A Queuing Model for Finite-Buffered Multistage Interconnection Networks*

Prasant Mohapatra and Chita R. Das
Department of Electrical and Computer Engineering
The Pennsylvania State University
University Park, PA 16802

Abstract

In this paper, we present a queueing model for performance analysis of finite-buffered multistage interconnection networks. The model captures network behavior in an asynchronous communication mode and is based on realistic assumptions. Throughput and delay are computed using the proposed model and the results are validated via simulation. Various design decisions using this model are drawn with respect to delay, throughput, and system power.

1 Introduction

Multistage interconnection networks (MINs) have been proposed as an efficient interconnection medium for multiprocessors. They have been used in various commercial and experimental systems. Behavior of the interconnection network plays an important role in the performance of multiprocessors. For an optimal design, it is necessary to analyze various configurations and constraints of the interconnection network.

Earlier research on MIN performance study have focussed on three types of network models: circuit switched, packet switched with infinite buffer, and packet switched with finite buffer. Study of circuit switched networks has gradually diminished since various packet switching techniques have become more prevalent. Infinite buffer analysis does not necessarily predict realistic behaviors of MINs under various workloads. Recent research effort therefore is directed towards analysis of finite-buffered MINs.

A model for finite buffered MINs should capture the following issues for predicting realistic performance.

- The processors in an MIMD mode operate independent of each other with occasional synchronization. Thus the network model should be based on *asynchronous* message transmission.
- The packets are normally of fixed size. Therefore, the time required for transferring a packet from one stage to the next stage is *deterministic*.
- Messages that can not be transmitted from one stage to the next due to the unavailability of buffer space should be *blocked* rather than rejected. Systems like Cedar use blocking of packets.

Prior work on finite-buffered MINs are mainly based on probabilistic models [1-5]. These analyses are valid when all the input/output operations happen at discrete stage cycles. These models do not capture asynchronous behavior especially when the service time of the switching elements (SEs) is more than one clock cycle. A queueing model for finite-buffered asynchronous MINs developed in [6] assumed non-blocking capability and exponential service time for switching elements.

None of the above models has considered all the design issues mentioned earlier. In this paper, we present a queueing model for performance analysis of MINs that considers asynchronous packet switching transmission, finite buffers, deterministic switch service time, and message blocking. The model has been validated via extensive simulation. Average message delay and throughput are used as performance measures to characterize a MIN. Variation of performance with input load and buffer length is discussed.

2 Model Assumptions

The model is based on the following assumptions.

(i) Each processor generates fixed-size messages independently at a rate λ and the intermessage times are exponentially distributed.

(ii) A memory request is uniformly distributed among all the MMs.

(iii) The SEs have deterministic service time (d cycles).

(iv) A packet is blocked at a stage if the destination buffer at the next stage is full. Packets arriving at the first stage of the MIN are discarded if the buffer is full.

*This research was supported in part by the National Science Foundation under grant MIP-9104485.

Fig. 2. A Queueing Model of an n-stage MIN

3.2 MIN Analysis

The notations used in Section 3.1 are also used for the MIN analysis with a few modifications as follows.

λ_i: packet arrival rate at stage i, $1 \leq i \leq n$.

$p_k^{(L)}(i)$: $p_k^{(L)}$ of stage i, $1 \leq i \leq n$.

ρ_i: traffic intensity at the server $= \lambda_i \cdot d$, $1 \leq i \leq n$.

x_i: blocking probability at stage i.

The basic model of a (4x4) MIN using (2x2) SEs is shown in Figure 1. The packet arrival and departure rates at each buffer are indicated in the figure. The departure rate is affected by the blocking probability as well as the service time distribution of the server. The uniform memory reference assumption makes all the servers of a particular stage statistically indistinguishable. It is therefore sufficient to analyze one buffer per stage of the MIN. A packet has to travel through a chain of n buffers in an n-stage MIN. A MIN is thus modelled as a chain of n queueing centers as shown in Figure 2.

Characterization of the interdeparture time distribution and hence the departure rate is necessary to analyze the MIN model. We therefore analyze the probability density function (*pdf*) of the interdeparture time of an $M/D/1/L$ queue. Let τ_i be a random variable which represents the time between departures from an $M/D/1/L$ queueing center of ith stage. Let ϕ be the event that the queue is empty after a departure. $f_{\tau_i}(t)$ represents the probability density function of τ_i and $f_{\tau_i|\phi}(t)$ denotes the probability density function of τ_i given that the queue is empty. $f_{\tau_i|\overline{\phi}}(t)$ denotes

the probability density of τ_i, given that the queue is non-empty. The state $p_0^{(L)}(i)$ denotes the probability that the queue is empty. Hence, the interdeparture probability density function is given by

$$f_{\tau_i}(t) = f_{\tau_i|\phi}(t)p_0^{(L)}(i) + f_{\tau_i|\overline{\phi}}(t)[1 - p_0^{(L)}]. \quad (4)$$

As the server has a deterministic service time of d cycles, there will be a departure every d cycles when the queue is not empty. Thus

$$f_{\tau_i|\overline{\phi}}(t) = \delta(t - d), \quad (5)$$

where $\delta(t)$ is an impulse function. When the queue is empty, the *pdf* is the density of the service time plus the arrival time. It can be derived as [7],

$$f_{\tau_i|\phi}(t) = \lambda_i e^{-\lambda_i(t-d)}U(t - d), \quad (6)$$

where, $U(t)$ is an unit step function.

Let $E[\tau_i]$ represents the expected value of the interdeparture time of packets from the queueing center. $E[\tau_i]$ can be obtained from equation (4) as

$$E[\tau_i] = \int_0^\infty t \cdot f_{\tau_i}(t)dt = d + \frac{p_0^{(L)}(i)}{\lambda_i}. \quad (7)$$

It is extremely difficult to accurately characterize the nature of interdeparture process. In order to keep the model tractable, we can approximate the interdeparture time distribution from one stage to the next as exponential with an average value of $\lambda_{i+1} = 1/E[\tau_i]$ requests/cycle. It will be shown in Section 4 that this assumption does not induces substantial difference between analytical and simulation results.

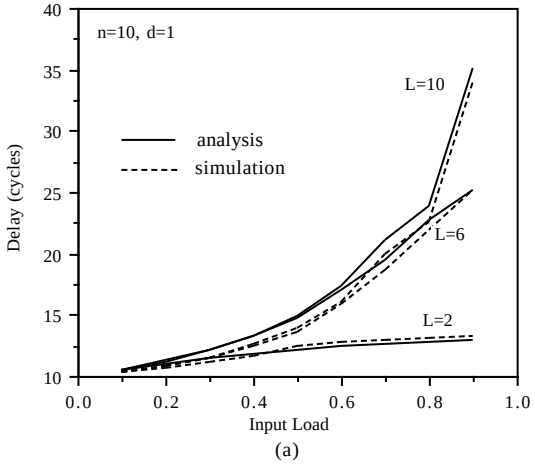
Based on our approximation, buffers at each stage of the MIN will have a Poisson arrival process and can thus be modelled as $M/D/1/L$ queueing centers. Using equation (7) and the blocking probability x_i, we get

$$\lambda_i = \begin{cases} \dfrac{\lambda_{i-1}(1 - x_{i-1})}{p_0^{(L)}(i) + \lambda_{i-1}(1 - x_{i-1})d}, & \text{for } 2 \le i \le n; \\ \lambda, & \text{for } i = 1. \end{cases} \quad (8)$$

The above expression is used to compute λ_i starting from $i = 1$ to n. The average time spent at each stage can be computed using equation (3). The average delay for a packet is obtained by summing up the delays of all the stages. The normalized throughput, X, is determined by the output of a buffer in the last stage of the MIN model, and is equal to λ_{n+1}.

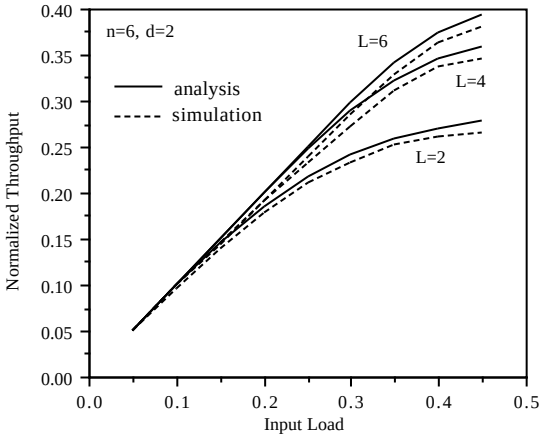
4 Performance Evaluation

A simulation model of a MIN was developed in which packets were generated randomly with an exponential distribution of interarrival time by each processor. A uniform random number generator was used to determine the destination memory. Throughput and delay were computed by counting the number of request completions and the average time taken to reach the output port, respectively. Comparisons between the analytical and simulation results for (64x64) and (1024x1024) systems using (2x2) SEs are shown in Figures 3 and 4. The difference between the analysis and the simulation results is within 7%. The curves indicate that the analytical results are fairly accurate.



(a)

Fig. 3. Delay of a (64x64) MIN



(b)

Fig. 4. Throughput of a (1024x1024) MIN

The effect of buffer length on delay of a 256-node MIN is depicted in Figure 5. It is mentioned in [1] that a small buffer length shows performance equivalent to

an infinite buffer. It can be inferred from Figure 5 that this is true only when the input load is less. The variation of delay is prominent until the buffer length is considerably high for heavy traffic. The model can be used to determine the minimum buffer length required to get a performance equivalent to the infinite buffer case.

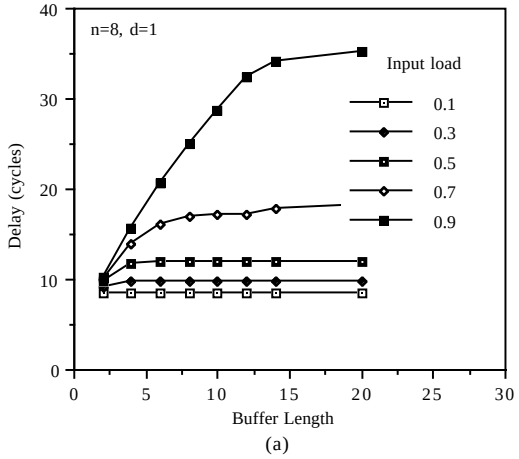


Fig. 5. Effect of Buffer Length on MIN Performance

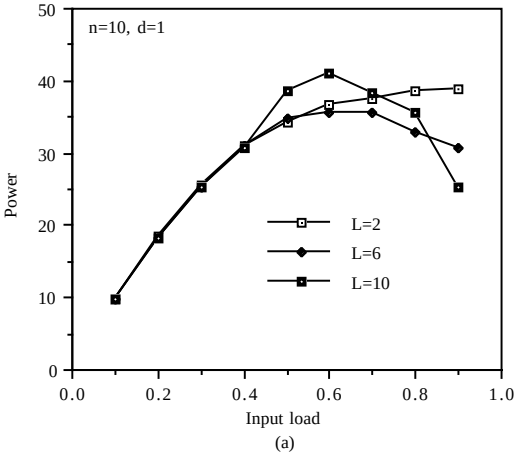


Fig. 6. Variation of System Power

Throughput and delay are not necessarily sufficient measures of system performance. It is observed that higher throughputs result in longer delay. A combined metric called *system power* is sometimes more meaningful. System power is defined as the ratio of throughput to delay. A higher power means either a higher throughput or lower delay. The variation of system power with respect to the input load is shown in Figure 6 for various buffer lengths. It is observed that system power increases with the input load for small buffers. For large buffer size, the throughput first increases with the input load until it saturates. On the other hand, delay increases monotonically. Thus, after a certain input load, the power reduces. The model can be used for predicting the optimum load to maximize the power of a MIN.

5 Concluding Remarks

A queueing model for evaluating performance of finite-buffered, asynchronous MINs is presented in this paper. The uniqueness of this model compared to previous finite-buffered analyses is that it captures asynchronous operations, deterministic service time of switches, and message blocking. Comparison with simulation results show that the analytical model is highly accurate. Various design alternatives based on performance requirements are discussed. It is difficult to come up with an optimal set of design parameters to satisfy all performance measures. The model can be used to compute suitable values of MIN parameters based on the priorities of performance metrics.

References

[1] H. Jiang, L. N. Bhuyan, and J. K. Muppala, "MVAMIN: Mean Value Analysis Algorithms for Multistage Interconnection Networks," Journal of Parallel and Distributed Computing, pp. 189-201, July 1991.

[2] D. M. Dias and J. R. Jump, "Analysis and Simulation of Buffered Delta Network," IEEE Trans. on Computers, pp. 273-282, Aug. 1981.

[3] D. L. Willick and D. L. Eager, "An Analytical Model of Multistage Interconnection Networks," ACM SIGMETRICS Conf. on Measurement and Modeling of Computer Systems, pp. 192-202, 1990.

[4] T. Lin and L. Kleinrock, "Performance Analysis of Finite-Buffered Multistage Interconnection Networks with a General Traffic Pattern," ACM SIGMETRICS Conf. on Measurement and Modeling of Computer Systems, pp. 68-78, May, 1991.

[5] H. Yoon, K. Y. Lee, and M. T. Liu, "Performance Analysis of Multibuffered Packet-switching networks in Multiprocessor Systems," IEEE Trans. on Comput., vol. C-39, no.3, pp. 319-327, Mar. 1990.

[6] T. N. Mudge and B. A. Makrucki, "An Approximate Queueing Model for Packet Switched Multistage Interconnection Networks," Int. Conf. on Distributed Computing Systems, pp. 556-562, Oct. 1982.

[7] P. Mohapatra, *Dependability and Performance Modelling of Parallel Computers,* Ph.D. Thesis, The Pennsylvania State University, Aug. 1993.