DESIGN AND ANALYSIS OF FLOW CONTROL PROTOCOLS FOR METROPOLITAN AREA NETWORKS

D. Towsley, S. Fdida and H. Santoso

COINS Technical Report 90-70
July, 1990

Design and Analysis of Flow Control Protocols for Metropolitan Area Networks[1]

Don Towsley*, Serge Fdida†, Harry Santoso†

*Department of Computer & Information Science, University of Massachusetts, Amherst, MA 01003, USA

†Laboratoire MASI Université Pierre et Marie Curie, 4, place Jussieu, 75252 Paris cedex 05, FRANCE

Abstract: In this paper we study the problem of flow control for LAN's interconnected through a high speed MAN. We consider several input bridges feeding data through the MAN to an output bridge and specifically concern ourselves with the avoidance or minimization of buffer overflow at these bridges. We study the behavior of four flow control policies that differ from each other according to the type of information passed aroundamong the bridges. The most complex protocols use queue length information whereas the simpler protocols use either no information or packet age information. We show that the protocols using queue length information are optimal in the sense that they minimmize buffer overflow for a broad class of systems. In addition we compare the performance of these policies through a combination of analysis and simulation. We observe that using age information, which is relatively inexpensive to acquire, yields half of the benefit of queue length information. Furthermore, if most of the buffers are allocated to the output bridge, then there is little difference between the behavior of these policies. This suggests that simple protocols may work well under such allocations. Last, we study the issue of fairness when input bridges are not identical. We observe that the policies based on queue length information provide fairer treatment when the performance metric is probability of loss and the simple policies provide fairer treatment in the case of mean packet delay.

Keywords: approximate analysis, bridges, flow control, metropolitan area networks, optimal control.

---

[1]The work of the first author was supported in part by the Office of Naval Research under grant ONR N00014-87-K-0304. It was performed while the author was on sabbatical at Laboratoire MASI, UPMC, Paris.

1 Introduction

We consider a metropolitan area network (MAN) that provides network interconnection services between Local Area Networks (LANs) (Figure 1). Several LANs are connected to the MAN through bridges having a finite amount of buffers for storing data. Hence one of the main responsibilities of the MAN, acting as network provider, is to handle the problem of buffer overflow at the bridges. The speed difference between a MAN (working at speed in excess of 100Mb/s) and LANs connected to it (speed in the range of 10Mb/s) can result in congestion problems at the output bridge if there are a number of input bridges acting as sources of traffic destined for the output LAN. Hence a flow control policy is required in order to avoid excessive lost packets at the bridges. Packets can be lost, either at the input buffer (that is a buffer holding the traffic sent from a LAN to the MAN) at a bridge, or at the output buffer (the buffer serving the traffic sent from a MAN to the receiving LAN).

The purpose of this paper is to design and evaluate several flow control strategies for handling the problem of buffer overflow when the MAN provides bandwidth between several input bridges and a single ouput bridge all connected to LAN's. Specifically we study the performance of four protocols that include two protocols that use buffer occupancy information, one that uses packet age information and a fourth that uses no information. We prove the optimality of the first two protocols for a variety of environments (here optimality means minimizing buffer overflow) and develop approximate analytical models for three of them. The influence of different system parameters, such as buffer size, burstiness in the arrival process, etc., is determined through the use of these approxiamate analytical models and simulation. The main observation from this study is that performance is best when most of the buffers are allocated to the output bridge and that, in this case, there is little difference in the performance of the four protocols thus suggesting that simple protocols will suffice.

The subject of flow control in networks has received considerable attention (see the survey by Gerla and Kleinrock [6]). However, little work has been done on design and analysis of flow control protocols in high speed MANs. Most of the work in this area has been devoted to improving the Media Access Protocol [1,5]. In the following paragraphs we discuss earlier work that deals with flow-control issues in network interconnections.

Bux and Grillo [3] simulated an end-to-end window flow control in multiple token rings interconnected through bridges. Their model focussed upon a single connection which used IEEE LLC2 as an end-to-end protocol between two communicating stations. The authors showed that the fixed window scheme is not adequate as it can degrade through-

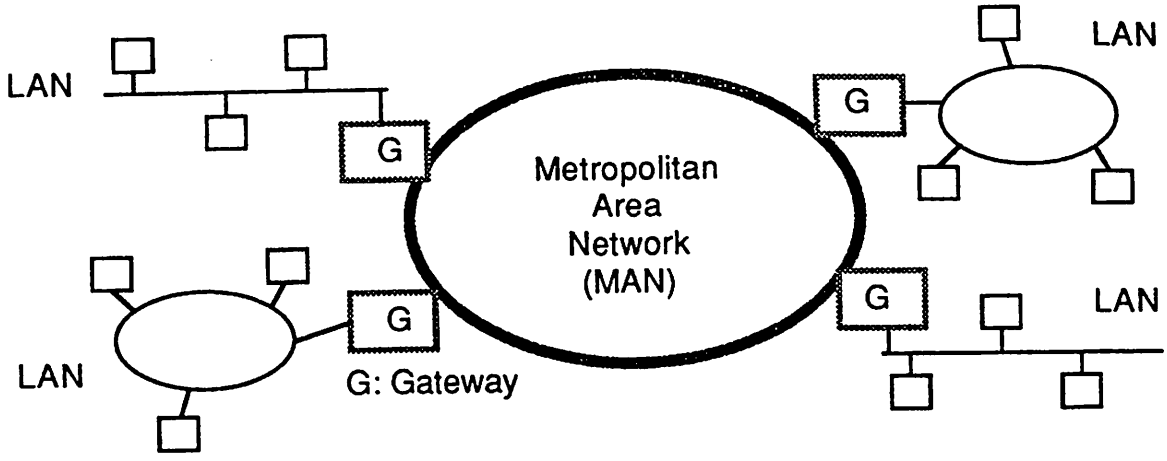

Figure 1: Interconnection of LANs through a MAN.

put due to congestion at the bridges by the successive retransmission of frames. To avoid bridge congestion, they suggested that the LLC2 protocol be modified to include a dynamic window mechanism. This protocol would change the window sizes in response to frame losses at the bridges. It would reduce the window to the size of one frame upon each loss, and increase it by one for every N positive acknowledgements received. Here N ($N < W$, W = max. window size) is an important parameter to be tuned. Jain [2] used similar ideas, but with a different strategy for increasing the window size.

We observe from these papers that, although the dynamic window adjustment can yield significant improvements over the fixed window scheme, the control decision is always made by individual end-to-end connections. Unfairness (i.e. in term of meeting user throughput requests) may arise due to differrent length of the end-to-end communication paths. Moreover, the dynamic window scheme is not a part of the IEEE LLC standard. In a recent paper, Wong and Schwartz [10] proposed a different approach to bridge flow control in MANs. Instead of relying on the end-user actions, they argue that flow control should be performed by the MAN nodes or bridges in charge of regulating the internet traffic.In order to do this, a scheduling policy for transmitting the internet packets that arrive at the different source bridges was developed. They set up the problem of determining the best policy as a Markov decicison process and solve it numerically. They observe that the optimal policy always delays transfers from the source bridges to the output bridge. In addition, they observe that when the source bridges are identical, the optimum

policy always transfers packets from the most heavily loaded source bridges. Our work is based on the approach of Wong and Schwartz. One contribution of our work is a proof that the optimal flow control policy exhibits the properties mentioned above. Last, the optimal control of a system containing two input bridges has been considered in [9].

The remainder of the paper is organized in the following way. The next section describes the four protocols that we study. Section 3 addresses the issue of optimality of policies that delay transfers from the source to output bridges and of policies that choose the most heavily loaded bridge. Approximate analyses of three of the four policies are given in section 4. Comparison of the different policies based on the approximate analyses and simulation are found in section 5 and a summary of the results is found in section 6.

2 Protocols

We consider two or more input bridges attached to low speed LAN's feeding data over a high speed MAN to a single output bridge attached to a low speed LAN. In this section we describe four protocols executed on the bridges that can be used to control the flow between the input bridges and the output bridge. The primary concern of these protocols is to reduce the loss of packets due to buffer overflow. Each protocol is characterized by a rule that determines what packet to transmit at the output bridge and when to transmit it and a second rule that determines what packet and when to transfer it from an input queue to the output queue. We will often refer to the first rule as the *service rule* and the second as the *transfer rule*. The protocols of interest to us are.

- *Largest Queue Delayed Transfer (LQDX)* - This protocol transfers a packet whenever the output buffer is empty and the bridge requires a new packet to serve or when an arriving packet finds a full input buffer and there is available space in the output buffer. In the first case, a packet is chosen from the input buffer with the largest queue length. Note that packet transfers are delayed until the last possible moment under this rule.
- *Largest Queue Earliest Transfer (LQEX)* - This protocol transfers packets when they arrive at an input bridge provided there is space at the output bridge. Once the ouput bridge fills up, packets are then held at the input bridges. Whenever space frees up at the output bridge and there are packets at one or more input bridge, a packet is transferred from the input bridge storing the largest number of packets.
- *Random Earliest Transfer (REX)* - This protocol uses the early transfer rule as in LQEX. However, when space comes available in the output bridge, and packets

reside at the input bridges, an input bridge is chosen randomly from which to transfer a packet.

- *Oldest Customer Earliest Transfer (OCEX)* - This policy uses the early transfer rule as in LQEX and REX. However, when space becomes available at the output bridge and there are packets present at the input bridges, the oldest packet is transferred. Note - this policy ensures that packets that are not lost are transmitted in FIFO.

Wong and Schwartz [10] studied both the LQDX and LQEX protocols. They observed in their study that in the case of a homogeneous system (i.e., identical arrival processes at the input bridges) their solution obtained by solving a Markov decision problem was always LQDX. In the next section we will show that under certain conditions, the LQDX and LQEX policies are optimal in the sense that they reduce the probability of packet loss due to buffer overflow.

We conclude this section with a discussion of some of the implementation issues that must be addressed in the choice of a flow control policy. A major factor that distinguishes these protocols from one another are their respective implementation complexities. There are two components to a flow control strategy. First it must obtain information from the input bridges (sources) and transfer it to the output bridge (sink). Second, a decision must be made when to transfer a packet to the sink and from which source it should come from. Consequently there are several important issues that must be addressed. There are several different types of information that can be used - queue lengths, packet ages - each of which poses certain problems. For example, if the information is packet age, then it is necessary that the bridges synchronize their clocks.

A second issue relates to the reliability of the information used by the sink. For example, if the information used is queue lengths, then its reliability depends on its age. If the propagation delay in the MAN is appreciable, then it may be very inaccurate. A third issue relates to the bandwidth requirements imposed by the protocol on the MAN. If the protocol is LQDX, then it requires every fluctuation in the source queue lengths to be transmitted. On the other hand, if the policy is LQEX, then queue length information is only transferred when the sink is full. The OCEX protocol can piggyback its information onto packets already being transferred from sources to sinks. Last, REX imposes a minimal bandwidth requirement on the MAN. A last issue relates to how buffers should be allocated within a bridge among different functions. Typically a bridge acts as a source for a number of different sinks and also a sink for its associated LAN. We will observe later that it is best to allocate most of the storage to the output function.

3 Optimality Results

We model the system as K input queues labelled $k = 1, \ldots, K$ feeding a single output queue labeled $k = 0$. The input queues are assumed to have capacity B and the output queue capacity B_0. Packets may be transferred from an input queue to an output queue at any time provided that there is sufficient room in the latter. Such transfers are assumed to take zero time. Let $0 < a_1 < \cdots < a_n < \cdots$ be the sequence of arrival times, i.e., the n-th customer arrives at time a_n, and let $\{\tau_n\}_{n=1}^{\infty}$ denote the interarrival times, $\tau_n = a_n - a_{n-1}$, $n = 1, \cdots$, $a_0 = 0$. Let $\{b_n\}_{n=1}^{\infty}$ be a sequence of r.v.s where b_n is the identity of the queue at which the n-th customer arrives, $b_n \in \{1, 2, \cdots, K\}$. Last, $\{\sigma_n\}_{n=1}^{\infty}$ is a sequence of r.v.'s that denote service times, i.e., the n-th customer to be served receives σ_n time units of service.

In addition to the four policies defined in the previous section, we are interested in the following classes of policies.

- Σ - The class of *non-idling* policies where the transfer and service rules are allowed to use any information regarding the system except service times of waiting customers. Here a non-idling policy is one that does not allow the output bridge to be idle whenever there are packets in any of the queues.
- Σ_{DX} - The subset of Σ that contains policies whose transfer policies behave in the following manner. A customer is transferred to Q_0 either when the bridge wants to transmit a packet and the output queue is empty or in order to avoid overflow at some input queue.
- Σ_{EX} - the subset of Σ that contain policies that always transfer packets from input queues to the output queue as soon as they arrive, provided there is space available.

Given a policy $\pi \in \Sigma$, we are interested in the number of customers that are lost by time $t > 0$, $L_\pi(t)$. Our results will be based on the following assumptions.

- **A1** Service times form an independent and identically distributed (i.i.d.) sequence of exponential r.v.'s and $\{\tau_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ are arbitrary sequences of r.v.'s.
- **A2** Service times form an independent and identically distributed (i.i.d.) sequence of exponential r.v.'s and $\{\tau_n\}_{n=1}^{\infty}$ is an arbitrary sequence of r.v.'s. $\{b_n\}_{n=1}^{\infty}$ is an i.i.d. sequence of r.v.'s with $\Pr[b_n = k] = 1/K$, $n = 1, \cdots$, $k = 1, \cdots, K$.

The first result that we establish is that the optimum non-idling policy, under assumption **A1** falls in the class of policies Σ_{DX}. The proof of this result is found in the Appendix.

Theorem 1 *Under assumption* **A1**, *for any policy* $\pi \in \Sigma$, *there exists a policy* $\gamma \in \Sigma_{DX}$ *such that* $L_\gamma(t) \leq_{st} L_\pi(t)$, $t > 0$ *provided that the initial states are the same under each policy.*

Remark. This result corroborates an observation made in [10] that the optimum policy appeared to always delay transfers from an input bridge to the output bridge.

The second result deals with the optimality of LQDX. Again, the proof is found in the Appendix.

Theorem 2 *Under assumptions* **A2**, $L_{LSDX}(t) \leq_{st} L_\pi(t)$, $\forall \pi \in \Sigma$ *provided the system starts in the same state under* π *and LSDX at* $t = 0$.

Remark. This result also corroborates the observation made in [10] that the optimum policy for systems with identical input bridges was LQDX.

—ast, we have a similar result for the class of policies Σ_{EX}. This class of policies is of interest because many existing network protocols, [], belong to it.

Theorem 3 *Under assumption* **A2**, *LQEX minimizes the number of customers lost over the class of policies* Σ_{EX},

$$L_{LQEX}(t) \leq_{st} L_\pi(t), \quad \forall \pi \in \Sigma_{EX}. \tag{1}$$

Proof. The proof is similar to that of theorem 2 and is omitted here. ∎

The results in this section can be extended to the case where the buffer capacity depends on the identity of the input queue. In this case the analog to LQDX is a policy that transfers not from the queue with the largest number of packets, but the one with the least available space.

4 Approximate Analysis

In this section we present simple approximate models for three of the policies described in section 2, LQEX, REX and OCEX. At this point in time we do not have an approximate analysis of LQDX and its evaluation will be performed by simulation in the next section.

All of our analyses assume that the arrival process to each input buffer is Poisson with a common parameter λ. We further assume that packet transmission times at the ouput bridge are exponentially distributed with mean $1/\mu$. Our first model yields a lower bound on the overflow probability for all EX policies. We follow this analysis with approximate analyses of the three EX policies.

4.1 A Lower Bound on the Probability of Overflow for EX Policies

The performance of all EX policies under the above assumptions can be bounded by a M/M/1/k queue with a queue length dependent arrival rate where all packets are allowed to enter so long as there are spaces for at least K packets in the system. Once the number of packets is $N > B_0 + K(B-1)$, no packets are allowed in to $N - B_0 - K(B-1)$ of the input buffers. Hence the queue length dependent arrival rate is

$$\lambda(n) = \begin{cases} K\lambda, & n \leq B_0 + K(B-1), \\ (B_0 + KB - n)\lambda, & B_0 + K(B-1) < n \leq B_0 + KB, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Let $p(n)$ denote the stationary distribution for this system, then the probability of overflow, P_{over} is

$$P_{over} = \sum_{k=0}^{K} p(B_0 + K(B-1) + k) \times k/K \tag{3}$$

4.2 The LQEX Policy

The performance of LQEX is close to the bound described in the previous subsection. This is not surprising as LQEX attempts to equalize the queue lengths at the input queues. The difference between the bound and the performance given by simulation is approximately 10%-20%. In this section we describe a simple approximation to the the behavior of LQEX that typically yields better accuracy. The approximation is given by a M/M/1/k queue where the capacity of the system is taken to be $B_0 + K(B_1)$ and the arrival rate is always $K\lambda$. Let $p(n)$ denote the stationary distribution for this system, then the probability of overflow, P_{over} is estimated to be $P_{over} = p(B_0 + K(B-1))$. We will observe in the next section that this quantity generally overestimates the loss probability for LQEX.

4.3 The REX Policy

We use an approximate analysis first introduced by Chlamtac and Ganz [4] in the context of analyzing finite population random access communication systems where each user has

a finite buffer. Consider the problem of inserting n distinct balls into k urns where each urn has the capacity to store B balls. Let $\vec{N} = (n_1, \ldots, n_k)$ be the resulting occupancy where n_l denotes the number of balls in the l-th urn, $1 \leq l \leq K$. This occcupancy must satisfy $0 \leq n_l \leq B,\ 1 \leq l \leq k,\ \sum_{l=1}^{k} n_l = n$. Let $\mathcal{S}(n,k)$ denote the set of *distinct* occupancies that satisfy these constraints. The number of distinct occupancies is (see [4] for details)

$$C(n,k) = \sum_{j=0}^{k} (-1)^j \left(\begin{array}{c} k \\ j \end{array} \right) \left(\begin{array}{c} n+k-j(B+1)-1 \\ k-1 \end{array} \right). \tag{4}$$

Our approximate analysis assumes that any time there are n customers in the K input queues, the conditional joint probability of the individual occupancies given that the total number is n can be approximated as

$$\Pr[\vec{N} = \vec{N} | \sum_{l=1}^{K} n_l = n] = 1/C(n,K),\ \ \vec{N} \in \mathcal{S}(n,K),\ 1 \leq n \leq KB. \tag{5}$$

Hence, the probability that queue K is full given that the number of packets in the input queues is $n \geq 0$ is

$$P^{full}(n) = \left\{ \begin{array}{ll} 0, & 0 \leq n < B_0, \\ C(n-B, K-1)/C(n,K), & B \leq n \leq KB. \end{array} \right. \tag{6}$$

Last, we model the system as a M/M/1/k queue with a queue length dependent arrival rate.

$$\lambda(n) = \left\{ \begin{array}{ll} K\lambda, & 0 \leq n < B + B_0, \\ (1 - P^{full}(n - B_0))K\lambda, & B + B_0 \leq n \leq KB + B_0, \\ 0, & \text{otherwise.} \end{array} \right. \tag{7}$$

Last, let $p(n)$ is the stationary queue length distribution, then the loss probability is

$$P_{over} = \sum_{n=B}^{KB} P^{full}(n) p(B_0 + n) \tag{8}$$

The expression in equation (4) contains terms with alternating signs. As this can sometimes cause numerical problems, we find the following recursion useful.

$$C(n,k) = \left\{ \begin{array}{ll} 1, & k = 1;\ 0 \leq n \leq B, \\ 0, & k = 1;\ n > B, \\ \sum_{j=0}^{\min(n,B)} C(n-j, k-1), & 2 \leq k \leq K. \end{array} \right.$$

4.4 The OCEX Policy

We use a similar approximation for OCEX as for REX. Briefly, we recognize that by modifying the system so that $B_0 = 0$ and the output bridge transmits a packet out of an input buffer that it belongs to, we obtain a system that can be modeled as a product form queueing system with K customer classes, [2], each with population B. Such a system has the following probability distribution,

$$\Pr[\vec{N} = \vec{N} | \sum n_k = n] = Norm \frac{n!}{n_1! n_2! \cdots n_K!}$$

where again $\vec{N} = (n_1, \ldots, n_K)$ where n_l denotes the number of packets in the l-th input buffer. This occcupancy must satisfy $0 \leq n_l \leq B$, $1 \leq l \leq K$, $\sum_{l=1}^{K} n_l = n$. Let $\mathcal{S}(n, K)$ denote the set of *distinct* occupancies that satisfy these constraints. We define

$$C(n, K) = \sum_{\vec{N} \in \mathcal{S}(n,K)} \frac{n!}{n_1! n_2! \cdots n_K!}$$

Then we have

$$\Pr[\vec{N} = \vec{N} | \sum n_k = n] = (C(n, K))^{-1} \frac{n!}{n_1! n_2! \cdots n_K!}$$

The following recursion can be used to obtain $C(n, k)$,

$$C(n,k) = \begin{cases} 1, & k = 1;\ 0 \leq n \leq B, \\ 0, & k = 1;\ n \geq B, \\ \sum_{j=0}^{\min(n,B)} \binom{n}{j} C(n-j, k-1), & 2 \leq k. \end{cases}$$

The probability that queue K is full given that the number of customers in the input queues is $n \geq 0$ is approximated by

$$P^{full}(n) = \begin{cases} 0, & 0 \leq n < B, \\ \binom{B}{n} C(n - B_K, K-1)/C(n, K), & B \leq n \leq KB. \end{cases} \tag{9}$$

Last, we model the system as a M/M/1/k queue with queue length dependent arrival rate.

$$\lambda(n) = \begin{cases} K\lambda, & 0 \leq n < B + B_0, \\ (1 - P^{full}(n - B_0))K\lambda, & B + B_0 \leq n \leq KB + B_0, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

P_{loss}	Method	$K=3$	$K=5$	$K=7$
$B=4$	Simulation	4.51	2.82	1.89
	Lower Bound	4.19	2.43	1.54
	Approximation	5.01	2.94	1.88
$B=5$	Simulation	3.24	1.85	1.14
	Lower Bound	2.97	1.59	0.93
	Approximation	3.47	1.88	1.12
$B=6$	Simulation	2.46	1.26	0.67
	Lower Bound	2.19	1.08	0.58
	Approximation	2.52	1.26	0.69

Table 1: Validation of approximate analysis for LQEX, $\rho = 0.95$

Once the stationary queue length distribution, $p(n)$, $0 \leq n \leq B_0 + KB$, is obtained, then the loss probability is

$$P_{over} = \sum_{n=B}^{KB} P^{full}(n) p(B_0 + n) \tag{11}$$

5 Results

In this section, we present the main performance results for the four flow control policies. We have presented approximate analyses of three flow control policies in the previous sections for homogeneous networks under Markovian assumptions. We will first validate these approximations, and where the model is sufficiently accurate, apply them along with simulation to study the sensitivity of probability of loss to a variety of parameters including traffic load, $\rho = K\lambda/\mu$, relative buffer allocation among the source and output bridges, and burstiness in the arrival and service processes.

Tables 1-3 compare the total loss probability of the LQEX, OCEX, and REX policies obtained by the approximate analysis and simulation. The 95% confidence intervals are not given as they are always within 2% of the mean. The results have been found to be very accurate for the OCEX and REX policies. The accuracy is not sensitive to either the buffer size or the number of input buffers. The approximation for the LQEX policy loses accuracy as the number of buffers at the source bridges approaches one. Interestingly enough, the loss probability for LQEX is typically within 10%-20% of the lower bound.

Figure 2 illustrates the behavior of the total loss probability as a function of the distribution of buffers between the input and output bridges. The results for LQDX and LQEX have been obtained through simulation and the results for OCEX and REX by the

P_{loss}	Method	$K = 3$	$K = 5$	$K = 7$
$B = 4$	Simulation	5.37	3.65	2.88
	Approximation	5.19	3.66	2.79
$B = 5$	Simulation	3.78	2.49	1.83
	Approximation	3.76	2.51	1.82
$B = 6$	Simulation	2.86	1.76	1.22
	Approximation	2.82	1.79	1.23

Table 2: Validation of approximate analysis for OCEX, $\rho = 0.95$

P_{loss}	Method	$K = 3$	$K = 5$	$K = 7$
$B = 4$	Simulation	5.60	4.17	3.31
	Approximation	5.78	4.34	3.48
$B = 5$	Simulation	4.14	3.06	2.32
	Approximation	4.33	3.14	2.46
$B = 6$	Simulation	3.40	2.22	1.70
	Approximation	3.35	2.37	1.81

Table 3: Validation of approximate analysis for REX, $\rho = 0.95$

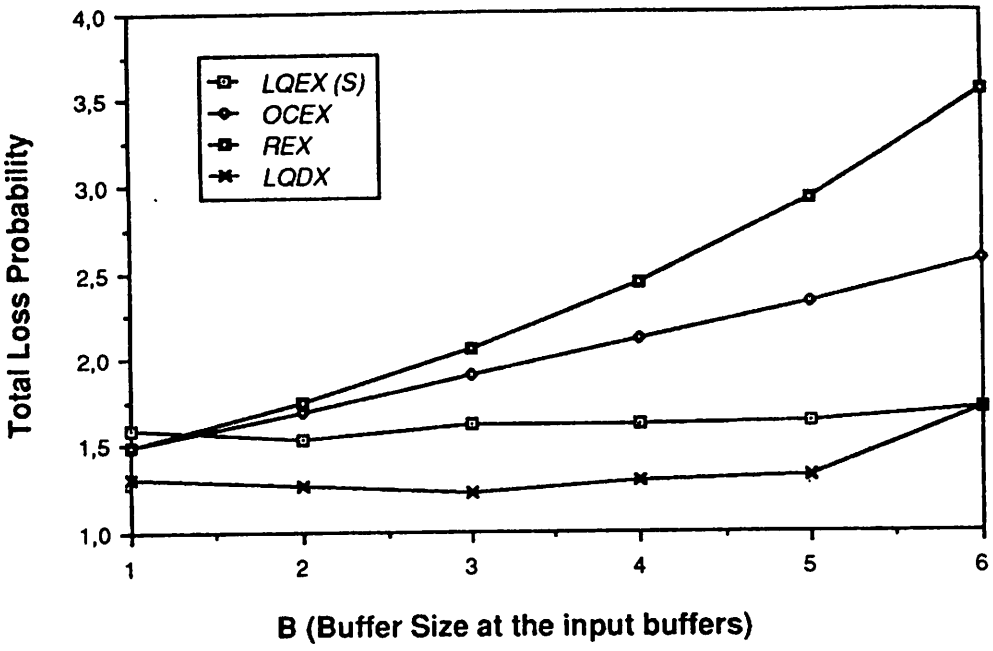

Figure 2: Total loss probability as a function of buffer distribution.

approximate models. The number of source bridges is taken to be five ($K = 5$), the offered traffic load is $\rho = 0.95$, and the total number of buffers distributed among all bridges is taken to be 31. The number of buffers at each source bridge is varied between one and 6 such that $5B + B_0 = 31$. The figure illustrates that there is little difference between the four policies when most of the buffers are allocated to the output bridge under our assumptions. This is because there are very few scheduling decisions made that exercise the rules that distinguish the different policies. As we increase the buffer space at the source bridges, we increase the number of scheduling instances at which the policies may differ from each other. We observe that the REX policy exhibits the worse performance and that OCEX exhibits a behaviour approximately halfway in between REX and LQEX. We also observe that neither LQEX nor LQDX are sensitive to the buffer allocation. Finally, it is of interest to observe that a simple policy (regarding implementation complexity) can lead to results close to the optimal policy (LQDX) provided that most of the buffers are allocated to the output bridge. We would recommmend that the minimal number of buffers be allocated to each input bridge as is required for error recovery. Depending on the error recovery mechanism (if any) used over the MAN, this may be as few as one or two. The remaining buffers should be allocated to the output bridge. In Figure 3, we present the total loss probability as a function of the offered traffic load, ρ, for the different policies for a system with 5 source bridges and storage capacity at each bridge of 4, $B = B_0 = 4$. In this case, the results for LQDX have been obtained through simulation

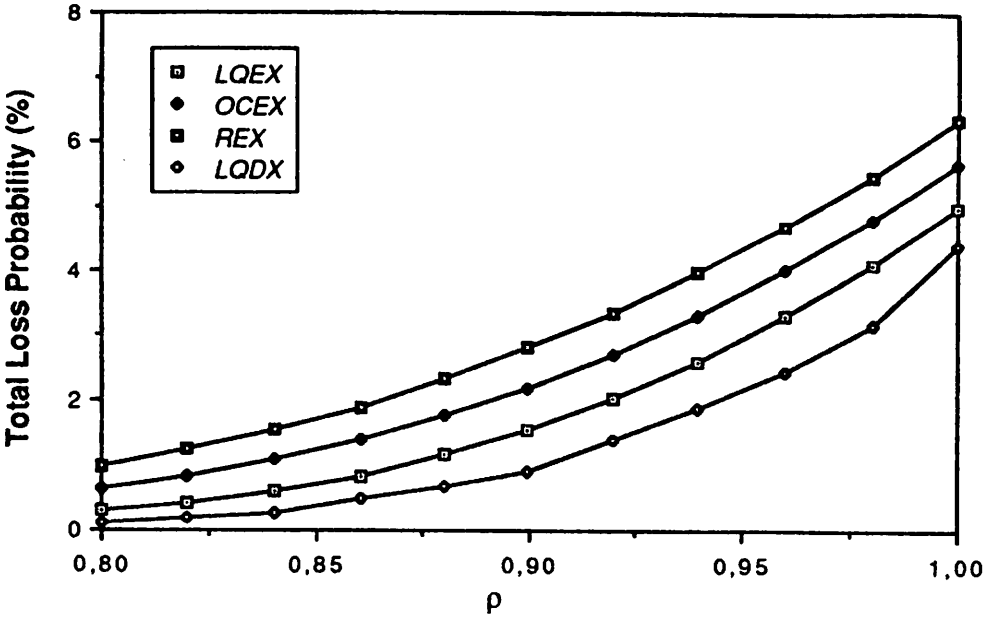

Figure 3: Total loss probability vs. offered load, $K = 5$, $B = B_0 = 4$.

and the results for the remaining policies through the approximate models reported in the last section. We observe a gradual falloff in performance as a function of the traffic load. We also observe that using packet age information provides half of the benefit of instantaneous queue length information over a purely random policy. The results presented so far have been obtained under the assumption of exponentially distributed interarrival times and service times. Figures 4 and 5 display the total loss probability for bursty arrivals (hyperexponential interarrival time distribution with a coefficient of variation equal to 2) and constant service times for the output bridge for the LQDX and OCEX policies. As before, the results are for 5 input bridges and a offered traffic load of $\rho = 0.95$. The iegend X/Y depicts the distribution of the arrival process at an input bridge (M or H2), and the service time distribution at the output bridge (M or D). Both policies show increasing loss probability when the variance of the input arrival or output service distribution increases. The impact of changing the distribution of either the arrival process or the service process is similar for both policies. Although not displayed here, the LQEX and REX policies show similar behavior. The results presented so far are for homogeneous networks. The next two figures illustrate the behavior of both the loss probabilities and the mean delays at each source bridge when there are 5 sources, storage capacity of 4 at each bridge, $B = B_0 = 4$, Poisson arrivals to the sources with a rate that depends on the input source. Specifically we assume that the arrival rate is a *geometric function* of the source bridge identity, $\lambda_k = \lambda_1 h^k$ where h is a parameter that can be

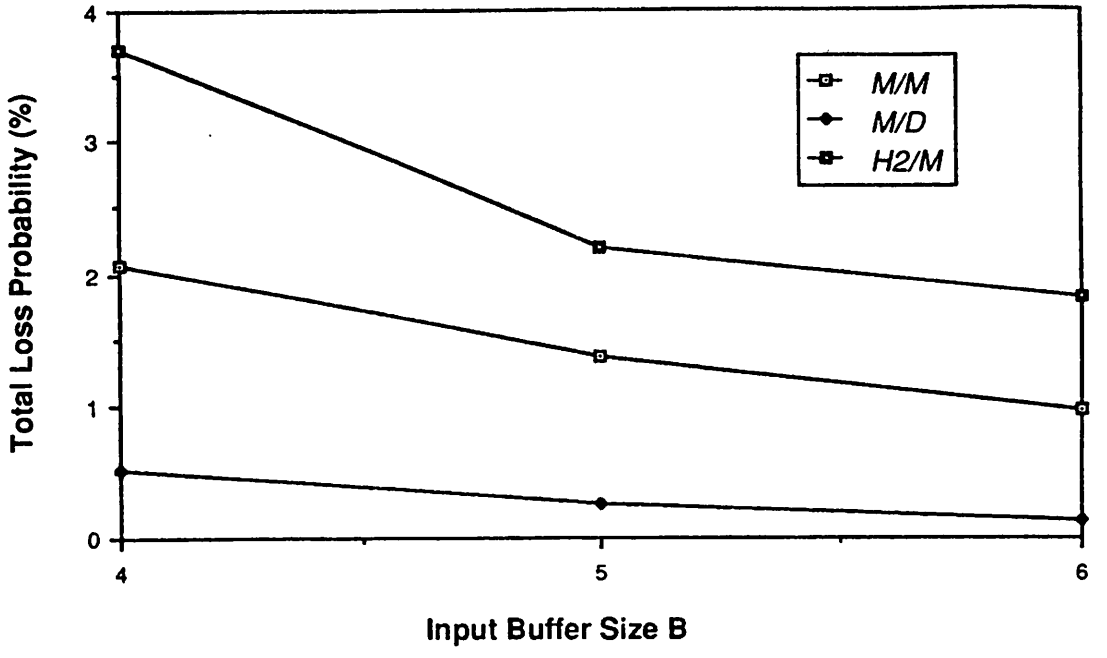

Figure 4: The effect of burstiness in the arrival process and determinism in the service process on loss probability for LQDX, $K = 5$, $\rho = 0.95$.

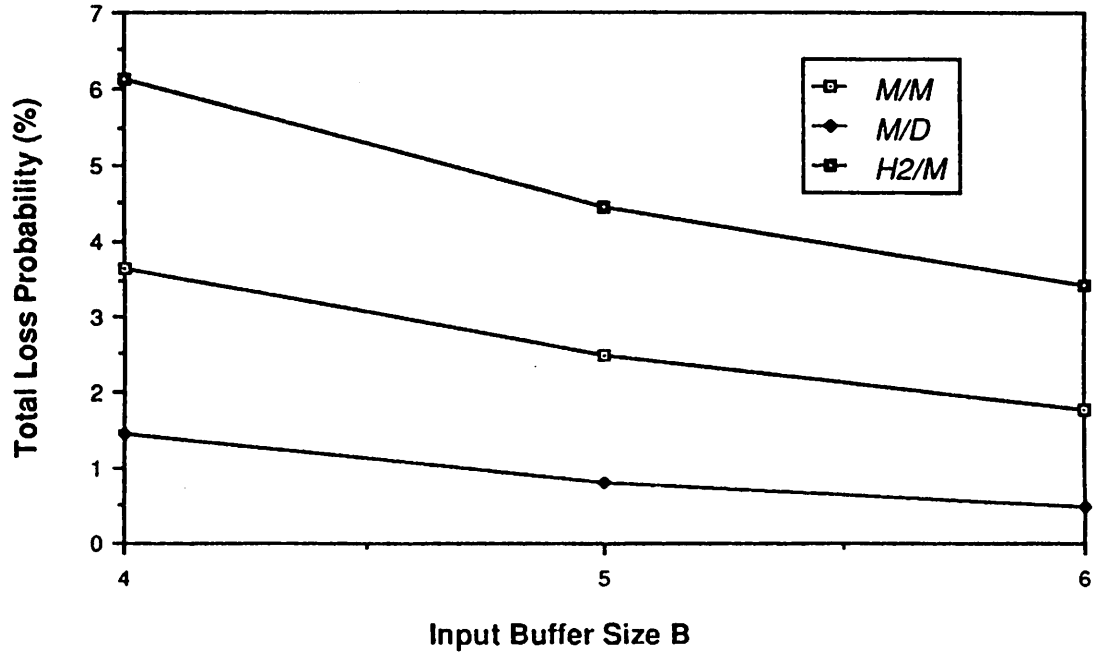

Figure 5: The effect of burstiness in the arrival process and determinism in the service process on loss probability for OCEX $K = 5$, $\rho = 0.95$.

Policy	Avg. Delay	Total Loss Prob.
LQDX	4.77	2.79
LQEX	2.31	4.16
OCEX	1.63	5.47
REX	1.43	6.20

Table 4: Aggregate performance for Heterogeneous system.

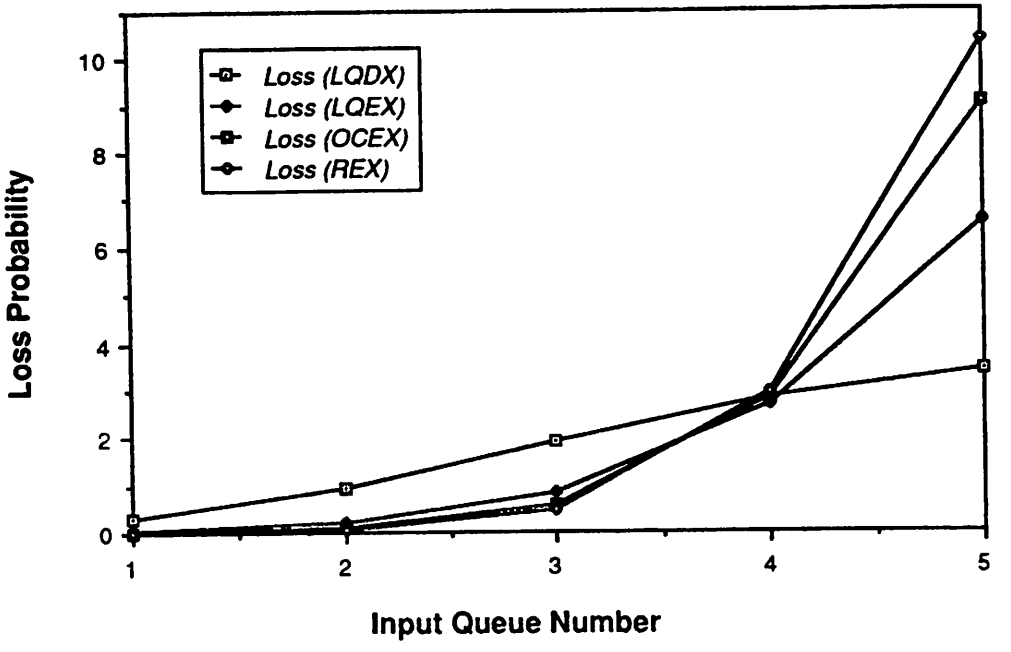

Figure 6: Loss probability vs. source bridge identity, $K = 5$, $B = B_0 = 4$.

chosen to control the degree of heterogeneity in the system. Figure 6 illustrates the total loss probability and figure 7 the mean packet delay for each source bridge for the different policies with $h = 2$ and $\sum_{k=1}^{5} \lambda_i/\mu = 0.95$. We observe that when the performance metric is loss probability, the best policy is LQDX followed by LQEX, OCEX and REX. On the other hand, when the metric is average delay, the best policy is REX followed by OCEX, LQEX, and LQDX (Table 4). Consequently, there is a mean delay/loss tradeoff that must be considered in the selection of an appropriate policy. However, it is our belief that a policy should be chosen based on loss probability - not average delay. It is also interesting to observe that LQDX is considerably fairer than the other policies with respect to loss probabilities among the queues.

6 Summary

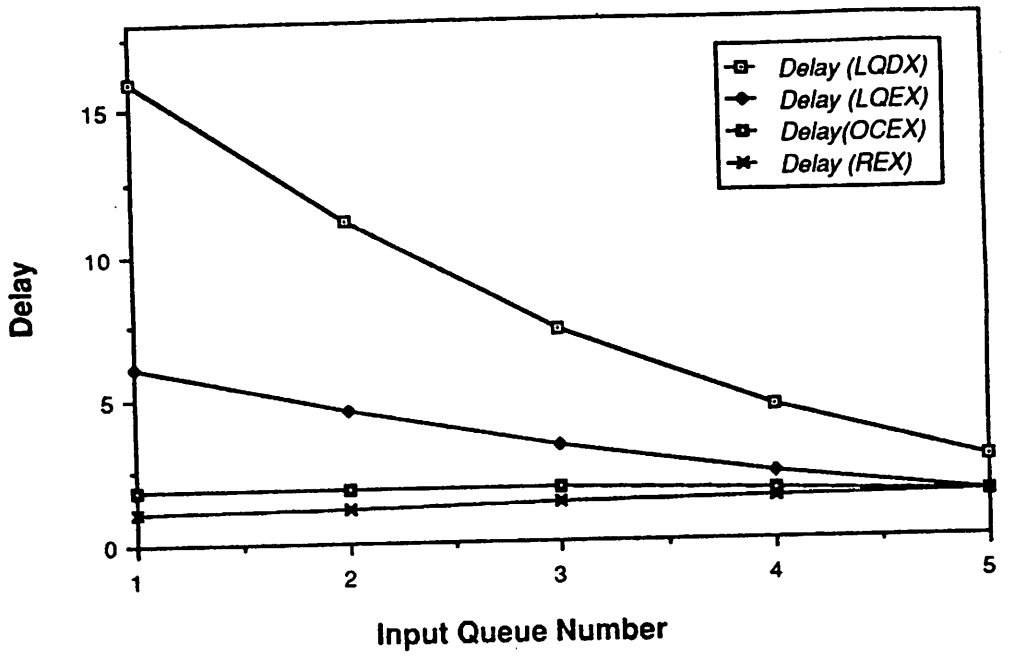

Figure 7: Avg. packet delay vs. source bridge identity, $K = 5$, $B = B_0 = 4$.

In this paper we have addressed the problem of flow control over a MAN that interconnects low speed LANs. We have studied the behavior of four policies that differ according to the type of information required. Using as a metric the probability of buffer overflow we find that policies that use queue length information perform best, followed by a policy that uses packet age information and, last, by a policy that uses no information. All of the policies perform best when most buffer space is allocated to the output bridge. In this case there is little difference between the policies in terms of their performance. We also observe that the two policies that use queue length information are relatively insensitive to the buffer allocation. Last, in the case of a heterogeneous system, we find that the ordering among the policies remain unchanged with respect to loss probability but are reversed with respect to mean packet delay. A number of issues remain to be addressed. Some of these are - non-negligible MAN delays, the effect of error recovery mechanisms on our results, and a more detailed understanding of the effects of heterogeneity.

References

1. *ANSI, FDDI Token RIng Media Access Control*, Draft Proposed American National Standard X3T9.5 (ISO/DIS 9314).

2. F. Baskett, K.M. Chandy, R.R. Muntz, F.G. Palacios, "Open, Closed and Mixed Networks of Queues with Different Classes of Customers", *J. ACM*, 22, pp. 248-260, 1975.

3. W. Bux, D. Grillo, "Flow Control in Local Area Networks of Interconnected Token Rings", *Advances in Local Area Networks*, (ed., K. Kummerle, J.O. Limb, F.A. Tobagi),

IEEE Press, 1987.

4. A. Ganz, I. Chlamtac, "A Linear Solution to Queueing Analysis of Finite Buffered Networks: Part I - Synchronous Communication Systems", *Proc. 2-nd Intntl. Workshop on Appl. Math. and Perf./Rel. Models of Comp./commun. Systems*, Rome Italy, 1987.

5. *Draft of proposed IEEE standard 802.6 DQDB MAN Media Access Control and Physical Layer Protocol Documents*, Aug. 1989.

6. M. Gerla and L. Kleinrock, "Flow Control: A Comparative Survey", *IEEE Trans. Communications*, **COM-28**, pp. 533-574, 1980.

7. R. Jain, "A Timeout-Based Congestion Control Scheme for Window Flow-Controlled Networks", *IEEE J. Sel. Areas on Commun.*, **SAC-4**, 7, pp. 1162-1167, Oct. 1986.

8. A.W. Marshall, I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, 1979.

9. Y.B. Suk, C.G. Cassandras, "Optimal Scheduling of Two Competing Queues with Blocking", *Proc. 27-th IEEE Conf. Decision and Control*, pp. 1102-1107, Dec. 1988.

10. Wong, M. Schwartz, "Flow Control in Metropolitan Area Networks", *Proc. INFOCOM'89*, 1989.

Appendix - Proof of Optimality Results

In addition, to the quantity $L_\pi(t)$, we are interested also in the following performance measures.

- $D_\pi(t)$ - the number of service completions by time t under policy π.
- $N_\pi(t,k)$ - the number of customers in the k-th buffer, $k = 0, 1, \cdots, K$.
- $N_\pi(t) = \sum_{k=0}^{K} N_\pi(t,k)$.
- $\hat{N}_\pi(t,k)$ - the number of customers in the input buffer with the k-th smallest number of customers, $k = 1, \cdots, K$.
- $\vec{N}_\pi(t) = (N_\pi(t,1), \ldots, N_\pi(t,K))$.
- $A_\pi(t,k)$ - the available space in the k-th buffer, $A_\pi(t,k) = B_i - N_\pi(t,k)$, $k = 0, 1, \cdots, K$.
- $\tilde{A}_\pi(t,k)$ - the space in the input buffer with the k-th most available space, $k = 1, \cdots, K$.
- $\vec{A}_\pi(t) = (A_\pi(t,1), \ldots, A_\pi(t,K))$.

Proof of theorem 1:

Theorem 1 *Under assumption* **A1**, *for any policy* $\pi \in \Sigma$, *there exists a policy* $\gamma \in \Sigma_{DX}$ *such that* $L_\gamma(t) \leq_{st} L_\pi(t)$, $t > 0$ *provided that the initial states are the same under each policy.*

Proof. Policy γ behaves in the following manner. First, it keeps track of the behavior of π, i.e., what the values of $N_\pi(t,k)$, $k = 0, 1, \cdots, K$ are given the sequence of arrivals and service times up until time t. Second, it transfers a customer from an input queue to the output queue either if the input queue is about to overflow (if there is space in the output queue) or when the server is idle and a customer is to be scheduled from one of the input queues (the output queue is empty). In the latter case, one of the following rules is used to transfer a customer to the output queue.

1. If $N_\pi(t,0) = 0$, then γ emulates π.
2. If $N_\pi(t,0) > 0$, then γ selects a customer from some input queue k such that $N_\pi(t,k) < N_\gamma(t,k)$. The existence of such a k will be shown below.

The reader should observe that, although π may not use queue length information, γ may be required to in order to emulate π.

We consider a given sequence of arrival times, queue selections, and service times and establish the following relations for that input sample,

$$L_\gamma(t) \leq L_\pi(t), \tag{12}$$

$$N_\gamma(t,k) \geq N_\pi(t,k), \quad k = 1, \cdots, K, \tag{13}$$

$$N_\gamma(t) \geq N_\pi(t). \tag{14}$$

The assumption that service times are i.i.d. exponentially distributed r.v.'s is required to couple service completions under both policies. The independence assumption is required because the coupling may require that two different packets may be assigned the same service times under the two policies and the exponential assumption is required because a new customer under π may receive the remaining service time of a customer already in service under γ.

The proof is by induction on the sequence of different events that can occur under both policies. Let $t_0 = 0, t_1, \cdots, t_n, \cdots$ denote the times of the events corresponding to arrivals and service completions. Clearly if relations (12)-(14) hold at time t_n, then they must hold for $t_n < t < t_{n+1}$, provided $t_n \neq t_{n+1}$, $n = 1, \cdots$.

Basis step. If the two policies begin in the same state, then the relations must hold for $t = t_0$.

Inductive Step. Assume that the relations hold for $t = t_n$. We show that they also hold for t_{n+1}. We consider each event separately.

i) Service completion. If there remain no customers in the system under π after the service ompletion, then relations (12)-(14) trivially hold for $t = t_{n+1}$. We assume that $N_\pi(t_n) > 1$. In this case it follows from the inductive hypothesis that $N_\gamma(t_n) > 1$. Hence it is possible to schedule a job under both policies. If $N_\gamma(t_n, 0) > 0$ and $N_\pi(t_n, 0) > 0$, then the queue length of the output queue decreases by one under both policies and relations (12)-(14) hold for $t = t_{n+1}$. Similarly, the relations hold if π selects a customer from the k-th input queue.

The last case occurs when $N_\pi(t_n, 0) > 1$ and $N_\gamma(t_n, 0) = 1$. In this case, relation (14) ensures that there is at least one input queue k such that $N_\pi(t_n, k) < N_\gamma(t_n, k)$. Hence, γ can select a customer from that queue and relations (12)-(14) will hold for $t = t_{n+1}$.

ii) Customer arrival. If there is either a loss under π, or no loss under both policies, then the relations hold for $t = t_{n+1}$. Hence the interesting case is a loss under γ. It follows that $N_\gamma(t, 0) = B_0$ and that $N_\gamma(t, 0) + N_\gamma(t, k) > N_\pi(t, 0) + N_\pi(t, k)$. This along with relation (13) guarantee that $L_\gamma(t_n) < L_\pi(t_n)$ and consequently, relations (12)-(14) hold for $t = t_{n+1}$.

This completes the inductive step and so relations (12)-(14) hold for $t > 0$. Removal of the conditioning on the interarrival times, queue selections, and the service times yields the desired result. ∎

Proof of theorem 2:

The proof of the optimality of LQDX requires the comparison of two vectors $\vec{X}, \vec{Y} \in I\!N^K$. Hence we define the concept of *weak majorization* which is discussed in full detail in [8].

Definition 1 *Vector* $\vec{X} = (X_1, \cdots, X_K)$ *is said to weakly majorize vector* $\vec{Y} = (Y_1, \cdots, Y_K)$ *(written* $\vec{Y} \prec_w \vec{X}$*) iff*

$$\sum_{i=1}^{k} \tilde{Y}_i \leq \sum_{i=1}^{k} \tilde{X}_i, \quad k = 1, \cdots, K$$

where the notation $\tilde{X}_i$ *is taken to be the* i*-th largest element of* $\vec{X}$.

Lemma 1 *If* $\vec{Y} \prec_w \vec{X}$, *then*

$$\begin{aligned}
(\tilde{Y}_1, \cdots, \tilde{Y}_{K-1}, \tilde{Y}_K + 1) &\prec_w (\tilde{X}_1, \cdots, \tilde{X}_{k-1}, \tilde{X}_k + 1, \tilde{X}_{k+1}, \cdots, \tilde{X}_K), \\
(\tilde{Y}_1, \cdots, \tilde{Y}_{K-1}, \tilde{Y}_K + 1) &\prec_w \vec{X} \text{ if } |\vec{X}| > |\vec{Y}|, \\
(\tilde{Y}_1, \cdots, \tilde{Y}_{k-1}, \tilde{Y}_k - 1, \tilde{Y}_{k+1}, \cdots, \tilde{Y}_K) &\prec_w (\tilde{X}_1, \cdots, \tilde{X}_{k-1}, \tilde{X}_k - 1, \tilde{X}_{k+1}, \cdots, \tilde{X}_K), \\
\vec{Y} &\prec_w (\tilde{X}_1, \cdots, \tilde{X}_{k-1}, \tilde{X}_k + 1, \tilde{X}_{k+1}, \cdots, \tilde{X}_K).
\end{aligned}$$

Proof. Properties 1, 2, and 4 follow in a straightforward manner from the defiinition of weak majorization. Property 3 corresponds to Lemma 5.D.2 in [8, p. 135]. ∎

Theorem 2 *Under assumptions* **A2**, $L_{LSDX}(t) \leq_{st} L_\pi(t)$, $\forall \pi \in \Sigma$ *provided the system starts in the same state under* π *and LSDX at* $t = 0$.

Proof. Given a specific sequence of interarrival times, queue selections, and service times, we establish the following relations,

$$\vec{A}_\gamma(t) \prec_w \vec{A}_\pi(t), \tag{15}$$

$$N_\pi(t) \leq N_\gamma(t), \tag{16}$$

The proof is by induction on the sequence of times corresponding to arrivals and service completions. Let these times be $t_0 = 0, t_1, \cdots, t_n, \cdots$. We will couple service completions in the same manner as the last theorem. In addition, we will couple arrivals to the queue with the k-th largest number of packets under both policies regardless of whether they correspond to the same physical queue. We are allowed to do so by the assumption that the queue selections form an i.i.d. sequence and that each input queue is equally likely to be chosen for an arrival.

Basis Step. The theorem is obviously true for t_0 since the policies begin with the same initial state.

Inductive Step. Assume that relations (15) and (16) hold for $t \leq t_n$. We have two cases according to the type of event.

i) Service Completion. If $N_\pi(t_n) = 1$, then relations (15) and (16) are trivially satisfied. Consider the event $N_\pi(t_n) > 1$. Then according to the inductive hypothesis, $N_\gamma(t_n) > 1$. There are four subcases according to the number of customers in Q_0 under both policies.

a) $N_\pi(t_n, 0) > 1$, $N_\gamma(t_n, 0) > 1$. In this case relations (15) and (16) are easily shown to

hold.

b) $N_\pi(t_n,0) > 1$, $N_\gamma(t_n,0) = 1$. Policy γ will select a customer from the input queue with the largest number of packets. In this case, property 2 of lemma 1 can be applied to show that relations (15) and (16) hold for $t = t_{n+1}$.

c) $N_\pi(t_n,0) = 1$, $N_\gamma(t_n,0) = 1$. In this case, property 1 of lemma 1 can be applied to show that relations (15) and (16) hold for $t = t_{n+1}$.

d) $N_\pi(t_n,0) = 1$, $N_\gamma(t_n,0) > 1$. Property 4 of lemma 1 ensures that relation (15) holds at $t = t_{n+1}$.

iii) Arrival. Assume that the arrival is to the queue with the k-th largest space available under both policies. This may not correspond to the same physical queue, however, assumption **A2** ensures that we can couple the arrivals to these queues. There are three subcases according to whether a buffer overflow occurs or not. Here overflow corresponds to the buffer being full. Whether a packet is lost or not depends on whether the output queue is full or not.

a) No buffer overflow. In this case, property 3 of lemma 1 ensures that relation $\vec{A}_\gamma(t_{n+1}) \prec_w \vec{A}_\pi(t_{n+1})$. Clearly $N_\pi(t_{n+1}) = N_\gamma(t_{n+1})$.

b) Buffer overflow under γ but not π. In this case $\hat{N}_\gamma(t_n,k) = B$ and $\hat{N}_\pi(t_n,k) < B$. The following relation follows from this and the inductive hypothesis,

$$\sum_{i=1}^{j} \tilde{A}_\pi(t_n,i) > \sum_{i=1}^{j} \tilde{A}_\gamma(t_n,i), \quad j = k,\cdots,K. \tag{17}$$

It follows from the addition of the customer to input queue k under π that $\vec{A}_\pi(t_{n+1}) \succ \vec{A}_\gamma(t_{n+1})$. If there is room in Q_0 for a customer from queue k under policy γ then clearly $N_\pi(t_{n+1}) \leq N_\gamma(t_{n+1})$. Suppose that Q_0 is full under γ at time t_n. Then it follows from equation 17 that $N_\pi(t_n) < N_\gamma(t_n)$ which again implies $N_\pi(t_{n+1}) \leq N_\gamma(t_{n+1})$.

c) Buffer overflow under π. Relations (15)-(16) are shown to trivially hold in this case for $t = t_{n+1}$.

d) Buffer overflow under π and γ. Again, similar arguments can be applied here, taking into account that there may or not be a customer loss under either policy.

This completes the induction step. Consequently relations (15)-(16) hold for $t > 0$ for a given input sample.

Let $C_\pi(t)$ denote the number of completions under policy $\pi \in \Sigma$. It follows from relations (15)-(16) that $C_\pi(t) \leq C_\gamma(t)$, $t > 0$. Now, $N_\pi(t) + L_\pi(t) + C_\pi(t)$ is independent of the policy π. Consequently, relations (15)-(16) along with this fact imply that $L_\pi(t) \geq L_\gamma(t)$.

Hence, the theorem holds when the conditioning is removed on the service times, interarrival times and queue selections. ∎