12

Discrete-time analysis of a finite capacity queue with an 'all or nothing policy' to reduce burst loss

Johan M. Karlsson
Department of Communication Systems, Lund Institute of Technology
P.O.Box 118, S-221 00 LUND, Sweden, johan@tts.lth.se
Hideaki Yamashita
Faculty of Business Administration, Komazawa University
1-23-1 Komazawa, Setagaya-ku, Tokyo 154, Japan, i38666@m-unix.cc.u-tokyo.ac.jp

abstract

Many new and old applications have to split the information into smaller units while transmitted through a network. If not the whole packet is able to get through to the destination the fraction transmitted is of no value. Several areas within the tele- and data communication field where this applies are pointed out. Further a discrete time model with bursty arrivals is introduced and analyzed. The result shows the advantage of the 'All or Nothing Policy' for the burst loss probability and the waiting time.

keywords

Discrete-time, burst arrival, multi-server, finite capacity queue, all or nothing policy

1 INTRODUCTION

The needs for tele- and data communications have evolved almost from the first day of computing. Communications on which all forms of distributed systems are built, are concerned with the different techniques that are utilized to achieve the reliable transfer of information between two distant devices. The length of the physical separation may vary, but the issue is however the same, to exchange information in the most efficient way using existing equipment in the networks. For business, governments, universities and other organizations these information exchanges have become indispensable. The importance of efficiently utilize the network becomes essential, since we deal with limited resources. One important factor is the size of the buffers within the network nodes or in other connected equipment, since these are a relative expensive part. The evolution of telecommunications is towards a multi-service network fulfilling all user needs for voice, data and video communications in an integrated way. These services are also many times real time applications, which means that there are no time for retransmissions of lost

S. Fdida et al. (eds.), *Data Communications and their Performance*

information. Several papers show a substantial delay using different ARQ-schemes, which confirm our opinion (Anagnostou, Sykas and Protonotarios, 1984). Retransmissions are mostly due to lost information in full buffers somewhere along the route from sender to destination.

2 MODEL DESCRIPTION

The model under consideration could be applied on several areas in the communication field. A few of those are outlined below. The general problem solved by this model is to utilize the buffer as efficient as possible. The traffic arriving to this system emanate from several sources, which all generate bursts with a randomly distributed length. The interarrival time for bursts from one source are also randomly distributed. A burst is a unity, which means that the individual packets constituting the burst are of no use single-handed. If we in advance could discard packets, belonging to a burst, that are not able to enter the common queue due to space limitations, we have gained a lot. The first packet of a burst carries a length indicator, which displays the total length of the burst. If not all packets, arriving in succession, of a burst have opportunity to enter the queue or the server(s) all of them are lost. Since it is of no use to waste queueing space and/or processing time on packets that are of no value for the receiver our proposed policy tries to minimize the burst loss probability. This *'All or Nothing Policy'* will be explored on below.

3 APPLICATIONS

The mentioned policy could be applied to several areas within the tele- and data communication field. A few of those are briefly discussed below, however there are many more which we hope that the reader will discover and be able to use this general model for performance and dimensioning studies on.

3.1 Video Coding

Many of the new services that are going to evolve and that already exists are using images and/or high quality sound, which demand high bandwidth. To reduce the amount of information that has to be transferred, different coding schemes are evaluated to obtain efficient techniques and algorithms. If we try to focus on some of these services, they would correspond to services like HDTV (50-100 Mbps), picture telephony (64-128 kbps (CCITT H.261)) , Hi-Fi sound and group 4 telefax (64 kbps) and some other services related to office based communication devices. Coding techniques are going to be essential for all graphic, image and video information services. Some already standardized by JPEG (Joint Photographic Experts Group) and MPEG (Motion Picture Experts Group), sponsored by ISO and CCITT, provide for compression ratios up to 1:200. For further information about line transmission of non-telephone signals, the reader is referred to (CCITT, 1988). In the case of MPEG (see International Organization for Standardization, Joint Technical Committee 1, Subcommittee 29, WG11), a new more efficient version called MPEG-2 is now used. In this technique the code is dependent not only on the current image, but also on the previous as well as the succeeding images, as shown in Figure 1. This means that if we have a buffer there is no use capturing part of a packet,

Figure 1 The principle of the MPEG-2 coding scheme.

i.e. we could as well discard the entire packet making space for future needs.

3.2 Intermediate Systems

An Intermediate System (IS) is a device to interconnect two or more systems. Depending on the services the IS has to perform they are divided into three categories;

- **Repeater** connects two identical networks, it just regenerates the signal. This means that even collisions or disturbed signals would be regenerated. A repeater operates on OSI layer 1.
- **Bridge** connects two homogeneous networks, i.e. two LANs. The bridge acts like as an address filter, picking up packets from one LAN that are intended for a destination on another LAN. The bridge operates at layer 2 of the OSI model.
- **Router** connects several networks that may or may not be similar. It has the capability of connecting more than two networks, which means that it has to have some sort of routing algorithm implemented to decide to which output port the packet should be directed. The router operates on OSI layer 3.

In such a device, many different traffic streams are merged together and share the same buffer. Different connections use the same intermediate system, for example several connectionless services could be routed through a Token Ring (IEEE 802.4) and a CSMA/CD (IEEE 802.3) using a interconnecting bridge. The packets have to be queued in the bridge waiting for access to the CSMA/CD network, in which no access is granted within a certain amount of time. In this cases using real time data, there is no time for the receiver to resequent or demand retransmissions of parts of a packet, i.e. if we loose part of a packet we could as well discard the whole packet in the IS.

3.3 IP Traffic

The IP (Internet Protocol) is intended for communication through several networks. It provides a connectionless delivery system for hosts connected to networks with the protocol implemented. The IP has to be implemented on the hosts constituting the OD pair as well. The connection is unreliable and on a best effort basis. The interface to other layers, shown in Figure 2 are to transport and network layers. The transport layer is

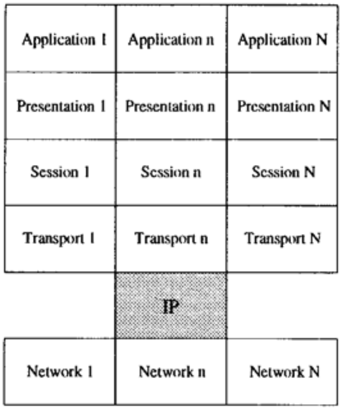

Figure 2 The seven layer protocol stack, showing the position of IP.

usually implemented as the Transmission Control Protocol (TCP), but other protocols could be used on this level. On the link layer we could have interfaces towards different LAN access protocols or others like SMDS. IP is also to be used over ATM networks and on ATM Local Area Networks (Chao, Ghosal, Saha and Tripathi, 1994). The basic unit for transfer is specified as a datagram. The datagram consists of a data field and a header, the different fields are shown in Figure 3 and a more specific explanation of each field could be found in (Comer, 1991). Parts of certain interest to our studies are the "Total Length" and "Fragment Offset" fields, described below:

- **Total Length** To identify the number of octets in the entire datagram. (Usually less than 1500 octets, which is the maximum packet length of Ethernet.) IP specifications sets a minimum size of 576 octets that must be handled by routers without fragmentation.
- **Fragment Offset** The field represents the displacement (in octets) of this segment from the beginning of the entire datagram. Since the datagrams may arrive out of sequence, this field is used to assemble the collection of fragments into the original datagram.

The datagram could during transmission be duplicated, lost, delayed or out of order. This is to permit nodes with limited buffer space to handle the IP datagram. In some error situations datagrams are discarded, while in other situations error messages are sent. If we have to discard information it is more efficient to discard information belonging to the same IP datagram since fractions are of no use.

4 ANALYTICAL MODEL

The queueing model under consideration is a discrete-time, multi-server, finite capacity queue with burst arrivals, shown in Figure 4. Once the first packet of a burst arrives at

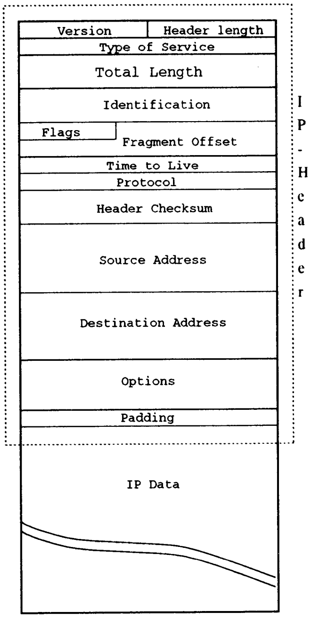

Figure 3 The IP datagram.

the queue, the successive packets will arrive on every time slot until the last packet of the burst arrives. The number of packets of the nth burst is denoted by S^n, which is assumed to be independent and identically distributed (*i.i.d.*) with a general distribution. We assume that there exists a positive number S_{max} such that $Pr[S^n > S_{max}] = 0$ and that we can know the value of S^n when the first packet of the nth burst arrives. The interarrival time between the nth and $(n+1)$st burst is denoted by T^{n+1}, which is assumed to be *i.i.d.* with a general distribution. We allow that T^n may take the value 0, i.e. the first packet of more than one burst may arrive on the same slot. There are m servers

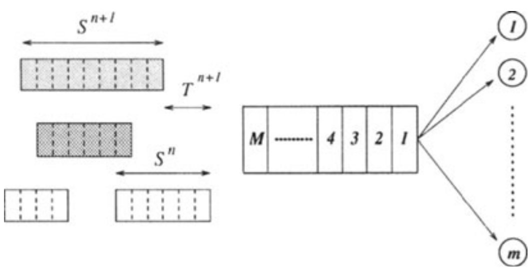

Figure 4 Queueing model of the analyzed system.

which are synchronized so that they start and end service at the same time. The service time of a packet is assumed to be equal to one slot. The packets arrive at the queue at the beginning of a slot and leave the queue at the end of a slot. The capacity of the shared buffer is finite, say M, excluding the service space. Arriving packets are allowed to enter the queue only when none of the packets belonging to the same burst are lost. At the arrival instant of the first packet of the nth burst, the system tries to reserve buffer space for all packets which belong to the nth burst, if it finds all servers busy, so that all packets of the nth burst can enter the queue. If this is not possible, all packets of the nth burst are lost. This *'All or Nothing Policy'* minimizes burst loss probabilities. The packets in the queue are served in a FIFO discipline on a burst basis. That is, the packets of the nth burst have higher priority than any packets of the (n+1)st, or higher numbered, burst whenever they arrive.

Indeed, if $T^{n+1} + t < S^{n+1}$, the tth packet of the $(n + 1)$st burst has already arrived at the arrival instant of the $(T^{n+1} + t + 1)$st packet of the nth burst. In this case, the tth packet of the $(n + 1)$st burst is served earlier than the $(T^{n+1} + t + 1)$st packet of the nth burst only if any server is idle.

In the following two sections, we propose an efficient numerical method to analyze the queueing model described above.

5 EMBEDDED MARKOV CHAIN

In this section, we construct a finite state embedded Markov chain, which will be useful for obtaining some stationary performance measures of the queue described in the previous section. First of all, let us consider an embedded Markov chain by giving attention to all active bursts, i.e., accepted bursts with remaining packets (which have not yet arrived). If we keep track of the number of remaining packets of each active burst and the number of packets in the buffer (the queue length) at the arrival instant of bursts, the process has a Markov property. It might be possible to obtain some stationary performance measures, e.g., the burst loss probability, the queue length distribution and the waiting time distribution from the steady state probability distribution of this process. However, the process becomes intractable as the number of active bursts increases. Therefore, it is important to construct an embedded Markov chain in order to efficiently obtain some performance measures such as a packet loss probability.

A similar model was analyzed and an effective embedded Markov chain was proposed by (Yamashita, 1994). We extend this methodology for the system with the *'All or Nothing Policy'*. The basic idea of the method is as follows: Let us consider the embedded point of the nth burst arrival instant. When the number of active bursts is equal to or less than m, we keep track of the number of remaining packets of every burst. However, when the number of active bursts is greater than m, we choose m bursts in decreasing order of the number of remaining packets and keep track of the numbers of remaining packets of these m bursts. The number of packets in the buffer never decreases while at least m bursts are active whenever the $(n + 1)$st burst arrives. Therefore, if we also keep track of the number of packets in the buffer on the last slot when at least m bursts are active, we can deside whether the $(n + 1)$st burst should be accepted or not.

Let v_i^n denote the ith largest number of remaining packets among active bursts at the arrival instant of the nth burst. In other words, v_i^n means the number of time slots with i arriving packets counting from the arrival instant of the nth burst , excluding the $(n+1)$st burst and all the bursts after $(n+1)$st. Note that $v_1^n \geq v_2^n \geq \cdots \geq v_m^n$. $v_{i-1}^n > v_i^n = 0$ means that only $(i-1)$ bursts are active at the arrival instant of the nth burst. $\boldsymbol{v}^n$ denotes the vector $(v_1^n, v_2^n, \cdots, v_m^n)$. Further, let w^n be the number of packets in the buffer on the v_m^nth slot counting from the arrival instant of the nth burst, excluding the $(n+1)$st burst and all the bursts after $(n+1)$st even if they have arrived already. w^n takes into account the arrival packets due to the other active bursts not included in the vecor $\boldsymbol{v}^n$.

Now, let us obtain the relationship between w^n and w^{n+1} given $\boldsymbol{v}^n$, T^{n+1}, and S^{n+1}. If $v_m^n \leq T^{n+1}$, then w^{n+1} represents the number of packets in the buffer on the T^{n+1}th slot counting from the arrival instant of the nth burst and is less than w^n since the packets in the buffer, if any, will be served after the v_m^nth slot. The ith server is capable of serving $(T^{n+1} - v_i^n)^+$ packets during the T^{n+1} slots, where $(N)^+ = \max(0, N)$. If $v_m^n > T^{n+1}$, on the other hand, w^{n+1} represents the number of packets in the buffer on the $\max(v_m^n, T^{n+1} + S^{n+1})$th slot counting from the arrival instant of the $(n+1)$st burst, and is greater than w^n since the number of packets increases on every slot by one from the T^{n+1}th to $\min(v_m^n, T^{n+1} + S^{n+1})$th slots counting from the arrival instant of the nth burst, as long as there is enough space in the buffer. If any arriving packets is not able to enter the buffer due to lack of space, all packets of the nth burst are rejected to enter the buffer. Then, the ith server will serve the packets in the buffer on every slot by one from $(v_i^n + 1)$st to T^{n+1}th slots (if $v_i^n < T^{n+1}$) counting from the arrival instant of the nth burst. From the above two discusions, we have

$$w^{n+1} = \begin{cases} [w^n - \sum_{i=1}^{m}(T^{n+1} - v_i^n)^+]^+, & \text{if } \ v_m^n - T^{n+1} \leq 0, \\ w^n + \min(v_m^n - T^{n+1}, S^{n+1}), & \text{if } \ 0 < \min(v_m^n - T^{n+1}, S^{n+1}) \\ & \qquad \leq M - w^n, \\ w^n, & \text{if } \ \min(v_m^n - T^{n+1}, S^{n+1}) \\ & \qquad > M - w^n. \end{cases} \tag{1}$$

Similarly, we can obtain the relationship between $\boldsymbol{v}_i{}^n$ and $\boldsymbol{v}_i{}^{n+1}$ given w^n, T^{n+1}, and S^{n+1}. When the $(n+1)$st burst is accepted, the last packet of the $(n+1)$st burst arrives on the $(T^{n+1} + S^{n+1})$th slot counting from the arrival instant of the nth burst. If $v_i^n \leq T^{n+1} + S^{n+1} < v_{i-1}^n$, then the ith largest number of remaining packets among active bursts at the arrival instant of the $(n+1)$st burst, v_i^{n+1}, is S^{n+1}. At the same time, $v_{i-1}^{n+1} = v_{i-1}^n - T^{n+1}$ and $v_{i+1}^{n+1} = v_i^n - T^{n+1}$. On the other hand, when the $(n+1)$st burst is rejected because of the *'All or Nothing Policy'*, no new burst arrives during T^{n+1} slots. Accordingly, we have the following relations:

$$v_i^{n+1} = \begin{cases} (v_{i-1}^n - T^{n+1})^+, & \text{if } \min(v_m^n - T^{n+1}, S^{n+1}) \le M - w^n, \\ & \quad v_{i-1}^n \le T^{n+1} + S^{n+1}, \\ S^{n+1}, & \text{if } \min(v_m^n - T^{n+1}, S^{n+1}) \le M - w^n, \\ & \quad v_i^n \le T^{n+1} + S^{n+1} < v_{i-1}^n, \\ v_i^n - T^{n+1}, & \text{if } \min(v_m^n - T^{n+1}, S^{n+1}) \le M - w^n, \\ & \quad T^{n+1} + S^{n+1} < v_i^n, \\ & \text{or } \min(v_m^n - T^{n+1}, S^{n+1}) > M - w^n. \end{cases} \quad (2)$$

where $i = 1, 2, \cdots, m$ and we define $v_0^n = \infty$ for $n = 1, 2, \cdots$.

$(\boldsymbol{v}^n, w^n)$ has the Markov property, because $(\boldsymbol{v}^{n+1}, w^{n+1})$ depends only on $(\boldsymbol{v}^n, w^n)$ given T^{n+1} and S^{n+1}. Let us denote the relationship by:

$$(\boldsymbol{v}^{n+1}, w^{n+1}) = f(\boldsymbol{v}^n, w^n, S^{n+1}, T^{n+1}).$$

Since $\boldsymbol{v}^n$'s are bounded by S_{max}, $(\boldsymbol{v}^n, w_1^n)$ is a finite state embedded Markov chain at the arrival instant of bursts with less than $(S_{max} + 1)^m(M + 1)$ states, i.e., $O(S_{max}^m M)$. $T^{n+1} \ge S_{max} + M$ is a sufficient condition for $v_1^{n+1} = S^{n+1}$, $v_2^{n+1} = \cdots = v_m^{n+1} = 0$, and $w^{n+1} = 0$. Therefore, it is sufficient to consider the case $T^{n+1} = 0, 1, \cdots, S_{max} + M$, $S^{n+1} = 1, 2, \cdots, S_{max}$ for every state $(\boldsymbol{v}^n, w^n)$ when we calculate the coefficients of the equilibrium equations using (1) and (2). That is, it requires $O(S_{max}^{m+1} M(S_{max} + M))$ time units to calculate. Once we calculate the coefficients of the equilibrium equations, we can get the steady state probability distribution of $(\boldsymbol{v}^n, w^n)$, denoted by $P(\boldsymbol{v}, w)$, by solving the system of stationary equilibrium equations:

$$P(\boldsymbol{v}, w) = \sum_{S=1}^{S_{max}} \sum_{T=0}^{S_{max}+M} P(S)P(T) \sum_{(\boldsymbol{v}', w') \in \Lambda(\boldsymbol{v}, w, S, T)} P(\boldsymbol{v}', w')$$

for all possible states $(\boldsymbol{v}, w)$, where

$$\Lambda(\boldsymbol{v}^{n+1}, w^{n+1}, S^{n+1}, T^{n+1})$$
$$= \{(\boldsymbol{v}^n, w^n) \mid (\boldsymbol{v}^{n+1}, w^{n+1}) = f(\boldsymbol{v}^n, w^n, S^{n+1}, T^{n+1})\},$$

and $P(S)$ and $P(T)$ denote the probability that the number of packets of a burst is S and the probability that the interarrival time between bursts is T, respectively.

In particular, when $m = 1$, we get,

$$P(v, 0) = \sum_{v'=1}^{S_{max}} \sum_{w'=0}^{M} Pr[S = v] \sum_{k=v'+w'}^{\infty} Pr[T = k] P(v', w')$$

$$+\Psi[v \geq M+1] \sum_{v'=v}^{S_{max}} \sum_{j=M+1}^{S_{max}} Pr[S=j]Pr[T=v'-v]P(v',0).$$

$$P(v,w) = \sum_{v'=1}^{S_{max}} \sum_{w'=\max(0,w-v',w-v)}^{M} Pr[S=v]Pr[T=v'+w'-w]P(v',w')$$

$$+\sum_{v'=v}^{S_{max}} \sum_{w'=\max(0,w-v+1)}^{w-1} Pr[S=w-w']Pr[T=v'-v]P(v',w')$$

$$+\Psi[v \geq M-w+1] \sum_{v'=v}^{S_{max}} \sum_{j=M-w+1}^{S_{max}} Pr[S=j]Pr[T=v'-v]P(v',w),$$

$$w = 1, 2, \cdots, M.$$

where $\Psi[\cdot]$ is an indicator function which takes 1 or 0.

We note that this method is still much more efficient than the straightforward way mentioned at the beginning of this section, though the process $(\boldsymbol{v}^n, w^n)$ becomes intractable as the number of servers increases.

6 PERFORMANCE MEASURES

In this section, we get the performance measures using the steady state probability distribution $P(\boldsymbol{v}, w)$ obtained in the previous section. We first calculate the burst loss probability, i.e., that one or more packets in the burst is rejected. As discussed in the previous section, when $\min(v_m^n - T^{n+1}, S^{n+1}) > M - w^n$ the $(n+1)$st burst is lost because of the *'All or Nothing Policy'*. Then the burst loss probability denoted by P_{loss}^{burst} is represented by

$$P_{loss}^{burst} = \sum_S \sum_T P(S)P(T) \sum_{v_m} \sum_{w > M - \min(v_m - T.S)} P(v_m, w), \tag{3}$$

where $P(v_m, w)$ is the marginal probability of $P(\boldsymbol{v}, w)$. Let us define the packet loss probability by the ratio between the average number of packets that are lost and the average number of packets that arrive in a burst. Similar arguments give the expression for the packet loss probability denoted by P_{loss}^{packet} as follows:

$$P_{loss}^{packet} = \sum_S \sum_T SP(S)P(T) \sum_{v_m} \sum_{w > M - \min(v_m - T,S)} P(v_m, w) / \sum_S SP(S). \tag{4}$$

Now, we get the waiting time distribution W, assuming FIFO discipline on a burst basis, that is, the packets of the nth burst have the higher priority than any packets of the (n+1)st burst whenever they arrive. We define the waiting time distribution so that it satisfies the following equation:

$$\sum_{k=0}^{\infty} P[W=k] + P_{loss}^{packet} = 1.$$

Accordingly, we suppose the nth burst is not rejected. We first consider the waiting time of the jth packet of the $(n+1)$st burst, denoted by W_j^{n+1}, given $\boldsymbol{v}_n$, w_n, and T^{n+1}. Note that any packets of the (n+1)st burst can not be served at least until the v_m^nth slot counting from the arrival instant of the nth burst, if $T^{n+1} \le v_m^n$. The jth packet of the $(n+1)$st burst arrives at the queue on $(T^{n+1}+j)$th slot counting from the arrival instant of nth burst. $W_j^{n+1} = k$ means that the number of packets in the buffer which should be served before the jth packet of the $(n+1)$st burst first becomes 0 at the $(T^{n+1}+j+k)$th slot counting from the arrival instant of the nth burst. The number of packets in the buffer which should be served before the jth packet of the $(n+1)$st burst at the $(T^{n+1}+j+k)$th slot counting from the arrival instant of the nth burst is equal to :

[queue length on the v_m^nth slot excluding all the bursts after nth: w_n]

\+ [# arrived packets of the $(n+1)st$ burst at the arrival instant of the jth packet: j]

\+ [# arrived packets of all the bursts before $(n+1)$st from the (v_m^n+1)st slot: $\sum_{i=v_m^n+1}^{T^{n+1}+j+k} e_i^n$]

− [# served packets from the (v_m^n+1)st slot: $(T^{n+1}+j+k-v_m^n)m$],

where e_i^n is the largest number which satisfies $v_{e_i^n}^n \ge i$, $(e_i^n = 0, 1, \cdots, m)$. In other words, e_i^n is the number of active bursts on the ith slot counting from the arrival instant of the nth burst, excluding all the burst after nth. Using

$$(T^{n+1}+j+k-v_m^n)m - \sum_{i=v_m^n+1}^{T^{n+1}+j+k} e_i^n = \sum_{i=1}^{T^{n+1}+j+k} (m-e_i^n),$$

we have the following relations:

$$W_j^n = \begin{cases} 0, & \text{if } w^n + j \le \sum_{i=1}^{T^{n+1}+j}(m-e_i^n), \\ \\ k, & \text{if } \sum_{i=1}^{T^{n+1}+j+k-1}(m-e_i^n) < w^n + j \le \sum_{i=1}^{T^{n+1}+j+k}(m-e_i^n). \end{cases}$$

Therefore, taking into account the condition that the burst is accepted, the waiting time distribution can be obtained by

$$Pr[W=0] = \sum_S \sum_T P(S)P(T) \sum_v \sum_{w \le M-\min(v_m-T,S)} P(v_m, w)$$

$$\sum_{j=1}^{S} \Psi[w+j \le \sum_{i=1}^{T+j}(m-e_i)] / \sum_S SP(S),$$

$$Pr[W=k] = \sum_S \sum_T P(S)P(T) \sum_v \sum_{w \le M-\min(v_m-T,S)} P(v_m, w)$$

$$\sum_{j=1}^{S} \Psi[\sum_{i=1}^{T+j+k-1}(m-e_i) < w+j \le \sum_{i=1}^{T+j+k}(m-e_i)] / \sum_S SP(S),$$

$k = 1, \cdots, S_{max} + \lfloor M/m \rfloor,$

where the summation of $\boldsymbol{v}$ extends over all possible states of $(v_1, v_2, \cdots, v_M)$ and e_i means e_i^n given $\boldsymbol{v}$.

We can now calculate the mean waiting time using the waiting time distribution. Then, we get the mean queue length $\bar{L}$ using Little's law [?], i.e.

$$\bar{L} = \bar{W}\bar{S}/\bar{T}, \tag{5}$$

where $\bar{S}$, $\bar{T}$, and $\bar{W}$ denote the first moments of S, T, and W, respectively. Though we assumed FIFO discipline on a burst basis when we derived the waiting time distribution, (3) $\sim$ (5) hold true for other service disciplines, e.g., FIFO on a packet basis.

7 NUMERICAL EXAMPLES

In this section, we present some numerical examples and demonstrate the advantage of the *'All or Nothing Policy'* for the burst loss probability and the waiting time. We consider two examples, single and double server queues. For both examples, we assume that the interarrival time between bursts is uniformly or binomially distributed from 1 to 15, with mean 8.0 and squared coefficient of variation 0.2917 or 0.06944. We also assume that the number of packets in a burst is uniformly or binomially distributed from 1 to 11, with mean 6.0 and squared coefficient of variation 0.2778 or 0.05469. The number of states of the embedded Markov chain $(\boldsymbol{v}^n, w^n)$ for each example are shown in Table 1.

We compare two systems, with and without the *'All or Nothing Policy'*. The packet loss probability, the burst loss probability, and the mean waiting time of packets are illustrated as a function of the buffer capacity in Figures 5 $\sim$ 10. In each example, we can find that the system with the *'All or Nothing Policy'* is superior to the one without it for the burst loss probability and the mean waiting time of packets. On the other hand, using the *'All or Nothing Policy'*, the packet loss probability increases, however the packet loss probability is not important in applications discussed in section 3.

Table 1 The number of states $(\boldsymbol{v}^n, w^n)$

Capacity of the buffer	1	3	5	7	9
$m = 1$	22	44	66	88	110
$m = 2$	152	304	456	608	760

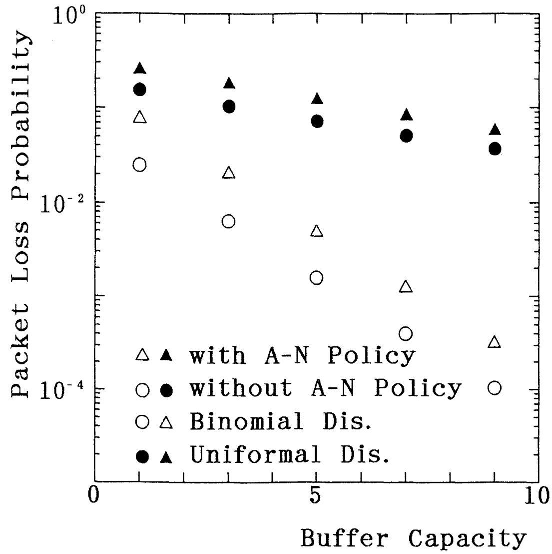

Figure 5 Packet Loss Probability ($m = 1$).

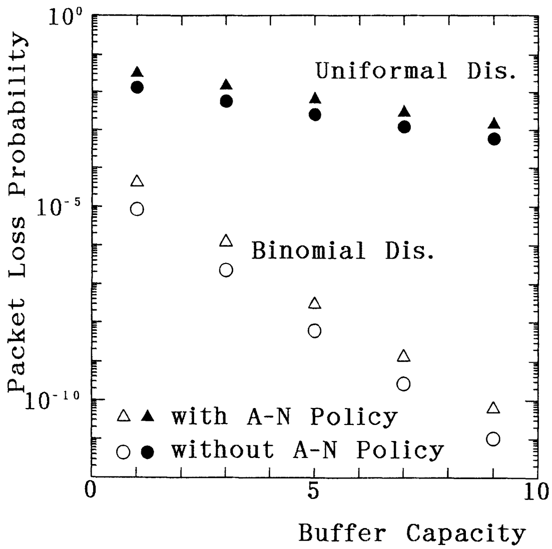

Figure 6 Packet Loss Probability ($m = 2$).

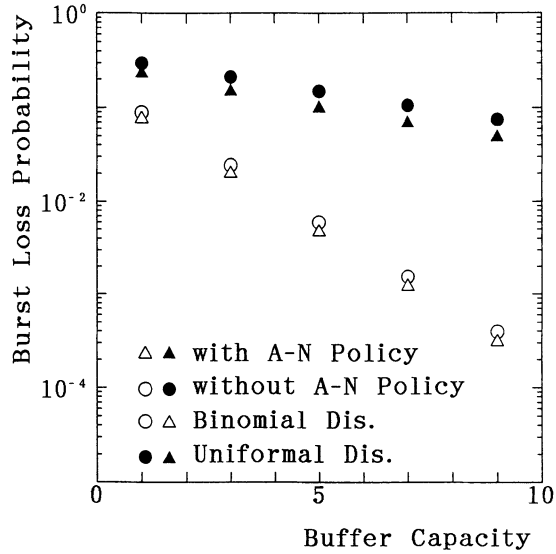

Figure 7 Burst Loss Probability ($m = 1$).

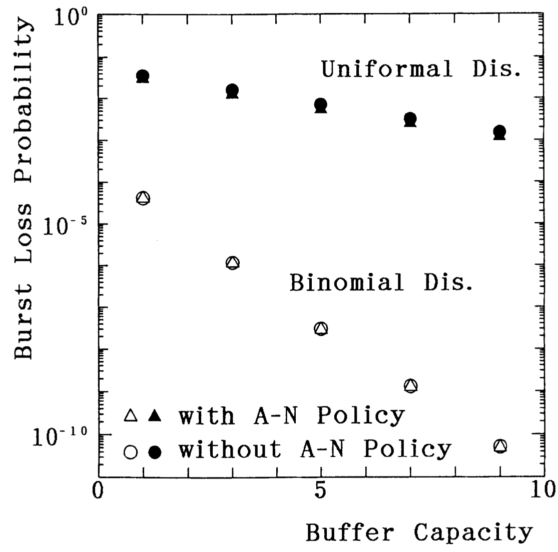

Figure 8 Burst Loss Probability ($m = 2$).

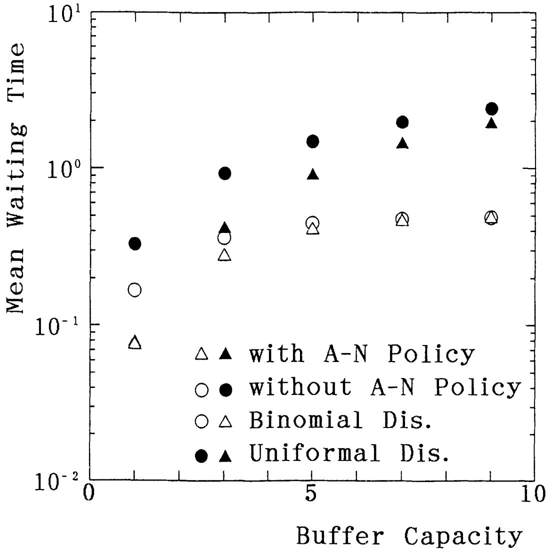

Figure 9 Mean Waiting Time ($m = 1$).

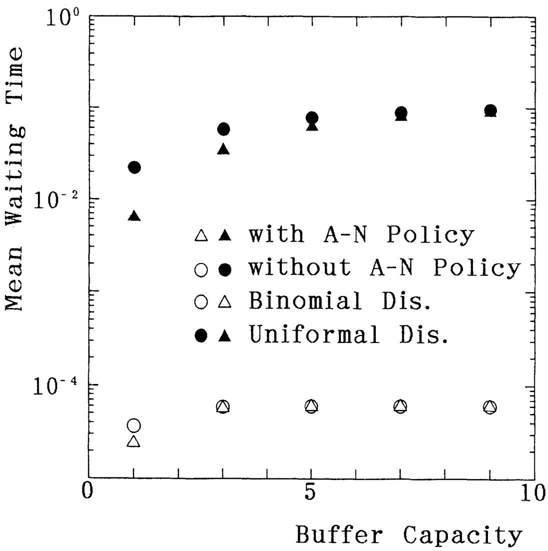

Figure 10 Mean Waiting Time ($m = 2$).

8 REFERENCES

Anagnostou, M. E., Sykas, E. D. and Protonotarios, E. D. (1984), Steady-state and transient delay analysis of ARQ protocols, *Computer Communications,* **7**, No.1.

CCITT (1988), *CCITT Recommendation,* H and J series, November.

Chao, H. J., Ghosal, D., Saha D. and Tripathi, S. K. (1994), IP on ATM Local Area Networks, *IEEE Communications Magazine,* August.

Comer, D.E. (1991), *Internetworking with TCP/IP, vol.I,* Prentice-Hall.

Little, J. D. C. (1961), A Proof of the Queueing Formula $L = \lambda W$, *Operations Research* **9**.

Yamashita, H. (1994), Numerical analysis of a discrete-time finite capacity queue with a burst arrival, *Annals of Operations Research,* **49**.

9 BIOGRAPHY

Johan M. Karlsson has been Associate Professor of the Department of Communication System, Lunt Institute of Technology, Sweden, since 1993.

Hideaki Yamashita received the B.S., M.S., and Ph.D. degrees in mechanical engineering from Sophia University in 1982, 1984, and 1987, respectively. He is currently Lecturer (faculty member) of the Faculty of Business Administration, Komazawa University, Tokyo, since 1995. As a Visiting Faculty in Computer Science, he stayed at the North Carolina State University, Raleigh, during 1989 - 1991. His research interests include the modeling and analysis of queueing networks and their applications to production systems and high speed networks. He is a member of the Operations Research Society of Japan, the Institute of Electronics, Information and Communication Engineers, and the Japan Society of Mechanical Engineers.