Performance Measures and Scheduling Policies in Ring Networks

Leandros Tassiulas, *Member, IEEE,* and Jinoo Joung

***Abstract*—A unidirectional ring network is considered. A node may transmit at most one packet per slot to its downstream neighbor. Potentially all nodes may transmit at the same slot. The achievable performance is studied and policies are proposed for both the evacuation mode and continual operation. In the evacuation mode each node has initially an amount of packets destined for every other node of the ring, and no more packets are generated later. It is shown that the furthest destination first (FDF) policy, that gives priority to the packet with longest way to go at each node, minimizes the time until every packet reaches its destination. Furthermore it is shown that the closest destination first (CDF) policy, that gives priority to the packet with the shortest way to go at each node, minimizes the average packet delivery time. A formula for the optimal evacuation time is obtained. The continual operation of the ring is considered then where packets are generated according to some arrival process. It is shown that, for any arrival sample path, the FDF maximizes the fraction of the time at which the ring is empty. The performance analysis of individual origin-destination traffic streams under FDF is facilitated based on the following. For each traffic stream, a single server priority queue is identified such that the average sojourn time of the traffic stream in the ring is equal to the aggregate transmission time plus the queueing delay of the low priority stream in the queue. Formulas for the sojourn time are obtained for i.i.d. arrivals. The performance of CDF and FIFO in continual operation is studied by simulation. It turns out that the CDF, minimum delay policy for the evacuation, has the worst performance in continual operation.**

I. Introduction

The RING is a fairly common communication network architecture for local area networks (LAN) as well as processor interconnection networks. As the transmission speed increases, simultaneous transmissions in nonoverlapping segments of the network (spatial reuse) should be facilitated in order to maintain certain levels of efficiency. Several new ring architectures have been proposed in order to provide the desired levels of spatial reuse. Those architectures include the slotted ring, register insertion ring, etc., [3], [4], [8]. There has been considerable activity lately on studying the stability and the achievable throughput in these networks [6], [7], [12]. Our goal in this paper is to study the interaction of different traffic streams and the effect on performance. We adopt a simple ring network model that captures these interactions. The achievable performance under different measures is investigated and certain policies are identified in evacuation mode as well as continual operation.

A unidirectional ring is considered with N nodes numbered $0, 1, \cdots, N-1$. Each node i may transmit to its downstream neighbor $i \oplus 1$[1] and receive from its upstream neighbor $i \ominus 1$ simultaneously. We assume that the packet length is fixed and the actual transmission time plus the propagation delay and communication startup overhead is equal to one time slot. The slot length is the same for all nodes and they are synchronized to begin transmission in the beginning of the slot. In one slot, each node may transmit at most one packet to its immediate downstream neighbor. A packet may be forwarded by one hop per slot. A transmission scheduling policy determines the transmission priorities at each node.

A number of packets x_{ij} is residing in node i at $t = 0$ to be delivered to node j. That is $x = (x_{ij}, , i, j = 0, \cdots, N-1)$ is the initial backlog vector. Let $D_i^{x,\pi}$ be the delivery time of the ith packet under policy π when the initial state is x and there are no further packet arrivals generated after $t = 0$. When there is no ambiguity, the superscripts x, π are dropped from the notation later. The performance of the policy for the evacuation of the network, is fully specified by the delivery time vector $D^{x,\pi} = (D_i^{x,\pi}, i = 1, \cdots, K)$ where $K = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} x_{ij}$ is the total number of packets in the system at $t = 0$.

One performance measure associated with a scheduling policy is the *maximum delivery time*

$$V(x, \pi) \stackrel{\text{def}}{=} \max_{i=1,\cdots,K} \{D_i^{x,\pi}\}$$

which is also called *schedule length, makespan* or *evacuation time*. The schedule length has been used widely to evaluate the performance of communication algorithms in parallel architectures. It is also referred to as the communication complexity of the algorithm in this context.

Certain standard communication tasks that are considered in the evaluation of the performance of communication algorithms in parallel architectures, involve the minimization of the evacuation time for certain types of initial conditions [1], [2], [5], [10]. The *scattering* problem, as it is considered in [2], [5], [10], is the evacuation of the network in minimum time when only one node has packets at the beginning of time. The *gathering* problem [2], [10] is the evacuation of the network in minimum time when all packets in the network have a

Manuscript received July 7, 1994; revised September 17, 1994 and January 24, 1995; approved by IEEE/ACM Transactions on Networking Editor R. Guerin. This work was supported in part by the NSF under Grants NCR-9211417, NCR-9406415, and AFOSR F49620-95-1-0061. This paper was presented in part at the 25th Conference on Information Sciences and Systems.

L. Tassiulas is with the Department of Electrical Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA.

J. Joung is with the Department of Electrical Engineering, Polytechnic University, Brooklyn, NY 11201 USA.

IEEE Log Number 9413192.

[1] We define $i \oplus j = (i + j)\text{mod}N$ and $i \ominus j = (i - j)\text{mod}N$.

1063-6692/95$04.00 © 1995 IEEE

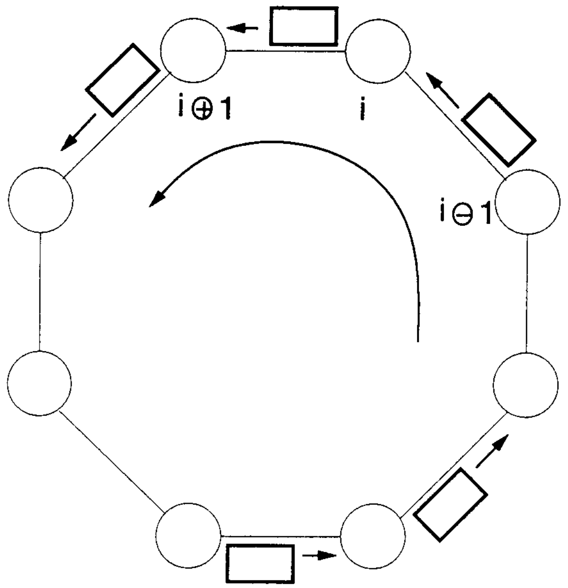


Fig. 1. A ring network with 8 nodes is depicted. At each slot, each node i may transmit at most one packet to its immediate downstream neighbor $i \oplus 1$.

common destination. The *broadcasting* problem, that is deliver a message from node i to every other node of the network, is equivalent to the problem of node i sending a message to itself, in the case of a unidirectional ring. The *total exchange* problem is equivalent to having each node sending a packet to every other node. A detailed account of all the different communication tasks can be found in [2], [10].

In this paper, in addition to the evacuation time, we consider another performance measure namely the *average delay* which is defined as

$$D(x,\pi) \stackrel{\text{def}}{=} \frac{1}{K}\sum_{i=1}^{K} D_i^{x,\pi}.$$

Initially we focus on the evacuation problem where there is an initial backlog and no new arrivals in the system. We show that the FDF and CDF policies minimize the evacuation time $V(x,\pi)$ and average delay $D(x,\pi)$ respectively, for every initial state. Note that the FDF policy has been considered in [10] for the scattering in a ring of processors. It was proven that it is optimal for the scattering operation in [5] where its complexity was also obtained.

Both the evacuation time and the average delay are sensible performance measures for the evacuation operation. Furthermore, as it is demonstrated later, the policy that is optimal with respect to one measure, may be suboptimal with respect to the other, for certain initial conditions. Hence there is no single policy that is optimal with respect to both measures for all initial conditions.

The continual operation of the system, where packets are generated in the nodes according to some arrival processes during the operation of the system, is considered next. The FDF policy is studied and it is shown that it retains its optimal evacuation time in continual operation as well, in the sense that it minimizes the fraction of time at which the ring is nonempty. Regarding the evaluation of the performance of FDF, we show that the average sojourn time of a source-destination session in the ring is equal to the total transmission time plus the average queueing delay experienced by a traffic stream in an equivalent single server queue with priorities. This result readily provides closed formulas for the sojourn times in the case of i.i.d. arrivals, while it considerably facilitates the performance evaluation for other arrival processes. The CDF policy is studied by simulation. In addition to FDF and CDF, the first in–first out (FIFO), first arrival–first out (FAFO) and maximum sojourn-time first (MSTF) policies (to be defined later) are studied by simulation.

The paper is organized as follows. In Section II the evacuation problem is considered and the FDF, CDF policies are studied for this problem. In Section III the FDF policy is considered in continual operation and some optimality properties are identified. In Section IV, performance analysis and comparisons of the CDF and the other policies are done.

II. Optimal Evacuation

The system starts from some initial state and there are no further arrivals. The following two problems are considered.

- **Minimum Evacuation Time:** Find a scheduling policy π^* that achieves minimum evacuation time for any initial state x. That is,

$$V(x,\pi^*) = V(x) \stackrel{\text{def}}{=} \min_{\pi} V(x,\pi).$$

- **Minimum Delay:** Find a scheduling policy π^* that achieves minimum average delay for any initial state x. That is,

$$D(x,\pi^*) = D(x) \stackrel{\text{def}}{=} \min_{\pi} D(x,\pi).$$

A. Evacuation Time and FDF Scheduling

Consider the furthest destination first policy where every node i gives priority to the packet the destination of which is furthest away from i (the packet with the largest cumulative remaining service time) among those residing in i. Packets with the same destination are served arbitrarily.

Theorem 1: For any initial state x, the FDF policy minimizes the evacuation time.

Proof: Theorem 1 follows from Theorem 4 which is a more general result shown for the continual operation of the system in Section III. Theorem 1 is stated separately here for the sake of completeness. ∎

If the ring has the cut-through capability, it has been shown in [6] that the evacuation time is the same for every work-conserving policy. The minimum evacuation time for the cut-through case is certainly smaller than the store-and-forward case considered here. The difference of the two evacuation times though cannot be larger than a quantity which is independent of the size of the backlog and depends only on the number of nodes. This is a consequence of a more general result shown in [11]. Hence, the evacuation time of any work conserving policy becomes asymptotically optimal in the store-and-forward case as well, when the backlog increases. It is

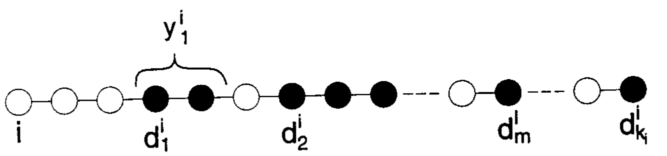


Fig. 2. The node d_j^i is the jth upstream node of $i$that contains packets waiting to cross i while $d_j^i \oplus 1$ does not contain any such packet. The number of packets in nodes $d_j^i, \cdots, d_{j+1}^i \oplus 2$ that need to cross i is y_j^i.

worth noting at this point that the average delay may vary for different work-conserving policies even if the ring has cut-through capability.

The optimal evacuation time $V(x)$, achieved under the FDF policy, is obtained in the following. For every node i consider all the packets that need to cross i in order to reach their destinations, or in other words all the packets with origin node $i \ominus l$ and destination node $i \oplus m$, $l \geq 0$, $m \geq 1$ such that $m + l \leq N - 1$ where N is the number of nodes in the ring. When for every node i all such packets cross i, the network is clearly empty. We obtain first the time at which all packets that need to cross some node i do so. Let d_j^i be the jth upstream node of i with the property that, at $t = 0$, node d_j^i has a packet that needs to cross i, while node $d_j^i \oplus 1$ has no such packet. If i is nonempty at $t = 0$ then by definition $d_1^i = i$. Assume that there are k_i such nodes. The notation is illustrated in Fig. 2. Let y_j^i be the number of packets that need to cross i and at $t = 0$ are residing in the nodes $d_j^i, d_j^i \ominus 1, \cdots, d_{j+1}^i \oplus 2$. Let V^i be the time at which all those packets will cross node i. The following lemma provides V^i.

Lemma 1: The time V^i is equal to $V_{k_i}^i$ where $V_j^i, j = 1, \cdots, k_i$ are defined recursively as follows.

$$V_1^i = (i \ominus d_1^i) + y_1^i,$$
$$V_j^i = \max\{V_{j-1}^i, (i \ominus d_j^i)\} + y_j^i. \tag{1}$$

Proof: Observe that under the FDF policy, any packet that has to cross i to reach its destination, at every node before i, has priority over any packet that does not have to cross i to reach its destination. Hence, V^i is equal to the evacuation time of the tandem network obtained if we cut the ring just after node i and we let in the initial state only those packets of the ring that need to cross i. Note that the evacuation time of the tandem is the same for all work-conserving policies.

Focusing on the tandem for node i now, if we assume that the packets are served FIFO, then V_j^i is interpreted as the time at which all the packets at the first j groups of contiguous nonempty nodes, reach their destination. The validity of formula (1) then follows inductively. Clearly the relation for V_1^i holds. The recursive formula for V_j^i follows if we observe that the last packet of group j will leave the network y_j^i slots after the first packet of this group leaves the network. The first packet of this group reaches node i after traveling the distance of $i \ominus d_j^i$ links and after all the packets in the group $j - 1$ leave the network, something that will happen at time V_{j-1}^i. Hence the last packet of group j will reach the destination at time $\max\{V_{j-1}^i, (i \ominus d_j^i)\} + y_j^i$. ∎

The following theorem provides the optimal evacuation time.

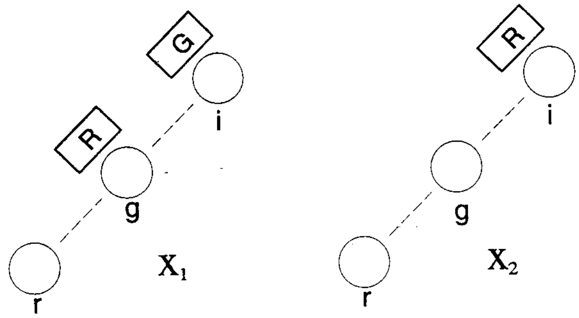


Fig. 3. The initial conditions for Lemma 3.

Theorem 2: The optimal evacuation time $V(x)$ from some initial state x is

$$V(x) = \max_{i=0,\cdots,N-1}\{V^i\}.$$

Proof: Note that at any time $t < V^i$ there is some packet, among those that need to cross i to reach their destination, that has not crossed i yet. Therefore

$$V(x) \geq \max_{i=0,\cdots,N-1}\{V^i\}. \tag{2}$$

At any time $t < V(x)$ there is some packet in the system that hasn't reached its destination yet. Hence there is a node i between its current node and its destination for which $V^i > t$. Therefore it is not possible that

$$V(x) > \max_{i=0,\cdots,N-1}\{V^i\} \tag{3}$$

since in that case there would be a time t such that $t < V(x)$, $t \geq \max\{V^i\}$ and this is a contradiction. ∎

B. Minimum Delay and CDF Scheduling

Consider the closest destination first policy where every node i gives priority to the packet the destination of which is closest to i, among those packets residing at i. Packets with the same destination may be served arbitrarily.

Theorem 3: For any initial state x, the CDF policy minimizes the average delay.

The proof of the theorem relies on the following lemma, shown by an interchange argument.

Lemma 2: Assume that for some initial state x_0 and some policy π, node i at time t_0 does not follow CDF. There is a policy π' identical to π in the slots $1, \cdots, t_0$, except that π' follows CDF at t_0, such that

$$D(x_0, \pi') \leq D(x_0, \pi). \tag{4}$$

In the proof of the lemma we need an intermediate result presented in the following. Consider two initial states x_1 and x_2 that are identical except of the following. Nodes g and i in state x_1 have packets R and G respectively, with corresponding destinations r and g. Node i in x_2 has packet R while packet G is not in the system. The order of the nodes in the direction of the ring is $i \to g \to r$. The configuration is depicted in Fig. 3. The following holds.

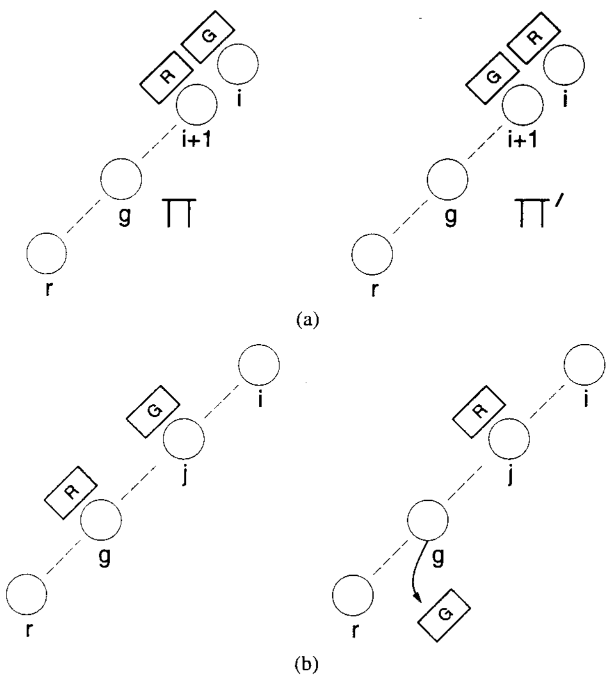


Fig. 4. The configuration of the packet under policies π and π' are depicted, at some time slots.

Lemma 3: For any policy π there is a policy π' such that when π acts on the ring with initial state x_1 and π' acts on the ring with initial state x_2 then the delivery times of the G and R under the two policies are related as

$$D_R^{\pi'} = \max\{D_G^{\pi} + (r \ominus g), D_R^{\pi}\}. \quad (5)$$

Furthermore the delivery times of all other packets under π' are less than or equal to the delivery times of the corresponding packets under π.

Proof: The policy π' is constructed based on π. Consider first the nodes $r, r \oplus 1, \cdots, g \ominus 1$. Every node among $r, r \oplus 1, \cdots, g \ominus 1$ at every slot t transmits under π', the packet that is transmitted under π, given that this packet is not G. If anyone of the nodes $r, r \oplus 1, \cdots, g \ominus 1$ transmits packet G under π, then under π' it transmits R. The construction of the schedule for nodes $g, g \oplus 1, \cdots, r \ominus 1$ is more elaborate as follows.

- Every node l among $g, g \oplus 1, \cdots, r \ominus 1$ transmits under π the same packet as under π' at every slot preceding the slot at which node l will transmit packet R under π.
- If at slot t^l node l transmits R under π then under π' it will transmit the first packet (among those residing in l at t^l) that is going to leave the node under π after t^l. If under π' node l is empty then it idles.
- If at some slot $t \geq t^l$ packet R arrives at node l under π' then it is to be forwarded to the next node in the next slot.
- At any slot $t \geq t^l$ that node l transmits some packet under π, policy π' attempts to transmit the same packet. If that packet does not reside at the node in the system under π', then the packet that will leave first under π after t (among those that reside already in the node) is transmitted at t under π'.

If π' is constructed as above there are two possibilities. At some slot t packets R under both policies will reside at the same node i among $g, g \oplus 1, \cdots, r \ominus 1$ in which cases the states match and

$$D_R^{\pi'} = D_R^{\pi}.$$

The other possibility is that after packet R under π' reaches node g, it will be forwarded by one node per slot in which case it will be delivered to r at slot

$$D_G^{\pi} + (r \ominus g).$$

Hence, the lemma follows. ■

Proof of Lemma 2: The policy π' will be constructed for $t \geq t_0$ such that (4) holds.

Assume that according to CDF, node i should transmit at time t_0 packet G with destination node g. Node i either idles at t_0 under π or transmits another packet R, with destination another node r which is further downstream from i than g is from i.

If node i idles under π then the construction of π' is easy. Just let every node under π' imitate the corresponding node under π until the time node i under π transmits packet G in which case node i idles under π'. At this slot the states of the systems under the two policies, match and from this slot onwards, we let π' be identical to π. It is clear that (4) holds either with equality or with strict inequality if it happens the destination of packet G to be node $i \oplus 1$.

When node i transmits packet G at t_0 under π, the configuration of the system state at t_0 is as depicted in Fig. 4(a).

Let t_1 be the time when packet R reaches node g under π. In the time slots $t_0 + 1, \cdots, t_1$ let the transmissions of a node under π' be identical with the transmissions of the same node under π except if the node under π transmits packet G or R in which case under π' it transmits R or G respectively. Hence at slot t_1 packet G will reach its destination under π' and will leave the system. The configuration of the packets will be as depicted in Fig. 4(b). From Lemma 3, the policy π' can be constructed from time t_1 and onwards such that for the delivery times of packet R under π' and of packets R, G under π the relation (5) holds. Note that for all packets that reached their destination until time t_1, the delivery times are the same under π and π'. All packets other than G and R, that reached their destination after t_1, have delivery times under π' smaller than or equal to those under π. From (5) and the fact that packet G under π' reached its destination at t_1, it can be easily checked that for the delivery times of packet R and G it holds

$$D_R^{\pi'} + D_G^{\pi'} \leq D_R^{\pi} + D_G^{\pi}.$$

Hence, the lemma follows. ■

Proof of Theorem 3: From Lemma 3 we can easily see that for any policy π_0 we can construct a sequence of policies π_i, $i = 1, 2, \cdots$ such that

$$D(x, \pi_{i+1}) \leq D(x, \pi_i) \quad i = 0, 1, \cdots \quad (6)$$

and π_{kN} acts identically to CDF at least during the first k slots. Since the network will empty within a finite time, all the policies π_i will be identical among themselves and with

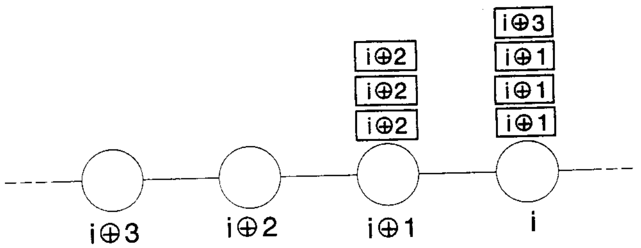


Fig. 5. An initial packet placement for which the FDF policy has strictly larger average delay than the minimum achieved by CDF. For this initial condition CDF has strictly suboptimal evacuation time.

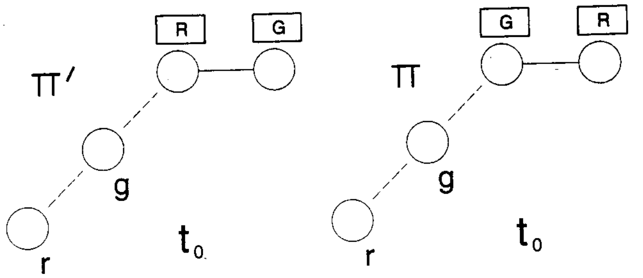


Fig. 6. The configuration under π and π' for the proof of Lemma 4 is depicted.

CDF, for $i \geq i_0$ where i_0 is an appropriately large integer. The theorem follows then from relation (6). ■

The CDF policy may be strictly suboptimal with respect to the evacuation time while the FDF policy may be strictly suboptimal regarding the average delay, as it is demonstrated for the initial condition depicted in Fig. 5. For this initial condition the delivery times under CDF are 1,1,2,2,3,3,6 and the average delay is 2.57. The delivery times under FDF are 1,2,3,3,3,4,4 and the average delay is 2.857. Notice also that the CDF has evacuation time equal to 6, which is strictly larger than the optimally achievable one under FDF, which is equal to 4.

III. The FDF Policy in Continual Operation

The minimum evacuation time property of the FDF policy holds in a stronger sense than that implied by Theorem 1. FDF evacuates the system in minimum time, even if there are additional arrivals generated after time $t = 0$. To express this property in the generality that it holds we introduce the following notion of domination.

A scheduling policy π_1 *dominates* a policy π_2, and it is written as $\pi_1 \succ \pi_2$, if the system is empty under π_1 at any slot t at which is empty under π_2, for any initial condition and arrival process common for both policies.

Theorem 4: The FDF policy dominates any other scheduling policy π.

$$\text{FDF} \succ \pi.$$

The theorem relies on the following lemma that is proved by an interchange argument.

Lemma 4: Consider an initial backlog state x_0, a sample path of arrivals and assume that node i under policy π does not follow FDF at some slot t_0. There is a policy π' which is identical to π in the slots $1, \cdots, t_0$, except that under π' node i acts like in FDF at t_0, such that if the system is empty at any slot t under π then it is empty under π' as well.

Proof: Clearly for all t, $1 \leq t \leq t_0$, the system is empty under π only if it is empty under π'. The policy π' will be constructed for $t > t_0$ such that the same holds for all $t > t_0$. At time t_0, let R be the packet at node i with the furthest destination, r. Under policy π' packet R will be transmitted. Under π node i either idles or transmits some packet G with destination node g which is closer to i than r is to i.

If under π node i idles at t_0 then π' is easy to construct. Let every node under π' act similarly as under π at all $t \geq t_0$ except if at t node i under π transmits packet R in which case under π' it idles. After that time t the state of the system under the two policies matches. Note that if $r = i \oplus 1$, the system may empty earlier under policy π' than under π.

If under π node i transmits packet G then the configuration of the packets under π and π' is as depicted in Fig. 6.

The policy π' acts as follows for t greater than t_0.

- All nodes that are transmitting neither packet G nor packet R under π are scheduled to do the same under π'.
- When a node transmits packet G or R under π it is scheduled to transmit R or G, respectively, under π'.
- If at some time t' packets G and R reside in the same node under π and π' then the states of the two systems match and let π' be identical to π from this time onwards. Otherwise let t_1 be the time at which packets G and R under π and π', respectively, reach node g.
- After time slot t_1 when a node transmits packet R under π it transmits packet G under π', until time t_2 at which packets G and R, under π' and π respectively, reach node g. Also packet R under π' is not transmitted from time t_1 until time t_2. At time t_2 the states of the systems under π and π' match and the systems evolve identically.

Clearly the system is empty at $t > t_0$ under π only if it is empty under π'. ■

Proof of Theorem 4: Given Lemma 4, the proof is similar to that of Theorem 3. ■

Note that Theorem 1 follows readily from the fact that FDF dominates any other policy in a system with no arrivals. Also Theorem 4 implies that in two systems with that same initial condition and arrival process, operated under FDF and π respectively, it holds that $V(\hat{x}(t)) \leq V(x(t))$ for all t, where $\hat{x}(t)$ and $x(t)$ are the queue length vectors at t under FDF and under the other policy, respectively. Furthermore, Theorem 4 implies that the FDF policy minimizes the busy cycle length where as busy cycle we understand the time period between two successive slots at which the ring is empty.

A. Sojourn Time Analysis of FDF

The long time average performance of FDF is analyzed in this section. We refer to the traffic from a specific origin node to a specific destination node as session in the following. Usually the performance of a network is difficult to analyze because as the traffic is forwarded from node to node, its characteristics change in a manner that cannot be modeled easily. When the relative priority of two sessions is the

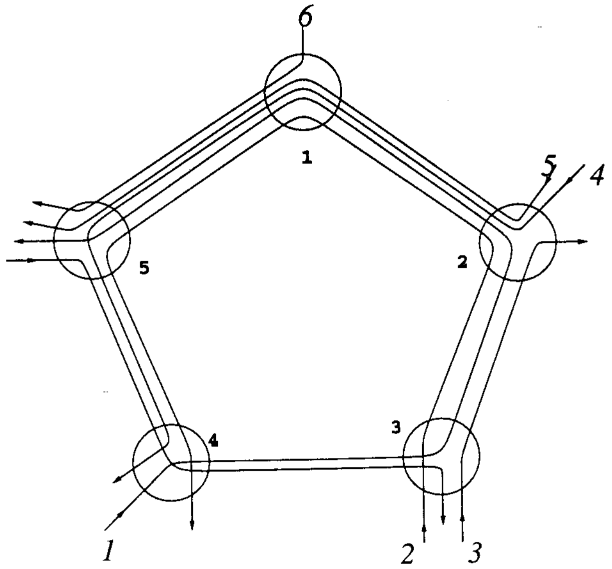


Fig. 7. A ring with 7 sessions is depicted.

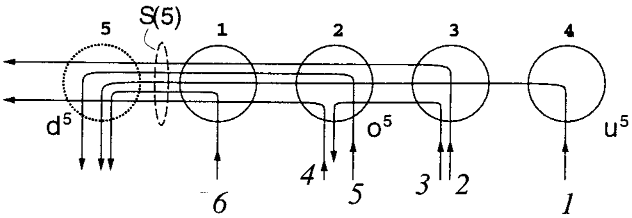


Fig. 8. The notation for session 5 of the ring in Fig. 7 is illustrated.

same in all network nodes then the performance analysis is considerably facilitated, even in arbitrary topology networks, since the highest priority stream can be analyzed independently from all the rest, the second highest priority stream can be analyzed based on the performance of the higher priority stream and independently of the rest lower priority streams and so on. In the FDF policy the packets at each node are served according to a fixed priority ordering. Even though the priorities are fixed at each node, they are not fixed throughout the network, and it is possible that the relative priorities of two sessions alternate from node to node. Despite this fact, there is a special version of FDF, that is analyzable. Consider the FDFO policy which is similar to FDF except that the packets with the same destination are given relative priorities based on their origin. They are served furthest-origin-first. Due to a local consistency property in the priority ordering under the FDFO we are able to analyze its performance by identifying an appropriate single server queue for each origin-destination pair, such that the queueing delay of the session is the same with the delay in the queue. We introduce some notation in order to state the result.

Denote by o^l and d^l the origin and destination nodes of session l. Let u^l be the furthest upstream node of node d^l, from which originates some session that exits the network at d^l or transverses d^l. Let finally $S(l)$ be the set of sessions for which one of the following holds : a) the session crosses node d^l; b) the session originates from a node in the segment of the ring from u^l to o^l and terminates at d^l. Note that $S(l)$ by

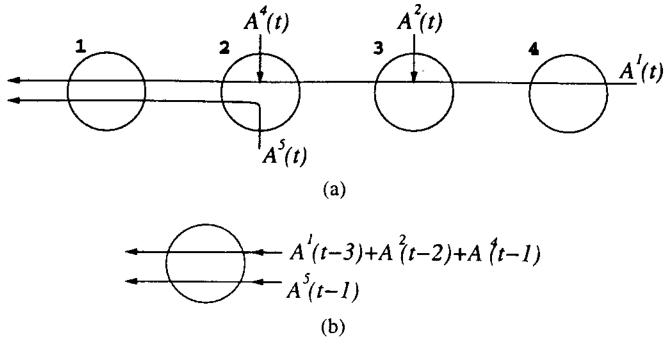


Fig. 9. The tandem and the equivalent node. The output processes of stream 5 in queue (b) and session 5 in the ring are identical.

definition contains exactly those sessions that may have higher priority than l at some node. The notation is illustrated in Fig. 8. Let $A^{ij}(t_1, t_2)$ (or $A^l(t_1, t_2)$) be the number of packets of the session l with origin node i and destination node j, that arrived in the network during the slots $t_1, t_1+1, \cdots, t_2$; let $A^{ij}(t) \stackrel{\text{def}}{=} A^{ij}(t,t)$ ($A^l(t) \stackrel{\text{def}}{=} A^l(t,t)$). It is assumed that the arrival streams are ergodic and the arrival rate of stream l is denoted by a^l. Traffic streams will be identified with superscripts the origin-destination nodes or the session symbols alternatively in the following. The departures from link i are denoted by the binary variables $D_i^l(t)$, where $D_i^l(t)$ is equal to 1 if at slot t a packet of session l is transmitted by link i and 0 otherwise. Associated with session l, consider a queue with two arrival streams $\{A_1^l(t)\}_{t=0}^{\infty}$ and $\{A_2^l(t)\}_{t=0}^{\infty}$ which are as follows:

$$A_1^l(t) = \begin{cases} A^l(t-(d^l \ominus o^l)+1) & t \geq d^l \ominus o^l \\ 0 & d^l \ominus o^l > t \geq 0 \end{cases}$$

$$A_2^l(t) = \sum_{m \in S(l)} A^m(t-(d^l \ominus o^m)+1).$$

The service discipline is work conserving with the stream 2 having strict priority over 1. Denote by $D^l(t)$ the binary variables that represent the departures of stream $\{A_1^l(t)\}_{t=1}^{\infty}$ in the queue. Assume for simplicity that the ring as well as the queue are initially empty. The following holds.

Theorem 5: The departures of stream 1 in the queue are identical to the departures of session l in the ring, that is

$$D^l(t) = D_{d^l}^l(t), t = 1, \cdots. \quad (7)$$

If the average queueing delay of stream 1 in the single queue exists and is equal to Q^l, then the average sojourn time J^l of session l exists as well and is

$$J^l = Q^l + (d^l \ominus o^l). \quad (8)$$

Note that $d^l \ominus o^l$ is the number of hops from the origin to the destination node of session l and therefore the cumulative transmission time for that session. Hence Q^l is the cumulative queueing delay experienced by session l packets.

The proof of the theorem is based on the following fact. For each traffic stream l we identify a unidirectional tandem network T^l in which there is strict priority between the streams and the traffic characteristic of session l in the ring coincide

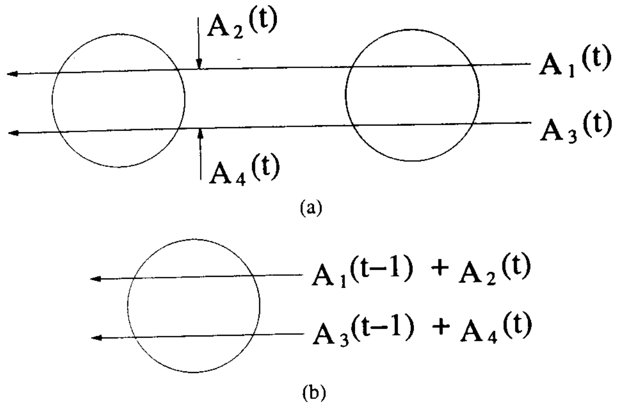


Fig. 10. A two node tandem queueing system and the equivalent single queue.

with those of a traffic session in the tandem. The tandem T^l has $N+1$ nodes. The origin and the destination end nodes correspond to node d^l in the ring. The rest are in one-to-one correspondence with the other nodes of the ring. All the traffic exits the tandem at the destination end node. There are two distinct traffic classes. Class 1 enters the network at node o^l and the arrival stream is identical to the arrival stream of session l in the ring. For class 2 there might be arrivals potentially at any node of the tandem. The class 2 arrival stream at node i is identical to the aggregation of the arrival streams of all sessions in $S(l)$ entering the network at node i, excluding session l if $i = o^l$. Fig. 9(a) illustrates the tandem T^l for the session $l = 5$ in the ring of Fig. 7. The traffic of class 2 has strict priority over the traffic of class 1 at all nodes of the tandem. The service is work conserving and arbitrary within each class. Let $\{\hat{D}_i(t)\}_{t=1}^{\infty}$ be the departure process of stream 1 from node i in the tandem. Then we have the following.

Lemma 5: For the streams l in the ring and 1 in the tandem it holds

$$\hat{D}_i(t) = D_i^l(t), t \geq 0, i = o^l, o^l \oplus 1, \cdots, d^l \ominus 1.$$

Proof: Observe that a session m in $S(l)$ that enters the ring in o^m has priority over any session not in $S(l)$ at every node $m, m \oplus 1, \cdots, d^l$. Furthermore the traffic of session l has lower priority than the traffic of any session m in $S(l)$ at all nodes $o^m, o^m \oplus 1, \cdots, d^l$. Therefore, as far as the session l is concerned, the ring is equivalent to the tandem. ∎

Corollary 1: The sojourn times of session l in the ring and 1 in the tandem are the same.

The analysis of the sojourn times of session l in the tandem T^l is reduced to the analysis of a single queue by the repetitive application of the reduction stated in Lemma 6. Consider the two and one node queueing systems depicted in Fig. 10. The packet length is fixed and the system is slotted. The streams 1 and 2 have strict service priority over the streams 3 and 4 both in (a) and (b). We have the following.

Lemma 6: The aggregate departure process of streams 1 and 2 in system (a) is identical with the aggregate departure process of streams 1 and 2 in system (b). Similarly, for the aggregate departure processes of the streams 3 and 4 in the two systems.

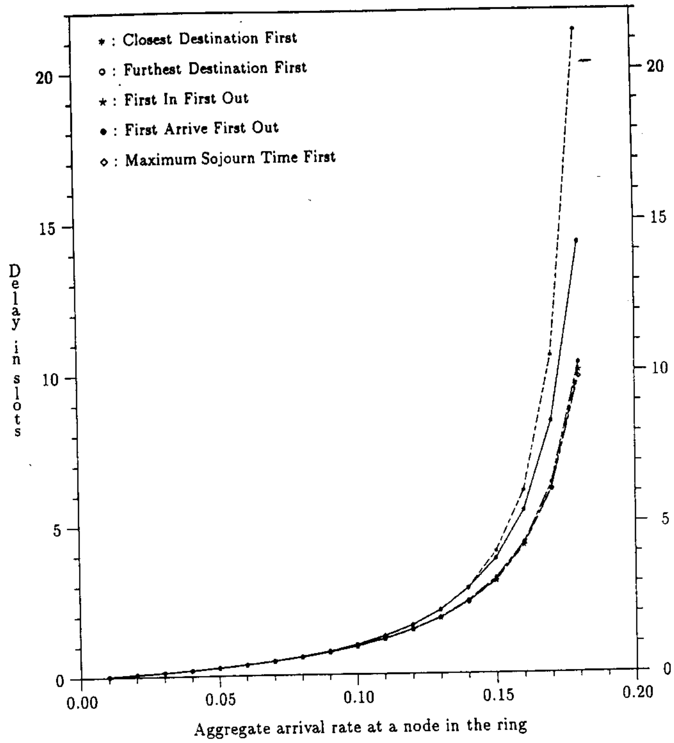


Fig. 11. Average queueing delay in a ring with 10 nodes, for $\lambda = 0.01$ to 0.18 by 0.01. Input traffic and destinations are assumed to be Bernoulli and uniformly distributed, respectively.

Proof: Since streams 1 and 2 have strict priority over 3 and 4, both the two node tandem in (a) and the queue in (b) operate as if the streams 3 and 4 do not exist for the streams 1 and 2. It is argued that the downstream queue in (a) is empty from packets of stream 1 or 2 if and only if the queue in (b) is empty of packets from steam 1 or 2. If the queue in (b) is empty from packets of streams 1 and 2, then the only packets of streams 1 and 2 that may be present in the tandem are the $A_1(t)$ packets arrived at slot t and all those will be residing in the upstream node. If the downstream queue in (a) is empty, then all the packets of stream 1 arrived in the system until slot $t-1$ will be out of the system. Therefore the queue in (b) will have no packets of stream 1 or 2. Because of the above, the aggregate departure process of streams 1 and 2 in systems (a) and (b) are identical.

For the streams 3 and 4, notice that if the downstream queue in (a) and the queue in (b) are nonempty from packets of streams 1 or 2 then clearly the aggregate departures of 3 and 4 is 0 at that slot in both system. If the downstream queue in (a) and the queue in (b) are empty from stream 1 and 2 packets then the queue in (b) has nonzero packets from streams 3 or 4 if and only if the downstream queue in system (a) has nonzero packets of streams 3 or 4. Therefore there will be a stream 3 or 4 packet departure at system (a) if and only if there will be a stream 3 or 4 departure at system (b). ∎

Proof of Theorem 5: By applying iteratively the reduction in Lemma 6 to the tandem associated with session l we reduce the tandem to the equivalent queue of Theorem 5. Relation (7) is an immediate consequence of Lemmas 5 and 6. For the relation (8) we argue as follows.

The amount of packets $J^l(t)$ of stream $l = (i, j)$ which are in the system at the beginning of slot t is

$$J^l(t) = \sum_{\tau=1}^{t} A^l(\tau - 1) - \sum_{\tau=2}^{t} D_j^l(\tau - 1). \tag{9}$$

Consider the equivalent queue for stream l and let $Q_1^l(t)$ be the number of packets of the stream that corresponds to l. Then we have

$$Q_1^l(t) = \sum_{\tau=1}^{t} A_1^l(\tau - 1) - \sum_{\tau=0}^{t} D^l(\tau - 1)$$
$$= \sum_{\tau=1}^{t-(j\ominus i)+1} A^l(\tau - 1) - \sum_{\tau=0}^{t} D_j^l(\tau - 1). \tag{10}$$

From (9) and (10) we have that

$$J^l(t) = Q_1^l(t) + \sum_{\tau=t-(j\ominus i)}^{t} A^l(\tau - 1). \tag{11}$$

From Little's Law, we have that the average sojourn time J^l of session l is

$$J^l = \frac{1}{a^l} \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} J^l(\tau). \tag{12}$$

From (11) we have

$$\frac{1}{t}\sum_{\tau=1}^{t} J^l(\tau) = \frac{1}{t}\sum_{\tau=1}^{t} Q_1^l(\tau) + \frac{1}{t}\sum_{\tau=1}^{t} \sum_{n=\tau-(j\ominus i)}^{\tau} A^l(n - 1)$$

and from (12)

$$J^l = \frac{1}{a^l} \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} Q_1^l(\tau) +$$
$$\frac{1}{a^l} \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} \sum_{n=\tau-(j\ominus i)}^{\tau} A^l(n - 1). \tag{13}$$

Note that the first term of the sum in the right side of (13) is the average waiting time in the equivalent system, that is the queueing delay Q^l plus 1. The second term, given the existence of average rates of the arrival processes, is equal to $j \ominus i - 1$. Hence the theorem follows from (13). ■

IV. Performance Analysis in Continual Operation

The performance of CDF was studied by simulation. In addition to CDF, three other policies were considered; the FIFO,FAFO and MSTF. In the FIFO policy, each node i applies the first in–first out rule locally, based on the arrival times of the packets at node i. In FAFO each node i applies the first in–first out rule based on the arrival time of the packets at the ring and not at i. In MSTF an estimated sojourn time is considered and is given priority to the packet with the largest estimated sojourn time. The estimated sojourn time for a packet is taken to be the remaining cumulative transmission time (the number of hops to the destination from the current node) plus the elapsed time since the arrival of the packet at the ring. Three different types of arrivals were considered.

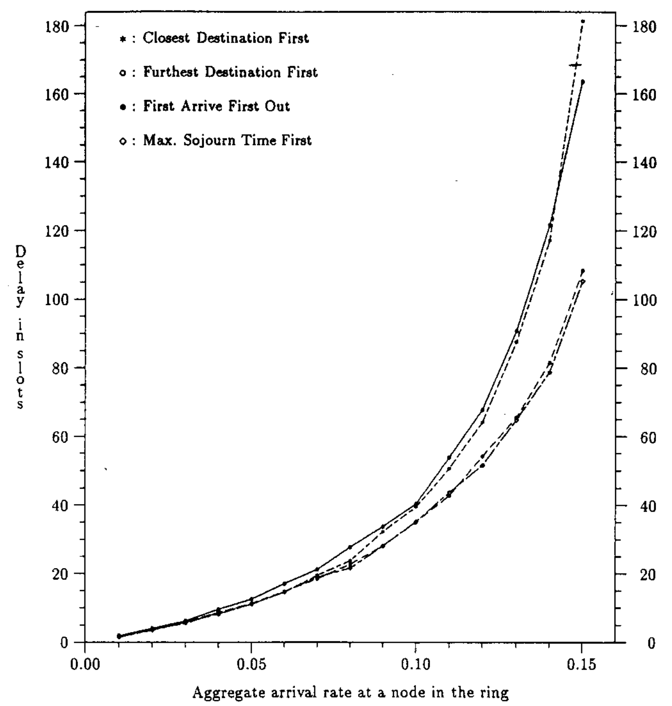


Fig. 12. Average queueing delay in a ring with 10 nodes, for $\lambda = 0.01$ to 0.15 by 0.01. Input traffic and output nodes are assumed to be on-off bursty with average burst length 30 and uniformly distributed, respectively.

In one case we had Bernoulli aggregate arrival process at each node and the destination of each packet were random with uniform distributions. The average packet delay, that is the end-to-end sojourn time minus the transmission time is depicted in Fig. 11. In Fig. 12 we see the delay for on-off binary arrivals process at each node with average burst length 30. Finally we considered the case where each origin-destination session had its own Bernoulli arrival process and the delay is depicted in Fig. 13. In the last case a formula can be obtained for the delay based on the equivalent queue result of Theorem 8. For a session $(j\ominus i, j)$ the equivalent queue has two traffic streams. One is the aggregation of $(N^2 - N - 2i)/2$ independent Bernoulli streams with rate q and the other is a Bernoulli stream with rate q. By using z-transforms [9] we can obtain the delay of a discrete time queue with M identical Bernoulli processes of rate q which is

$$\bar{W} = \frac{(M-1)q}{2(1 - Mq)}. \tag{14}$$

Since the stream 1 has priority over stream 2 the average delay W_1 for stream 1 is given from (14) for $M = (N^2 - N - 2i)/2$ while the average delay W_2 of all traffic is given by (14) for $M = 1 + (N^2 - N - 2i)/2$. Hence, the average delay W of stream 2 is

$$W = (N^2 - N - 2i)(W_2 - W_1)/2 + W_1.$$

Note that the CDF policy, which minimizes the average delay when there are no arrivals, loses this property in continual operation. The FIFO,FAFO and MSTF have approximately the same performance, better than both CDF and FDF.

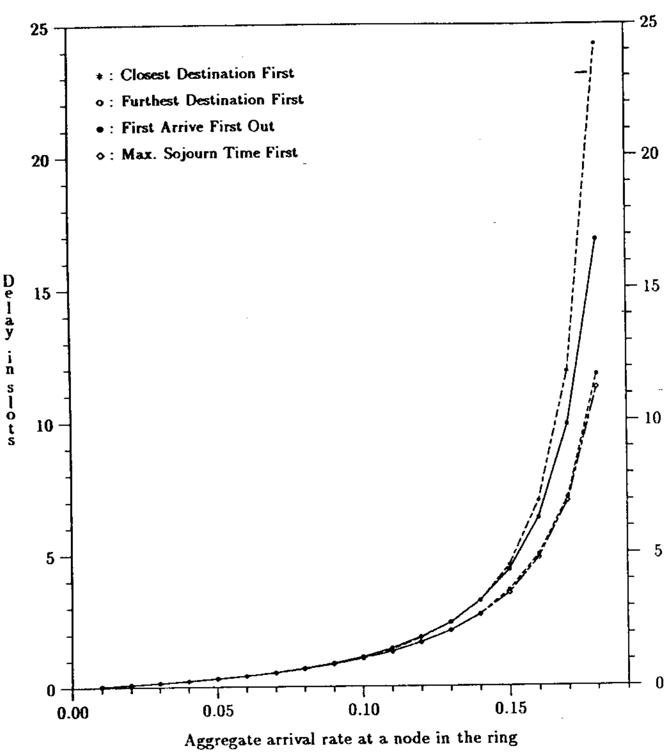


Fig. 13. Average queueing delay in a ring with 10 nodes, for $\lambda = 0.01$ to 0.18 by 0.01. Input traffic for each input-output session is i.i.d. with rate $\lambda/9$. Output nodes are uniformly distributed.

REFERENCES

[1] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation.* Englewood Cliffs, NJ: Prentice-Hall, 1989.

[2] S. N. Bhatt, G. Pucci, A. Renade, and A. L. Rosenberg, "Scattering and gathering messages in networks of processors," *IEEE Trans. Comput.*, vol. 42, pp. 938-949, Aug. 1993.

[3] I. Cidon and Y. Ofek, "MetaRing—A full-duplex ring with fairness and spatial reuse," *IEEE Trans. Commun.*, vol. 41, pp. 110-120, 1993.

[4] R. M. Falconer and J. L. Adams, "Orwell: A protocol for an integrated service local network," *Brit. Telecom Technol. J.*, vol. 3, no. 4, pp. 21-35, Apr. 1985.

[5] P. Fraigniaud, S. Miguet, and Y. Robert, "Complexity of scattering on a ring of processors," *Parallel Computing*, vol. 13, pp. 377-383, 1990.

[6] L. Georgiadis, R. Guerin, and I. Cidon, "Throughput properties of fair policies in ring networks," *IEEE/ACM Trans. Networking*, pp. 718-728, Dec. 1993.

[7] L. Georgiadis, W. Szpankowski, and L. Tassiulas, "A scheduling policy with maximum stability region in ring networks with spatial reuse," in *Queueing System Theory and Applications*, vol. 19, pp. 131–148, 1995.

[8] A. Hopper and R. M. Neddham, "The Cambridge fast ring networking system," *IEEE Trans. Comput.*, vol. 37, pp. 1214-1223, Oct. 1988.

[9] L. Kleinrock, *Queueing Systems, Vol. 1: Theory.* New York: Wiley, 1975.

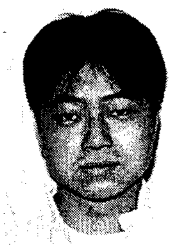
[10] Y. Saad and M. H. Schultz, "Data communication in parallel architectures," *Parallel Computing*, vol. 11, pp. 131-150, 1989.

[11] L. Tassiulas, "Cut-through switching, pipelining and scheduling for network evacuation," Polytechnic Univ., Tech Rep. CATT–94–80, 1994.

[12] L. Tassiulas and L. Georgiadis, "Any work-conserving policy stabilizes the ring with spatial reuse," *IEEE/ACM Trans. Networking*, to be published.

Leandros Tassiulas (S'89–M'91) was born in 1965 in Katerini, Greece. He received the Diploma in electrical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1987, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park in 1989 and 1991, respectively.

From September 1991 to June 1995 he was an Assistant Professor in the Department of Electrical Engineering, Polytechnic University, Brooklyn, NY. Since July 1995 he has been an Assistant Professor in the Department of Electrical Engineering, University of Maryland, College Park. His research interests are in the field of computer and communication networks with emphasis on wireless communications and high-speed networks, in control and optimization of stochastic systems and in parallel and distributed processing. He coauthored a paper which received the INFOCOM'94 best paper award.



Jinoo Joung was born in Seoul, Korea, on January 29, 1970. He received the B.S. degree in electronics engineering from Korea Advanced Institute of Science and Technology in 1992 and the M.S. degree in electrical engineering from Polytechnic University, Brooklyn, New York in 1994.

He is currently a Research Assistant at the Center for Advanced Technology in Telecommunications, Polytechnic University, and is working toward the Ph.D. degree at the Polytechnic University. His main research interests are in the area of high-speed communication networks and their performance analysis.