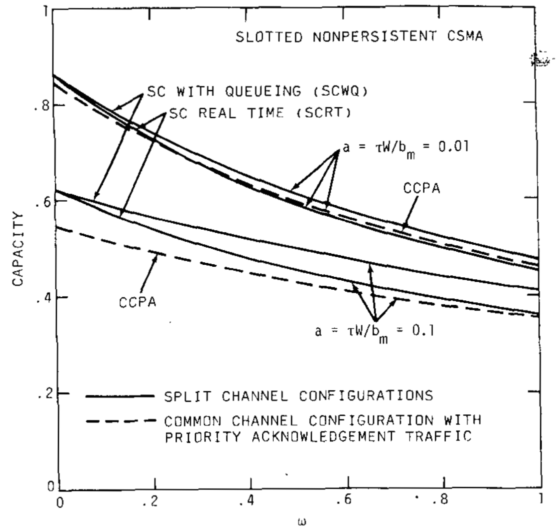


Figure 13. Nonpersistent CSMA: Capacity versus ω for common and split channel configurations.

SCWQ. The same conclusion prevails for CSMA except when we are in the presence of systems with a large a for which CCPA loses slightly due to the particular implementations of priority used here. Finally we note that the ratio ω of acknowledgment length to message packet length is an important system parameter affecting channel capacity, and its value must be kept as small as possible.

APPENDIX A

SLOTTED ALOHA: CCPA, $\omega < 1$

By allowing terminals to readjust slot boundaries following the transmission of an acknowledgment packet, several situations (similar to those depicted in Fig. 9 for 1-persistent CSMA) are encountered for an idle period which (as with 1-persistent CSMA) render P_s^1 dependent on P_s^{last} and on the way the busy period starts. With arguments similar to those made in section (II.B) for 1-persistent CSMA, Eq. (15) is now written as

$$P_s^i = (1 - P_s^{i-1}) \frac{q_1}{1-q_0} + P_s^{i-1} \frac{q_1}{1-q_0}(1 - e^{-\omega G})$$
$$= c + c' P_s^{i-1} \tag{A.1}$$

with

$$c = \frac{q_1}{1-q_0} \tag{A.2}$$

$$c' = -\frac{q_1}{1-q_0}(1 - e^{-\omega G}). \tag{A.3}$$

The solution of P_s^i in terms of P_s^1 is given by Eq. (45); P_s^{last} is expressed in terms of P_s^1 as in Eqs. (46) through (48) where c and c' are as defined in (A.2) and (A.3) above. On the other hand, as in Eq. (49) we have

$$P_s^1 = (1 - P_s^{\text{last}}) \frac{q_1}{1-q_0} + P_s^{\text{last}} \left[e^{-\omega G} \frac{q_1}{1-q_0} + \omega G e^{-\omega G} \right]$$
$$= \mathrm{C} + \mathrm{D} P_s^{\text{last}} \tag{A.4}$$

with

$$\mathrm{C} = \frac{q_1}{1-q_0} \tag{A.5}$$

$$\mathrm{D} = \omega G e^{-\omega G} - \frac{q_1}{1-q_0}(1 - e^{-\omega G}). \tag{A.6}$$

P_s^1 and P_s^{last} are thus expressed as in Eq. (52) and Eq. (61) respectively. $\bar{U}$ and $\bar{A}/\omega$ are given by Eq. (57) and Eq. (58) respectively. The average busy and idle periods are given by

$$\bar{B} = \frac{1}{q_0} + \bar{A} \tag{A.7}$$

$$\bar{I} = (1 - P_s^{\text{last}}) \frac{1}{1-q_0} + P_s^{\text{last}} \left[e^{-\omega G} \left(\omega + \frac{1}{1-q_0} \right) + (1 - e^{-\omega G})\omega \right]. \tag{A.8}$$

The throughput is finally obtained as the ratio of $\bar{U}$ to the average cycle length $\bar{B} + \bar{I}$. It can easily be checked that with $\omega = 0$ we get $S = Ge^{-G}$ (slotted ALOHA without acknowledgment traffic) and with $\omega = 1$, the throughput equation reduces to Eq. (13) obtained in section (II.A).

REFERENCES

[1] L. Kleinrock and S. Lam, "Packet-switching in a slotted satellite channel," in *Nat. Comput. Conf., AFIPS Conf. Proc.*, vol. 42. Montvale, N.J.: AFIPS Press, 1973, pp. 703-710.

[2] L. Kleinrock and F. Tobagi, "Packet switching in radio channels: Part I–Carrier sense multiple access modes and their throughput delay characteristics," *IEEE Trans. Commun.*, vol. COM-23, pp. 1400-1416, Dec. 1975.

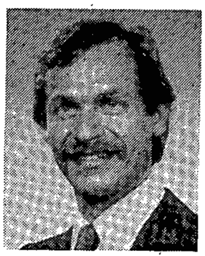
[3] L. Roberts, "ARPANET Satellite System," Notes 8 (NIC Document 11290) and 9 (NIC Document 11291), available from the ARPA Network Information Center, Stanford Research Institute, Menlo Park, Calif.

[4] F. Tobagi, "Random access techniques for data transmission over packet switched radio networks," Ph.D. dissertation, Comput. Sci. Dep., School of Eng. and Appl. Sci., Univ. of California, Los Angeles, rep. UCLA-ENG 7499, Dec. 1974.

[5] F. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part II–The hidden terminal problem in carrier sense multiple access and the busy tone solution," *IEEE Trans. Commun.*, vol. COM-23, pp. 1417-1433, Dec. 1975.

[6] F. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part III–Polling and (dynamic) split-channel reservation multiple access," *IEEE Trans. Commun.*, vol. COM-24, pp. 832-844, Aug. 1976.

★



Fouad A. Tobagi (M'77) was born in Beirut, Lebanon on July 18, 1947. He received the Engineering Degree from Ecole Centrale des Arts et Manufactures, Paris, France, in 1970 and the M.S. and Ph.D. degrees in computer science from the University of California, Los Angeles, in 1971 and 1974, respectively.

From 1971 to 1974 he was with the University of California, Los Angeles, where he participated in the ARPA Network Project as a Postgraduate Research Engineer and did research on packet radio communication. During the summer of 1972 he was with the Communications Systems Evaluation and Synthesis Group, IBM J. Watson Research Center, Yorktown Heights, N.Y. Since December 1974 he has been a Research Staff Project Manager with the ARPA project, Computer Science Department, University of California, Los Angeles working on the analysis and measurements of packet radio systems. His current research interests include computer communication networks, and packet switching over radio and satellite networks. From 1967 to 1970 he held a scholarship from the Ministry of Foreign Affairs of the French government. During the academic year 1972-1973 he held an Earl Anthony Fellowship. In July 1978, Dr. Tobagi will join the faculty of the Department of Electrical Engineering at Stanford University, Stanford, California.

Leonard Kleinrock (S'55–M'64–SM'71–F'73) received the B.E.E. degree from City College of New York, New York, NY in 1957 and the M.S.E.E. and Ph.D.E.E. degrees from Massachusetts Institute of Technology, Cambridge, in 1959 and 1963, respectively.

He was an Assistant Engineer at the Photobell Co. in New York from 1951 to 1957 and a Staff Associate at the MIT Lincoln Laboratory from 1957–1963. In 1963 he joined the faculty of the School of Engineering and Applied Science, University of California, Los Angeles, where he is now Professor of Computer Science. His research spans the fields of computer networks, computer systems modeling and analysis, queueing theory and resource sharing, and allocation in general. At UCLA, he directs a large group in advanced teleprocessing systems and computer networks. He is the author of three major books in the field of computer networks: *Communication Nets: Stochastic Message Flow and Delay* (New York: McGraw Hill, 1964; also New York: Dover, 1972); *Queueing Systems, Vol. I: Theory* (New York: Wiley-Interscience, 1975); and *Queueing Systems, Vol. II: Computer Applications* (New York: Wiley-Interscience, 1976). He has published over 85 articles and contributed to several books. He serves as consultant for many domestic and foreign corporations and governments and he is a referee for numerous scholarly publications and a book reviewer for several publishers.

Dr. Kleinrock was awarded a Guggenheim Fellowship for 1971–1972 and is an IEEE Fellow "for contributions in computer-communication networks, queueing theory, time-shared systems, and engineering education." He was the co-winner of the 1976 Lanchester Prize for his book *Queueing Systems, Vol. II: Computer Applications* (J. Wiley, 1976).