Chapter 7

Route Division Multiple Access Performance Evaluation

7.1 Introduction

7.1.1 The Problem

This chapter addressed the problem of evaluating the performance of RDMA when compared with more traditional switching techniques: Packet Switching (PS) and Circuit Switching (CS). The study is carried using both analytical and simulation models.

7.1.2 Main Results

Analytical approximations. The time-dependent behavior of RDMA complicates the performance study. Usual analytical performance analysis techniques ignore time-dependent (or transient) behaviors when simplifying models in search for tractable solutions. This chapter presents approximate closed-form formulas for end-to-end delay and loss in RDMA-, RDMA+, and RDMA++. The analytical approximations are very accurate when checked with the simulation results.

Simulation studies. The natural rescue when analytical techniques fail is simulation. Even though completely realizable, simulation studies must be carefully implemented to

avoid extremely long executions and big state-spaces. This chapter presents simulation results on RDMA-, RDMA+, and RDMA++ performance.

Performance study. RDMA is compared favorably with respect to PS and CS for Poisson arrivals and bursty on-off arrivals. For low input loads, PS outperforms RDMA and CS because there is no admission delay waiting for a band or a circuit. When the load increases, frame header processors become saturated and then RDMA outperforms PS. CS always performs worst than RDMA because of the penalty incurred waiting for a circuit that is larger than the one waiting for a band. QoS parameters such as minimal delay and no jitter can be guaranteed with RDMA by using priority bands. Finally, when the propagation delay is large (in wide area communications), the time waiting for a band is negligible when compared with the propagation time. In this scenario, the performance advantages of PS with respect to RDMA become negligible.

7.1.3 Chapter Organization

The reminder of this chapter is organized as follows. Section 7.2 analyzes the performance of RMDA for a single node. Section 7.3 simulates the performance of RDMA for a single node. Section 7.4 simulates the performance of RMDA for a network of nodes. Section 7.5 describes future work. Finally, Section 7.6 summarizes.

7.2 Analysis: Contention for Outgoing Links

This section studies how RDMA resolves contention for an outgoing link. The situation is described in Figure 7.1 where a set of incoming links is competing for an outgoing link. The incoming links are partitioned in classes such that two links are in the same if and

only if they are part of the same routing tree. In the figure, there are T routing trees and tree i , $1 \le i \le T$, contains n_i links. Each link is labeled L[x,y], where x is the respective routing tree and y is the link order within the tree. The outgoing link needs to be time divided among the T trees. The goal of this section is to build a performance understanding of this scenario using analytical techniques.

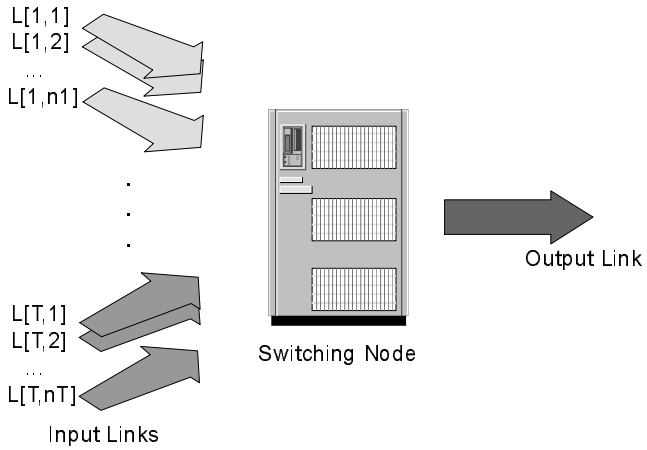

Figure 7.1: Contention for an outgoing link in RDMA

One interesting characteristic of RDMA is that each tree behaves independently of the others. From the perspective of a particular tree, the output link is available to it for a given period U and then becomes unavailable during a period V. The link is effectively a server that can service frames at a given link rate. The server cycle between two states, an operational state of duration U and a vacation state of duration V. The cycle is $C = U + V$. From this point onwards, the study will concentrate on one such independent trees.

It is assumed that the tree under study contains n links. The arrival process is Poisson with mean $\frac{\lambda}{n}$ in each link. It can be thus concluded [Kleinrock 75] that the merged input

stream has a Poisson distribution $A(t)$ with mean λ. The Poisson assumption in each link can be justified by noticing that incoming traffic to a contention band is usually random and does not require QoS guarantees.

The frame sizes are distributed according to a generic distribution $B(x)$ with mean μ. The whole tree can be modeled as an M/G/1 system with input rate λ and service rate μ that cycles between an active period of size U and a vacationing period of size V. The system utilization is $\rho = \frac{\lambda}{\mu}$. Notice that, for the system to be stable, $\rho < \frac{U}{U+V}$.

7.2.1 The RDMA- System

Under RDMA- regimen, it is assumed that sources will generate traffic according to the Poisson distribution during each band. If the traffic sources are not Poisson in nature, a filter can be placed between the sources and the input to the system to decrease the chances of burst losses. The filter will output traffic according to the Poisson distribution from any generic input traffic. The filter is thus a G/M/1 system and can be analyzed accordingly [Kleinrock 75] in order to find its contribution in the overall end-to-end frame delay.

If the band size U is large compared to the mean transmission rate μ, then one can assume that the RDMA- system operates like a Type I Counter described in [Karlin and Taylor 75] and summarized here. The Type I Counter registers instantaneous signals. Signals arrive according to a generic arrival distribution $A(t)$ and take some time to be recorded also generic with distribution $B(x)$. While a signal is being recorded, other signals are lost. The similarity with RDMA- is apparent if one replaces the transmission process

for the recording process. While a packet is being transmitted in RDMA-, other arriving packets are lost.

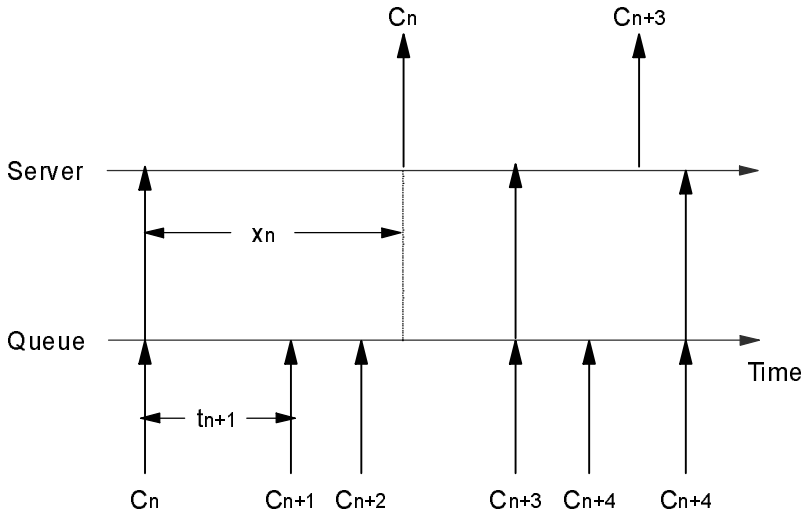

Figure 7.2: RDMA- model

Figure 7.2 illustrates the model for the RDMA- systems. C_n is the n-th frame to arrive to the system, t_{n+1} is the interval between the arrival of frames C_n and C_{n+1}, and x_n is the time to transmit frame n. The figure depicts three stages of the process: frames arriving to the queue, frames leaving the queue and immediately entering the server (that is, starting transmissions), and frames leaving the server. The arrows represent frames and the horizontal lines the queue and the server. If the link is not transmitting any information, the arriving frame is immediately serviced. Otherwise, it is lost. In the figure, only frames C_n, C_{n+3}, and C_{n+4} are not lost.

Following [Karlin and Taylor 75], let z_i denote the interval of time between the i-th and (i+1)-th successful transmissions, γ_t be the residual life of t_i at time t, $S_k = \sum_{i=1}^{k} t_i$ be

the time of the *k*-th arrival, and $N(t) = \sup\{n: S_n \le t\}$ be the number of arrivals to the system by time *t*. The residual life γ_t of t_i is the time remaining for the arrival of the next frame at time *t*. Then z_1 can be computed as:

$$z_1 = x_1 + \gamma_{x_1} = S_{N(x_1)+1} \tag{7.1}$$

That is, the interval between the first and second successful transmissions is equal to the time to transmit the first frame plus the residual time for the arrival of the next frame after the first finished being transmitted.

Example 7.1: Distribution of the inter-departure time when the service has a deterministic distribution.

If the service time is always $\dot{x}$, then $B(x) = \begin{cases} 1 \text{ if } x \ge \dot{x} \\ 0 \text{ if } x < \dot{x} \end{cases}$, where $\frac{1}{\mu} = \dot{x}$.

Then, if $z > \dot{x}$:

$$\Pr\{z_1 \le z\} = \int_0^{z-\dot{x}} \lambda e^{-\lambda x} dx = 1 - e^{-\lambda(z-\dot{x})}$$

and if $z \le \dot{x}$:

$$\Pr\{z_1 \le z\} = 0.$$

This example shows the inter-departure distribution when ATM [De Prycker 93, Le Boudec 92] cells are being transmitted. The time to transmit an ATM cell is $\dot{x}$.

Since the processes $\{t_n\}$ and $\{x_n\}$ are independent, the law of total probabilities gives:

$$\Pr\{z_1 \le z\} = \int_0^z \Pr\{x + \gamma_x \le z | x_1 = x\} dB(x) \tag{7.2}$$

Defining the distribution of γ_t as $\hat{A}_w(t) = \Pr\{\gamma_t > w\}$ (an explicit formula for it can be found in [Karlin and Taylor 75]):

$$\Pr\{z_1 \le z\} = \int_0^z \{1 - \hat{A}_{z-x}(x)\} dB(x) \tag{7.3}$$

Since the arrival process is Poisson, the memory-less property of the interarrival exponential distribution leads to the conclusion that γ_t follows an exponential distribution. Thus for RDMA-:

$$\Pr\{z_1 \le z\} = \int_0^z B(z-x)\lambda e^{-\lambda x} dx \tag{7.4}$$

and:

$$\Pr\{z_1 > z\} = 1 - \int_0^z B(z-x)\lambda e^{-\lambda x} dx \tag{7.5}$$

But:

$$\int_0^z \lambda e^{-\lambda x} dx = 1 - e^{-\lambda z} \Leftrightarrow \int_0^z \lambda e^{-\lambda x} dx + e^{-\lambda z} = 1 \tag{7.6}$$

Replacing the number 1 in Equation 7.5 by the number 1 in Equation 7.6:

$$\Pr\{z_1 > z\} = \int_0^z \lambda e^{-\lambda x} dx + e^{-\lambda z} - \int_0^z B(z-x)\lambda e^{-\lambda x} dx \tag{7.7}$$

$$\Pr\{z_1 > z\} = e^{-\lambda z} - \int_0^z \{1 - B(z-x)\}\lambda e^{-\lambda x} dx = e^{-\lambda z} - \int_0^z \overline{B}(z-x)\lambda e^{-\lambda x} dx \tag{7.8}$$

Finally, one can conclude that Equation 7.8 is valid for any z_n because $\{z_n\}$ are the inter-arrival times of a renewal process [Karlin and Taylor 75].

Examples 7.1 and 7.2 illustrate how Equation 7.4 can be used to compute the distribution of the inter-departure times of successfully transmitted frames.

Example 7.2: *Distribution of the inter-departure time when the service has a Poisson distribution.*

In this case, $B(x) = 1 - e^{-\mu x}$.

Then:

$$\Pr\{z_1 \le z\} = \int_0^z (1 - e^{-\mu(z-x)})\lambda e^{-\lambda x} dx = \int_0^z \lambda e^{-\lambda x} dx - \lambda e^{-\mu z} \int_0^z e^{(\mu-\lambda)x} dx$$

$$\Pr\{z_1 \le z\} = 1 - e^{-\lambda z} - (\lambda e^{-\mu z} \frac{e^{(\mu-\lambda)z} - 1}{\mu - \lambda}) = 1 - e^{-\lambda z} - \lambda \frac{e^{-\mu z} - e^{-\lambda z}}{\lambda - \mu}$$

$$\Pr\{z_1 \le z\} = 1 - \frac{\lambda e^{-\mu z} - \mu e^{-\lambda z}}{\lambda - \mu}$$

This example shows the inter-departure distribution when ATM cells are being transmitted in batches of random sizes.

The reminder of this section uses Equation 7.8 to derive the mean loss rate in RDMA-. One can observe that frames are being transmitted at rate $\frac{1}{E[z]}$ while they are being received at rate $\frac{1}{E[x]} = \lambda$. Thus, the packet loss rate L is:

$$L = 1 - \frac{1/E[z]}{1/E[x]} = 1 - \frac{1}{\lambda E[z]} \tag{7.9}$$

The expectation can be computed as:

$$E[z] = \int_0^\infty \Pr\{z_1 > z\} dz = \int_0^\infty e^{-\lambda z} dz + \int_0^\infty \int_0^z \overline{B}(z-x)\lambda e^{-\lambda x} dx dz \tag{7.10}$$

But:

$$\int_0^\infty e^{-\lambda z} dz = \frac{1}{\lambda} \tag{7.11}$$

and:

$$\begin{aligned}\int_0^\infty \int_0^z \overline{B}(z-x)\lambda e^{-\lambda x} dx dz &= \int_0^\infty \int_x^\infty \overline{B}(z-x)\lambda e^{-\lambda x} dz dx \\ &= \int_0^\infty \lambda e^{-\lambda x} \int_x^\infty \overline{B}(z-x) dz dx \\ &= \int_0^\infty \lambda e^{-\lambda x} \frac{1}{\mu} dx \\ &= \frac{1}{\mu}\end{aligned} \tag{7.12}$$

Replacing Equations 7.11 and 7.12 in 7.10, one obtains:

$$E[z] = \frac{1}{\lambda} + \frac{1}{\mu} = \frac{\lambda + \mu}{\lambda\mu} \tag{7.13}$$

Finally:

$$L = 1 - \frac{1}{\lambda(\frac{\lambda+\mu}{\lambda\mu})} = \frac{\lambda}{\lambda+\mu} = \frac{\rho}{\rho+1} \tag{7.14}$$

An interesting conclusion that can be derived from Equation 7.14 is that under RDMA- policy, when $\rho \to 1$, $L \to 0.5$. That is, the upper bound on the loss rate for Poisson arrivals and generic service time is 50%.

The delay in the node in RDMA- is equal to the service time of a frame. The mean delay D is thus:

$$D = \frac{1}{\mu} \tag{7.15}$$

7.2.2 The RDMA++ System

RDMA++ is a challenging system to analyze because of its time dependent behavior. From the point of view of an end node, the server is the routing tree. It is active during its band and then goes on vacation during other bands. A complete state description of such a

system must include the current server status and how long it will last. References [Federgruen and Green 86, Ott 87] study systems with deterministic time dependent vacations and can provide only limited insight on the system behavior.

This section presents an approximation for the average delay and loss in the RDMA++ system when the queueing discipline is M/G/1 and the system load is low to medium.

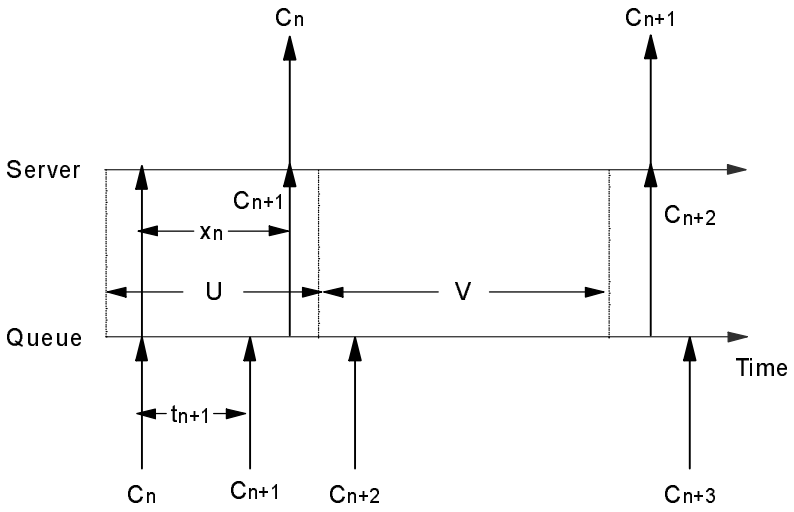

Figure 7.3: RDMA++ model

Figure 7.3 illustrates the model for the RDMA++ system. The terminology is the same as in Figure 7.2. Frame C_n arrives to an empty system and is serviced immediately. C_{n+1} waits until C_n is finished and then begins to be serviced. During its service time, the server goes on vacation and then returns to complete the service. Only then is C_{n+2} admitted into service.

Following [Bertsekas and Gallager 92], let w_i be the waiting time in the queue for the i-th frame, q_i be the number of frames found waiting in the queue by the i-th frame upon arrival, and r_i the residual service time seen by the i-th arriving frame. The residual service