DIFFUSION PROCESS APPROXIMATION FOR THE QUEUEING DELAY IN CONTENTION PACKET BROADCASTING SYSTEMS¹

Hideaki Takagi and Leonard Kleinrock

Computer Science Department
University of California, Los Angeles
Los Angeles, California 90024
U.S.A.

The average packet delay (including queueing and randomized retransmission delays) for a finite number of random access users of a channel with infinite buffers is studied. For a class of contention-type memoryless protocols (including ALOHA and nonpersistent CSMA), a diffusion process approximation for the joint queue length distribution is formulated, and on the basis of its stationary solution, two approximate mean delay formulas are proposed and examined against simulation.

1 Introduction

An important performance measure in packet broadcasting communication systems such as ground packet radio networks and local-area computer networks is the average packet delay at a given throughput value. When the channel access protocol falls in the class of random access schemes, this delay versus throughput performance for a finite user population has been studied mainly by use of linear feedback models; for example, ALOHA-type schemes are studied in [Lam75], [Carl75] and [Davi80], and carrier-sense-multiple-access (CSMA)-type protocols are studied in [Toba77], [Toba80b], [Hans79] and [Heym82]. In a linear feedback model, a Markov or semi-Markov process is formulated for a finite population of statistically identical users each being capable of storing at most one backlogged packet. (The system state is usually the number of backlogged packets.) This model may be realistic for a system of users who can actually have at most one outstanding request (like interactive terminals), or a system of so many users that traffic per user must be held at a sufficiently low level in order for the system to be stable.

To analyze the throughput-delay relationship for a group of users with capability of storing more than one packet, some extension of the linear feedback model has been attempted in [Toba80a]. One of the conclusions obtained in this study is expressed by the phrase in [Toba80a] that the (optimally controlled) system is mostly channel bound as opposed to storage bound. This statement is drawn from an observation that the improvement (in the throughput-delay performance) brought about by increasing the number of packet buffers from 2 to 3 is not as significant as that by increasing it from 1 to 2. Studying the performance in the case of users each having an infinite buffer motivates us not only on its own value but also with interest in comparing the difference between the finite-buffer and infinite-buffer cases.

In [Taka83], we introduced a notion of memoryless protocols assuming that all users have packets (of constant length) ready for transmission at all times. This memoryless property is defined such that when-

¹This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense under Contract MDA 903-82-C-0064. Hideaki Takagi is also an employee of IBM Japan, Ltd.
ever each user experiences an idle (non-transmitting) period (whose duration is assumed to be exponentially or geometrically distributed) he renews his action independently of the past happenings. Random-access protocols as pure and slotted ALOHA, and slotted and unslotted CSMA and CSMA-CD (with collision detection), in a single-hop (fully-connected or hidden-terminal) environment, can fall into this class. It has been shown in [Taka83] that the memoryless protocol systems have independent and identically distributed packet inter_departure times (i.e., the intervals between two successive successful transmissions).

In this paper, we study the mean packet delay (which includes the queuing and randomized retransmission delays) in a finite population of users each of whom has an independent (rewindable) packet arrival process and an infinite capacity of storing outstanding packets. When the channel access protocol is slotted ALOHA, this problem has been addressed in several papers. For example, Tobagi and Kleinrock [Tobs76] showed simulation results. Kleinrock and Yemini [Klei80,Yemi80] developed a Wiener-Hopf technique in the case of two users. Saadawi and Ephremides [Saad81] proposed an iterative approximation method using the notion of user and system Markov chains. Finally, Sidi and Segall [Sidi83] found an explicit expression for the mean delay in the case of two identical users. The present paper continues these efforts by newly introducing a diffusion process approximation to this problem which can handle all memoryless protocols.

The organization of the following sections is as follows. In Section 2, we develop a diffusion process approximation to the joint queue length distribution for a finite population of users of one of the contention-type protocols. Based on the stationary solution to the diffusion equation with reflecting boundary conditions, two approximate mean packet delay formulas are proposed. In Section 3, we discuss the accuracy of our formulas by comparing them with simulation in several example cases. Concluding remarks are given in Section 4.

2 Diffusion Process Approximation for a Contention System

In this section, we present a diffusion process approximation to the joint queue length distribution in an open contention system. We assume that a population of $M$ users, each having an independent arrival stream of packets, contend for a communication channel which can administer service to one user at a time. If more than one user demands service simultaneously, none of them get a successful service by the channel (the case called collision). In the case of a successful service, one packet is removed from the originating user’s queue. The time of next service request is scheduled independently by each user according to a given randomization procedure. It should be noted that because of contention the queue lengths in all users are dependent on each other. A diffusion process approximation is applied to the joint process of queue lengths in all users.

The time may be slotted (discrete) or unslotted (continuous); its unit is chosen to be the constant packet transmission time. (In a slotted-time system, the slot size may equal this unit time (for ALOHA) or its fraction (for CSMA).) Let the $M$ users be indexed as $1,2,\ldots,M$. Let $1/\lambda_i$ and $1/\sigma_i$ be the mean and the coefficient of variation, respectively, of the packet interarrival time at user $i$ ($i=1,2,\ldots,M$). Likewise, since there can be no more than one successful service (called departure) in a unit time, let $1/S$ and $C^2$ be the mean and the coefficient of variation, respectively, of the system interdeparture time. Note that $S$ is equivalent to the channel throughput. Furthermore, we assume that a successful transmission is achieved by user $i$ with probability $q_i$ ($i=1,2,\ldots,M$), where $\sum_{i=1}^{M} q_i = 1$. If we define $1/S$, and $C^2$ as the mean and the coefficient of variation, respectively, of the packet interdeparture time from user $i$, then it can be shown (see [Taka83] for derivation) that
Diffusion approximation in packet broadcasting systems

\[ S_i = q_i S ; \quad 1 - C_i^2 = q_i (1 - C^2) \quad i = 1, 2, \ldots, M \]  

We note that \( S_i \) and \( C_i^2 \) have been calculated in [Taka83] for a number of contention-type memoryless protocols (exactly for fully-connected systems, and approximately for hidden-terminal environments).

2.1 Diffusion Equation for a Contention System

Let us choose the time origin \( t = 0 \) arbitrarily and let \( A_i(t) \) be the number of packet arrivals at user \( i \) during interval \([0, t]\) \((i = 1, 2, \ldots, M)\). Similarly, let \( D_i(t) \) denote the number of departures from user \( i \) during the same interval \([0, t]\). Our approximation is based on the assumption that all users are nonempty at all times. It follows that \( Q_i(t) \), the number of packets existing in user \( i \) at time \( t \), is given by

\[ Q_i(t) = Q_i(0) + A_i(t) - D_i(t) \quad i = 1, 2, \ldots, M \]  

Therefore, the change in \( Q_i(t) \) during an interval \([t, t+\Delta]\) is expressed as

\[ Q_i(t+\Delta) - Q_i(t) = [A_i(t+\Delta) - A_i(t)] - [D_i(t+\Delta) - D_i(t)] \]

which we write as

\[ \Delta Q_i(t) = \Delta A_i(t) - \Delta D_i(t) \quad i = 1, 2, \ldots, M \]  

We consider an \( M \)-dimensional process

\[ \Delta Q(t) = [\Delta Q_1(t), \Delta Q_2(t), \ldots, \Delta Q_M(t)] \]

Note that \( A_i(t) \) and \( A_j(t) \) are independent for \( i \neq j \). We assume that \( A_i(t) \) and \( D_i(t) \) (possibly \( i = j \)) are also independent due to the assumption that all users are always nonempty (so that the arrival process does not affect the departure process). Then, it follows from Eq. (3) that

\[ \Delta Q_i(t) = \Delta A_i(t) - \Delta D_i(t) \]

\[ \Var[\Delta Q_i(t)] = \Var[\Delta A_i(t)] + \Var[\Delta D_i(t)] \]

\[ \Cov[\Delta Q_i(t), \Delta Q_j(t)] = \Cov[\Delta D_i(t), \Delta D_j(t)] \quad i \neq j \]

We are now to find the quantities on the right-hand sides of Eq. (4) by approximation. Our approximation replaces the integer-valued variables \( \Delta A_i(t) \) and \( \Delta D_i(t) \) by the corresponding continuous-valued Gaussian variables.

If \( \Delta \) is sufficiently large that we observe many arrivals and departures during \([t, t+\Delta]\), then on the basis of the central limit theorem, we can approximate each \( \Delta A_i(t) \) by a Gaussian variable such that

\[ \Delta A_i(t) \sim \mathcal{N} \left( \mu_i \Delta, \sigma_i^2 \Delta \right) \quad i = 1, 2, \ldots, M \]  

Similarly, the number of departures from all users in \([t, t+\Delta]\)

\[ \Delta D(t) \sim \mathcal{N} \left( \sum_{i=1}^{M} \Delta D_i(t) \right) \]

\[ \Var[\Delta D(t)] = SC^2 \Delta \]  

\[ \Delta D(t) = \mathcal{N} \left( 0, SC^2 \Delta \right) \]
The $M$-dimensional process $\{\Delta D_1(t), \Delta D_2(t), \cdots, \Delta D_M(t)\}$ is also approximated by a multivariate Gaussian process because $\Delta D_i(t)$ and $\Delta D_j(t)$, $i \neq j$, are dependent. It can be readily shown (see [Tak83]) that

$$\Delta D_i(t) = q_i \Delta D(t),$$

$$\text{Cov}[\Delta D_i(t), \Delta D_j(t)] = q_i q_j \left[ \text{Var}[\Delta D(t)] - \Delta D(t) \right] + \delta_{ij} q_i \Delta D(t),$$

$$i, j = 1, 2, \cdots, M$$

where

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Substituting Eqs. (1), (5), (7) and (8) into Eq. (4), we may determine the coefficients for the diffusion equation which is given shortly:

$$m_i, \Delta \Delta Q_i(t) = (\lambda_i - S_i) \Delta,$$

$$\alpha_i \Delta \text{Cov}[\Delta Q_i(t), \Delta Q_j(t)] = [\delta_{ij} (\lambda_i C_i^2 - S_i) - S_i (1 - C_i^2)] \Delta,$$

$$i, j = 1, 2, \cdots, M$$

Since $\Delta Q(t)$ has been defined as a linear combination of the two independent multivariate Gaussian processes by Eq. (3), it can also be approximated by a multivariate Gaussian process whose means and covariances are given by Eq. (9).

Let $p(x,t)$ be the joint probability density function of $\Delta Q(t)$, where $x = [x_1, x_2, \cdots, x_M]$. It satisfies the $M$-dimensional forward diffusion equation

$$\frac{\partial p(x,t)}{\partial t} = \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \alpha_{ij} \frac{\partial^2 p(x,t)}{\partial x_i \partial x_j} - \sum_{i=1}^{M} m_i \frac{\partial p(x,t)}{\partial x_i}$$

If there is no boundary condition imposed on $\Delta Q(t)$, then it is an $M$-dimensional Brownian motion with drift. However, since we have assumed that all users are nonempty at all times (i.e., $\Delta Q_i(t) > 0$), each of $M$ boundaries $x_i = 0$ ($i = 1, 2, \cdots, M$) should act as a reflecting barrier such that no probability mass can collect at $x_i = 0$. The reflecting boundary condition is given by

$$\left. \frac{1}{2} \sum_{j=1}^{M} \alpha_{ij} \frac{\partial p(x,t)}{\partial x_j} - m_i p(x,t) \right|_{x_i = 0} = 0$$

$$i = 1, 2, \cdots, M$$

2.2 Stationary Solution to the Diffusion Equation

The stationary solution to Eq. (10) where the time derivative is set to zero which satisfies the reflecting boundary condition in Eq. (11) is given by

$$p(x) = \prod_{i=1}^{M} (\omega_i) \exp(\omega_i x_i)$$

Here the column vector $\omega = [\omega_i]^T$ is computed from the column vector $m = [m_i]^T$ and matrix $\sigma = [\sigma_{ij}]$ by

$$\omega = 2 \sigma^{-1} m$$

where $\sigma^{-1}$ is an inverse matrix of $\sigma$. 
Diffusion approximation in packet broadcasting systems

From Eq. (9), it can be shown that

\[
\det (\sigma) = \left| 1 - \sum_{k=1}^{M} \frac{S_k (1 - C^2)}{\lambda_k C^2 + S_k} \right| \prod_{i=1}^{M} (\lambda_i C^2 + S_i),
\]

\[(\sigma^{-1})_{ij} = \frac{1}{\det (\sigma)} \prod_{(k,m) \neq (i,j)} (\lambda_k C^2 + S_k) \left[ \delta_{ij} \left| 1 - \sum_{k=1}^{M} \frac{S_k (1 - C^2)}{\lambda_k C^2 + S_k} \right| + \frac{S_i (1 - C^2)}{\lambda_i C^2 + S_i} \right] \]

\[i,j = 1,2,\ldots,M \quad (14)\]

Therefore, \( \omega \) can be computed through Eq. (13).

Let us now consider the special case of statistically identical users such that

\[
\lambda_i = \lambda; \quad C_i^2 = C^2; \quad q_i = \frac{1}{M}
\]

\[S_i = \frac{S}{M}; \quad C_i^2 = 1 - \frac{1}{M} (1 - C^2) \quad i = 1,2,\ldots,M \quad (15)\]

In such a case, from Eq. (9) we have

\[
m_i = \lambda - s, \quad \sigma^{-1}_{ij} = \delta_{ij} (\lambda C^2 + s) - s (1 - C^2) \quad i,j = 1,2,\ldots,M \quad (16)\]

It follows from Eq. (16) that

\[
\det (\sigma) = (\lambda C^2 + s)^{M-1} (\lambda C^2 + sC^2),
\]

\[(\sigma^{-1})_{ij} = \frac{(\lambda C^2 + s)^{M-2}}{\det (\sigma)} \left[ \delta_{ij} (\lambda C^2 + sC^2) + s (1 - C^2) \right] \quad i,j = 1,2,\ldots,M \quad (17)\]

Therefore, from Eq. (13), we obtain

\[
\omega_i = \frac{2 (\lambda - s)}{\lambda C^2 + sC^2} \quad i = 1,2,\ldots,M \quad (18)\]

We note that this is of a similar form to what we meet in the diffusion process approximation to the queue size of a GI/G/1 queuing system. The only difference here is the second term of the denominator of Eq. (18). The coefficient of variation \( C^2 \) for the whole system (instead of \( C^2 \)) accounts for the interaction among users.

2.3 Proposed Mean Delay Formulas

Eq. (12) has a product form for the marginal probability density functions. Therefore, the mean queue length in each user can be calculated separately. In the context of the diffusion process approximation, there seem to be several ways to properly evaluate the mean queue length as shown in [Gele80]. A straightforward way is to simply calculate the mean of an exponential distribution in each term of Eq. (12). Let \( \bar{Q}_i \) be the mean queue length in user \( i \) calculated in this way:

\[
\bar{Q}_i = \frac{1}{\omega_i} \quad i = 1,2,\ldots,M
\]
Another way is to first discretize the exponential distributions in Eq. (12) and then take the average of the resulting discrete distribution. Let us follow [Koba74] which shows this technique. The discretized distribution of the queue length at user $i$ is given by

$$p_n(i) \triangleq \text{Prob} \{ Q_i = n \} = \int_n^{n+1} (-w_i) \exp(w_i x) \, dx = (1 - \rho_i)(\rho_i)^n \quad n = 0, 1, 2, \ldots$$

(20)

where

$$\rho_i = \exp(w_i) \quad i = 1, 2, \ldots, M$$

(21)

From Eq. (20), the average queue length in user $i$ (now denoted by $\bar{Q}_i^{**}$) is given by

$$\bar{Q}_i^{**} = \sum_{n=0}^{\infty} n p_n(i) = \frac{\rho_i}{1 - \rho_i} \quad i = 1, 2, \ldots, M$$

(22)

It is reasonable to assume that if the arrival rates at all users were negligibly small (let us denote this situation by $\lambda \to 0$) then the packets would be delayed only due to the randomized time before the first transmission (and the transmission time). Let us denote by $D_i^{(0)}$ the expected packet delay in user $i$ when $\lambda \to 0$ ($i = 1, 2, \ldots, M$). They depend on the channel access protocol and transmission parameters. Noting that

$$w_i = -\frac{2}{C^2} \quad \text{as} \quad \lambda \to 0 \quad i = 1, 2, \ldots, M$$

(23)

we propose the following two formulas for the mean packet delay in accordance with the two expressions in Eqs. (19) and (22):

$$D_i^* = D_i^{(0)} \frac{1 - 1/w_i}{1 + C_i^2/2} \quad i = 1, 2, \ldots, M$$

(24)

$$D_i^{**} = D_i^{(0)} \frac{1 - \exp(-2/C^2)}{1 - \rho_i} \quad i = 1, 2, \ldots, M$$

(25)

3 Discussion of the Numerical Examples

Let us look at some example systems for which we can examine the accuracy of our proposed formulas in Eqs. (24) and (25). We first confine our concern to the cases of statistically identical users (i.e., the same arrival processes and transmission parameter values for all users); these are the cases where the diffusion approximation is expected to work since all the users then tend to saturate in the same fashion. We then look at a case of nonidentical users. Our discussion refers to Figures 1 through 5. It is noteworthy in these figures that Eqs. (24) and (25) give similar numerical values despite their different appearance.

Our first and second examples are for slotted ALOHA with Bernoulli arrivals. For a system of $M$ users each with a transmission probability $p$ and an arrival probability $\lambda$ in any slot, we have [Taka83]

$$S = Mp(1-p)^{M-1} ; \quad C_i^2 = 1 - S ; \quad C_i^2 = 1 - \lambda ; \quad D_i^{(0)} = 1/p$$

Thus, from Eqs. (18), (24) and (25), we can easily calculate $D^*$ and $D^{**}$. Our first example is a system of 3 users of slotted ALOHA each with $p = 0.4$. The mean delay values $D^*$ and $D^{**}$ by Eqs. (24) and
Figure 1. Mean delay for a system of 3 identical users of slotted ALOHA.

(25), respectively, are shown in Figure 1. The two upper bounds $D_{UT}$ and $D_U$ given in [Taka83] and the lower bounds which are numerically obtained by assuming a finite buffer for each user are also displayed. Since the numerical solution for the 9-buffer case is expected to be close to the true values (so long as the imposed throughput is not near the allowable maximum), the accuracy of our formulas seems good.

The second example is a similar system of 10 users of slotted ALOHA each with $p = 0.1$. In Figure 2, our approximation $D^*$ and $D^*$ are compared with simulation results. We have here simulated 10,000 slots and shown the results by the bars centered at their sample means and with width equal to twice their
Figure 2. Mean delay for a system of 10 identical users of slotted ALOHA (with simulation results).

sample standard deviations. We see that our approximation deviates further from the sample means (in the direction that gives larger values of mean delay than the simulation results) as the throughput is increased.

In the third example shown in Figure 3, we consider a system of $M = 5$ users of pure ALOHA each with a Poisson arrival stream. The interval between two successive transmissions at each nonempty user is assumed to be exponentially distributed with mean $\frac{1}{\mu}$ where $\mu = 1.1$ in this example. The values of $S$ and $C^2$ for such a pure ALOHA system with parameters $M$ and $G = gM$ can be approximately calculated by use
of a procedure given in [Taka83]. In the present case, we have \( S = 0.19 \) and \( C^2 = 0.74 \). Because of the exponentially distributed interarrival times at each user, we use \( \gamma = 1 \) in Eq. (18). The packet delay at zero input is given by \( D^{(0)} = 1 + (1/\rho) = 11 \). The values of our approximate mean delays \( D^* \) and \( D'' \) in Eqs. (24) and (25) are plotted together with the simulation results (only the sample means are shown by circles) for 2,000 packets (i.e., 2,000 successful transmissions). Here too, our diffusion approximation appears to overestimate the mean delay as the arrival rates are increased.

The fourth example deals with a symmetric hidden-user environment for a population of \( M = 20 \) unslotted nonpersistent CSMA users each of whom can hear only \( m = 17 \) other users's transmission. We
assume zero propagation delay and a Poisson arrival process at each user. The time until the next transmission is started by any user (who is not sensing a busy channel) is assumed to be exponentially distributed with mean $1/\mu$. According to our analysis in [Taka83] (approximate but validated against simulation), for $G = \mu M = 1.778$ which nearly makes $S$ maximum, we have $S = 0.45$ and $C^2 = 0.46$. Again, $D^{100} = 1 + (1/\mu) = 12.25$. In Figure 4, we show the values of $D^*$ and $D^{**}$ in Eqs. (24) and (25) with simulation results (sample means only) for 10,000 packets. The agreement is remarkable.

The last example illustrates a case where the agreement of our proposed formula and the simulation results is not so good for an asymmetric system configuration. Figure 5(a) shows a hearing graph (each
Diffusion approximation in packet broadcasting systems

Figure 5(a). The hearing configuration of a wall configuration \((M=10)\). A node represents a user, and an edge is drawn between nodes \(i\) and \(j\) if users \(i\) and \(j\) hear each other) for a system of \(M=10\) users forming a wall configuration. We assume again unslotted nonpersistent CSMA with a zero propagation delay, and Poisson arrivals at each user who schedules his next transmission at an exponentially distributed interval (with mean \(1/\lambda\)). When \(G = gM = 1\) is chosen, an analysis in [Taka83] yields approximately \(S = 0.3943, C^3 = 0.4507, q_1 = q_{10} = 0.0811, q_2 = q_3 = 0.0896, q_4 = q_8 = 0.0990, q_5 = q_9 = 0.1094, q_6 = q_7 = 0.1209\). The \(S_i\)'s and \(C_i^3\)'s are then calculated by Eq. (1). In Figure 5(b), \(D_i^3\) \((i = 1, 3, 5)\) computed by using these values are plotted together with the corresponding simulation results for 100,000 successful packet transmissions. (The values of \(D_i^3\) are not so different from those of \(D_i\).)

We see here that for users 3 and 5 our delay approximation gives lower values than the simulation results. This indicates a limitation for the applicability of our approximation.

5 Conclusion

In this paper, we have formulated a diffusion process approximation to the joint queue length distribution for a finite population of memoryless-protocol users with infinite buffers. Based on the stationary solution to the diffusion equation, we have proposed the mean packet delay formulas, and examined them against simulation in a few cases.

Our diffusion approximation can be applicable to any single-hop system (including hidden-user configurations) for which we can calculate the first two moments of the distribution of the packet interdeparture times when all the users are assumed to be nonempty. However, the accuracy of this approximation appears to be good only for the case of statistically identical users (i.e., the same arrival process and transmission parameters) since they then saturate in a similar manner.
Figure 5(b). Comparison of the diffusion approximation with simulation results in wall-configuration, unslotted CSMA.
References


