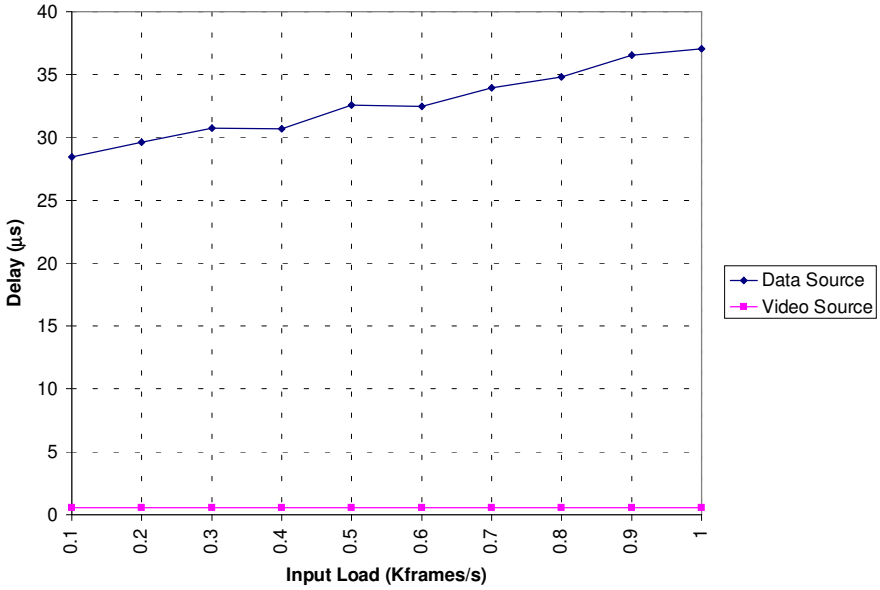

Figure 7.20: Mean frame delay comparison (in µs) for Poisson and deterministic arrivals and deterministic service under RDMA++ (the frame size is 53 bytes, the link speed 2.4 Gb/s, the band size 41.67 µs, the cycle period 125 µs)

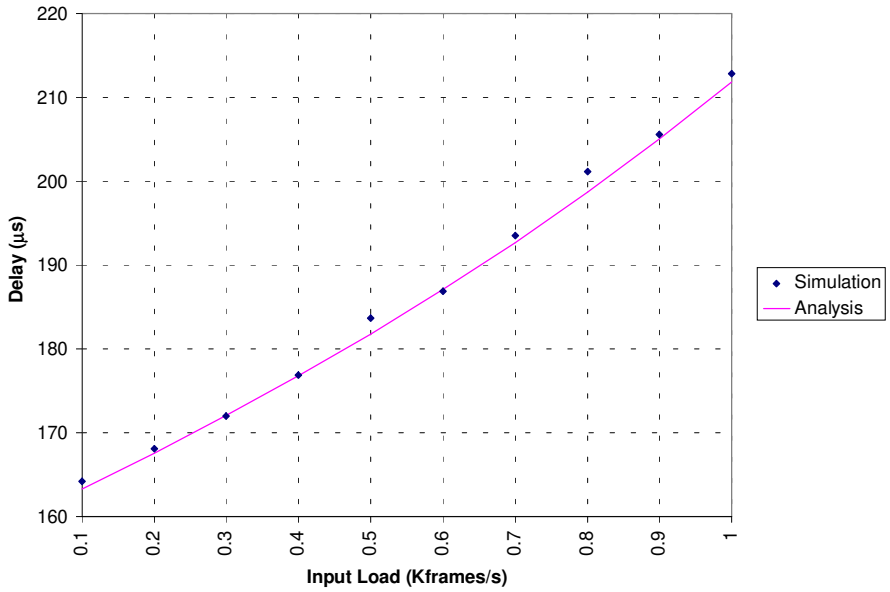

Figure 7.21: Mean frame delay (in µs) for Poisson arrivals and deterministic service under RDMA++ (the frame size is 53 bytes, the link speed 2.4 Gb/s, the band size 41.67 µs, the cycle period 125 µs)

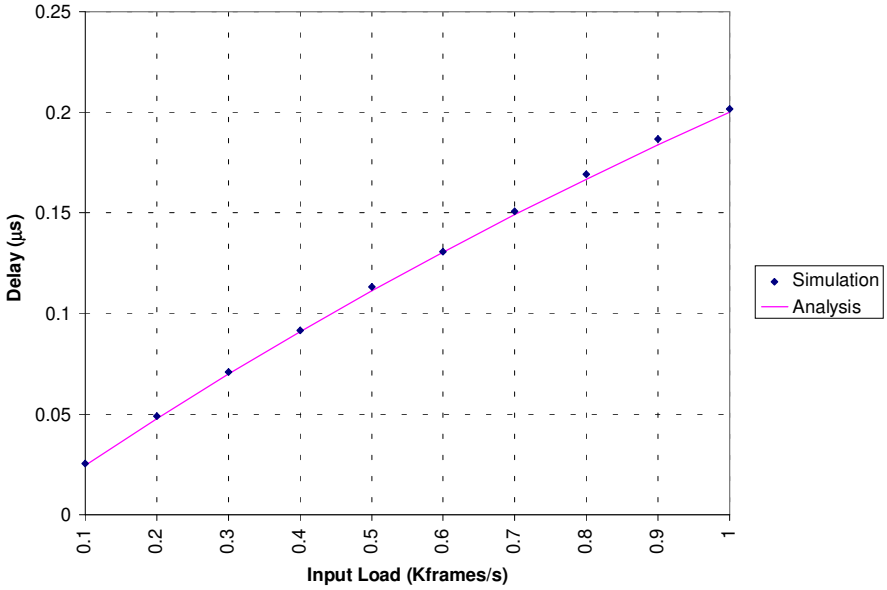

Figure 7.22: Mean loss rate for Poisson arrivals and deterministic service under RDMA- (the frame size is 53 bytes, the link speed 2.4 Gb/s, the band size 41.67 µs, the cycle period 125 µs)

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 7.22 compares the simulation and analysis results for the RDMA- mean packet loss rate. Each tree is active only 1/3 of the cycle. Thus the input load in the analysis formulas varies between 0 and 1/3 . This is why the loss rate never reaches the 50% loss upper bound.

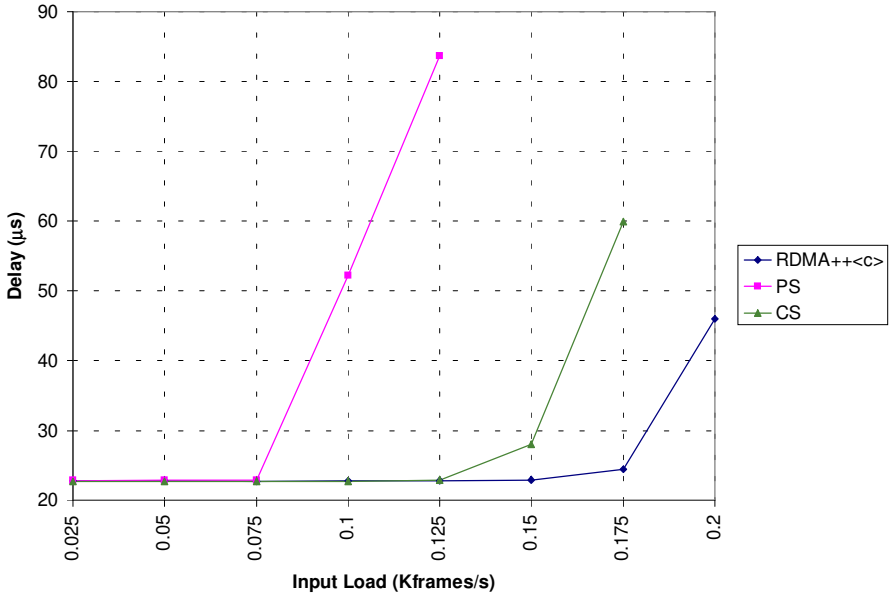

Figure 7.23: Mean frame delay comparison (in µs) for Poisson arrivals and deterministic service (the frame size is 53 bytes, the link speed 2.4 Gb/s, the band size 17.8571 µs, the cycle period 125 µs)

Finally, Figure 7.23 depicts the mean delay when the NSF T3 backbone network operates under RDMA++. The propagation delays are approximated since the exact measures are not available. The link speeds were upgraded to 2.4 Gb/s. The CPU speed was 100 MIPs for the PS simulation and some 50 instructions were necessary to process each ATM cell. It is important to notice that the topology of the NSF backbone is not particularly suitable for RDMA since only two trees can coexist in each band. Nevertheless, the performance advantage of RDMA is clear in the figure.

An important observation that can be derived from the experiment in Figure 7.23 is that, when applied to wide area networks, the time incurred waiting for a particular band is negligible when compared to the propagation delays. For instance, the waiting time for the band in the simulated NSF backbone is at most 125µs (a complete cycle), but the cross-country propagation delay is of the order of 30 ms (240 times larger). Thus, the time

waiting for a band in RDMA++ is a negligible component of the total frame delay, which in this case is very close to the PS frame delay for low loads.

7.5 Future Work

The initial performance assessment of RDMA in this chapter concentrated on understanding basic RDMA performance and on comparing it with PS and CS. The next step is to study policies to allocate RDMA bands and to study their performance. One possible outcome of such study, given specific demands, is the best allocation policy and what is the performance that it can deliver.

The analysis in this chapter had to settle for approximations due to the difficulties in finding closed-form solutions. A natural extension for such analysis is to use mechanisms that rely on numerical algorithms to resolve expressions numerically. Potentially, the same mechanisms could be used to analyze higher moments of the delay and the loss in the network.

7.6 Summary

This chapter provided initial analysis of mean RDMA delay and loss. The goal of the analysis was to understand RDMA performance and compare it with PS and CS.

The initial step was to provide an analytical understanding of RDMA behavior with Poisson arrivals on a single node. A closed form solution was found for RDMA- with generic frame size distribution. It was shown that the loss rate will never exceed 50%. The mean delay for RDMA++ was approximated for low to medium input loads. The mean

delay for RDMA+ could be approximated as well. The loss rate in RDMA+ was shown to be negligible when the band size is large compared to the mean frame size.

All the analytical results were validated through simulations. The simulations were also used to compare RDMA performance with PS and CS in two network topologies. It was found that PS has better performance for low loads. When the load in the system saturates the frame header processors, RDMA outperforms PS. It is important to notice that the PS simulation did not incur admission control delays that may deteriorate PS performance. Additionally, it was shown that, if the propagation delay in the network is large, the time waiting for bands in RDMA is negligible. In such scenario, the final end-to-end delay is comparable to PS even for low system loads. RDMA always outperforms CS because the mean time waiting for a band is smaller than the mean time waiting for a circuit. Additionally, the destination is allocated to a tree and not a circuit, which has the potential of further increasing RDMA performance when compared to CS.