Simulation of Network Delays

Larisa Dvorkina

08/20/96

Abstract

We compare performance of two types of queuing networks: First-Come-First-Served (FCFS) and Processor Sharing (PS). Both have Poisson arrivals; both have deterministic (constant time) servers. However, in the first type the service order is FCFS, and in the second it is PS. We investigate which factors affect the relationship between the average delay of packets in the FCFS and the PS networks. Knowing how to calculate the average delay in a PS network, we try to come up with a systematic way to estimate the delay of packets in the corresponding network with FCFS servers.

1 Introduction

Many computer applications require packets to be routed in a network. To be able to estimate performance of a particular network we need to have a way to calculate average time by which packets in that network are delayed. Since many networks traditionally use FCFS service order, it would be beneficial to have an exact formula for calculating delays of packets in such networks. Unfortunately, no such formula exists for most FCFS networks. On the other hand, there do exist such formulas for the average delay in the corresponding networks with PS service order. In this paper we study the relationship between the PS and FCFS average packet delay for a few specific networks. From our experiments we try to extract a few properties, or characteristics of networks which affect this relationship. Our overall goal is to be able to approximate the average delay in a FCFS network by knowing the delay in the corresponding PS network.

1.1 Terminology

We assume that when a packet arrives at a network, it has a route associated with it which it follows. The total time that a packet spends on its route consists of the time during which a packet is receiving service and the time spent waiting to be serviced.

Network Delay of a packet p is the time p spends waiting (not receiving service) along its route.

Average Delay is defined as the average *Network Delay* over all packets.

In FCFS and PS networks delay is accumulated in different ways.

• Servers of a FCFS network are *non-preemptive*, which means that once a server starts serving a packet it can not be interrupted and start serving another packet until the first packet has been completely serviced. Since a server can serve only one packet at a time, a waiting queue forms. When a new arrival p occurs at a server which is busy serving another packet, p is placed in the waiting queue. Before p can start receiving service it has to wait for the packet currently at the server plus any other packets which arrived to this server earlier and are in the waiting queue in front of p to complete. Once p gets to serve it receives dedicated service. Thus, the only delay p accumulates is while waiting in the queue.

• In a PS network servers are *preemptive*, i.e. whenever a new packet p arrives, the server starts serving it immediately along with any other packets which it is already serving. Thus, there is no delay associated with waiting in a queue. However, p does not receive dedicated service; it only gets a fraction of service which the

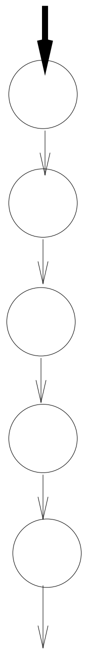

Figure 1: Chain network with 5 servers. All servers have service time 1. All packets arrive at the top (head) of the chain, serve at each server in order, and depart.

server gives to all packets currently at this server. For example if there are 3 packets being served at a server with service rate 1 packet per second, each packet gets served at rate 1/3. Observe that p is also slowed down by future arrivals which occur during p's service.

1.2 Experimental Design

All our networks have servers with rate 1 packet per second. We investigate the relationship between *average delays* in FCFS and PS networks with respect to 3 parameters:

- *load in system* - rate of outside arrivals to the network as a fraction of the maximum possible steady-state arrival rate.
- *route length* - average number of servers on all packet routes.
- *length of intersection* - number of servers two routes have in common.

We look at how each of these factors influence the relationship between delay in two networks. We consider both the difference and the ratio of the delays. We denote the average delay in the PS network by $\mathcal{D}_{PS}$ and the average delay in the corresponding FCFS network by $\mathcal{D}_{FCFS}$.

Ratio between average delay in PS and FCFS networks is $\frac{D_{PS}}{D_{FCFS}}$.

Difference between average delay in PS and FCFS networks is $D_{PS} - D_{FCFS}$.

In all our experiments the average delay is measured over one million packets.

2 Description of Examples and Results

2.1 Route Length

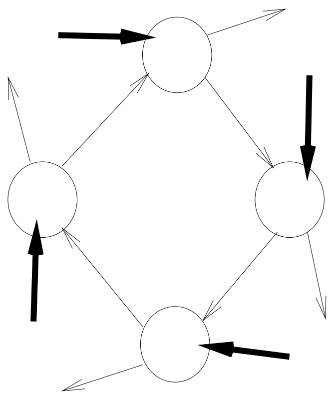

Figure 2: Ring network with 4 servers. All servers have service time 1. All packets move in clockwise direction. The outside arrival rates are the same at each server. Each packet is born with a route. A packet is equally likely to depart after serving at only the first server, first two servers, first three servers, or first four servers.

The *chain* (see Figure 1) is a completely tractable network for which the following relationship between $\mathcal{D}_{PS}$ and $\mathcal{D}_{FCFS}$ is known to be true:

$$D_{PS} = 2 \cdot RouteLength \cdot D_{FCFS}.$$

Thus, the **Ratio** is twice the route length irrespective of the load in the system. Table 1 verifies this result experimentally. The same table also shows that $\mathcal{D}_{FCFS}$ in the chain network is constant with respect to the route length. Since all the servers in our networks have constant service time of 1 second, after the packets serve at the first server in the chain they arrive at all of the subsequent servers 1 second apart. Thus, in the chain network of a FCFS service order a packet acquires delay only at the first server. However, in the PS service order network packets can receive service concurrently and, therefore, they are not necessarily spaced out by 1 second. So, as the route length increases, $\mathcal{D}_{PS}$ grows. Consequently, the **Ratio** increases proportionally to the route length and the **Difference** increases as well.

We try to determine if the same simple relationship holds in the *ring* network (see Figure 2). Our ring allows packets to have different route lengths, so we examine delay as a function of the average route length. Table 2 displays the results of this experiment. Analogously to the chain network, in the ring the **Ratio** increases with the increase in the route length. However, it does not happen as quickly as in the chain network. For the ring, the **Ratio** appears to be between 1 and 1.5 times the average route length:

$$D_{PS} = k \cdot RouteLength \cdot D_{FCFS}, where\ 1 < k < 1.5.$$

$$(k\ decreases\ as\ the\ average\ route\ length\ increases.)$$

Furthermore, $\mathcal{D}_{FCFS}$ is not constant for various route lengths. Since in the ring network the outside arrivals occur at each server, unlike the case of the chain, packets are not necessarily separated in time by 1 second. The interference between the packets produces possible delay at each server in the network; therefore, $\mathcal{D}_{FCFS}$ is higher in ring networks with longer routes.

This experiment suggests some generalizations about the effect of the route length on the relationship between the average delays in the PS and the corresponding FCFS networks. If the outside arrivals occur at one server only, the average delay in the network of a FCFS service order is constant with respect to the average route length; likewise, the **Ratio** in such networks is proportional to the average route length. However, once some interference is introduced at each server of the network, the simple relationship between $\mathcal{D}_{PS}$ and $\mathcal{D}_{FCFS}$ does not hold any longer. The only conclusion we can make then is that both the **Ratio** and the **Difference** grow as a function of the average route length, but they can not be described by a simple formula.

See Figures 3 through 8 for the plots of $\mathcal{D}_{PS}$, $\mathcal{D}_{FCFS}$, the **Ratio**, and the **Difference** as functions of the route length.

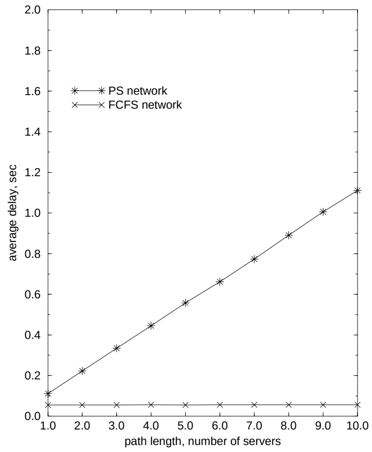

Figure 3: Average delays in the FCFS and the corresponding PS chain networks as a function of route length.

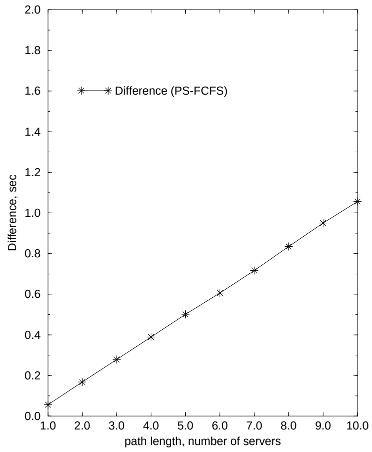

Figure 4: **Difference** between average delays in the FCFS and the corresponding PS chain networks as a function of route length.

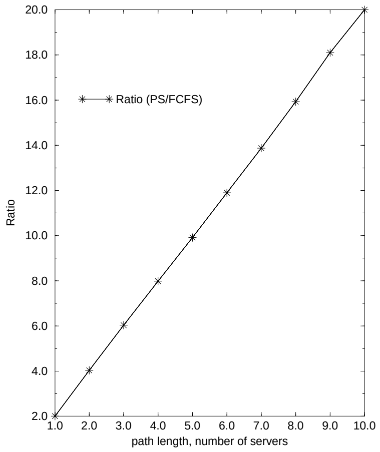

Figure 5: **Ratio** of average delays in the FCFS and the corresponding PS chain networks as a function of route length.

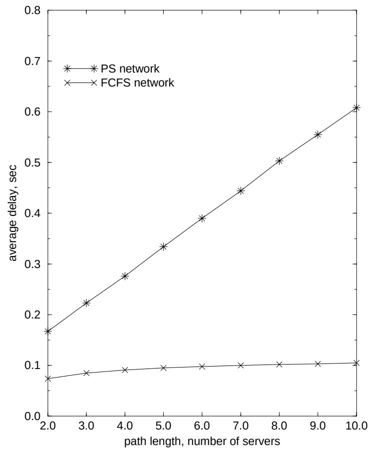

Figure 6: Average delays in the FCFS and the corresponding PS ring networks as a function of route length.

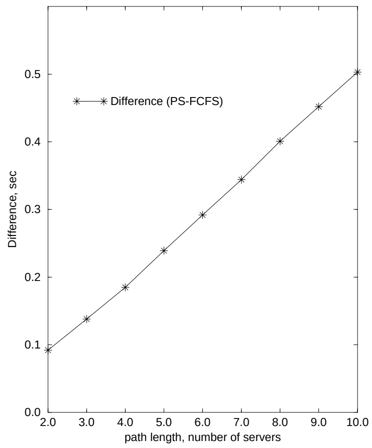

Figure 7: **Difference** between average delays in the FCFS and the corresponding PS ring networks as a function of route length.

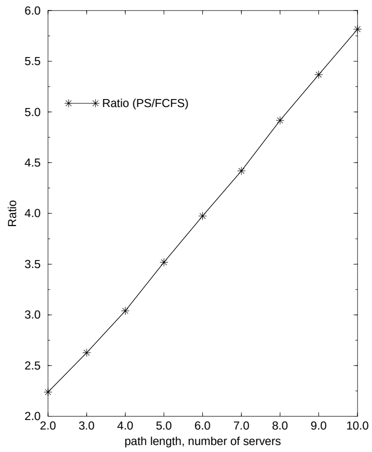

Figure 8: **Ratio** of average delays in the FCFS and the corresponding PS ring networks as a function of route length.

Route Length	1	2	3	4	5	6	7	8	9	10
$\mathcal{D}_{PS}$	.111	0.223	0.334	0.445	0.558	0.662	0.773	0.891	1.006	1.111
$\mathcal{D}_{FCFS}$	0.055	0.055	0.055	0.056	0.055	0.056	0.056	0.056	0.056	0.056
Ratio	2.008	4.031	6.028	7.978	9.918	11.898	13.879	15.938	18.107	20.0
Difference	0.056	0.168	0.279	0.389	0.501	0.606	0.717	0.835	0.950	1.056

Table 1: Average Delay for PS and FCFS networks for various route lengths. (Chain network)

Average Route Length	1.5	2	2.5	3	3.5	4	4.5	5	5.5
$\mathcal{D}_{PS}$	0.167	0.223	0.276	0.334	0.390	0.444	0.503	0.555	0.608
$\mathcal{D}_{FCFS}$	0.074	0.085	0.091	0.095	0.098	0.100	0.102	0.103	0.105
Ratio	2.239	2.627	3.039	3.518	3.975	4.419	4.916	5.367	5.816
Difference	0.092	0.138	0.185	0.239	0.292	0.344	0.401	0.452	0.503

Table 2: Average Delay for PS and FCFS networks for various route lengths. (Ring network)

2.2 Load in System

Since the *chain* network is analytically tractable, it can be proven that the load in system does not have any effect on the relationship between $\mathcal{D}_{PS}$ and $\mathcal{D}_{FCFS}$. The results of our experiment on the chain are shown in Table 3. The **Ratio** was nearly constant, dropping slightly at a very high load. One possible explanation for this negligible decline is that when the load in the system approaches the maximum possible steady-state load, it takes longer for the system to reach its steady state. In all our experiments the average delay was calculated over 1 million of packets regardless of the incoming rate; therefore, at the very high load the system might not have been stabilized.

Having established the fact that the **Ratio** in the chain network is constant, we then decided to analyze what happens to the relationship between the delays in the PS and the corresponding FCFS *ring* network. As the Table 4 shows, a similar relationship to that observed in the chain network holds for the ring network as well: surprisingly, the **Ratio** is basically constant as a function of load, although decreasing steadily as the incoming rate increases. We attribute the slight decline to the fact that while the ring network is similar to the chain network (all packets go from one server to the next in one direction only), in the ring network, unlike the chain, there are outside arrivals occurring at each server. Hence, in the FCFS ring network packets are not spaced out by 1 second. As the incoming rate increases, there is more and more interference at each server, $\mathcal{D}_{FCFS}$ increases faster, so the **Ratio** decreases. Thus, load in the system has some effect on the relationship between average delays in the ring networks of PS and FCFS service order.

For both the ring network and the chain network we found that the ratio was somewhat constant as a function of load. It would be interesting to investigate whether this property extends to more general networks as well.

See Figures 9 through 14 for the plots of $\mathcal{D}_{PS}$, $\mathcal{D}_{FCFS}$, the **Ratio**, and the **Difference** as functions of the system load.

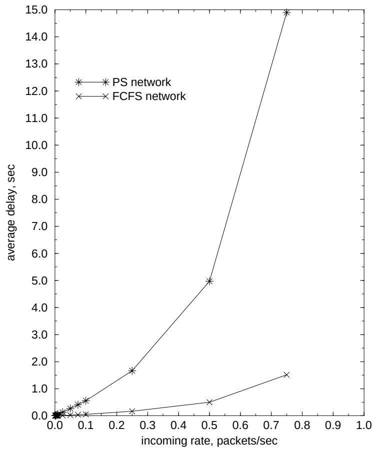

Figure 9: Average delays in the FCFS and the corresponding PS chain networks as a function of system load.

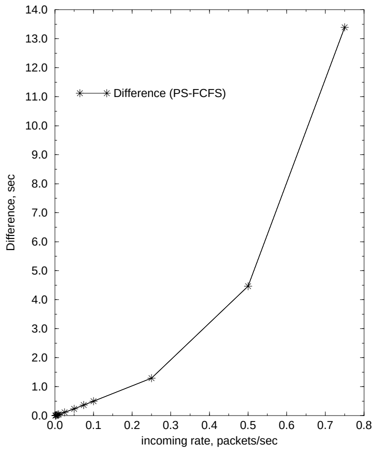

Figure 10: **Difference** between average delays in the FCFS and the corresponding PS chain networks as a function of system load.

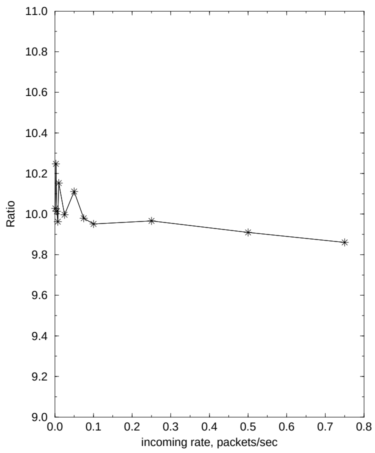

Figure 11: **Ratio** of average delays in the FCFS and the corresponding PS chain networks as a function of system load.

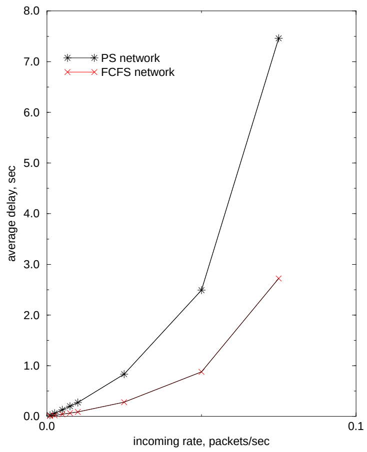

Figure 12: Average delays in the FCFS and the corresponding PS ring networks as a function of system load.

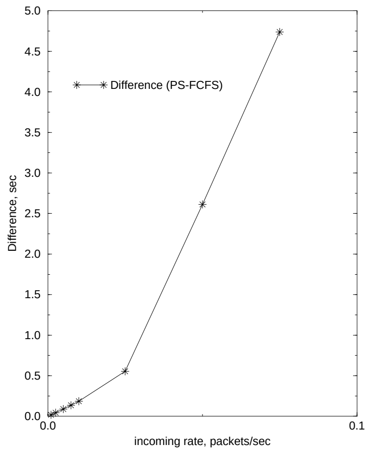

Figure 13: **Difference** between average delays in the FCFS and the corresponding PS ring networks as a function of system load.

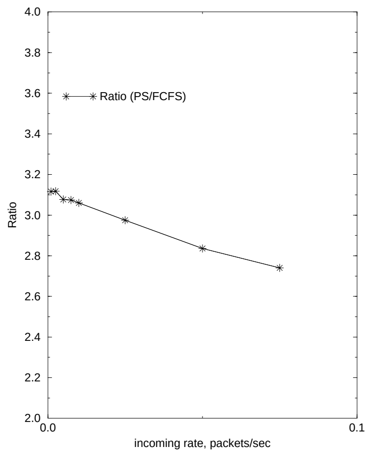

Figure 14: **Ratio** of average delays in the FCFS and the corresponding PS ring networks as a function of system load.

System Load	0.001	0.0025	0.005	0.0075	0.01	0.025	0.05	0.075	0.1	0.25	0.5	0.75
D_{PS}	0.005	0.013	0.025	0.038	0.051	0.128	0.264	0.408	0.556	1.664	4.971	14.907
D_{FCFS}	0.0005	0.001	0.002	0.004	0.005	0.013	0.026	0.041	0.056	0.167	0.502	1.512
Ratio	10.028	10.247	10.014	9.963	10.153	9.997	10.111	9.980	9.951	9.966	9.910	9.860
Difference	0.005	0.011	0.022	0.034	0.046	0.115	0.238	0.367	0.500	1.497	4.470	13.395

Table 3: Average Delay for PS and FCFS networks for various loads. (Chain network)

System Load	0.001	0.0025	0.005	0.0075	0.01	0.025	0.05	0.075
D_{PS}	0.025	0.064	0.132	0.203	0.276	0.835	2.491	7.460
D_{FCFS}	0.008	0.021	0.043	0.066	0.090	0.281	0.878	2.722
Ratio	3.116	3.118	3.077	3.074	3.060	2.975	2.836	2.740
Difference	0.017	0.043	0.089	0.137	0.186	0.554	2.613	4.738

Table 4: Average Delay for PS and FCFS networks for various loads. (Ring network)

2.3 Length of Intersection

The analysis of the ring network in the previous experiments led us to believe that the relationship between average delays in the FCFS and the corresponding PS networks is influenced not only by the route length, but also by the interference among the packets. In this section we examine the effect of path intersections on **Ratio** and **Difference**. We analyze a network in which packet routes intersect causing extra interference. This *intersection network* is shown in the Figure 15. It consists of two intersecting routes; the number of servers common to both routes is gradually increased from 1 to 10.

As shown in Table 5, increasing the number of common servers increased the **Ratio** and the **Difference**. To understand why this is the case, consider the following explanation. In the intersection network the incoming rate into each route is kept constant; however, the result of the intersection is that all of the servers which are common to both routes receive double load. As explained in Section 2.1, the average delay in the FCFS network does not depend on the route length, so increasing the number of common servers does not affect $\mathcal{D}_{FCFS}$; to verify this, see that in Table 5 $\mathcal{D}_{FCFS}$ is roughly constant throughout the experiment. However, in the PS network, delay is incurred at every server. As the length of intersection of the 2 routes increased, the number of servers receiving the double load is greater. Thus, $\mathcal{D}_{PS}$ also increases, so the **Ratio** goes up.

See Figures 16 through 18 for the plots of $\mathcal{D}_{PS}$, $\mathcal{D}_{FCFS}$, the **Ratio**, and the **Difference** as functions of the length of intersection.

3 Conclusion

We set out to determine whether it is possible to estimate the average delay of packets in the First-Come-First-Served network given the average delay in the corresponding Processor Sharing network. We observed that the relationship between average delays in the PS and the corresponding FCFS networks can be measured by the

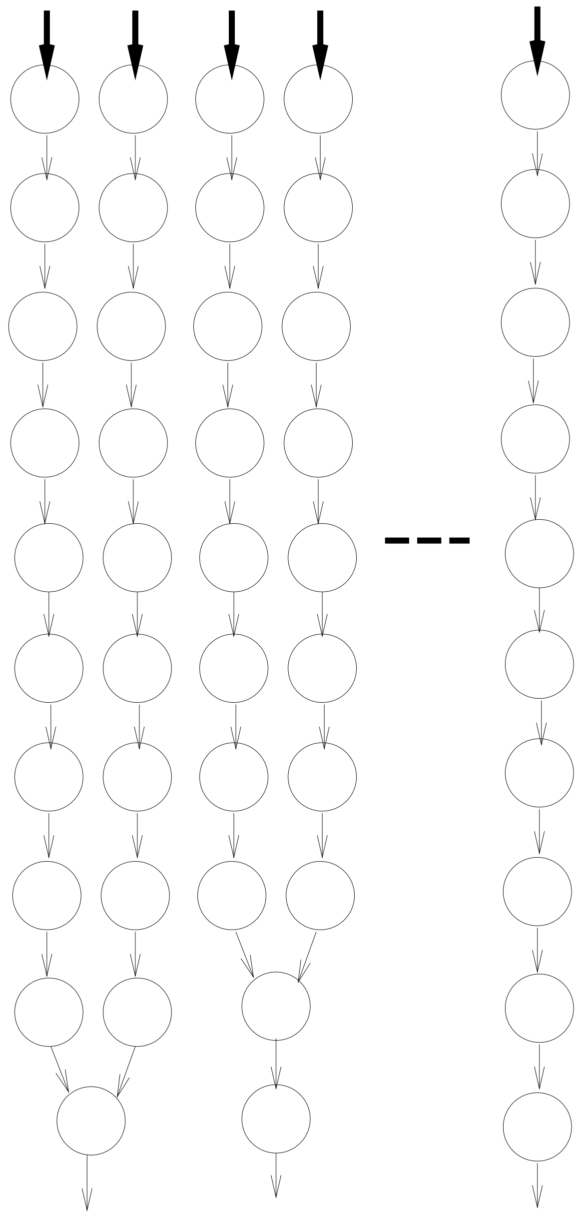

Figure 15: Network with two intersecting routes. Each route's length is 10 servers. The length of intersection varies from 1 to 10 servers. All servers have service time 1. All packets arrive at the top (head) of each route, serve at each server on that route in order, and depart.

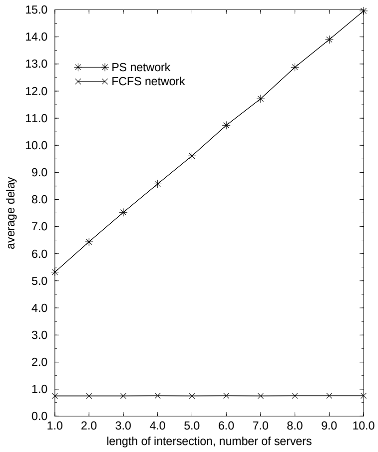

Figure 16: Average delays in the FCFS and the corresponding PS networks as a function of length of intersection.

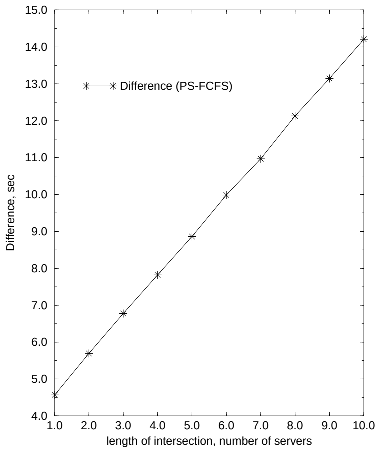

Figure 17: **Difference** between average delays in the FCFS and the corresponding PS networks as a function of length of intersection.

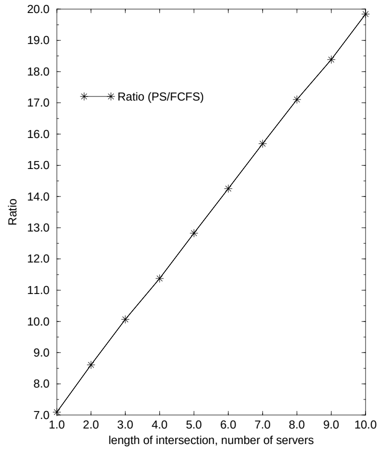

Figure 18: **Ratio** of average delays in the FCFS and the corresponding PS networks as a function of length of intersection.

Intersection Length	1	2	3	4	5	6	7	8	9	10
D_{PS}	5.321	6.441	7.522	8.574	9.611	10.737	11.716	12.881	13.901	14.962
D_{FCFS}	0.750	0.748	0.747	0.755	0.749	0.753	0.746	0.753	0.756	0.754
Ratio	7.097	8.607	10.065	11.377	12.826	14.255	15.695	17.106	18.385	19.844
Difference	4.571	5.692	6.775	7.820	8.862	9.984	10.970	12.128	13.145	14.208

Table 5: Average Delay for PS and FCFS networks for various number of servers in intersection.

Difference, $\mathcal{D}_{PS}$ - $\mathcal{D}_{FCFS}$, and the **Ratio**, $\frac{\mathcal{D}_{PS}}{\mathcal{D}_{FCFS}}$. We found that there are several network properties which influence this relationship. These include the *system load*, the *average packet route length*, and the *length of intersection.* We only measured 3 networks; however, on our narrow sample space, we were able to extract a few preliminary conclusions:

- *Route length* matters to both the **Ratio** and the **Difference**. Given $\mathcal{D}_{PS}$, to get $\mathcal{D}_{FCFS}$, approximately divide $\mathcal{D}_{PS}$ by k times the average Route Length, where $k = 2$ for the chain network, $1 < k < 1.5$ for the ring network.
- *System Load* doesn't matter with respect to the **Ratio**, but does matter with respect to the **Difference**. We haven't been able to come up with an approximation of $\mathcal{D}_{FCFS}$ given $\mathcal{D}_{PS}$ and the **Difference** with respect to the load.
- *Length of Intersection* influences both the **Ratio** and the **Difference**. Given $\mathcal{D}_{PS}$, if the network has a lot of intersecting routes, $\mathcal{D}_{FCFS}$ is even smaller than one would expect for a non-intersecting network with the given average route length.