Bounding Delays in Packet-Routing Networks

Mor Harchol-Balter[*] David Wolfe[†]

Computer Science Division
University of California at Berkeley
Berkeley, California 94720

Journal version of STOC '95 paper

Abstract

Consider the problem of computing the average packet delay in a general dynamic packet-routing network with Poisson input stream, during steady-state.

Any packet-routing network can be formulated as a queueing network, where each server has a constant service time. If each server had exponentially-distributed service time, queueing theory techniques could be used to determine the expected packet delay. However, it is not known how to compute the average packet delay for all but the simplest networks with constant time servers.

It has been conjectured that to get an upper bound on expected packet delay in the constant service network, one can simply replace each constant time server with an exponential server of equal mean service time.

This paper shows that for a large class of networks, this conjecture is true, but that surprisingly there exists a network for which it is false. This large class of networks is all queueing networks with Markovian routing. Queueing networks with Markovian routing are important because they include many packet-routing networks where the packets are routed to random destinations.

1 Introduction

Many parallel and distributed applications require packets to be routed in a network. As packets move along their routes, they are delayed by other packets. In computing performance bounds for a given network and routing scheme, it is useful to be able to determine the time by which the average packet is delayed.

There are two general classifications of packet-routing networks: static and dynamic. Static packet-routing refers to the case where the packets to be routed are all present in the network when the routing commences. In dynamic packet-routing, packets arrive at the

[*] Supported by National Physical Science Consortium (NPSC) Fellowship. Also supported by NSF grant number CCR-9201092. harchol@cs.berkeley.edu, http://http.cs.berkeley.edu/~harchol/

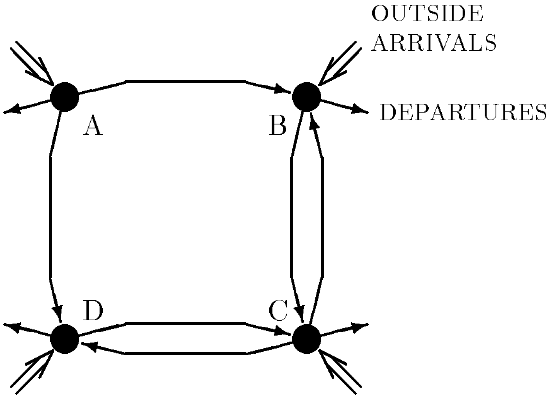
[†] wolfe@cs.berkeley.edu, http://http.cs.berkeley.edu/~wolfe/

network at random times and the routing proceeds in a continuous fashion. In this paper we will be interested in the dynamic case, in steady state, with Poisson input stream.

Most theoretical research has concentrated on analyzing delays in the static case. The dynamic case appears more difficult to deal with using conventional techniques. The most commonly used technique for bounding the delay in packet-routing networks is to use Chernoff bounds to bound the maximum number of packets which could possibly need to traverse a given edge during a window of time (w.h.p.). Examples of research on static packet-routing networks are [Lei90], [Lei92], [VB81], [Val82], [Ale82], [Upf84], [GL85], [ALMN90], [CS86]. All of these are specific to a particular network and a particular routing scheme. They mostly concentrate on the problem of permutation routing, and use the Chernoff bound approach. Some research on static packet-routing networks applies to general networks (see [LMR88] [PU87]). This research concentrates on worst-case bounds. There are very few theoretical results for dynamic packet-routing networks. A few are [Lei90],[KL95], and [CS86]. [Lei90] and [KL95] assume a discrete Poisson arrival steam (a new packet is born at each node of the network at every second with probability p). [CS86] assume a new permutation arrives every T seconds. *Both these results are network and routing scheme specific*, and although their bounds are very strong, the analysis is very involved. Lastly, since in most of the above routing schemes packets are first sent to intermediate random destinations, there's been a lot of research which concentrates on computing delays for the case where the final destinations are random (see for example [Lei90], [Val82], [Lei92], [KL95]). Again, except for [Lei90] and [KL95] all these results are for the static case.

For computing delays in dynamic packet-routing networks, queueing theory provides a huge body of useful results which apply to *any network configuration and routing scheme.* Unfortunately, these results rely on a few unrealistic assumptions (the most unrealistic being independent exponentially distributed service times), and therefore people are reluctant to make use of them. In this paper we discuss *when* the assumption of independent exponential service times can be replaced by (the much more realistic) constant service times. More specifically (see Section 1.3.3):

- We show packet-routing networks may be described as queueing networks with constant-time servers, where each server in the queueing network corresponds to an edge (or any other bottleneck) in the packet-routing network.
- We prove that if the queueing network has Markovian routing and the contention resolution protocol at the servers is FCFS (first-come-first-served), then replacing constant time servers with exponentially-distributed time servers of equal mean service time increases average packet delay. Since queueing networks with FCFS, exponential time servers are easily solvable for average delay, we therefore have an upper bound on the average delay in any packet-routing network. An almost immediate consequence of our proof is a sufficient condition for the stability of any queueing networks with Markovian routing and constant time servers. Our result easily generalizes to a broader class of contention resolution protocols.
- We show there exists a queueing network (with Non-Markovian routing) for which



Possible packet path	Average rate of birth
A →B →C →	1 every 10 sec
A →B →C →B →	1 every 15 sec
B →C →B →C →	1 every 20 sec
1 ⋮	⋮

Figure 1: *A packet-routing network*

replacing constant time servers with exponential time servers (of equal mean service time) decreases average packet delay.

1.1 Definition of Packet-Routing Network

A packet-routing network consists of nodes with wires connecting the nodes, as shown in Figure 1. Packets arrive continuously from outside the network at the nodes of the network. Each packet is born with a path. For example, in the routing scheme of Figure 1, packets with path $A \longrightarrow B \longrightarrow C \longrightarrow$ are born at a rate of one every 10 seconds, and packets with path $B \longrightarrow C \longrightarrow B \longrightarrow C \longrightarrow$ are born at a rate of one every 20 seconds, etc. Most literature considers the edges (wires) of the packet-routing network to be the bottlenecks. Specifically, it takes some constant time to traverse an edge (this constant may be different for each edge), and only one packet may traverse the edge at a time. The packets traverse the edge in a non-preemptive order. This causes a packet to be delayed when it arrives at an edge that is currently being used. The nodes of the network serve only to route the packets from one edge to the next. In our analysis, it is equally easy to assume the nodes of the network also form bottlenecks (in the same way as the edges).

In this paper we'll be interested in computing the time an average packet is delayed by waiting in queues.

Although we haven't explicitly mentioned transmission delays and propagation times in our definition of a packet-routing network, we observe below that our definition is general

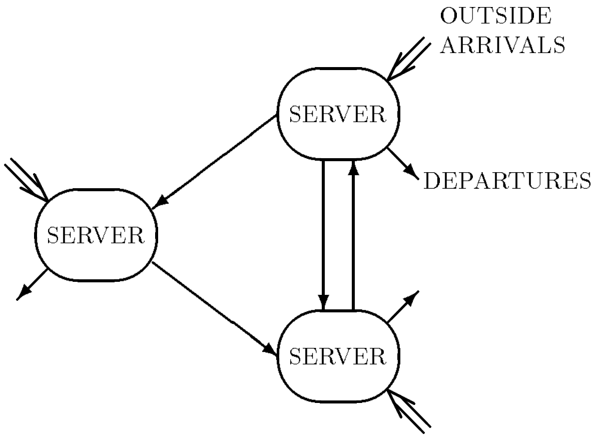


Figure 2: *A queueing network*

enough to capture all these properties of a real network with fixed-size packets (as in an ATM [Vet95] network). We will use $\mathcal{R}$ to denote a "real" packet-routing network, and $\mathcal{P}$ to denote a packet-routing network of the type in our definition above. In a real packet-routing network, $\mathcal{R}$, more than one packet may be on a given link, however the packets must be separated by the transmission time, t, namely the time necessary to load the whole packet onto the link (note that since the packets all have the same size, t is the same for all packets). Let T be the time for a packet to traverse a given link (this is a function of the length of the link). Then we can replace each link in $\mathcal{R}$ by a chain of T/t short links in $\mathcal{P}$, where each of these short links requires time t to traverse and only one packet may traverse each short link at a time. Note that $\mathcal{P}$ now exactly models $\mathcal{R}$.

1.2 Queueing Network Definitions and Background

A *queueing network* $\mathcal{N}$ consists of servers with edges connecting the servers, as shown in Figure 2. It's behavior is very similar to our definition of a packet-routing network, except that time is only spent at the servers, and no time is spent on the edges. (Thus packets queue up at the servers of $\mathcal{N}$). Packets arrive continuously at the servers of $\mathcal{N}$, and each packet has a path associated with it which it follows. A queueing network is defined by 3 parameters:

service-time distribution The service time associated with a server is a random variable from a distribution. (Note the distribution — or just its mean — may be different for each server).

contention resolution protocol The order in which packets are served in case of conflict at a server.

outside arrival process In this paper, whenever we speak of a queueing network, we will assume that outside arrivals occur at each server according to a Poisson process.

We say that a queueing network has *Markovian routing* if when a packet finishes serving at a server i, the probability that it next moves to some server j (or leaves the network) depends only on where the packet last served and is independent of its previous history (or route). In this case the packets appear indistinguishable. Thus a queueing network with Markovian routing can simply be described by a directed graph with probabilities on the edges.[1]

Given a queueing network $\mathcal{N}$, we define $\mathcal{N}_{C,FCFS}$ (respectively, $\mathcal{N}_{E,FCFS}$) to be queueing network $\mathcal{N}$ where each server has a constant (respectively, exponentially distributed) service time with the same mean as the corresponding server in $\mathcal{N}$, and the packets are served in a First-Come-First-Served order.

1.2.1 Recasting a packet-routing network $\mathcal{P}$ as a queueing network $\mathcal{N}_C$

Observe that every packet-routing network $\mathcal{P}$ may be described as a queueing network of type $\mathcal{N}_C$ as follows: Corresponding to each bottleneck in $\mathcal{P}$, we create a server in $\mathcal{N}$. For example, the edges of the packet-routing network are bottlenecks, so we create one server in $\mathcal{N}$ corresponding to each edge in $\mathcal{P}$. We set the service time at each server to be constant, equal to the time required to traverse the corresponding edge in $\mathcal{P}$. Since only one packet at a time may traverse an edge in $\mathcal{P}$, we restrict the contention resolution protocol in $\mathcal{N}$ to one where only one packet at a time serves at a server. In the case where edge contention in $\mathcal{P}$ is specifically FCFS, $\mathcal{P}$ can be represented by a queueing network of type $\mathcal{N}_{C,FCFS}$.

Thus from now on, we will never refer to packet-routing networks again, but rather we will only address how to compute delays in queueing networks of type $\mathcal{N}_C$. In this section and the next, we will look specifically at networks of type $\mathcal{N}_{C,FCFS}$. (In Section 3, we will consider more general contention resolution protocols.) Unfortunately, it is not known how to compute the average packet delay for all but the simplest $\mathcal{N}_{C,FCFS}$ networks, since $\mathcal{N}_{C,FCFS}$ networks don't have product-form. However, $\mathcal{N}_{E,FCFS}$ is a product-form network (more specifically it's a classed Jackson queueing network, see [BS93]) and therefore the average packet delay is easy to determine for networks of this type (see, for example, [Wal89] [BS93]).

The objective is therefore to bound the average delay of $\mathcal{N}_{C,FCFS}$ (which we care about) by the average delay of $\mathcal{N}_{E,FCFS}$ (which we know how to compute).

1.3 In This Paper We Show ...

1.3.1 Overall Goal

Our overall goal is to identify for which $\mathcal{N}$

$$\text{AvgDelay}(\mathcal{N}_{C,FCFS}) \leq \text{AvgDelay}(\mathcal{N}_{E,FCFS}) \tag{1}$$

[1] Note an equivalent, but more elegant, way to define a queueing network $\mathcal{N}$ is to say that each outside arrival to $\mathcal{N}$ is associated with some class. A packet of class ℓ moves from server i to server j next with probability p_{ij}^{ℓ}. The special case of a *Markovian* network $\mathcal{N}$ is defined as a network with only one class of packets.

1.3.2 Previous Work on Goal

All previous work seems to indicate (1) holds for all queueing networks $\mathcal{N}$.

For example, the average packet delay is an increasing function of the variance in the service time distribution for each of the following single queue networks: the M/G/1 queue, the M/G/1 queue with batch arrivals, the M/G/1 queue with priorities, and the M/G/k queue, [Whi83] [Whi80] [Ros89, pp. 353–356].

With respect to networks of queues, [ST94] showed that for all *layered* (i.e. acyclic) networks $\mathcal{N}$with Markovian routing, $\mathcal{N}_{E,FCFS}$ has greater average packet delay than $\mathcal{N}_{C,FCFS}$. Unbeknownst to [ST94] and to us, [RS92] earlier proved a stronger result for all networks with Markovian routing, namely that $\mathcal{N}_{E,FCFS}$ has greater average packet delay than $\mathcal{N}_{ILR,FCFS}$, where ILR denotes any service time distribution which has an increased likelihood ratio. However, [RS92]'s proof requires specialized cross-coupling and conditioning arguments, and therefore we choose to present our own elementary proof. There are also simulation studies of several non-Markovian networks $\mathcal{N}$ (i.e. general classed networks) which find the average packet delay to be greater for $\mathcal{N}_{E,FCFS}$ than for $\mathcal{N}_{C,FCFS}$ (see [HBB94] [MC86] [HC86]) .

With respect to how tight this upper bound is, in all of the above simulations the average delay in $\mathcal{N}_{E,FCFS}$ was never greater than that of $\mathcal{N}_{C,FCFS}$ by more than a factor of 3 (this included networks loaded to 99% of capacity and having 100 servers). However, since the difference increases both with the load *and* with the number as

servers (see for example Section 2.2 and also [KL95]), this ratio could be greater for large networks.

The above results have led to a general belief that greater variance in service times leads to greater average packet delay [Whi84] [Wal94] [Fer94] [Kle94]. In Section 2.2, we give some intuition for this. Counterexamples to this theory have only been found in the case where arrivals are *not* Poisson [Wol77] [Ros78]. For example Figure 3 indicates why counterexamples can be found which use *batch Poisson* arrivals such as those in [Wol77]. The final thing we do in this paper is to demonstrate a counterexample for the case of Poisson arrivals.

1.3.3 Main Results

- (Section 2) We give an easy proof showing that all queueing networks $\mathcal{N}$ with Markovian routing,

  $$\mathrm{AvgDelay}(\mathcal{N}_{C,FCFS}) \leq \mathrm{AvgDelay}(\mathcal{N}_{E,FCFS})$$

  and $\mathcal{N}_{C,FCFS}$ is stable under the same conditions as $\mathcal{N}_{E,FCFS}$.

  Significance of this result: Recall that computing delays in packet-routing networks when the packets have random destinations is important because most randomized routing algorithms consist of two random routing problems (see the third paragraph of the introduction). Queueing networks with Markovian routing are important because they include many packet-routing networks in which the packets have random destinations. A couple common examples are the mesh network with greedy routing (packets are first routed to the correct column and then to the correct row) and the hypercube network

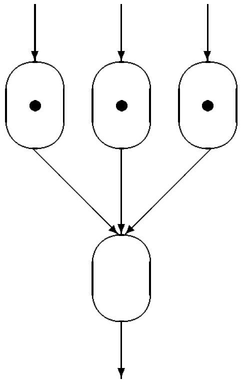


Figure 3: *Non-Poisson (in this case batch-Poisson) arrivals can favor more variance in service distributions. For example, if three packets arrive in a batch (serving in the top three servers above), they'll collide at the next server unless their service-completions are staggered.*

with canonical routing (packets cross each dimension if needed in order). When the packets have random destinations, rather than choosing the random destination when the packet is born, we can view the random destination as being decided a little at a time, by flipping a coin after each server.[2] The above result tells us that we can easily compute an upper bound on the average delay for any packet-routing network which can be modeled by a queueing network with Markovian routing. Also, Section 1.3.2 cites evidence that this upper bound is not far from tight in practice.

- (Section 4) We demonstrate (a non-Markovian) network $\mathcal{N}$, s.t.

$$\text{AvgDelay}(\mathcal{N}_{C,FCFS}) \; > \text{AvgDelay}(\mathcal{N}_{E,FCFS})$$

Significance of this result: The counterexample disproves the widely believed conjecture that for all networks $\mathcal{N}_{C,FCFS}$ has better average delay than $\mathcal{N}_{E,FCFS}$ (see Section 1.3.2).

2 Upper Bounding Average Delay in Markovian Queueing Networks

In this section we will prove the following theorem:

Theorem 1 *For all Markovian queueing networks $\mathcal{N}$,*

$$AvgDelay(\mathcal{N}_{C,FCFS}) \leq AvgDelay(\mathcal{N}_{E,FCFS})$$

[2] Observe that since the server in the queueing network represents an edge in the packet-routing network, all we need to know to determine the probabilities is the server (edge) at which the packet just finished serving.

Our proof is modeled after [ST94] who proved this result for layered Markovian networks. Whereas their proof uses induction on the levels of the network, we induct on time, thereby simplifying the proof and obviating the need for a layered network.

Define $\mathcal{N}_{C,PS}$ to be the queueing network $\mathcal{N}$ where each server has a constant service time with the same mean as the corresponding server in $\mathcal{N}$ and the service order is Processor Sharing. (In Processor Sharing, the server is shared equally by all the packets currently waiting at the server. So, for example if the service time at the server is 2, and there are 3 packets waiting at the server, each packet is being served at a rate of $\frac{1}{6}$). By [BCMPG75] and [Kel75], we know that the average packet delay in $\mathcal{N}_{C,PS}$ is equal to the average packet delay in $\mathcal{N}_{E,FCFS}$ for all $\mathcal{N}$.[3] It is therefore sufficient to prove that for any queueing network $\mathcal{N}$ with Markovian routing,[4]

$$\text{AvgDelay}(\mathcal{N}_{C,FCFS}) \leq \text{AvgDelay}(\mathcal{N}_{C,PS}) = \text{AvgDelay}(\mathcal{N}_{E,FCFS})$$

We start by proving the inequality for a single server network.

Claim 1 *If the sequence of arrivals to a (single server) FCFS queue is no later than the arrivals to a PS queue, then the i^{th} departure from the FCFS queue occurs no later than the i^{th} departure from the PS queue.*

Proof: In both queues, each packet must wait for all packets with earlier arrivals to depart, but only in the PS queue must a packet also wait while later arrivals get service. ■

To generalize the statement from the single server to the network, we'll use a coupling argument. Consider the behavior of the two networks when coupled to run on the same *sample point* consisting of:

1. the sequence of *outside* inter-arrival times at each server, and
2. the choices for where the j^{th} packet served at each server proceeds next.

Note the above quantities are all independent for a Markovian network. Also, the j^{th} packet to complete at a particular server in the two networks may not be the same packet.

Claim 2 *For a given sample point, the j^{th} service completion at any server of the FCFS network occurs no later than the j^{th} service completion at the corresponding server of the PS network.*

Proof: Assume the claim is true at time t. We show it's true at time $t' > t$, where t' is the time of the next service completion. We distinguish between *outside arrivals* to a server (packets arriving from outside the network) and *inside arrivals* to the server (service completions), and make the following sequence of observations:

[3] This powerful theorem is also described more recently in [Wal89] and [Kle76].

[4] Our proof of the inequality is valid for *any* sequence of outside arrivals, not just a Poisson arrival stream.

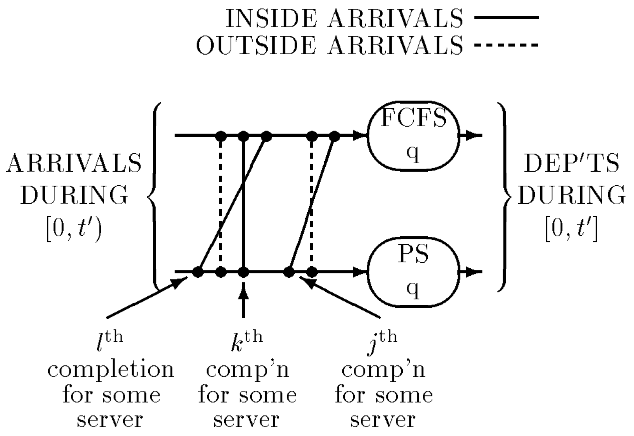

INSIDE ARRIVALS
OUTSIDE ARRIVALS
ARRIVALS DURING $[0,t')$
FCFS q
PS q
DEP'TS DURING $[0,t']$
l^{th} completion for some server
k^{th} comp'n for some server
j^{th} comp'n for some server

Figure 4: *Illustration of proof of Claim 2. We consider the same server q in the FCFS network and the PS network. The arrival stream into PS server q is delayed relative to the arrival stream into FCFS server q. Of course, the order of arrivals may be different for the two queues, but for Markovian routing, packets are indistinguishable with respect to routing anyway.*

- During $[0,t')$, Claim 2 is true.
- During $[0,t')$, every arrival at any PS server, q, must have already occurred at the corresponding FCFS server, q (see Figure 4). (This is true for inside arrivals because any inside arrival at PS server q is, say, the j^{th} service completion at some PS server q', and by the previous observation, the j^{th} service completion at FCFS server q' is at least as early. By definition of the sample point, the observation is also true for outside arrivals.)
- Therefore, during $[0,t')$, the i^{th} packet to arrive at any server of the FCFS network arrives no later than the i^{th} arrival at the corresponding server of the PS network.
- Hence, by Claim 1, we see Claim 2 holds during $[0,t']$. This includes the current service completion.

■

By Claim 2, it follows that for any sample point, the i^{th} departure from $\mathcal{N}_{C,FCFS}$ occurs no later than the i^{th} departure from $\mathcal{N}_{C,PS}$. This implies that

$$\text{Number of packets in } \mathcal{N}_{C,FCFS} \text{ at time t} \leq_{st} \text{Number of packets in } \mathcal{N}_{C,PS} \text{ at time t}$$

which implies that

$$\mathbf{E}\left\{\text{Number of packets in } \mathcal{N}_{C,FCFS} \text{ at time t}\right\} \leq \mathbf{E}\left\{\text{Number of packets in } \mathcal{N}_{C,PS} \text{ at time t}\right\}$$

So by Little's Law [Wol89, p. 236] we have shown that

$$\text{AvgDelay}(\mathcal{N}_{C,FCFS}) \leq \text{AvgDelay}(\mathcal{N}_{C,PS})$$

and therefore proved Theorem 1 above.

2.1 Stability of $\mathcal{N}_{C,FCFS}$

A sufficient condition for the stability of a queueing network is that the expected time between which all the queues empty out is finite.

Stability is a well understood issue for any network of type $\mathcal{N}_{E,FCFS}$. A sufficient condition for stability is that the average arrival rate into each server is less than the service rate at that server because (by the product-form distribution of $\mathcal{N}_{E,FCFS}$) under this condition the probability that all the queues are empty is non-zero.

Since networks of type $\mathcal{N}_{C,FCFS}$ don't satisfy product-form, it is harder to prove sufficient conditions for their stability. However observe that the stochastic ordering in the proof of Theorem 1 immediately implies that for any queueing network $\mathcal{N}$ with Markovian routing, $\mathcal{N}_{C,FCFS}$ is stable whenever $\mathcal{N}_{E,FCFS}$ is. For let $\mathcal{N}$ be any network with Markovian routing and assume that the average arrival rate into each server is less than the service rate at that server. Then:

$$\begin{aligned}
& \mathbf{Pr}\{\text{all queues of } \mathcal{N}_{C,FCFS} \text{ are empty}\} \\
= \; & 1 - \mathbf{Pr}\{\text{tot. num. packs in queue in } \mathcal{N}_{C,FCFS} > 0\} \\
\geq \; & 1 - \mathbf{Pr}\{\text{tot. num. packs in queue in } \mathcal{N}_{E,FCFS} > 0\} \\
= \; & \mathbf{Pr}\{\text{all queues of } \mathcal{N}_{E,FCFS} \text{ are empty}\} \\
> \; & 0.
\end{aligned}$$

2.2 How much worse is PS than FCFS?

In section 1.3.2 we stated that simulations indicate that the average delay in $\mathcal{N}_{C,PS}$ is always within a factor of 3 of the average delay in $\mathcal{N}_{C,FCFS}$. However, in this section we will show that when the number of servers, n, in a network is very large, this difference might be much greater. Consider a queueing network $\mathcal{N}$ consisting of only a single line of n servers, each with service time 1. Packets arrive only at the first server, and leave the network after serving at the nth server. $\mathcal{N}_{C,FCFS}$ (respectively, $\mathcal{N}_{C,PS}$) is the network $\mathcal{N}$ where the service resolution protocol is FCFS (respectively, Processor-Sharing). To determine the average delay in each network, consider the delay experienced by a newly-arriving packet p. In both $\mathcal{N}_{C,FCFS}$ and $\mathcal{N}_{C,PS}$, p is delayed by the packets it finds queued up at the first server (in $\mathcal{N}_{C,PS}$ later arrivals also cause p to be delayed, but we ignore them). The difference is that in $\mathcal{N}_{C,FCFS}$, these packets only each delay p by 1 (after that initial delay the packets are spread out and move in lockstep), whereas in $\mathcal{N}_{C,PS}$, these packets each delay p by n (since the packets all move in a "clump" down the network).

3 Some Easy Generalizations

Observe that the proofs of Claims 1 and 2 do not depend on serving the packets in a FCFS order. For example, packets could be born with priorities, with servers serving higher-priority packets first, in a non-preemptive fashion. In fact, for a given sample point, the time of the j^{th} service completion at a server is the same for every non-preemptive contention resolution protocol which serves one packet at a time.

Naturally, we still require that the packet priorities are independent of the packet route, for otherwise we cannot perform the coupling argument required in the proof of Claim 2.

The stability claim thus also generalizes, even when the contention resolution protocol intentionally tries to starve a particular packet.

4 A Non-Markovian Counterexample

In this section, we demonstrate an $\mathcal{N}$ for which

$$\text{AvgDelay}(\mathcal{N}_{C,FCFS}) > \text{AvgDelay}(\mathcal{N}_{E,FCFS})$$

More specifically, defining $\mathcal{N}_{C,PS}$ as in Section 2, we will demonstrate a network $\mathcal{N}$ for which

$$\text{AvgDelay}(\mathcal{N}_{C,FCFS}) > \text{AvgDelay}(\mathcal{N}_{C,PS}) = \text{AvgDelay}(\mathcal{N}_{E,FCFS})$$

For some insight into why it is counterintuitive that such a network $\mathcal{N}$ exists, see Section 2.2.

4.1 Network Description

Let $\mathcal{N}$ be the queueing network shown in Figure 5. The servers in $\mathcal{N}$ either have service time 1 or ϵ, as shown. The only outside arrivals are into the top server. Packets arrive from outside $\mathcal{N}$ according to a Poisson Process with rate $\lambda = \frac{1}{2e^2n^2}$, where n is the number of servers of mean service time 1 in $\mathcal{N}$. Half the arriving packets are of type *solid* and half are of type *dashed* (by "type" we mean class). Packets of type *solid* are routed straight down, only passing through the time 1 servers. Packets of type *dashed* are routed through the dashed edges, i.e. through all the ϵ servers and through every other 1-server.

4.2 Intuition

We will compare the average delay in $\mathcal{N}_{C,FCFS}$ with the average delay in $\mathcal{N}_{C,PS}$, as shown in Figure 6, by comparing the average delay experienced by an arriving packet p at $\mathcal{N}_{C,FCFS}$ and $\mathcal{N}_{C,PS}$. Throughout our argument, we implicitly use PASTA (Poisson Arrivals See Time Averages).

The intuition behind the analysis is as follows: Since λ is so low, usually for either network, p will see no other packets during its traversal of the network. In this case $\mathcal{N}_{C,FCFS}$ behaves identically to $\mathcal{N}_{C,PS}$. With some probability, however, one other packet will be present in the network during p's traversal of the network. The expected delay on p in this case is greater

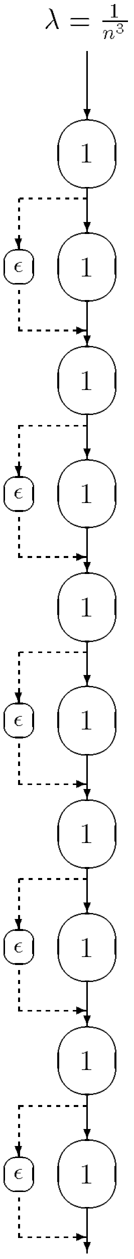


Figure 5: *Counterexample network $\mathcal{N}$ with n servers of mean service time 1 and $n/2$ with mean service time ϵ. Packets arrive at the top; half follow the dashed route, while half follow the solid route.*

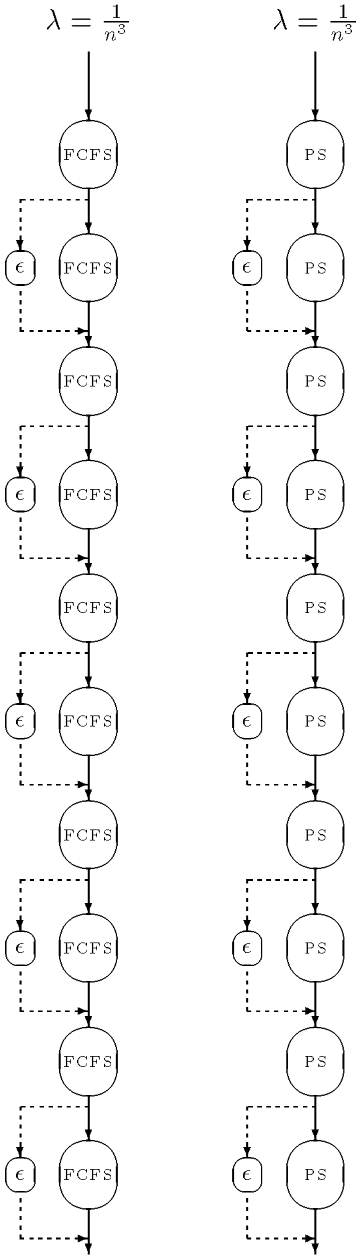


Figure 6: $\mathcal{N}_{C,FCFS}$ *and* $\mathcal{N}_{C,PS}$ *networks.*

for the $\mathcal{N}_{C,FCFS}$ network than for the $\mathcal{N}_{C,PS}$ network. Figure 7 shows us why: Consider first $\mathcal{N}_{C,FCFS}$. Suppose q is of type *solid* and some packet p of type *dashed* enters $\mathcal{N}_{C,FCFS}$ within $\frac{n}{2}$ seconds after q. Then p will eventually catch up to q, and from this point on, q will delay p by one second at every other server throughout the rest of the $\mathcal{N}_{C,FCFS}$. That is, p will be delayed by $\Theta(n)$ seconds. Now observe that the same scenario would only cause a delay of at most 2 seconds in $\mathcal{N}_{C,PS}$, because when p catches up to q, it will only interfere with q for two servers and then p will pass q forever. A worse situation for $\mathcal{N}_{C,PS}$ is the case where p meets up with another packet of the same type as p during its traversal (since in that case p is clearly delayed by $\Theta(n)$). Observe, however, that this scenario can only happen if the two packets both arrived at $\mathcal{N}_{C,PS}$ within a second of each other. This occurs with such low probability for our choice of small λ that the scenario's affect on average delay is negligible.

Lastly, we have to consider the case that two or more packets are present in the network during p's traversal of the network. The expected delay on p in this case is greater for the $\mathcal{N}_{C,PS}$ network than for the $\mathcal{N}_{C,FCFS}$ network, but this case occurs with such low probability that its effect on p's delay is also negligible.

4.3 The details

By PASTA, the expected delay a newly arriving packet experiences is equal to the average packet delay for the network. We will compute an upper bound on the delay an arrival experiences in $\mathcal{N}_{C,PS}$ and a lower bound on the delay an arrival experiences in $\mathcal{N}_{C,FCFS}$. We will show

$$\text{lowerbound}\left(\mathbf{E}\left\{ \text{ Delay on arrival in } \mathcal{N}_{C,FCFS} \right\}\right) > \text{upperbound}\left(\mathbf{E}\left\{ \text{ Delay on arrival in } \mathcal{N}_{C,PS} \right\}\right).$$

4.3.1 Upperbound E{Delay on arrival in $\mathcal{N}_{C,PS}$}

Let p represent an arriving packet in $\mathcal{N}_{C,PS}$. Clearly, p may only be delayed by packets which are in $\mathcal{N}_{C,PS}$ during the time p is in $\mathcal{N}_{C,PS}$. Note that if i packets are in $\mathcal{N}_{C,PS}$, they may take up to time in to clear the system. So, denoting p's arrival time by 0, if packet p is delayed, at least one of the following must occur:

- at least 1 other packet arrives during $(-n, n)$.
- at least 2 other packets arrive during $(-2n, 2n)$.
- at least 3 other packets arrive during $(-3n, 3n)$.
- etc.

Define

E_i' : the event that *at least* i packets arrive during $(-in, in)$

E_i : the event that *exactly* i packets arrive during $(-in, in)$

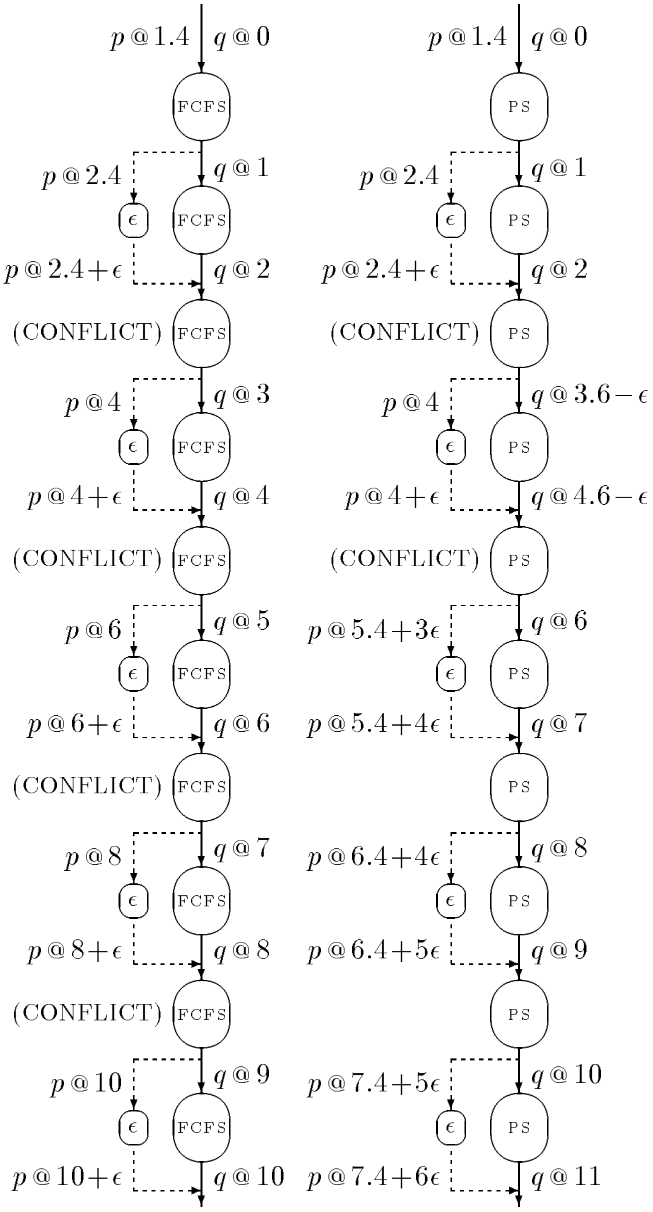


Figure 7: *Example illustrating how a packet, p, of type dashed and a packet, q, of type solid clash repeatedly in $\mathcal{N}_{C,FCFS}$, but only twice in $\mathcal{N}_{C,PS}$.*

Now p can only be delayed is at least one of the E_i' occur, i.e., if $\bigcup E_i'$ is true. However $\bigcup E_i'$ can only occur if $\bigcup E_i$ occurs (See footnote[5] for a proof of this subtle point), so p can only be delayed if at least one of the following events occurs:

- exactly 1 other packet arrives during $(-n, n)$.
- exactly 2 other packets arrive during $(-2n, 2n)$.
- exactly 3 other packets arrive during $(-3n, 3n)$.
- etc.

We will compute the expected delay on p due to each of the above events, and then we'll sum these. This will be an overcount, but that's o.k. because we're just upperbounding.
Let

$$P_i = \mathbf{Pr}\left\{\text{exactly } i \text{ arrivals during time } (-in, in)\right\}$$

Let

$$D_i^{PS} = \mathbf{E}\left\{\text{delay on } p \text{ due to } i \text{ arrivals during } (-in, in) \text{ in } \mathcal{N}_{C,PS}\right\}$$

So

$$\begin{aligned} \mathbf{E}\left\{\text{delay on } p \text{ in } \mathcal{N}_{C,PS}\right\} &\leq P_1 D_1^{PS} + P_2 D_2^{PS} + P_3 D_3^{PS} + \ldots \\ &\leq P_1 D_1^{PS} + P_2(2n) + P_3(3n) + \ldots \end{aligned}$$

where the last inequality is an over-estimate, since we are assuming the worst case where all the packets continually run into each other over and over again during their entire time in the network.

By definition of the Poisson Process,

$$P_i = \frac{e^{-\lambda \cdot 2in}\left(\lambda \cdot 2in\right)^i}{i!}$$

For $i \geq 2$, we can express P_i in terms of P_1 as follows:

$$\begin{aligned} P_i(i \geq 2) &= \frac{e^{-\lambda \cdot 2in}\left(\lambda \cdot 2in\right)^i}{i!} \\ &= \frac{i^i}{i!} \cdot e^{-\lambda \cdot 2in}\left(\lambda \cdot 2n\right)^i \\ &< e^i \cdot e^{-\lambda \cdot 2n}\left(\lambda \cdot 2n\right)^i \\ &= P_1 \cdot \left(\lambda \cdot 2n\right)^{i-1} \cdot e^i \end{aligned}$$

[5] Let $x(i)$ denote the number of arrivals during $(-in, in)$. Observe that $x(i)$ is a non-decreasing integer-valued function of i. Let $\mathcal{L}$ be the line $x(i) = i$. Since $\mathbf{E}\left\{x(i)\right\}$ is less than 1, if x(i) is ever above $\mathcal{L}$, with probability one it must eventally cross $\mathcal{L}$ and come below $\mathcal{L}$ (by the Law of Large Numbers). Thus if $\bigcup E_i'$ is true, then so is $\bigcup E_i$.

Substituting $\lambda = \frac{1}{2e^2n^2}$, we have:

$$P_i(i \geq 2) < P_1 \cdot \frac{1}{n^{i-1}}$$

Now, substituting P_i, $i \geq 2$ into the formula for the expected delay on p, we have:

$$\begin{aligned}
\mathbf{E}\{\text{delay on } p \text{ in } \mathcal{N}_{C,PS}\} &\leq P_1 D_1^{PS} + P_2(2n) + P_3(3n) + \ldots \\
&< P_1 D_1^{PS} + P_1 \cdot 2 + P_1 \cdot \frac{3}{n} + P_1 \cdot \frac{4}{n^2} + \ldots \\
&< P_1 \left(D_1^{PS} + 6 \right) \ (\text{for } n \geq 2)
\end{aligned}$$

Since

$$\begin{aligned}
D_1^{PS} &= \mathbf{E}\Big\{ \text{ Delay on } p \text{ caused by 1 other packets arriving in } (-n, n) \Big\} \\
&= \mathbf{E}\Big\{ \text{ Delay on } p \text{ caused by 1 packet of same type arriving in } (-n, n) \Big\} \\
&\quad + \mathbf{E}\Big\{ \text{ Delay on } p \text{ caused by 1 packet of opposite type arriving in } (-n, n) \Big\} \\
&= \mathbf{Pr}\Big\{ \text{ same type arrival } \Big\} \cdot \mathbf{E}\Big\{ \text{Delay } \Big| \text{ same type arrival } \Big\} \\
&\quad + \mathbf{Pr}\Big\{ \text{ opp. type arrival } \Big\} \cdot \mathbf{E}\Big\{ \text{Delay } \Big| \text{ opp. type arrival } \Big\} \\
&= \frac{1}{2} \cdot \Theta\left(\frac{1}{n} \cdot n \right) \quad (\text{delayed by } n \text{ only if packet arrived in } (-1, 1)) \\
&\quad + \frac{1}{2} \cdot \Theta(1) \qquad (\text{opposite type packet causes at most delay of } \Theta(1)) \\
&= \Theta(1)
\end{aligned}$$

We have

$$\begin{aligned}
\mathbf{E}\{\text{delay on } p \text{ in } \mathcal{N}_{C,PS}\} &= P_1 \left(D_1^{PS} + 6 \right) \\
&= P_1 \cdot \Theta(1)
\end{aligned}$$

4.3.2 Lowerbounding $\mathbf{E}\{\text{Delay on arrival in } \mathcal{N}_{C,FCFS}\}$

To derive a simple lower bound for the expected delay in $\mathcal{N}_{C,FCFS}$, again let p represent an arriving packet in $\mathcal{N}_{C,FCFS}$. Assume p arrives at $\mathcal{N}_{C,FCFS}$ at time 0. To lowerbound the $\mathbf{E}\{\text{Delay on } p \text{ in } \mathcal{N}_{C,FCFS}\}$, we consider only the delay on p caused by 1 packet arriving during $(-n, n)$.

$$\mathbf{E}\{\text{delay on } p \text{ in } \mathcal{N}_{C,FCFS}\} \geq P_1 D_1^{FCFS}$$

$$\begin{aligned}
D_1^{FCFS} &= \mathbf{E}\left\{ \text{ Delay on } p \text{ caused by 1 other \ packets arriving in } (-n,n) \ \right\} \\
&= \mathbf{E}\left\{ \text{ Delay on } p \text{ caused by 1 packet \ of same type arriving in } (-n,n) \ \right\} \\
&\quad +\mathbf{E}\left\{ \text{ Delay on } p \text{ caused by 1 packet \ of opposite type arriving in } (-n,n) \ \right\} \\
&= \mathbf{Pr}\left\{ \text{ same type \ arrival } \right\} \cdot \mathbf{E}\left\{ \text{Delay} \ \middle| \ \text{same type \ arrival } \right\} \\
&\quad +\mathbf{Pr}\left\{ \text{ opp. type \ arrival } \right\} \cdot \mathbf{E}\left\{ \text{Delay} \ \middle| \ \text{opp. type \ arrival } \right\} \\
&= \frac{1}{2} \cdot \Theta(1) \quad \text{(see intuition section)} \\
&\quad +\frac{1}{2} \cdot \Theta(n) \qquad \text{(see intuition section)} \\
&= \Theta(n)
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbf{E}\{\text{delay on } p \text{ in } \mathcal{N}_{C,FCFS}\} &> P_1 \cdot D_1^{FCFS} \\
&= P_1 \cdot \Theta(n)
\end{aligned}$$

5 Conclusion and Future Work

We started this paper by formulating any dynamic packet routing network as a queueing network of type $\mathcal{N}_{C,FCFS}$. Since queueing theory only provides us with results on $\mathcal{N}_{E,FCFS}$, our goal became to bound the average delay of $\mathcal{N}_{C,FCFS}$ by the average delay of $\mathcal{N}_{E,FCFS}$:

$$\text{AvgDelay}(\mathcal{N}_{C,FCFS}) \ \leq \text{AvgDelay}(\mathcal{N}_{E,FCFS}) \tag{2}$$

We first proved that (2) holds for all queueing networks wtih Markovian routing. This result was significant because many packet-routing networks where the packets have random destinations can be formulated as queueing networks with Markovian routing. We then gave a counterexample showing that (2) does not always hold, contrary to popular belief.

There are three natural open questions raised by these results. Let $\mathcal{S}$ be those networks for which (2) holds. The first is *"How large is the set $\mathcal{S}$?"* We know $\mathcal{S}$ contains more than just Markovian networks. For instance it's easy to prove that $\mathcal{S}$ contains the network $\mathcal{N}$ which consists of just a single server, where each incoming packet serves once, goes back to the end of the queue, and then serves a second time. Also, simulations suggest $\mathcal{S}$ contains many other non-Markovian networks (see Section 1.3.2). In fact, the difficulty in constructing a network not in $\mathcal{S}$ leads us to speculate that almost all networks are in $\mathcal{S}$.

This leads us to the second question of *"How tight an upper bound is $\mathcal{N}_{E,FCFS}$ on $\mathcal{N}_{C,FCFS}$ with respect to average delay?"*, both in practice and theoretically.

Lastly, *"For the networks not in $\mathcal{S}$, how far off is the AvgDelay($\mathcal{N}_{C,FCFS}$) from the AvgDelay($\mathcal{N}_{E,FCFS}$)?"*

6 Acknowledgements

We thank Jean Walrand for many helpful discussions and for deflecting us away from dead ends.

References

[Ale82] R. Aleliunas. Randomized parallel communication. In *ACM-SIGOPS Symposium on Principles of Distributed Systems*, pages 60–72, 1982.

[ALMN90] Bill Aiello, Tom Leighton, Bruce Maggs, and Mark Newman. Fast algorithms for bit-serial routing on a hypercube. In *2nd Annual ACM Symposium on Parallel Algorithms and Architectures*, pages 55–64, July 1990.

[BCMPG75] F. Baskett, K.M. Chandy, R.R. Muntz, and F. Palacios-Gomez. Open, closed, and mixed networks of queues with different classes of customers. *Journal of the Association for Computing Machinery*, 22:248–260, 1975.

[BS93] John A. Buzacott and J. George Shanthikumar. *Stochastic Models of Manufacturing Systems.* Prentice Hall, 1993.

[CS86] Yukon Chang and Janos Simon. Continuous routing and batch routing on the hypercube. In *Proceedings of the 5th Annual ACM Symposium on Principles of Distributed Computing*, pages 272–278, 1986.

[Fer94] Dominico Ferrari, 1994. Personal Communication.

[GL85] Ronald I. Greenberg and Charles E. Leiserson. Randomized routing on fat-trees. In *26th Annual Symposium on Foundations of Computer Science*, pages 241–9, October 1985.

[HBB94] Mor Harchol-Balter and Paul E. Black. Queueing analysis of oblivious packet-routing algorithms. In *Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 583–592, January 1994.

[HC86] Bruce Hajek and Rene L. Cruz. Delay and routing in interconnection networks. In *Flow Control of Congested Networks*, pages 235–242, 1986.

[Kel75] F. P. Kelly. Networks of queues with customers of different types. *Journal of Applied Probability*, 12:542–554, 1975.

[KL95] Nabil Kahale and Tom Leighton. Greedy dynamic routing on arrays. In *Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 558–566, January 1995.

[Kle76] Leonard Kleinrock. *Queueing Systems Volume II: Computer Applications.* John Wiley and Sons, New York, 1976.

[Kle94] Len Kleinrock, 1994. Personal Communication.

[Lei90] Tom Leighton. Average case analysis of greedy routing algorithms on arrays. In *2nd Annual ACM Symposium on Parallel Algorithms and Architectures*, pages 2–10, July 1990.

[Lei92] Tom Leighton. Methods for message routing in parallel machines. In *24th Annual ACM Symposium on Theory of Computing*, pages 77–96, May 1992.

[LMR88] Tom Leighton, Bruce Maggs, and Satish Rao. Universal packet routing algorithms. In *29th Annual Symposium on Foundations of Computer Science*, pages 256–69, October 1988.

[MC86] Debasis Mitra and R. Cieslak. Randomized parallel communications. In *International Conference on Parallel Processing*, pages 224–230, 1986.

[PU87] David Peleg and Eli Upfal. The generalized packet routing problem. *Theoretical Computer Science*, 53(2–3):281–93, 1987.

[Ros78] Sheldon M. Ross. Average delay in queues with non-stationary Poisson arrivals. *Journal of Applied Probability*, 15:602–609, 1978.

[Ros89] Sheldon M. Ross. *Introduction to Probability Models.* Academic Press, Inc., Boston, 1989.

[RS92] Rhonda Righter and J. George Shanthikumar. Extremal properties of the fifo discipline in queueing networks. *Journal of Applied Probability*, 29:967–978, 1992.

[ST94] George D. Stamoulis and John N. Tsitsiklis. The efficiency of greedy routing in hypercubes and butterflies. *IEEE Transactions on Communications*, 42(11):3051–3061, November 1994.

[Upf84] Eli Upfal. Efficient schemes for parallel communication. *Journal of the ACM*, 31:507–517, 1984.

[Val82] L. G. Valiant. A scheme for fast parallel communication. *SIAM Journal on Computing*, 11(2):350–61, May 1982.

[VB81] L.G. Valiant and G.J. Brebner. Universal schemes for parallel communication. In *13th Annual ACM Symposium on Theory of Computing*, pages 263–277, May 1981.

[Vet95] Ronald J. Vetter. Atm concepts, architectures, and protocols. *Communications of the ACM*, 38(3):31–38, February 1995.

[Wal89] Jean Walrand. *Introduction to Queueing Networks.* Prentice Hall, New Jersey, 1989.

[Wal94] Jean Walrand, 1994. Personal Communication.

[Whi80] Ward Whitt. The effect of variability in the $GI/G/s$ queue. *Journal of Applied Probability*, 17(4):1062–1071, 1980.

[Whi83] Ward Whitt. Comparison conjectures about the $M/G/s$ queue. *Operations Research Letters*, 2(5):203–209, December 1983.

[Whi84] Ward Whitt. Minimizing delays in the $GI/G/1$ queue. *Operations Research*, 32(1):41–51, 1984.

[Wol77] R. W. Wolff. The effect of service time regularity on system performance. In K. M. Chandy and M. Reiser, editors, *International Symposium on Computer Performance Modeling, Measurement, and Evaluation*, pages 297–304, Amsterdam, 1977.

[Wol89] R. W. Wolff. *Stochastic Modeling and the Theory of Queues.* Prentice Hall, Englewood Cliffs, NJ, 1989.