THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 61, No. 8, October 1982
Printed in U.S.A.

An Analysis of the Carrier-Sense Multiple-Access Protocol

By D. P. HEYMAN

(Manuscript received February 16, 1982)

In this paper we analyze the throughput and delay characteristics of the Carrier-Sense Multiple-Access protocol with a queueing model. The effects of finite buffer size, bursty arrivals, and collision detection with exponential rescheduling are examined.

The Carrier-Sense Multiple-Access protocol could be used on a bus in a packet switch. It works by sending a signal to all sources when the bus is occupied. A source postpones transmission to the bus when this signal is heard. Since the signal takes a positive amount of time to reach the sources, two or more sources occasionally will attempt to use the bus at about the same time. When this occurs, all the packets are destroyed and rescheduled.

We conclude that: (i) it is important to choose the mean rescheduling time correctly, and (ii) performance degrades significantly when compound Poisson arrivals (peaked traffic) replace Poisson arrivals (smooth traffic).

I. INTRODUCTION

The Carrier-Sense Multiple-Access (CSMA) protocol has been proposed for resolving conflicts when several sources attempt to use a single channel. In this paper we investigate the throughput and delay characteristics of CSMA.

1.1 Background

When several sources attempt to use a single channel, a protocol (in this context, protocol is synonymous with queue discipline) is required to allocate the channel among the sources. When the channel is occupied, a signal is sent indicating that state of affairs. When a source wants to use the channel, it first listens for this signal. If the signal is not heard, the source starts transmitting. If the signal is heard, the source postpones transmitting and tries again at another time. The advantage of this system is that a device to control the sources is not required. The disadvantage is that occasionally messages will be destroyed because two sources will transmit at the same time. This is a consequence of the fact that signals travel at finite speeds, so there is a delay between the epoch when one source seizes an idle channel and the epochs when the other sources can first hear the busy-channel signal. Thus, soon after a source seizes the channel, another source may sense that the channel is free even though the channel is busy. When this occurs, both messages are destroyed because their bits have been merged. At the end (earlier if collision detection is used) of each transmission, the source determines whether the transmission was successful. If it was, the source goes about its business; if it was not, each message is rescheduled as if the channel were sensed as busy.

The CSMA protocol is used in the Ethernet* local-area distribution system and has been considered for part of a digital switch. Since CSMA is envisioned as a protocol for packet networks, we will refer to the arrivals as packets and the channel as a bus.

1.2 Relation to other work

The first study of CSMA is Kleinrock and Tobagi.[1] In addition to the version of CSMA studied here, which they call nonpersistent CSMA, they investigated various forms of persistent CSMA and slotted CSMA that are not alluded to in this paper. Their paper implicitly assumes that the rescheduling delay is infinite. This means that packets which find the bus occupied or which are destroyed in a collision are abandoned. For the most part, we assume, as their paper does, that packets arrive according to a Poisson process. In a subsequent paper by Kleinrock and Tobagi,[2] the rescheduling times are modelled as geometrically distributed random variables. They used a finite number of sources and assumed the arrivals to be quasi-random (i.e., finite-source Poisson). Also, they introduced an unnatural discretization of the time scale, which causes some small distortions in the results. A continuous version of this model is described and solved in Halfin.[3] Our analysis shares many features with the analysis found in Ref. 2, particularly the exploitation of the regeneration epochs.

Tobagi and Hunt[4] add collision detection to the model of Kleinrock and Tobagi.[2] Collision detection is a feature that informs a source that its packet has been destroyed soon after the collision occurs. Our model of collision detection is based on the model presented in Ref. 4.

Rappaport[5] and Rappaport and Bose[6] describe an elaborate model of a protocol similar to CSMA.

* Ethernet is a trademark of Xerox Corporation.

1.3 Summary of results

The models considered in this paper can be easily solved numerically. The following conclusions were reached by considering many numerical examples.

(*i*) It is important to choose the mean rescheduling time correctly; performance is not sensitive to changes in this number near its best value.

(*ii*) If the mean rescheduling time is based on a designed load but the realized load is different, the realized performance is almost as good as if the mean rescheduling time were based on the realized load.

(*iii*) When the traffic intensity is no larger than 0.7, the throughput is insignificantly lower than the traffic intensity.

(*iv*) Performance is significantly degraded when compound Poisson arrivals (peaked traffic) replace Poisson arrivals (smooth traffic).

(*v*) When the traffic intensity is no larger than 0.7, collision detection lowers the average waiting time significantly and reduces the sensitivity of the throughput to the mean rescheduling time. For higher traffic intensities, collision detection significantly increases the throughput and decreases the average delay.

1.4 Outline of this paper

Our model is described in detail in Section II, and the method of solution is outlined. The details of the solution are presented in Section III. Numerical examples and empirical conclusions from the model are given in Section IV. In Section V, we consider the effects of bursty traffic by introducing compound Poisson arrivals. We return to Poisson arrivals in Section VI and examine the benefits of collision detection. Appendix A records the transition probabilities omitted from the text, and Appendix B lists the most important symbols used in the text.

II. MODEL AND OVERVIEW OF THE SOLUTION METHOD

Our setting is a queue of the $M/D/1/K$ type. Let λ be the rate of the Poisson arrival process. There is a single server with a constant service time per packet. At most K packets may be in the system (queue plus server) at any time. Packets that arrive when the system is full are lost forever and have no effect on the system. Our use of constant service times reflects the assumption that all packets have the same length.

The parameter K has two potential interpretations. One is to suppose there is a buffer that can hold $K-1$ packets, so the maximum number of packets in the queue is $K-1$. The other is to suppose there are K ports and a packet that cannot seize a port is lost. We will show that proper limiting probabilities will not exist without this bound on the number of packets in the system.

When a packet is on the bus (i.e., when it enters service), a signal is sent that warns other packets that the bus is occupied. (Equivalently, a signal indicating the bus is free is turned off when a packet is on the bus.) Let h denote the one-way propagation delay for this signal. It is common to all potential users. The propagation time between a source and the bus depends on the distance between them. If we calculate h for the source that is farthest from the bus, the assumption that h, the one-way propagation delay, is common to all users will underestimate the performance measures. This propagation delay has two effects. The first effect is that for the first h time units after a packet occupies a previously idle bus, other packets that arrive and want to use the bus will do so because the signal that the bus is occupied has not reached them. When this happens, we say that a "collision" has occurred and the first h time units of a service time are called the "vulnerable" period. The second effect is that the signal that the bus is free is not received until the bus has been free for h time units.

When a collision occurs, the bits in both packets are scrambled, and so both packets must be retransmitted. We assume that the collision is detected at the end of the unsuccessful transmission. Then each source involved in the collision draws a number from an exponential distribution with mean $1/\alpha$, which is the length of time until the source next attempts to use the bus. Furthermore, all sources that attempt to transmit a packet and hear the signal that the bus is occupied make their next attempt in this fashion. We call α the "retry rate."

Our objective is to express the throughput (equivalently, the asymptotic departure rate or the asymptotic proportion of time that the bus is successfully transmitting packets) and average waiting time of a packet (i.e., the departure time minus the arrival time) to the system parameters λ, K, h, and α.

2.1 Outline of solution

Choose time units so that the processing time of a packet is unity. Consider a packet, $\mathscr{P}$, that gets on an empty bus at time t. This packet will relinquish the server at $t + 1$ and the signal that $\mathscr{P}$ is on the bus will cease being sensed at some random time τ with $t + 1 + h \leq \tau \leq t + 1 + 2h$. If no packets arrive during the vulnerable period, then $\tau = t + 1 + 2h$.

Suppose we assume that $\mathscr{P}$ holds the bus for a (nonrandom) length of time ν, with $1 \leq \nu \leq 1 + 2h$, at the end of which $\mathscr{P}$ relinquishes the server and the signal that $\mathscr{P}$ is on the bus ceases. Retries that occur during the vulnerable period destroy the messages involved in the collision. These retries are not considered to have seized the bus.

When $\nu = 1$, these assumptions will provide an upper bound for the throughput and a lower bound for the average number of customers present (in a buffer or on the bus) in the steady state. When $\nu = 1 + 2h$, a lower bound for the throughput and an upper bound for the average number of customers present is obtained. Little's theorem (the queueing formula $L = \lambda W$) implies that setting $\nu = 1$ provides an upper bound for the average waiting time and that setting $\nu = 1 + 2h$ provides a lower bound. Therefore, we will solve a model with constant service times ν.

Since h is typically much smaller than one (we use $h = 0.01$ in our examples), we expect that the bounds will be close together. That is the case in our numerical examples. Therefore, either of the bounds is a good approximation to the true measure of performance.

Let $X(t)$ be the number of packets present at time t. We want to obtain

$$p_i = \lim_{t \to \infty} P\{X(t) = i\} \tag{1}$$

for each $i = 1, 2, \cdots, K$. These "limiting probabilities" are easily shown to be independent of the distribution of the initial state $X(0)$. To obtain $\{p_i\}$, we embed a discrete Markov chain at customer "ejection" epochs, which are defined as follows. Let C_m be the epoch where the mth packet to get on an empty bus leaves the bus. This may be the end of a successful transmission where the packet leaves the system or it may be the end of a destroyed transmission where the packet rejoins the queue. To encompass both types of events, we call C_m the mth ejection epoch.

Let $Y_m = X(C_m^+) \triangleq \lim_{s \downarrow 0} X(C_m + s)$; it represents the number of customers present just after C_m. Since the arrivals are Poisson and retrials are governed by an exponential distribution, it is easy to see that $\{Y_m;\, m = 0, 1, \cdots\}$ is a Markov chain. It is also easy to see that this chain is irreducible and aperiodic (details are omitted) so that the limits

$$\pi_i \triangleq \lim_{m \to \infty} P\{Y_m = i\}, \qquad i = 0, 1, \cdots, K,$$

exist and are independent of Y_0, are positive, and sum to one.

Three steps are used to obtain $\{p_i\}$ from $\{\pi_i\}$. The first step is to relate $\{\pi_i\}$ to $\{p_i'\}$ where

$$p_i' \triangleq \lim_{n \to \infty} P\{Y_n = i \text{ and transmission is successful}\},$$

$$i = 0, 1, \cdots, K - 1. \tag{2}$$

The relationship is obtained by calculating the state-dependent probability of having a collision. The second step is to use a "rate up equals rate down" argument to show

$$\lambda p_i = \zeta p_i', \qquad i = 0, 1, \cdots, K - 1, \tag{3}$$

where ζ is the ejection rate. The third step is to obtain p_0 by a renewal-reward argument. Then eq. (3) is used to calculate ζ and the remaining p_i's.

III. SOLVING THE MODEL

In this section we give details of the solution of the model described in Section II.

3.1 Collision probabilities

Let c_n be the probability that n packets arrive in an interval of length, ν; then

$$c_n = e^{-\lambda\nu}\frac{(\lambda\nu)^n}{n!}, \qquad n = 0, 1, \cdots.$$

A well-known property of Poisson arrival processes (e.g., Corollary 5-13 in Heyman and Sobel[7]) is that the arrival epochs are iid and uniform over $[0, \nu]$ when it is given that $n > 0$ arrivals occurred during $[0, \nu]$. Thus, for a service interval that starts with i packets in the queue, we have

$$\begin{aligned}\eta_n &\triangleq \text{P}\{\text{no arrivals prior to } h \,|\, n \text{ arrivals in a} \\ &\qquad \text{service interval that starts with } i \text{ in queue}\} \\ &= \left(\frac{\nu - h}{\nu}\right)^n, \qquad \text{n} = 0, 1, \cdots \text{ and any } i. \end{aligned} \tag{4}$$

Use the memoryless property of the exponential distribution to obtain

$$\begin{aligned}\delta_i &\triangleq \text{P}\{\text{no retries prior to } h \,|\, n \text{ arrivals in a} \\ &\qquad \text{service that starts with } i \text{ in queue}\} \\ &= e^{-i\alpha h}, \qquad i = 0, 1, \cdots \text{ and any } n. \end{aligned} \tag{5}$$

Since a collision is avoided if, and only if, there are no arrivals and no retries during the vulnerable period, we have

$$\begin{aligned}\bar{d}_n(i) &\triangleq P\{\text{no collision and } n \text{ arrivals in a service} \\ &\qquad \text{interval} \,|\, \text{start with } i \text{ in queue}\} \\ &= \eta_n c_n \delta_i, \; i, n \geq 0, \end{aligned} \tag{6}$$

and

$$\begin{aligned}d_n(i) &\triangleq P\{\text{collision and } n \text{ arrivals in a service} \\ &\qquad \text{interval} \,|\, \text{start with } i \text{ in queue}\} \\ &= (1 - \eta_n \delta_i) c_n, \; i, n \geq 0. \end{aligned} \tag{7}$$

Equations (6) and (7) are used to calculate the transition probabilities of the embedded Markov chain $\{Y_m; m = 0, 1, \cdots\}$.

3.2 Transition probabilities

Recall from Section 2.1 that Y_m is the number of packets in the system just after the mth ejection epoch. All of these packets must be in the queue because of the propagation delay in broadcasting that the bus is available. Let

$$p_{ij} \triangleq P\{Y_{m+1} = j \mid Y_m = i\}. \tag{8}$$

In this section we present formulas for computing p_{ij}, $0 \le i, j \le K$. It will be convenient to call both exogenous packet arrivals and retries "arrivals"; the former are "outside" arrivals. Then, for $2 \le i \le K - 1$ and $0 \le i + n \le K - 2$, we may write

$$\begin{aligned} p_{i,i+n} &= P\{\text{next arrival is from outside} \mid i \text{ in queue}\} \\ &\times [P\{n \text{ outside arrivals during service and no collision} \mid \text{service starts with } i \text{ in queue}\} \\ &+ P\{n-1 \text{ outside arrivals during service and a collision} \mid \text{service starts with } i \text{ in queue}\}] \\ &+ P\{\text{next arrival is a retry} \mid i \text{ in queue}\} \\ &\times [P\{n+1 \text{ outside arrivals during service and no collision} \mid \text{service starts with } i-1 \text{ in queue}\} \\ &+ P\{n \text{ outside arrivals during service and a collision} \mid \text{service starts with } i-1 \text{ in queue}\}]. \end{aligned}$$

Use the memoryless property of the exponential distribution to obtain

$$P\{\text{next arrival is from outside} \mid i \text{ in queue}\} = \frac{\lambda}{\lambda + i\alpha}.$$

Set $\beta = \lambda/\alpha$ for notational simplicity; then

$$p_{i,i+n} = \frac{\beta}{i+\beta}[\bar{d}_n(i) + d_{n-1}(i)] + \frac{i}{i+\beta}[\bar{d}_{n+1}(i-1) + d_n(i-1)]. \tag{9}$$

To obtain the remaining entries of (p_{ij}) we need to account for boundary conditions. The details and the formulas are given in Appendix A.

It is easy to see that (p_{ij}) is irreducible and aperiodic. Therefore, there is a unique stationary distribution that is also the limiting distribution. Furthermore, since the continuous-time process $\{X(t); t \ge 0\}$ starts from scratch (i.e., it regenerates) each time the Y-process reaches zero, the X-process is regenerative and the mean time between regeneration points is finite. This implies the limits in eq. (1) exist. A similar argument establishes that the limits in eq. (2) exist.

We can use eq. (9) to explain why the finite capacity, K, is essential for our analysis. Suppose $K = \infty$; then eq. (9) is valid for $i \ge 2$ and every $n \ge 0$. Equation (9) is also valid for $n = -1$ if we set $d_j(i) = 0$ whenever $j < 0$. Define

$$g_i = \sum_1^{\infty} np_{i,i+n} - p_{i,i-1};$$

it is the expected size of a jump out of state i in the Y-process. Intuitively, if $\lim_{i\to\infty} g_i > 0$, the Y-process is tending to infinity and $\lim_{m\to\infty} P\{Y_m \le j\} = 0$ for all j. From eq. (9) we can calculate (details are omitted)

$$g_i = \lambda\nu + \frac{\beta(1 - e^{-i\alpha h}e^{-\lambda h})}{i+\beta} - \frac{i[e^{-\alpha(i-1)}(e^{-\lambda\nu} - e^{-\lambda h}) + e^{-\alpha i}e^{-\lambda\nu}]}{i+\beta};$$

and so, if $h > 0$,

$$\lim_{i\to\infty} g_i = \lambda\nu. \tag{10}$$

This phenomenon can be described physically. When j is large, $e^{-j\alpha h}$ is small so the probability that an ejection is a completion is very small. This means that once state j is reached (as it must be because the process is irreducible), the number of customers present grows monotonically with very high probability.

The corollary in Kaplan[8] states that an irreducible and aperiodic Markov chain for which eq. (10) holds is not ergodic if $p_{ij} = 0$ whenever $j < i - k$ for some k that is independent of i. At most, one departure can occur at a time, so $p_{ij} = 0$ whenever $j < i - 1$. Therefore, the Y-process is not ergodic when $K = \infty$.

3.3 Solving the balance equations

Instead of solving $\pi = \pi P$ to obtain $\{\pi_i\}$, we will solve balance equations between two sets of states. Let

$$p_i(\ge j) \triangleq P\{Y_{m+1} \ge j \mid Y_m = i\} = \sum_{k=j}^{K} p_{ik}.$$

In the steady state, the rate of transitions between states $\{0, 1, \cdots, i\}$ and states $\{i+1, i+2, \cdots, K\}$ must be the same in both directions, so*

$$\pi_{i+1}p_{i+1,i} = \sum_{k=0}^{i} \pi_k p_k(\ge i+1), \qquad i = 0, 1, \cdots, K-1. \tag{11}$$

We can solve eq. (11) by replacing π_i by x_i and setting $x_0 = 1$, and using eq. (11) to obtain x_{i+1} from $x_0, \cdots, x_i$. By setting

$$\pi_i = x_i \Big/ \sum_{j=0}^{K} x_j$$

we obtain a nonnegative solution of eq. (11) that sums to one.

---

* The basic idea was developed by Robert Morris and Eric Wolman. The details for Markov chains are given in Theorem 7–13 of Heyman and Sobel.[7]

3.4 Departure point probabilities

Recall that C_m is the mth ejection epoch. Let us write $C_m = D$ when C_m is a departure epoch. Now

$$P\{C_{m+1} = D,\ Y_{m+1} = j\}$$
$$= \sum_{i=0}^{K} P\{C_{m+1} = D,\ Y_{m+1} = j \,|\, Y_m = i\} P\{Y_m = i\}$$

for every j and m. We will see below that $P\{C_{m+1} = D,\ Y_{m+1} = j \,|\, Y_m = i\}$ does not depend on m, so denote it by q_{ij}. Then

$$p'_j = \lim_{m \to \infty} P\{C_{m+1} = D,\ Y_{m+1} = j\} = \sum_{i=0}^{K} q_{ij}\pi_i,$$
$$j = 0, 1, \cdots, K-1. \tag{12}$$

Observe that, for $1 \le j \le K-2$ and $1 \le i \le j-1$, $P\{C_{m+1} = D,\ Y_{m+1} = j \,|\, Y_m = i\}$

$= P$ {next arrival is from outside$|i$ in queue}
$\times\ P$ {$j - i$ outside arrivals during service and no collision|service starts with i in queue}
$+\ P$ {next arrival is a retry$|i$ in queue}
$\times\ P$ {$j - 1 + 1$ outside arrivals during service and no collision|service starts with $i - 1$ in queue}

$$= \frac{\beta}{\beta + i}\,\bar{d}_{j-i}(i) + \frac{i}{\beta + i}\,\bar{d}_{j-i+1}(i-1).$$

The remaining entries are obtained by accounting for boundary behavior. The results are:

$$q_{00} = c_0\eta_0 = c_0, \tag{13a}$$

$$q_{10} = \frac{c_0}{1+\beta}, \tag{13b}$$

and

$$q_{i0} = 0 \quad \text{for} \quad i \ge 2. \tag{13c}$$

For $1 \le j \le K-2$,

$$q_{0j} = \bar{d}_j(0), \tag{14a}$$

$$q_{ij} = \frac{\beta}{\beta + i}\,\bar{d}_{j-i}(i) + \frac{i}{\beta + i}\,\bar{d}_{j-i+1}(i-1), \qquad 1 \le i \le j, \tag{14b}$$

$$q_{j+1,j} = \frac{j+1}{\beta + j + 1}\,\bar{d}_0(j), \tag{14c}$$

and

$$q_{ij} = 0, \qquad i \geq j + 2. \tag{14d}$$

When $j = K - 1$ we obtain

$$q_{0,K-1} = \sum_{K-1}^{\infty} \bar{d}_n(0), \tag{15a}$$

$$q_{i,K-1} = \frac{\beta}{\beta+1} \sum_{K-1-i}^{\infty} \bar{d}_n(i) + \frac{i}{\beta+i} \sum_{K-i}^{\infty} \bar{d}_n(i-1),$$

$$1 \leq i \leq K - 1 \tag{15b}$$

$$q_{K,K-1} = \delta_{K-1}.$$

In eq. (46) we show that the infinite sums in eq. (15) have representations as finite sums containing no more than K terms. Since $C_{m+1} = D$ and $Y_{m+1} = K$ cannot occur simultaneously, $q_{iK} = 0$ for every i.

Notice that (the subscript ∞ denotes a limit)

$$n_c \triangleq \sum_{0}^{K-1} p_j' = P\{C_\infty = D,\ Y_\infty \leq K - 1\}$$

$$= P\{\text{an ejection epoch is a departure epoch}\}$$

$$= P\{\text{no collision}\}, \tag{16}$$

where the last two probabilities are steady-state quantities and can be interpreted as long-run proportions.

3.5 Converting $\{p_i'\}$ to $\{p_i\}$

Let $E(t)$ be the number of ejections by time t. Since $\{E(t);\ t \geq 0\}$ regenerates when an ejection leaves the system empty (i.e., when $Y_m = 0$ for some m),

$$\zeta \triangleq \lim_{t\to\infty} E(t)/t$$

exists. It is the ejection rate. It is also the rate at which packets gain access to an idle bus, so ζ is the arrival rate of new and rescheduled packets, which is denoted by G in Kleinrock and Tobagi.[1]

The rate at which the number of packets present jumps from i to $i + 1$ is λp_i. The rate at which it jumps from $i + 1$ to i is $\zeta p_i'$. In the steady state these rates must be equal so we have

Lemma 1: $\lambda p_i = \zeta p_i', \qquad i = 0, 1, \cdots, K - 1.$ (17)

Lemma 1 can be proved rigorously. The proof uses standard methods and is omitted.

There are $K + 1$ unknowns in eq. (17) and K equations. We will find

p_0 by an independent argument; the remaining p_i's and ζ are obtained from eq. (17).

Let T_0 be the amount of time $X(\cdot)$ is zero in an arbitrary cycle of the X-process, and let M be the expected length of a regeneration cycle. A basic property of regenerative processes (see, e.g., Theorem 6–7 in Heyman and Sobel[7]) is that

$$p_0 = E(T_0)/M. \tag{18}$$

Since $X(t) = 0$ if, and only if, t is in an idle period,

$$E(T_0) = 1/\lambda. \tag{19}$$

When $Y = i$, let ψ_i be the average time between en ejection epoch and the next ejection epoch. We have

$$\psi_i = \begin{cases} \nu + \dfrac{1}{i\alpha + \lambda} & \text{if} \quad i < K \\ \nu + \dfrac{1}{K\alpha} & \text{if} \quad i = K, \end{cases} \tag{20}$$

because ψ_i is ν plus the expected time to the next arrival after the ejection epoch. Let m_i be the mean number of visits of the Y-process to state i during an arbitrary regeneration cycle of the X-process. Thus, m_i is the mean number of visits of the Y-process to state i between visits of the Y-process to state zero, because the X-process regenerates whenever the Y-process enters state zero. It is easy to show (see Exercise 7-78 in Heyman and Sobel[7]) that

$$m_i = \pi_i/\pi_0, \qquad i = 0, 1, \cdots, K. \tag{21}$$

Using eqs. (20) and (21) we obtain

$$\begin{aligned} M = \sum_{i=0}^{K} \psi_i m_i &= \frac{1}{\pi_0}\left[\sum_{0}^{K-1} \pi_i \left(\nu + \frac{1}{i\alpha + \lambda}\right) + \pi_K \left(\nu + \frac{1}{K\alpha}\right)\right] \\ &= \frac{1}{\pi_0}\frac{1}{\lambda}\left(\rho + \sum_{0}^{K-1} \frac{\beta\pi_i}{\beta + i} + \frac{\beta\pi_K}{K}\right), \end{aligned} \tag{22}$$

where $\rho = \lambda\nu$.

Substituting eqs. (19) and (22) into (18) yields

$$p_0 = \pi_0 \bigg/ \left(\rho + \sum_{0}^{K-1} \frac{\beta\pi_i}{\beta + i} + \frac{\beta\pi_K}{K}\right). \tag{23}$$

Obviously [and formally from eqs. (12) and (13)], we find that

$$p_0' = \pi_0. \tag{24}$$

Combining eq. (17) with $i = 0$ and eqs. (23) and (24) yields

$$\zeta = \lambda p_0/p_0' = \lambda \bigg/ \left(\rho + \sum_0^{K-1} \frac{\beta\pi_i}{\beta + i} + \frac{\beta\pi_K}{K}\right). \tag{25}$$

Equation (25) shows that ζ can be computed when the balance equations are solved.

3.6 The throughput, occupancy, and average waiting time

The throughput, θ, is the asymptotic departure rate. Since the departure process regenerates whenever the ejection process regenerates, θ is well defined. The asymptotic departure rate equals the asymptotic rate at which packets are accepted, because all accepted packets eventually depart and the number of packets present is no larger than K; therefore

$$\theta = \lambda \sum_0^{K-1} p_i. \tag{26}$$

Combining eq. (26) with eqs. (16) and (17) yields

$$\theta = \zeta n_c. \tag{27}$$

This equation states that the departure rate equals the ejection rate multiplied by the asymptotic proportion of ejections that do not suffer a collision. Equation (27) illuminates the essential trade-offs involved in using CSMA. One expects that ζ increases and n_c decreases as λ and α increase. This means that for each λ, there is a value of α that maximizes θ. Let $\theta^*(\lambda)$ be the largest value of θ that can be achieved when λ is specified. We also expect that $\theta^*(\lambda)$ will first increase with λ and then start decreasing.

Let ϕ be the long-run proportion of time that the bus is occupied. The regenerative arguments used above can be used to prove that ϕ is well defined and that

$$\phi = \nu \sum_0^K m_i/M. \tag{28}$$

Using eqs. (22) through (25) in conjunction with eq. (28) yields

$$\phi = \nu\zeta. \tag{29}$$

This equation can be obtained from Little's theorem by regarding the bus as "the system." Then ζ is the arrival rate (since only packets that depart could have arrived) and ν is the expected time a packet is on the bus. Little's theorem asserts that $\zeta\nu$ is the average number of packets on the bus, which is the proportion of time that the bus is occupied. Combining eqs. (27) and (29) yields

$$\phi = \nu\theta/n_c. \tag{30}$$

Suppose that we attempt to increase θ by increasing the arrival rate, and K and α are adjusted to keep n_c constant (i.e., a fixed proportion of collisions). Equation (27) shows that increases in θ must be accompanied by decreases in the occupancy of the bus, and that doubling the throughput would halve the occupancy.

From eqs. (17) and (25), p_i for $1 \leq i \leq K - 1$ is obtained in the obvious way and p_K is obtained from $p_K = 1 - \sum_0^{K-1} p_i$. Then

$$L = \sum_1^K i p_i$$

is the average number of packets present in the steady state. Since θ is the arrival rate of packets that enter the system, Little's theorem yields

$$W = L/\theta, \tag{31}$$

where W is the average length of time that a packet is in the system.

3.7 Relation to the model of Kleinrock and Tobagi

In this section we explain how the "basic equation for the throughput" [eq. (3) in Ref. 1] can be obtained from the model in this paper. This will clarify the similarities and differences between the two models.

In Ref. 1 Assumption 1 states that the average retry interval is large compared with the packet transmission time. In our notation, this assumption is that $1/\alpha$ is large compared with ν. Since there is no parameter corresponding to α in eq. (3) of Ref. 1, it appears that they have set $\alpha = 0$. Let us do that. This means that every transmission that is destroyed by a collision stays in the queue forever. Consequently, Assumption 2 in Ref. 1, which states that the interarrival times of the point process consisting of packet arrival epochs and retry epochs is a Poisson process, is valid because there are no retry epochs and the arrival epochs form a Poisson process.

Another consequence of $\alpha = 0$ is that without a finite buffer, $\lim_{t\to\infty} P\{X(t) = i\} = 0$ for all i because each packet transmission has a positive probability of being destroyed by a collision. Therefore, we make these two assumptions.

Assumption 3. There is a finite buffer that can hold $K - 1 \geq 0$ packets.

Assumption 4. If a packet transmission is destroyed when the buffer is full, that packet is flushed from the system.

The next assumption is required to replicate the collision process in Ref. 1.

Assumption 5. Packets that arrive when the buffer is full and a packet is being transmitted will destroy that transmission.

With these assumptions and $\alpha = 0$, the arguments and formulas in Sections 3.2 through 3.5 can be used to obtain the results given below, but we will bypass that roundabout route and give a direct argument (which is essentially the argument in Ref. 1).

In the steady state, whenever a packet seizes the bus, there are $K - 1$ packets in the buffer. The mean time between entries to state $K - 1$ is, via eq. (20), $\nu + 1/\lambda$, and $1/\lambda$ is the mean length of stay in state $K - 1$, so

$$p_{K-1} = \frac{1/\lambda}{\nu + 1/\lambda} = \frac{1}{1 + \lambda\nu}. \tag{32}$$

The probability that a transmission is not destroyed by a collision is $e^{-\lambda h}$, so

$$\theta = \lambda p_{K-1} e^{-\lambda h}. \tag{33}$$

Equation (7) of Ref. 1 and $\alpha = 0$ show that

$$\nu = 1 + 2h - (1 - e^{-\lambda h})/\lambda. \tag{34}$$

Substituting eqs. (32) and (34) into eq. (33) yields

$$\theta = \frac{\lambda e^{-\lambda h}}{\lambda(1 + 2h) + e^{-\lambda h}},$$

which is eq. (3) in Ref. 1.

We can conclude that eq. (3) in Ref. 1 is valid when $\alpha = 0$ and the derivation in Ref. 1 requires assumptions 3, 4, and 5. Our model does not use assumptions 4 and 5 so it should produce a greater throughput when all other factors are the same; this is demonstrated numerically in Section IV. Since $K = 1$ is allowed, this derivation explains why eq. (3) in Ref. 1 becomes the single-server Erlang loss formula when there are no collisions, i.e., why $h \downarrow 0$ yields eq. (9) in Ref. 1.

IV. NUMERICAL RESULTS

A Fortran program to solve the equations in Section III was implemented in double-precision arithmetic on a PDP-1170. Because of memory limitations and the design of the program, $K \leq 30$ was required. The running time on a CRT display terminal with a 9600-baud line is less than a twinkling of an eye.

In all the numerical examples we use $h = 0.01$. This is the value used by other authors. It is a reasonable value to use for a system such as Ethernet with loops of about 300 meters carrying 128-byte packets at 10 megabits/sec. In general,

$$h\nu = \frac{lr}{sc},$$

where

l = loop length in meters,

r = transmission rate in bits/second,

s = packet size in bits, and

c = speed of light in m/second.

4.1 The upper and lower bounds are close

In Table I we present various cases that demonstrate that the lower and upper bounds for θ and W are close. The subscript "avg" stands for average and the entries are obtained by setting $\nu = 1 + h$. In all these cases, $K = 20$.

Table I suggests the approximations

$$W_{\text{avg}} = \tfrac{1}{2}(W_{lb} + W_{ub}) \quad \text{and} \quad \theta_{\text{avg}} = \tfrac{1}{2}(\theta_{lb} + \theta_{ub}).$$

We make no claim for the general validity of these approximations. Based on the encouraging results in Table I, and on several dozen other examples not reported here, we will henceforth use the case $\nu = 1 + h$ as an approximation; there will be no explicit mention that our numerical results are approximate.

4.2 The effects of changing α

The first two pairs of columns in Table I show that for fixed λ, different values of α yield different values of the performance measures. In this section we explore the consequences of changing α.

When K and λ are given, the maximum achievable throughput is attained when there are no lost service times from collisions and instantaneous retries, i.e., by an $M/D/1/K$ queue. Let the throughput of the $M/D/1/K$ queue be denoted by $\theta_{\max}$. When $K = 20$ and $\lambda = 0.7$, $\theta_{\max} = 0.700$ to three decimal places. In Table II we show that good choices of α will achieve $\theta = 0.699$, but the throughput for poor choices of α have a much lower value. In particular, very small values of α do

Table I—Evidence that the bounds for θ and W are close

λ	0.7	0.7	0.5	0.9	1.0	3.0
α	0.01	3.0	0.5	1.0	1.0	2.0
θ_{ub}	0.459	0.673	0.818	0.803	0.798	0.560
θ_{avg}	0.457	0.667	0.812	0.796	0.790	0.555
θ_{lb}	0.455	0.660	0.806	0.788	0.782	0.549
W_{ub}	42.1	11.4	19.6	19.9	22.7	35.8
W_{avg}	41.9	10.2	19.1	19.3	22.3	35.5
W_{lb}	41.7	9.1	18.7	18.7	21.9	35.1
avg n_c	0.991	0.828	0.923	0.861	0.824	0.570
avg ϕ	0.466	0.814	0.888	0.933	0.947	0.984

not do well. This confirms the observation given in Section 3.7 that the model in Ref. 1, in which $\alpha = 0$, underestimates the throughput.

The explanation for the qualitative properties shown in Table II is as follows. When α is very small, a packet that upon arrival finds another packet in service will spend a long time waiting to retry. This makes collisions rare but tends to keep the buffer positions full, which makes the throughput and bus utilization low. As α is increased, the retries are more frequent, and although there are more collisions, this is more than offset by the shorter times spent waiting to retry. When α is made sufficiently large, the retries occur so rapidly that too many packets are destroyed and performance degrades.

A significant feature of the data in Table II is that θ does not vary much for $0.5 \leq \alpha \leq 1.6$. The variation within the interval $0.5 \leq \alpha \leq 1.4$ is in the fourth decimal place. The value of α that achieves the lowest value of W also achieves a high value of θ, but the value of α that achieves the highest value of θ has a value of W about 9 percent larger than the best value of W we have found.

Let θ^* denote the largest value of θ that is found for fixed values of λ, h, and K. Numerical results not reported here have shown that θ^*/θ_{max} increases as λ decreases. This is not surprising because lower input rates result in fewer collisions. Since $\theta^* = 0.6993$ and $\theta_{max} = 0.700$, θ^*/θ_{max} is very close to one. This suggests that CSMA will provide good throughput performance when $\lambda \leq 0.7$, and that collision detection schemes will not increase throughput very much when $\lambda \leq 0.7$.

4.3 *The effects of changing λ*

Take $K = 20$ and $h = 0.01$. Choose α so that the throughput is maximized. The effects of changing λ are shown in Table III.

Notice that θ increases with λ for $\lambda \leq 2$ and then θ decreases very slightly when $\lambda = 3$. This indicates that the phenomenon of lower throughput with higher arrival rate will not occur in the normal operating range of $\lambda \leq 1$. Table III also shows that α should not increase as λ increases. This property also appeared in every example we tried.

Table II—Performance measures vs. α when K = 20 and $\lambda = 0.7$

α	0.001	0.01	0.1	0.5	0.8	1.0
θ	0.362	0.457	0.660	0.6989	0.6993*	0.6992
W	53.9	41.9	22.8	8.34	6.51	6.51
n_c	0.993	0.991	0.979	0.968	0.963	0.963
ϕ	0.368	0.468	0.681	0.729	0.734	0.734
α	1.4	1.6	2.0	3.0	4.0	5.0
θ	0.6986	0.6980	0.696	0.667	0.556	0.423
W	5.53	5.52	5.87	10.2	24.1	42.1
n_c	0.949	0.943	0.927	0.828	0.612	0.437
ϕ	0.743	0.747	0.758	0.814	0.917	0.977

The average waiting times shown in Table III are about the same as the corresponding quantities for 1000 sources shown in Fig. 3 of Ref. 2.

Suppose we design a system for a nominal arrival rate and from time to time the actual arrival rate differs from the nominal arrival rate. In Table II we see that performance suffers if a poor value of α is chosen. Now we show that moderate changes in λ will not cause much degradation in the throughput rate or the average waiting time.

We assume that $K = 10$ and $h = 0.01$ are held fixed. The nominal arrival rate is 0.7 and $\alpha = 1.6$ yields the largest throughput rate. We also assume that when λ changes, the new value is maintained long enough to ensure that steady-state performance measures are adequate. If we achieve good performance with arrival rates λ_1 and λ_2, it is reasonable to suppose that we will achieve reasonable performance when λ is changing from λ_1 to λ_2.

We will use the notations θ^* and W^* for the best values of θ and W we have found by varying α. The entries marked θ and W are for $\alpha = 1.6$.

Table IV shows that, for the values chosen, deviations from the nominal load will not cause serious performance degradation if α is kept fixed.

4.4 The effects of changing K

As K is increased, more packets can be stored, so the throughput should increase. This cannot be carried too far because as $K \to \infty$, $\lim_{n\to\infty} Y_n = \infty$ (as shown in Section 3.2), which causes the probability of a collision to approach 1 and $\theta \to 0$. We have not found a value of K to demonstrate this numerically.

Table III—Performance measures vs. λ when K = 20 and h = 0.01

λ	0.7	0.9	1.0	2.0	3.0
α	0.8	0.6	0.5	0.5	0.4
θ	0.699	0.813	0.817	0.818	0.817
θ_{max}	0.700	0.898	0.975	1.00	1.00
W	6.51	18.9	21.4	23.9	24.1
n_c	0.963	0.911	0.914	0.913	0.905
ϕ	0.734	0.901	0.904	0.905	0.912

Table IV—θ and W vs. λ when α = 1.6, K = 10, and h = 0.01

λ	0.5	0.6	0.7	0.8	0.9	1.0
θ	0.500	0.599	0.692	0.764	0.801	0.812
θ^*	0.500	0.599	0.692	0.764	0.801	0.815
W	2.30	3.07	4.37	6.22	8.06	9.39
W*	2.18	2.65	4.18	6.09	8.05	9.38

In Table V below we use the following notations:

$$\theta^* = \text{largest value of } \theta \text{ found}$$

$$\alpha_\theta^* = \text{value of } \alpha \text{ that achieves } \theta^*$$

$$W^* = \text{smallest value of } W \text{ found}$$

$$\alpha_W^* = \text{value of } \alpha \text{ that achieves } W^*$$

$$W(\alpha_\theta^*) = W \text{ when } \alpha = \alpha_\theta^*$$

$$\theta(\alpha_W^*) = \theta \text{ when } \alpha = \alpha_W^*$$

$$\theta_{\max} = \text{throughput of } M/D/1/K \text{ queue.}$$

In Table V, θ^* and W^* increase with K; α_W^* and α_θ^* decrease with K; $\alpha_W^* \geq \alpha_\theta^*$, and the difference is small and gets smaller as K increases; $W(\alpha_\theta^*)$ is not much bigger than W^*; and $\theta(\alpha_W^*)$ is not much smaller than θ^*.

4.5 Conclusions

From the data presented in this section, and from dozens of unreported sets of calculations, we conclude that:

(*i*) The bounds on θ and W are close, and choosing $\nu = 1 + h$ yields a good approximation.

(*ii*) It is important to choose a good value of α, and performance is not sensitive to changes in α near the best value.

(*iii*) The value of α that minimizes W is close to the value of α that maximizes θ.

(*iv*) The best value of α decreases as λ and K increase.

(*v*) For $\lambda \leq 0.7$, θ^* is essentially the same as $\theta_{\max}$ when $K \geq 5$.

(*vi*) Although α should vary with λ, θ and W do not significantly differ from their best values if α is held fixed while λ varies over a moderately wide interval.

V. BATCH ARRIVALS

In this section we replace the assumption that packets arrive according to a Poisson process with the assumption that packets arrive

Table V—Performance measures vs. K when $\lambda = 0.9$ and $h = 0.01$

K	5	10	15	20	30
$\theta_{\max}$	0.842	0.885	0.895	0.898	0.900
θ	0.771	0.801	0.810	0.813	0.814
$\theta(\alpha_W^*)$	0.771	0.798	0.808	0.811	0.814
α_θ^*	3.0	1.4	0.8	0.6	0.4
α_W^*	4.0	1.8	1.0	0.7	0.4
W^*	3.61	8.05	13.2	18.8	30.5
$W(\alpha_\theta^*)$	3.66	8.12	13.4	18.9	30.5

according to a compound Poisson process. In Fuchs and Jackson,[9] statistical analysis of call arrival times are given. Two of their conclusions are as follows: The exponential distribution is a reasonably good approximation of the time between bursts, and the size of a burst (measured in various ways) has a geometric distribution. The purpose of this investigation is to find out how sensitive the performance measures are to the assumption of Poisson arrivals. We will see that in this model, the throughput can be significantly lower with bursty arrivals than with Poisson arrivals with same rate.

Specifically, we assume that bursts arrive according to a Poisson process with rate λ_b, and the bursts $B_1, B_2, \cdots$ are iid with

$$P\{B_1 = i\} = (1 - \xi)\xi^i, \qquad i = 1, 2, \cdots. \tag{35}$$

It would be nice to interpret this process as one where messages arrive according to a Poisson process with rate λ_b and the jth message consists of a random number of packets with a geometric distribution. Unfortunately, we cannot do that in the context of the model described in Section II. This arrival process does not conform to the finite buffer interpretation of the model because only the current lead packet in each message would compete for the bus; we, on the other hand, assume that all packets in the buffer compete for the bus. This arrival process does not conform to the interpretation that there are K ports because that would require that if the first packet (i.e., the message) finds a free port, then all the packets in the message would enter the system. We assume that only those packets that find a port gain access to the system.

5.1 Details of the arrival process

Let $A(t)$ be the number of arrivals during $(0, T)$. Since $A(t)$ is a compound Poisson process, we know that there are constants M and V, such that

$$E[A(t)] = Mt$$

and

$$\mathrm{Var}[A(t)] = Vt$$

with $V \geq M$. We call M the arrival rate. From eq. (35) we obtain

$$E(B_1) = \frac{1}{1 - \xi} \quad \text{and} \quad \mathrm{Var}(B_1) = \frac{\xi}{(1 - \xi)^2}.$$

Standard calculations yield

$$E[A(t)] = \lambda_b t E(B_1) = \frac{\lambda_b t}{1 - \xi}$$

and

$$\operatorname{Var}[A(t)] = \lambda_b t \operatorname{Var}(B_1) + \lambda_b t [E(B_1)]^2$$
$$= \frac{\lambda_b t(\xi + 1)}{(1 - \xi)^2}.$$

In an obvious way we obtain

$$\lambda_b = \frac{2M^2}{M + V} \quad \text{and} \quad \xi = \frac{V - M}{M + V}.$$

Letting $z = V/M \geq 1$ yields

$$\lambda_b = \frac{2M}{1 + z} \quad \text{and} \quad \xi = \frac{z - 1}{z + 1}. \tag{36}$$

Equation (36) relates the parameters we might obtain from measurements, M and z, to the parameters of the model, λ_b and ξ. Let

$$b(n, j) = P\{B_1 + B_2 + \cdots B_j = n\};$$

it is the probability that n packets are contained in j bursts. The sum of iid geometric random variables has a negative binomial distribution,

$$b(n, j) = \begin{cases} 0 & \text{if} \quad n < j \\ \binom{n-1}{n-j} (1 - \xi)^j \xi^{n-j} & \text{if} \quad n \geq j \end{cases}$$

for all nonnegative integers j and n. We can compute $b(n, j)$ by recursion from

$$b(0, 0) = 1, \qquad b(1, 1) = 1 - \xi,$$
$$b(j, j) = (1 - \xi) b(j - 1, j - 1), \qquad j \geq 2,$$

and

$$b(j + k, j) = \frac{j + k - 1}{k} \xi b(j + k - 1, j), \qquad k \geq 1 \quad \text{and} \quad j \geq 0.$$

5.2 Changes in the Poisson model

In this section we describe the changes in the equations in Section III that are caused by compound Poisson arrivals.

Let $c_b(j)$ be the probability that j batches arrive in an interval of length ν; therefore,

$$c_b(j) = e^{-\lambda_b \nu} \frac{(\lambda_b \nu)^j}{j!}. \tag{37}$$

Let

$\bar{d}_n(i, j) = P\{$no collision and n arrivals in a service interval$|j$ batches arrive and start with i in queue$\}$;

then

$$\bar{d}_n(i) = \sum_{j=0}^{n} \bar{d}_n(i, j)c_b(j). \tag{38}$$

Expand on the argument used to obtain eq. (6) to conclude that

$$\bar{d}_n(i, j) = \delta_i \eta_j b(n, j). \tag{39}$$

Similarly,

$$d_n(i) = \sum_{j=0}^{n} (1 - \delta_i \eta_j) b(n, j) c_b(j). \tag{40}$$

The transition probabilities described in Section 3.2 and Appendix A are given in terms of the d and $\bar{d}$ functions, so they do not change form. The infinite sums, used in Section A.1 of Appendix A, have to be recomputed; this is done in Section A.2 of Appendix A.

The steady-state balance equations depend only on the transition probabilities, so everything in Section 3.3 is still valid. No changes are required for the equations in Section 3.4.

Since the arrivals do not form a Poisson process, time-average probabilities need not equal arrival epoch probabilities and we cannot use the arguments in Section 3.5 to obtain the steady-state probability that i packets are present. This means we have not obtained L and W for this model. We can obtain the throughput by using the arguments in Section 3.5.

Let ζ be the ejection rate and $p_0 = \lim_{t \to \infty} P\{X(t) = 0\}$. In the steady state, $\lambda_b p_0$ is the rate at which transitions leave state 0 and $\zeta \pi_0$ is the rate at which transitions enter state 0. Since these rates are equal, we have

$$\zeta = \lambda_b p_0 / \pi_0. \tag{41}$$

Equations (18) through (25) are valid for compound Poisson arrivals when λ is replaced by λ_b. The throughput is obtained from eq. (27).

5.3 Numerical examples

Let θ_z be the throughput when z is the variance to mean ratio. In Tables VI and VII we use the value of α that maximizes θ, and h = 0.01.

In Tables VI and VII we see that batch arrivals can significantly degrade throughput, and the degradation increases with z and decreases (except, possibly, for very large λ) with λ.

In Section IV we saw that θ is strongly influenced by α. That observation suggests that the throughput values in Tables I and II might improve if α varied with z. Our numerical experience suggests that, in the vicinity of the best α, θ_z is not sensitive to changes in α.

Table VI—Throughput vs. M for K = 5 and h = 0.01

M	0.1	0.5	0.7	0.9	1.0	2.0	5.0
α	20.0	4.0	4.0	3.0	3.0	2.0	2.0
θ_1	0.100	0.497	0.667	0.771	0.798	0.828	0.819
θ_2	0.069	0.382	0.535	0.663	0.709	0.819	0.813
θ_5	0.035	0.205	0.304	0.403	0.451	0.741	0.804
θ_2/θ_1	0.689	0.769	0.868	0.860	0.888	0.989	0.993
θ_5/θ_1	0.349	0.412	0.456	0.523	0.565	0.895	0.982

Table VII—Throughput vs. M for K = 10 and h = 0.01

M	0.1	0.5	0.7	0.9	1.0	2.0	5.0
α	10.0	1.5	1.6	1.4	1.2	0.9	0.6
θ_1	0.100	0.500	0.692	0.801	0.815	0.821	0.816
θ_2	0.069	0.400	0.590	0.737	0.775	0.816	0.805
θ_5	0.035	0.235	0.371	0.512	0.576	0.798	0.788
θ_2/θ_1	0.690	0.800	0.853	0.920	0.951	0.994	0.987
θ_5/θ_1	0.350	0.470	0.536	0.639	0.707	0.972	0.966

The largest improvement that we could achieve by changing α was 0.003.

VI. COLLISION DETECTION

Suppose that a time units after the vulnerable period ends, the bus is examined for a collision. When a collision is detected, the packet transmission is aborted and that packet is returned to the buffer. Collision detection reduces the time spent transmitting garbage, so it will increase throughput and reduce delays.

6.1 *Changes in the model without collision detection*

Two changes in the equations in Section III are required to describe collision detection. The first change is in eq. (7). Let

$$\eta_n^a = P\{\text{no outside arrivals prior to } h | n \text{ outside arrivals prior to } a + h\};$$

then

$$\eta_n^a = \left(\frac{a}{a+h}\right)^n, \qquad n = 0, 1, \cdots. \tag{42}$$

Let c_n^a be the probability than n outside arrivals occur in an interval of length $a + h$, i.e.,

$$c_n^a = e^{-\lambda(a+h)} \frac{(a+h)^n}{n!}, \qquad n = 0, 1, \cdots;$$

then

$$d_n^a(i) \triangleq P\{\text{collision and } n \text{ outside arrivals by } a + h | \text{service starts with } i \text{ in queue}\}$$

$$= (1 - \eta_n^a \delta_i\} c_n^a, \qquad i, n \geq 0, \tag{43}$$

where δ_i is given by eq. (5).

The equations in Sections 3.2, 3.3, and 3.4 are given in terms of the functions d and $\bar{d}$, so no changes are required. In Section 3.5 we need to change eq. (20) because the expected length of time that a packet is on the bus is not ν. Let

$$p_c(i) = P\{\text{a collision occurs} \,|\, \text{service starts with } i \text{ in the queue}\}.$$

There are no collisions if, and only if, no arrivals (inside or outside) occur within a time interval of length h. Thus,

$$1 - p_c(i) = \delta_i c_0^a$$

or

$$p_c(i) = 1 - \delta_i c_0^a, \qquad i = 0, 1, \cdots, K-2. \tag{44a}$$

When $i = K - 1$, no outside arrivals are permitted, so

$$p_c(K-1) = 1 - \delta_{K-1}. \tag{44b}$$

Let T_i be the transmission time of a packet, given that i packets were present at the last ejection epoch, and let $\nu_i = E(T_i)$; then

$$T_i = \begin{cases} 1 + h \text{ if no collision} \\ a + h \text{ if a collision} \end{cases},$$

and so

$$\nu_i = \frac{\lambda}{i\alpha + \lambda}\{(1+h)[1 - p_c(i)] + (a+h)p_c(i)\}$$

$$+ \frac{i\alpha}{i\alpha + \lambda}\{(1+h)[1 - p_c(i-1)] + (a+h)p_c(i-1)\}$$

$$= 1 + h - (1-a)\frac{\beta p_c(i) + i p_c(i-1)}{i + \beta}, \; i = 0, 1, \cdots, K-1. \tag{45}$$

We replace the ν in eq. (20) by ν_i, which induces some obvious changes in eqs. (22) through (25).

6.2 Numerical examples

In Tables II and III we saw that when $\lambda = 0.7$, the throughput in the vicinity of the best α (0.8) is very close to 0.7; but when $\alpha \leq 0.1$ or $\alpha \geq 3.0$, the throughput can be much smaller than 0.7. With collision detection, with $a = 0.02$, throughputs of about 0.7 can be achieved with α as large as 5.0. Collision detection reduces the average waiting time by about one-third. This example suggests that when a throughput close to the maximum possible can be achieved without collision

Table VIII—Performance measures vs. λ when K = 20, h = 0.01, and a = 0.02

λ	0.9	1.0	2.0	3.0
α	4.5	3.0	2.5	2.5
θ	0.891	0.935	0.943	0.942
W	8.53	14.9	20.6	20.9
n_c	0.718	0.673	0.628	0.619
ϕ	0.910	0.958	0.969	0.970

Table IX—Performance measures without collision detection

λ	0.9	1.0	2.0	3.0
α	4.5	3.0	2.5	2.5
θ	0.434	0.569	0.606	0.602
W	44.2	33.3	32.4	32.9
n_c	0.444	0.585	0.625	0.620
ϕ	0.989	0.983	0.980	0.981

detection, collision detection will lower the average waiting time and make the throughput less sensitive to α.

Now we investigate the effects of collision detection when the throughput is significantly smaller than the arrival rate.

By comparing Tables III and VIII we can see the effects of collision detection. Throughput increases significantly for each λ and W decreases. An indirect effect of collision detection is that larger values of α are best. This reduces the time spent waiting for a retry, makes the probability of no collision smaller, and yields a larger occupancy for the bus. If the values of α shown in Table VIII were used without collision detection, performance would degrade significantly, as shown in Table IX. From the last two rows of Table IX we deduce that the bus wastes a lot of time transmitting packets that have been destroyed.

6.3 Acknowledgment

I would like to thank S. Halfin for his valuable observations.

REFERENCES

1. L. Kleinrock and F. A. Tobagi, "Packet Switching in Radio Channels: Part I—Carrier Sense Multiple Access Modes and Their Throughput—Delay Characteristics," IEEE Trans. on Commun., *23,* No. 12 (December 1975), pp. 1400–16.
2. L. Kleinrock and F. A. Tobagi, "Packet Switching in Radio Channels: Part IV—Stability Considerations and Dynamic Multiple Control in Carrier Sense Multiple Access," IEEE Trans. on Commun., *25,* No. 10 (October 1977), pp. 1103–19.
3. S. Halfin, unpublished work.
4. F. A. Tobagi and V. B. Hunt, "Performance Analysis of Carrier Sense Multiple Access with Collision Detection," Proc. of the LACN Symp., (May 1979), pp. 217–44.
5. S. S. Rappaport, "Demand Assigned Multiple Access Systems Using Collision Type Request Channels: Traffic Capacity Comparisons," IEEE Trans. on Commun., *27,* No. 9 (September 1979), pp. 1325–31.

6. S. S. Rappaport and S. Bose, "Demand-Assigned Multiple-Access Systems Using Collision-Type Request Channels: Stability and Delay Considerations," IEEE Proc., *128,* Pt. E, No. 1 (January 1981), pp. 37–43.
7. D. P. Heyman and M. J. Sobel, *Stochastic Models in Operations Research,* Vol. I, New York: McGraw-Hill, 1982.
8. M. Kaplan, "A Sufficient Condition for the Nonergodicity of a Markov Chain," IEEE Trans. on Information Theory, *25,* No. 4 (July 1979), pp. 470–1.
9. E. Fuchs and P. E. Jackson, "Estimates of Distributions of Random Variables for Certain Computer Communications Traffic Models," Commun. ACM, *13,* No. 12 (December 1970), pp. 752–7.

APPENDIX A

Transition Probabilities

In Appendix A we record the transition probabilities omitted from the text.

A.1 Transition probabilities for Section 3.2

Let

$$S_{\bar{d}}(j, i) = \sum_{n=j}^{\infty} \bar{d}_n(i)$$

and

$$S_d(j, i) = \sum_{n=j}^{\infty} d_n(i).$$

The former is the probability that more than $j - 1$ packets arrive in a service interval and there are no collisions when i packets are in the buffer at the start of the service interval. The latter is the corresponding probability when there is a collision. For computational purposes it is important to represent these infinite sums as finite sums. Observe that

$$S_{\bar{d}}(j, i) = S_{\bar{d}}(0, i) - \sum_{n=0}^{j-1} \bar{d}_n(i), \qquad j \geq 1,$$

and

$$S_d(j, i) = S_d(0, i) - \sum_{n=0}^{j-1} d_n(i), \qquad j \geq 1;$$

then

$$S_{\bar{d}}(0, i) = \sum_{n=0}^{\infty} \bar{d}_n(i) = \delta_i \sum_{0}^{\infty} \left(\frac{\nu - h}{\nu}\right)^n e^{-\lambda\nu} \frac{(\lambda\nu)^n}{n!} = \delta_i e^{-\lambda h}. \tag{46}$$

Since $S_{\bar{d}}(0, i) + S_d(0, i) = 1$ is easily established,

$$S_d(0, i) = 1 - e^{-\lambda h}; \tag{47}$$

this equation also can be obtained by calculating the sum explicitly.

To modify eq. (9) when $i = 0$ we have to delete terms with negative arguments. When $i + n = K - 1$ we need to recognize that state $K - 1$ is reached if $n \geq K - 1 - i$ packets arrive and there is no collision. When $i + n = K$, we need to recognize that state K is reached if, and only if, at least $K - 1 - i$ packets arrive and there is a collision.

The following transition probabilities are obtained:

$$p_{0,0} = d_0(0)$$

$$p_{0,j} = \bar{d}_j(0) + d_{j-1}(0) \qquad 1 \leq j \leq K - 2$$

$$p_{0,K-1} = \sum_{K-1}^{\infty} \bar{d}_n(0) + d_{K-2}(0) = S_{\bar{d}}(K - 1, 0) + d_{K-2}(0)$$

$$p_{0,K} = \sum_{K-1}^{\infty} d_n(0) = S_d(K - 1, 0)$$

$$p_{1,0} = \frac{1}{1 + \beta} \bar{d}_0(0)$$

$$p_{1,1} = \frac{\beta}{1 + \beta} \bar{d}_0(1) + \frac{1}{1 + \beta} [\bar{d}_1(0) + d_0(0)]$$

$$p_{1,j} = \frac{\beta}{1 + \beta} [\bar{d}_{j-1}(1) + d_{j-2}(1)] + \frac{1}{1 + \beta} [\bar{d}_j(0) + d_{j-1}(0)], \qquad 2 \leq j \leq K - 2$$

$$p_{1,K-1} = \frac{\beta}{1 + \beta} [S_{\bar{d}}(K - 2, 1) + d_{K-3}(1)] + \frac{1}{1 + \beta} [S_{\bar{d}}(K - 1, 0) + d_{K-2}(0)]$$

$$p_{1,K} = \frac{\beta}{1 + \beta} S_d(K - 2, 1) + \frac{1}{1 + \beta} S_d(K - 1, 0).$$

For $2 \leq i \leq K - 2$,

$$p_{i,i-1} = \frac{i}{i + \beta} \bar{d}_0(i - 1)$$

$$p_{i,i} = \frac{\beta}{i + \beta} \bar{d}_0(i) + \frac{i}{i + \beta} [\bar{d}_1(i - 1) + d_0(i - 1)]$$

$$p_{i,j} = \frac{\beta}{i + \beta} [\bar{d}_{j-i}(i) + d_{j-i-1}(i)] + \frac{i}{i + \beta} [\bar{d}_{j-i+1}(i - 1) + d_{j-i}(i - 1)], \qquad i + 1 \leq j \leq K - 2,$$

$$p_{i,K-1} = \frac{\beta}{i+\beta}[S_{\bar{d}}(K-1-i, i) + d_{K-2-i}(i)]$$

$$+ \frac{i}{i+\beta}[S_{\bar{d}}(K-i, i-1) + d_{K-1-i}(i-1)]$$

$$p_{i,K} = \frac{\beta}{i+\beta} S_d(K-i-1, i) + \frac{i}{i+\beta} S_d(K-i, i-1)$$

$$p_{K-1,K-2} = \frac{K-1}{K-1+\beta} \bar{d}_0(K-2)$$

$$p_{K-1,K-1} = \frac{\beta}{K-1+\beta} S_{\bar{d}}(0, K-1) + \frac{K-1}{K-1+\beta}[S_{\bar{d}}(1, K-2)$$

$$+ d_0(K-2)]$$

$$p_{K-1,K} = \frac{\beta}{K-1+\beta} S_d(0, K-1) + \frac{K-1}{K-1+\beta} S_d(1, K-2)$$

$$p_{K,K-1} = \delta_{K-1}$$

$$p_{K,K} = 1 - \delta_{K-1}.$$

A.2 The effect of batch arrivals

Equations (46) and (47) have to be modified for the compound arrival process. From eqs. (38) and (39) we obtain

$$S_{\bar{d}}(0, i) = \sum_{n=0}^{\infty} \bar{d}_n(i) = \sum_{n=0}^{\infty} \sum_{j=0}^{n} \bar{d}_n(i, j) c_b(j)$$

$$= \sum_{j=0}^{\infty} c_b(j) \sum_{n=j}^{\infty} \bar{d}_n(i, j)$$

$$= \sum_{j=0}^{\infty} c_b(j) \sum_{n=j}^{\infty} \delta_i \eta_j b(n, j)$$

$$= \delta_i \sum_{j=0}^{\infty} c_b(j) \sum_{n=j}^{\infty} \left(\frac{\nu - h}{\nu}\right)^j b(n, j). \quad (48)$$

To evaluate the double sum, let

$$a_n = \sum_{j=0}^{n} b(n, j) c_b(j);$$

it is the probability that n customers arrive during a service interval. Let

$$\hat{A}(z) = \sum_{n=0}^{\infty} z^n a_n = \sum_{n=0}^{\infty} z^n \sum_{j=0}^{n} b(n, j) c_b(j). \quad (49)$$

Comparing eqs. (48) and (49) we see that the double sum in eq. (48) is $\hat{A}(\cdot)$ evaluated at $z = (\nu - h)/\nu$.

Standard generating-function arguments yield

$$\hat{A}(z) = \hat{C}[\hat{B}(z)], \tag{50}$$

where

$$\hat{B}(z) = \sum_{k=0}^{\infty} z^k P\{B_1 = k\} = \frac{z(1-\xi)}{1-\xi z}$$

and

$$\hat{C}(z) = \sum_{j=0}^{\infty} z^j c_b(j) = e^{-\lambda_b \nu (1-z)}.$$

Substitution into eq. (50) yields

$$\hat{A}(z) = \exp\left(\lambda_b \nu \frac{z-1}{1-z\xi}\right). \tag{51}$$

Evaluating eq. (51) at $z = (\nu - h)/\nu$ and substituting the result into eq. (48) yields

$$S_{\bar{d}}(0, i) = \delta_i \exp\left[\frac{-\lambda_b \nu h}{\nu - (\nu - h)\xi}\right].$$

As before,

$$S_d(0, i) = 1 - S_{\bar{d}}(0, i)$$

is obtained easily.

APPENDIX B

List of Symbols

The following is a list of symbols and their definitions as used in this paper.

α	=	retry rate
h	=	one-way propagation delay
K	=	system capacity in packets
λ	=	Poisson arrival rate
M	=	compound Poisson arrival rate
n_c	=	$P\{$no collision$\}$
ν	=	service time (constant)
ϕ	=	proportion of time bus is occupied
θ	=	throughput
θ_{max}	=	throughput of an $M/D/1/K$ queue
W	=	average wait (in queue plus in service) of a packet
z	=	variance to mean ratio of compound Poisson process

ζ = ejection rate of packets from the bus.

An asterisk on a symbol means the best value found. The subscripts *lb*, *ub*, and avg stand for lower bound, upper bound, and average, respectively.