Figure 7.9 repeats an experiment similar to the one in Figure 7.8 but with a faster link speed of 24 Gb/s and a mean frame size of 530 bytes. One possible interpretation for this experiment is that arrivals are bursty with mean burst size of 10 ATM cells. The figure displays the mean delay per burst.

Figure 7.10 depicts an experiment similar to the one in Figure 7.9 but with a bigger band size of 100 μs, or 80% of the cycle. Again, one can see the same trend as in the other experiments.

Finally, Figure 7.11 depicts the case of a CS system with fixed size ATM cells and 2.4 Gb/s maximum link rate. Each circuit lasts 2.5 μs and the cycle period is 125 μs. One can see that again the approximation matches the simulation for low to medium loads.

7.3.2 RDMA-

Figure 7.12 depicts the loss rate in RDMA- when the cycle period is 125 μs, the band size 25 μs, the arrival is Poisson within a band, and the frame size fixed at 53 bytes. Similar results can be obtained for exponentially distributed frame sizes. The input load axis indicates the ratio of the total input load in each input link during a band. For example, the ratio 0.8 indicates that the input links are transmitting at 1.92 Gb/s. One can see that the simulation closely matches the analysis and that the maximum loss in the system is 50%, as expected.

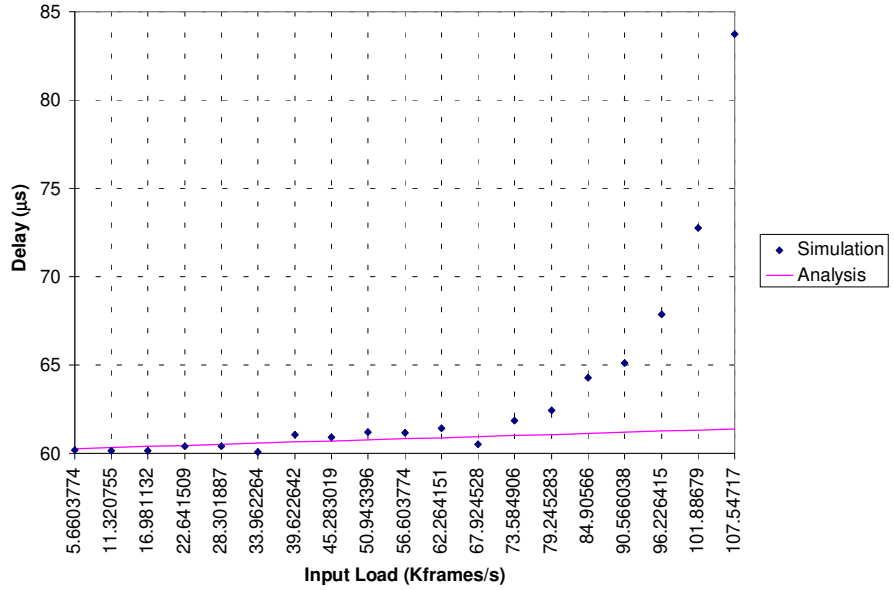

Figure 7.11: Mean frame delay (in µs) for Poisson arrivals and deterministic service under CS (the frame size is 53 bytes, the link speed 2.4 Gb/s, the circuit size 2.5 µs, the cycle period 125 µs)

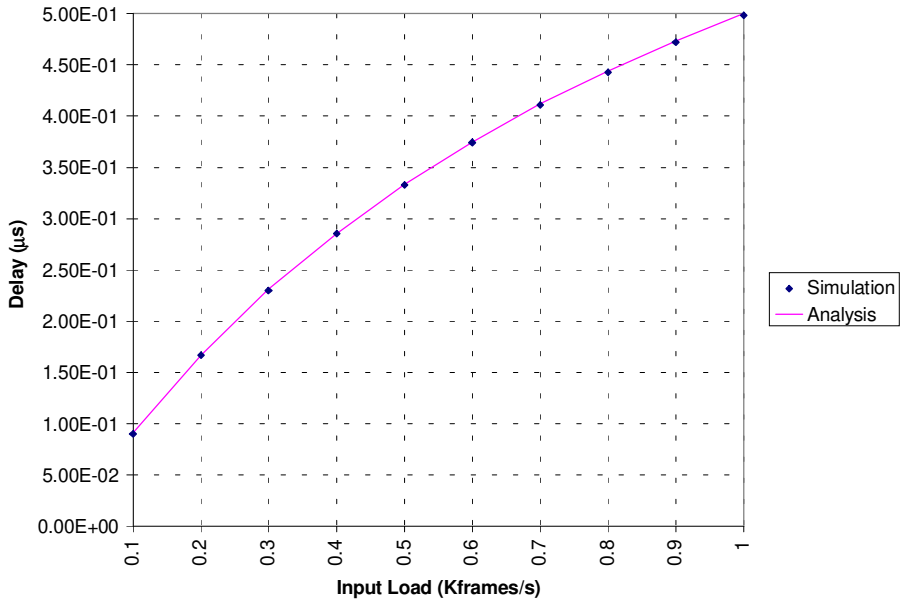

Figure 7.12: Mean loss rate for Poisson arrivals and deterministic service under RDMA- (the frame size is 53 bytes, the link speed 2.4 Gb/s, the band size 25 µs, the cycle period 125 µs)

7.3.3 RDMA+

Figure 7.13 illustrates the mean delay for RDMA+ with Poisson arrivals within a band. One can see that the analysis is a good approximation of the delay in the system, especially at low loads.

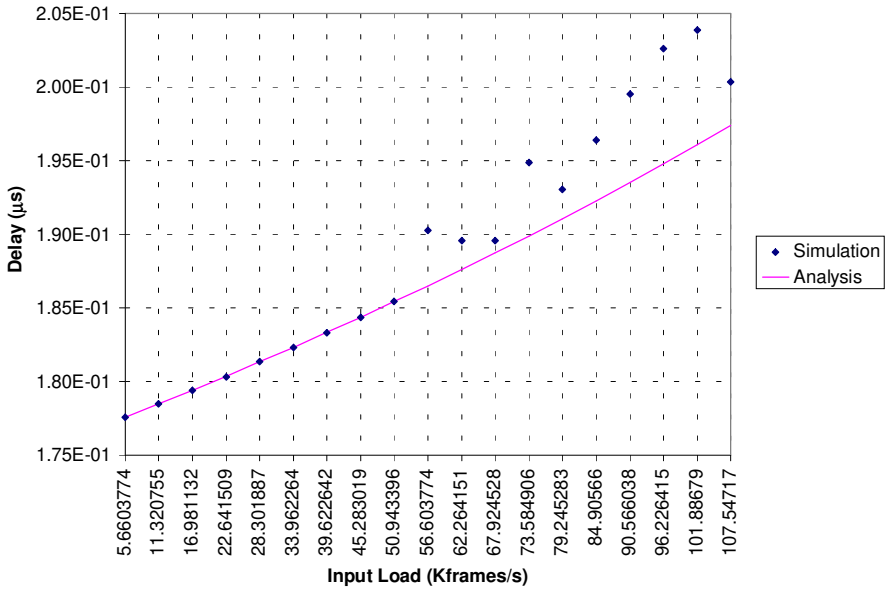

Figure 7.13: Mean frame delay (in µs) for Poisson arrivals and deterministic service under RDMA+ (the frame size is 53 bytes, the link speed 2.4 Gb/s, the band size 25 µs, the cycle period 125 µs)

7.3.4 Comparison

The goal of this section is to compare RDMA++ performance with PS and CS.

The CS simulation works as follows. There is a cycle of 125 µs within which end-to-end connections can be allocated. Each connection lasts 2.5 µs. This ensures a fair comparison with RDMA++ where a band of size 25 µs is shared among 10 sources.

CS is expected to perform always worst than RDMA++ because the time waiting for an end-to-end connection to arrive is bigger than the time waiting for a band. In the par-

ticular case being simulated, for instance, each band lasts 25 μs and is repeated every 125 μs while each end-to-end circuit lasts 2.5 μs and is also repeated every 125 μs.

The PS simulation works as follows. There is a processor operating at the same nominal speed of one of the input links. This means, for example, that if 50 instructions are necessary to process each ATM cell, the simulated processor operates at 283 MIPS, which is an optimistic assumption. Additionally, there could be contention for processor use from frames crossing the node to other destinations. Such interference does not happen in RDMA or CS. It is reflected in the simulation by making all input traffic from each link incur some delay to be processed at the node. The number of possible destinations for each input frame is 10, and so the overall input traffic to be processed is 10 times larger than the traffic that will eventually follow the correct output link in Figure 7.1.

It is expected that PS will perform better than RDMA until its processing resources are exhausted at which point RDMA will outperform PS.

The input load axis is actually the load from each input link to the output link. For the PS simulation, the input load to the processor from each input link is 10 times this value. The reason is that the PS simulation generates traffic to all 10 possible destinations and all frames need to cross the processor to be switched.

Figure 7.14 and Figure 7.15 show the mean delay for the three systems. In the first, the link speed is 2.4 Gb/s and the frame size fixed at 53 bytes. In the second, the link speed is 24 Gb/s and the frame size exponentially distributed with mean 530 bytes. As it can be verified, both experiments are in accordance with the expectations.

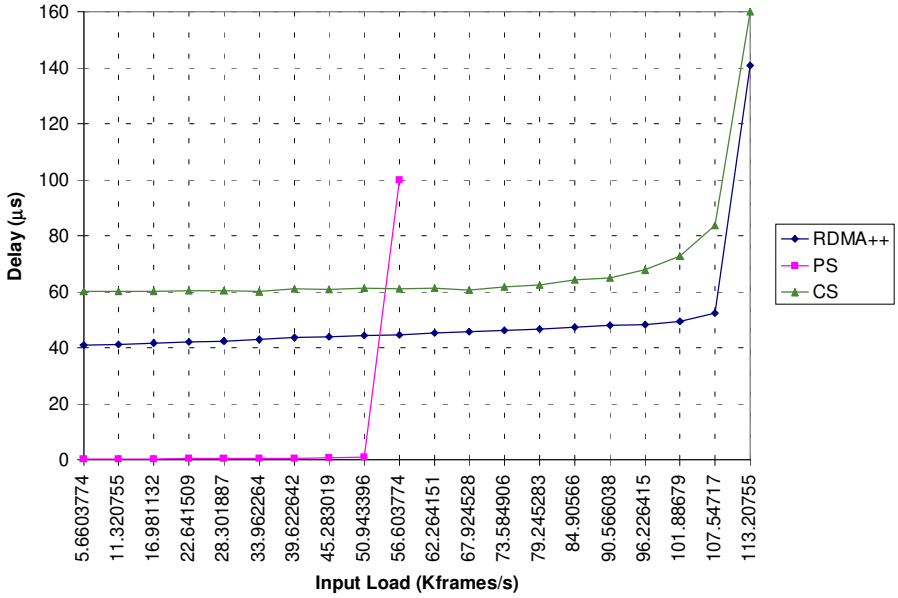

Figure 7.14: Mean frame delay comparison (in µs) for Poisson arrivals and deterministic service (the frame size is 53 bytes, the link speed 2.4 Gb/s, the band size 25 µs, the cycle period 125 µs)

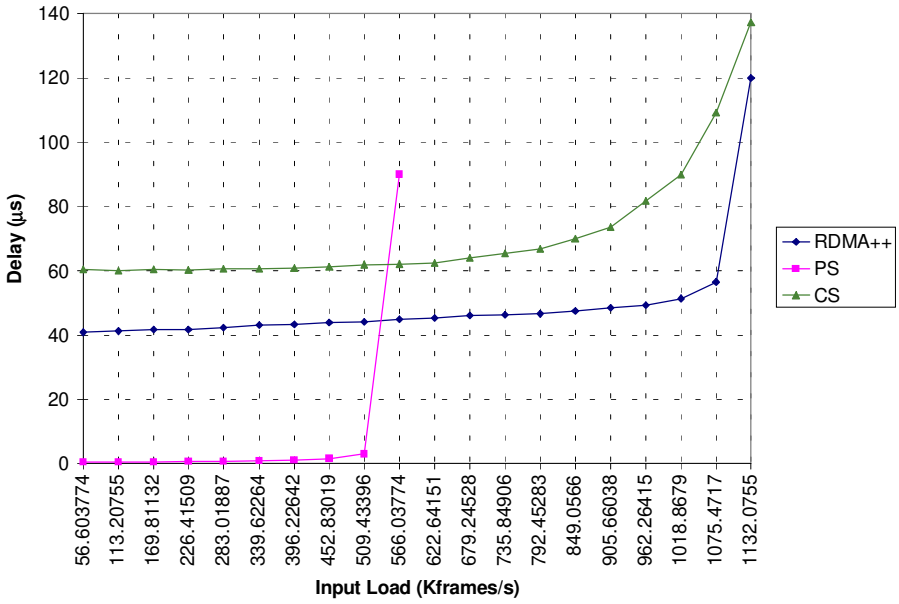

Figure 7.15: Mean frame delay comparison (in µs) for Poisson arrivals and Poisson service (the frame size is 530 bytes, the link speed 24 Gb/s, the band size 25 µs, the cycle period 125 µs)

Notice that PS becomes unstable in both simulations when the input load in each link is about 10% of its maximum capacity. At this point, the traffic to the output link in Figure 7.1 is 566.03774 Kframes/s (as depicted in Figure 7.14 and Figure 7.15) or 1% of the maximum capacity. The remaining 9% are directed to other destinations. The overall input to the processor from the 10 links is 100% of the processor capacity. That is, only 10% of the traffic is directed to the output link, but the remaining 90% still consume processor cycles to be switched. CS and RDMA++ become unstable at 2% of the input link capacity at which point the aggregate traffic for the output link is 20%, which is also the portion of the cycle allocated to the output link.

7.4 Simulation of Networks

This section extends the performance studies of RDMA to a network of nodes. As it will be seen, the results are similar to the ones obtained in the previous section.

The topology studied is depicted in Figure 7.16. It is a symmetric configuration that allows the overlapping of 3 non-interfering trees. An example of 3 non-interfering trees to destinations 1, 6 and 8 is depicted in Figure 7.17. These destinations thus can share a band. Two additional bands are sufficient to serve the 6 trees of the other destination nodes.

The simulation model works as follows. Each node generates ATM cells according to a Poisson process. Destinations are assigned to cells according to a uniform distribution. The link speed is 2.4 Gb/s, that is, each cell needs 177 ns to be transmitted. The cycle period is 125 μs. The propagation delay in each link is negligible (equivalent to 1 cell transmission delay). The bands to all destinations are of the same size. Each cell waits for the

proper destination band at the source nodes and then moves through the network down the respective tree. The goal is to give a comparison of PS, CS, and RDMA++ performance.

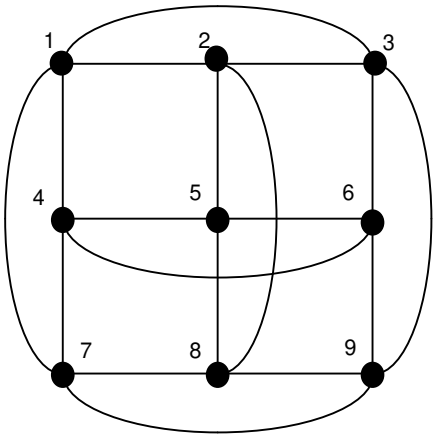

Figure 7.16: Simulated network topology

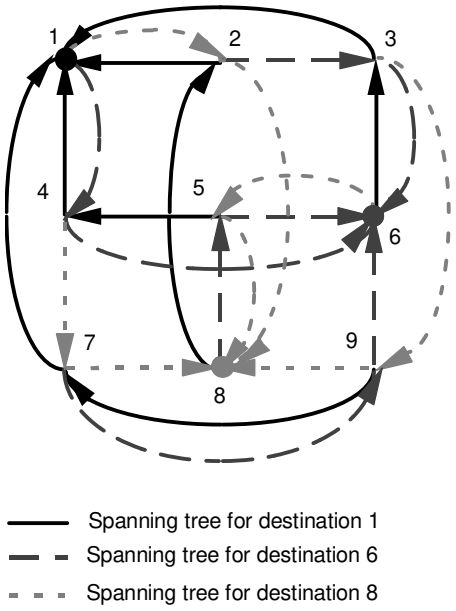

Figure 7.17: Allocation of trees in one band

The PS and CS simulations operate similarly to the ones in Section 7.3, use the same trees allocated for RDMA++ (see Figure 7.16), and cut-through technology [Kermani 78].

The CPU at each switch operates at the same rate of one input link. This means, for example, that if 50 instructions are necessary to process each ATM cell, the simulated processor operates at 283 MIPS, which is an optimistic assumption.

The simulation was run for periods where 10,000 ATM cells per input node were generated. After each of these periods, all statistics were saved and reset. Two RDMA++ experiments were conducted. In the first, all traffic had the same priority (RDMA++<c>). In the second, priority traffic was generated as follows. Each band was equally partitioned into priority sub-bands, one for each input node (RDMA++<p>).

Figure 7.18 depicts the mean packet delay (in μs) for the experiments. The input traffic load is given as a percentage of the 2.4 Gb/s maximum input rate at each node. PS has a steady performance until the input load 50% saturates the CPU capability at the nodes with delays growing unbounded. CS has a similar behavior. The unstable point is 30%. The reason is that this CS simulation allocates all nodes in its path (including the destination). It becomes unstable before RDMA++ because the latter allocates the destination node for the whole tree and not for a single circuit. An alternative CS operation could potentially be improved to allocate only links (and not nodes). In this case, the comparison is the same reported in Figure 7.11.

RDMA++ has a stable performance. Both RDMA++<c> and RDMA++<p> have the same mean packet delay characteristics, as expected from queueing analysis [Kleinrock 75], and thus overlap in the figure. The “Priority” curve plots the mean delay for priority traffic generated for the RDMA++<p> experiment. Priority was assigned randomly to ATM cells according to a uniform distribution. Priority traffic was scheduled to access the

network during its priority band, thus not incurring admission delays. The delay incurred by the priority traffic is only the propagation delay and contention with other cells scheduled at the beginning of the priority band.

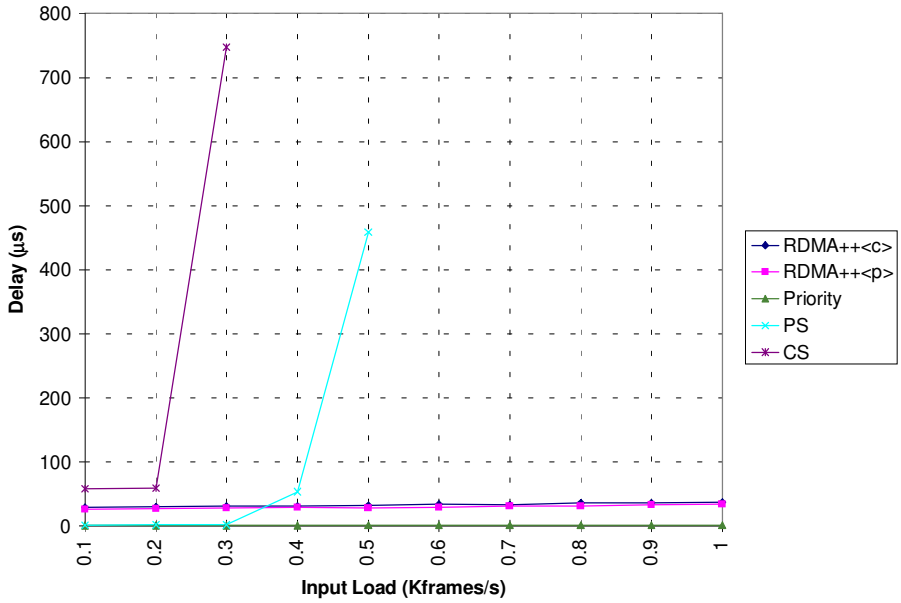

Figure 7.18: Mean frame delay comparison (in µs) for Poisson arrivals and deterministic service (the frame size is 53 bytes, the link speed 2.4 Gb/s, the band size 41.67 µs, the cycle period 125 µs)

Figure 7.19 shows the network behavior when sources generate bursty traffic according to an on/off model, where the on and off periods are geometrically distributed in the number of cells. The mean on period is 10 ATM cells.

Figure 7.20 displays a multimedia experiment. Source 9 sends motion picture frames to Destination 1. All other sources send normal data traffic generated according to a Poisson process at the load specified in the load axis. The video traffic is scheduled to be generated during the source's priority band, which is of size 10 cells every 125µs cycle. As it can be seen, the network provides high-quality service to the video source while non-QoS

demanding traffic proceeds normally. The isolation of both traffic types can be accomplished by simple band tuning (in microseconds).

Figure 7.21 shows the comparison of analysis and simulation results for the RDMA++ mean packet delay. As it can be seen, the results are in good agreement.

One can observe in Figure 7.17 that each tree has 4 links arriving at the final destination. Additionally, the final destinations in the simulation did not incur transfer delays. For this reason, one of the trees seen as a server has a throughput 4 times larger than the transmission rate at each link. This fact must be input into the formulas of the analysis by making the server 4 times faster than the maximum links rate. Additionally, the maximum load to any of the trees is 25% (and thus low). This is the reason why the analysis (that approximates better the system for low loads) closely matches the simulation results.

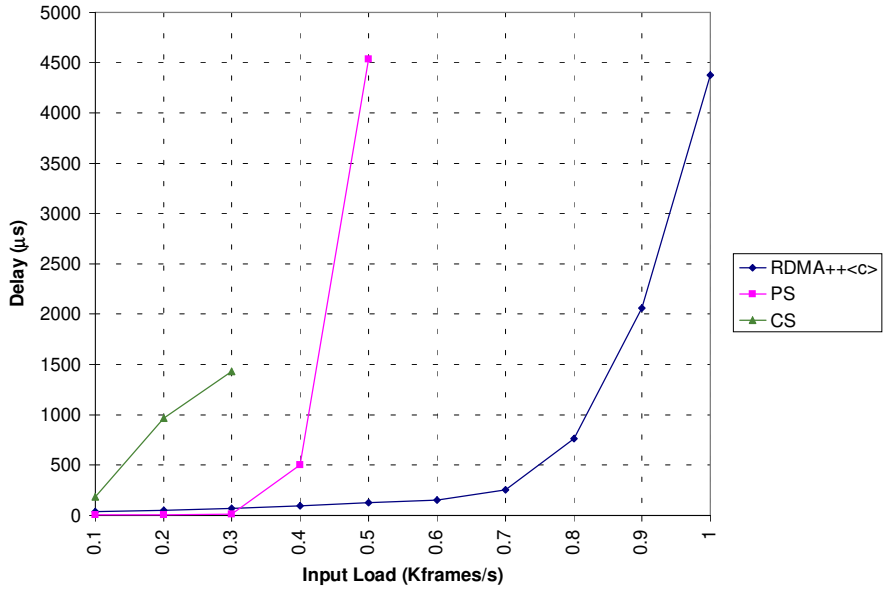

Figure 7.19: Mean frame delay comparison (in µs) for Poisson arrivals and Poisson service (the frame size is 530 bytes, the link speed 2.4 Gb/s, the band size 41.67 µs, the cycle period 125 µs)