Cost-Performance Trade-Offs in Manhattan Street Network Versus 2-D Torus

Tein Y. Chung, Nita Sharma, and Dharma P. Agrawal

***Abstract*—Two recent topologies—Manhattan Street Network (MSN) and HR^4-Net (or 2-D Torus)—proposed for metropolitan areas are primarily mesh connected networks. In this brief contribution, we compare the average distance, the diameter, the terminal reliability, and the cost. Simulation results indicate that they are fairly close in performance, while the MSN is observed to be more cost-effective than the HR^4-Net.**

***Index Terms*— Cost-performance ratio, HR^4-Net, Manhattan Street Network, reliability routing distribution.**

I. Introduction

Manhattan Street Network (MSN) [9] is a directed mesh connected network (Fig. 1), with its links resembling the one-way streets and avenues of Manhattan. The HR^4-Net [2] (or a 2-dimensional Torus proposed as a metropolitan area network) is similar to the MSN except that all its links are undirected (bidirectional). Both these networks are considered useful for commercial applications, such as an integrated network environment for LAN's at different sites of a city.

In a typical regularly connected network, the degree of a node determines the cost and capacity of the network, while the diameter and average distance are generally considered as the static properties of a network, as they are measured in a uniform traffic environment in which all links are assumed identical. The diameter is defined as the maximum value of the shortest distance between two nodes of the network, and the average distance is the average shortest path length from a node to any other nodes of the network. For a square grid MSN or HR^4-Net, the diameter and average distance is shown [1], [3] to be $O(\sqrt{N})$. These static properties of the MSN's are seen to be very close to that of HR^4-Net (or two-dimensional (2-D) Torus), although the degree of an MSN is just half of a 2-D Torus. This encourages us to compare the performance of the MSN to the 2-D Torus on a broader basis.

This brief contribution is organized as follows. Section II introduces some topological information of the MSN and 2-D Torus. In Section III, the routing schemes for the MSN are introduced. In Section IV, parameters for comparison are defined. Section V discusses the simulation result and the average terminal reliability. Finally, Section VI concludes the brief contribution.

II. Topological Description

In an MSN, the network size N is expressed as $m_1 * m_2$ and the address of each node is represented by a two-tuples (a_1, a_2), where $0 \le a_1 \le (m_1 - 1)$ and $0 \le a_2 \le (m_2 - 1)$. Given a node with address (a_1, a_2), it is connected to nodes (b_1, a_2) and (a_1, b_2), where $(b_1, a_2) = \{[(a_1 + / - 1) \bmod m_1], a_2\}, +/-$ when a_2 is even/odd, and $(a_1, b_2) - \{a_1, [(a_2 + / - 1) \bmod m_2]\}, +/-$ when a_1 is even/odd. Fig. 1 shows an example of the address assignment of a 4 * 4 MSN.

Manuscript received January 27, 1992; revised September 24, 1992. This work was supported by the U.S. Army under Research Contract DAAG-29-85-k-0236, by the NSF under Contract MIP-8912767, and by the SRC Fellowship Fund.

T. Y. Chung is with the Yuan-Ze Institute of Technology, CES Department, Tao Yuan, Taiwan, ROC.

N. Sharma is with nCUBE, 919 E. hillsdale Blvd., Foster City, CA 94404.

D. P. Agrawal is with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695-7911.

IEEE Log Number 9209046.

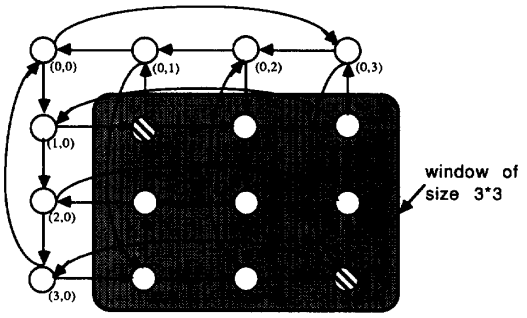

Fig. 1. Manhattan Street Network (MSN) of size $4 * 4$.

TABLE I
Performance Patterns for the MSN and 2-D Torus of Size $N = \sqrt{N} * \sqrt{N}$ and $\sqrt{N/2}$ = Even

2-D Torus Diameter = √N Average distance= $N^{3/2}/(2N-2)$		MSN Diameter = √N+1 Average distance = $(N^{3/2}/2 + N-4)/(N-1)$	
h=# of hops away from source	n= no. of nodes at distance h	h= no. of hops away from source	n= no. of nodes at distance h
$1 \le h < \sqrt{N}/2$	4*h	1	2
		$2 \le h < \sqrt{N}/2$ and h≠4	4*(h–1)
		4	11
$h = \sqrt{N}/2$	4*h – 2	$h = \sqrt{N}/2+1$	2√N–4
$\sqrt{N}/2 < h \le \sqrt{N}$	4(√N–h)	$\sqrt{N}/2 +2 < h \le \sqrt{N}$	4(√N–h+1)
$h = \sqrt{N}$	1	$h = \sqrt{N}+1$	2

In this brief contribution, we only consider a regular $\sqrt{N} * \sqrt{N}$ square grid MSN and Torus, with $\sqrt{N}$ = even, for the sake of simplicity. A two-dimensional Torus with $\sqrt{N} * \sqrt{N}$ nodes has diameter of $\sqrt{N}$ and average path length of $N^{3/2}/(2N - 2)$; while for MSN, the corresponding values are marginally larger as $\sqrt{N} + 1$ and $(N^{3/2} + 2N - 8)/(2N - 2)$ [3], respectively. The number of nodes at a given distance from a fixed source for MSN and multiple shortest paths between any two nodes in 2-D Torus are listed in Table I. For $N = 256$, if a uniform message routing distribution and shortest path routing are assumed, the relative frequency of a message transmission length for the Torus and MSN is shown in Fig. 2.

III. Routing Scheme

To compare the MSN to the 2-D Torus, we need to take the simplicity of their routing schemes into consideration, too. It is obvious that the routing scheme for a 2-D Torus is simple and easy to implement. Therefore, in this section, we focus on the routing scheme for the MSN. Based on the cyclic structure and the relative addressing space concept, several routing schemes have been proposed for the MSN [3], [4], [9], including the adaptive routing schemes, the fixed routing scheme, and the broadcasting routing scheme. A simple routing scheme proposed by Maxemchuk [9] first maps the destination address to (0, 0) and computes the relative address of the current node (a_2, a_1), which falls in one of the four quadrants. Then, the message is transmitted to the outgoing link that matches the preferred routing direction. If neither of the two outgoing links of the node matches the preferred routing directions, a message is transmitted to a randomly

0018–9340/94$04.00 © 1994 IEEE

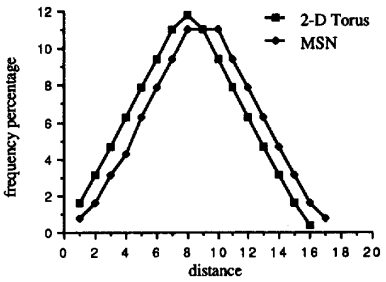

Fig. 2. Relative frequency for the message transmission length for the MSN and the 2-D Torus of size 16 * 16.

selected outgoing link.

We can see that the routing scheme for an MSN is simple and can be easily implemented by a VLSI chip. Although routing for MSN is relatively more complicated that the 2-D Torus, it is still fairly efficient and will not become the network performance bottleneck. On the other hand, unidirectional networks are becoming increasingly popular [2]. The use of paths in only one direction has also been recommended to avoid deadlocks in the Torus networks [5].

IV. Cost–Performance Ratio Comparison

A. Routing Distribution

A routing distribution basically specifies the probability that different network nodes exchange messages, thereby reflecting the application dependent feature of the network [6]. Since the networks behave differently for various message routing, we investigate three types of routing distribution.

1) Uniform Message Routing: A message routing distribution is said to be uniform if the probability of node i sending a message to node j is the same for all i and $j, i \neq j$ and $i, j \in V(G)$. Here, we exclude the case of nodes sending messages to themselves, s this is never done in actual practice, and we are interested in message transfers throughout the network. It has been observed [10] that most data transfer occurs in an area of interest; the routing distribution should exhibit some measure of communication locality. Thus, the uniform routing distribution could be said to lead to an upper bound of the mean internode message distance.

2) Sphere of Locality: Suppose the uniform message routing assumption were relaxed. We would expect a node to exchange messages more frequently with the nodes in the area of interest, or in close physical proximity. One abstract way of representing this idea is to make each node as a center of locality sphere with radius L, expressed in terms of the number of hops. A node sends message to the other nodes, with some (usually high) probability φ inside its sphere of locality, and with probability $(1-\varphi)$ to nodes outside the sphere. This model reflects the communication locality typical in the area of interest. In the MSN and 2-D Torus, it is appropriate to define this sphere of locality as a window of size $2 * L + 1$, with the source node as the center of the window. Thus, a window of size 3 consists of a square grid of 9 nodes, as shown by an example in Fig. 1 by the shaded area.

3) Decreasing Probability Message Routing: The definition of sphere of locality is useful, if the area of interest is small as compared to the size of network, and the probability of visiting the locality is relatively high. There are, however, many cases wherein the region of interest cannot be clearly defined in the form of a sphere. An alternate intuitive notion of locality, that sounds appealing, is the one with the probability of sending a message to a node to be an inverse function of the distance of the destination node from the source node. The distribution function $F(d) = \text{Norm}(D) * \theta^d$ is considered appropriate, where D is the network diameter and d is the distance between source and destination. Here, Norm(D) is a normalizing constant for the probability F, chosen such that the sum of all probabilities is one. θ is called the decay coefficient and has value ranging from 0 to 1, and the message transmission length is proportional to the value of θ. When θ is small, the probability of exchanging messages between nodes decreases dramatically as their distance increases. In other words, the message transmission length decreases. On the other hand, as θ approaches 1, routing distribution approaches the uniform distribution.

B. Cost

We introduce a cost function model that incorporates the economy of the available scale, reflecting the degree per node and the link bandwidth. The use of economy of scale (α) indicates that up to a certain point, a link with twice the bandwidth does not cost twice as much. Economy of scale accounts for the cost of high-capacity channels and is modeled to reflect the dependency between the port cost and the link bandwidth. The network cost function could be defined as [7]

$$S = C_p * N + k * N * C_c * B^{\alpha} \tag{1}$$

where k is the degree of a node, B is the bandwidth of each communication link in Mb/s, C_p and C_c are the processor and channel cost, respectively, N is the number of processors in the network, and α is an economical scaling constant, $0 \leq \alpha \leq 1$, reflecting a dependence of cost on the bandwidth.

Because fiber optics can be used either as a bidirectional or unidirectional transmission medium, the total link cost is the same for the MSN and the Torus. However, for each port, a transmitter and a receiver are needed, while their costs are proportional to the link capacity. This cost function is used in a later section to understand the tradeoffs between the network performance and the cost.

C. Average Packet End-to-End Delay

Assume that the input traffic rate from node i to node j is equal to γ_{ij}. Then, the input traffic load to the network becomes

$$\gamma = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \gamma_{ij}. \tag{2}$$

Suppose that the traffic, $\gamma - ij$, encounters a number of hops along the path from the source node i to the destination node j, and such a path length is equal to d_{ij}. As γ_{ij}/γ of the traffic traverses a path of length d_{ij}, the mean internode packet transmission distance could be given by

$$d_{\text{avg}} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{\gamma_{ij}}{\gamma} * d_{ij}. \tag{3}$$

Once d_{avg} is known, the network load is obtained as $\gamma * d_{\text{avg}}$. If we assume that the traffic load in the network is evenly distributed, then the link traffic rate for the network is

$$\lambda_i = \frac{\gamma * d_{\text{avg}}}{N_k} \tag{4}$$

where, N_k is the total number of links of the network. For the MSN, $N_k = 2 * N$, while for the Torus, $N_k = 4 * N$. The network load balancing assumption is reasonable because both the target networks are symmetrical, and each node is assumed to encounter the same type of routing distribution.

To simplify the computation of end-to-end delay for a packet, we assume that the packet input process is a Poisson process, and the packet size is exponentially distributed with mean $1/\mu$. Then, the mean packet end-to-end delay is obtained as

$$T = d_{\text{avg}}/(\mu B - \lambda_i) \tag{5}$$

where B is the link capacity/bandwidth. Under the balanced traffic assumption, the network is saturated when any link becomes saturated,

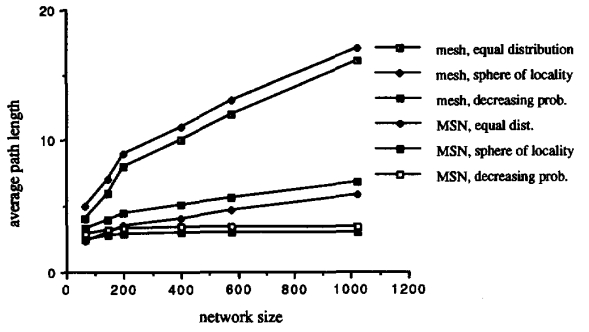

Fig. 3. Average distance for different size MSN and 2-D Torus (mesh).

i.e., $\lambda_i = \mu \cdot B$. By substituting $\lambda_i = \mu \cdot B$ in (4), it can be obtained

$$\mu \cdot B = N * \gamma_i * d_{\text{avg}}/N * k \tag{6}$$

where $N * \gamma_i = \gamma$, and γ_i is the input traffic rate to node i. Therefore, the network is saturated when

$$\gamma_i = (k * \mu * B)/d_{\text{avg}} \tag{7}$$

where k is the node degree. Therefore, the network capacity is proportional to the node degree, while it is inversely proportional to the network average path length.

D. Terminal Reliability

Terminal reliability, a commonly used measure of connectivity, is the probability that a processor, called the "source node," can successfully communicate with another processor, called the "destination node," in the network. To compute the terminal reliability in a given network, the system is modeled as a graph whose nodes represent the processing elements and whose edges represent the links of the network. The following assumptions are made:

1) the graph does not have any self-loops,
2) failure of a link is independent of other link failures,
3) a link has only two states—operational or faulty, and
4) the processors are assumed to be fault-free.

The terminal reliability is computed by finding all possible paths from the source to the destination node. The reliability expression is obtained by making the reliability terms corresponding to each of the paths disjoint with respect to each other [8]. The average terminal reliability is obtained by fixing a source node and averaging the terminal reliabilities obtained for all possible destination nodes in the network. The average terminal reliability is another parameter used to compare the MSN and 2-D Torus network from an availability perspective.

V. Performance Comparison

The average distance for the MSN and 2-D Torus of different sizes is given in Fig. 3. Fig. 4 compares the average distance of these two networks as a function of the locality probability φ and the window size. The networks could also be compared based on the routing decay coefficient and is shown in Fig. 5. The two networks perform almost the same when the decay coefficient is small and have a constant gap in average distance when the decay coefficient is close to 1.

If we assume the implementation cost for both the MSN and the 2–D Torus to be the same, then C_p and C_c will be the same. It can been seen that for a given network cost, S, the relationship between the link bandwidths of the 2-D Torus and the MSN is as follows:

$$B_{\text{MSN}} = B_{2-D\,\text{Torus}} * (2)^{1/\alpha} \tag{8}$$

Assuming that $1/\mu = 1$ kb/s and $B_{2-D\,\text{Torus}} = 1$ Mb/s, $C_p = 500$ units, and $C_c = 100$ units. Fig. 6 shows the cost-performance comparison between the 2-D Torus and MSN under different traffic conditions. At low α (less expensive link bandwidth cost), it is obvious that the MSN can carry more traffic load than the 2-D Torus.

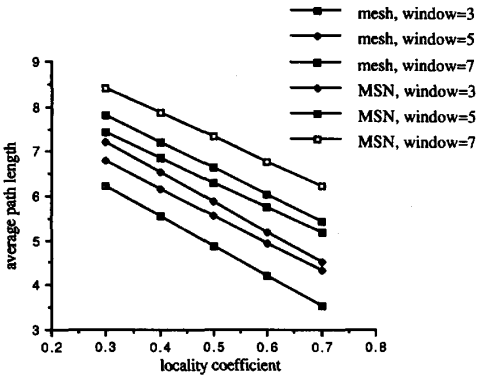

Fig. 4. Average distance as a function of locality probability in the MSN and 2-D Torus (mesh).

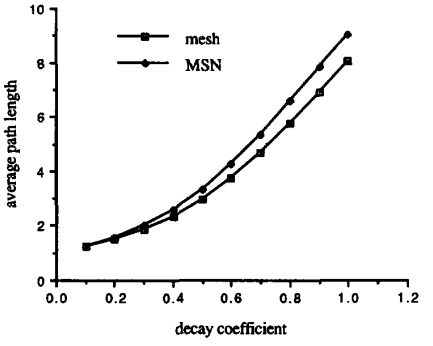

Fig. 5. Average distance as a function of decay coefficient for the MSN and 2-D Torus (or mesh).

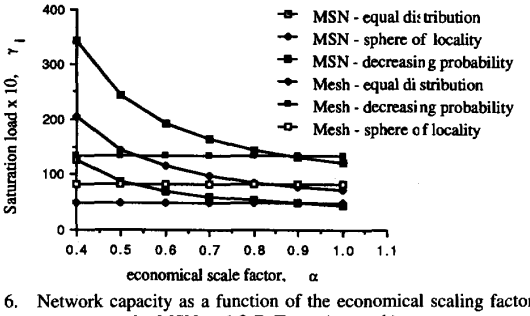

Fig. 6. Network capacity as a function of the economical scaling factor for the MSN and 2-D Torus (or mesh).

Even at higher α values, the 2-D Torus can carry only slightly higher traffic than the MSN. Table II shows the average terminal reliabilities for the MSN and the 2-D Torus networks, with an assumption that each link probability is 0.9. The difference in the two figures is relatively small to indicate that the MSN is comparable to the 2-D Torus from a reliability point of view.

VI. Conclusion

This brief contribution compares the cost-performance between the MSN and 2-D Torus based on 1) the routing distribution, 2) the cost function, 3) the network capacity, and 4) the average terminal reliability. A cost function has been defined to estimate the network implementation cost as a function of the number of transmitters and receivers and the link capacity. From the index employed in evaluating the cost-performance, it is seen that the MSN performs fairly close to the 2-D Torus, even though it has only half of the number of transmitters and receivers. As the increase in bandwidth

TABLE II
AVERAGE TERMINAL RELIABILITY
(PROBABILITY OF WORKING LINK = 0.9)

Network	Number of Nodes			
	8 (=2*4)	16 (=4*4)	24 (=4*6)	36 (=6*6)
MSN	.9509	.9165	.9160	.9099
2-D Torus	.9918	.9627	.9557	.9201

is not too large, the MSN actually has a higher network capacity cost ratio than the 2-D Torus, while maintaining almost the same value of the reliability.

REFERENCES

[1] F. Borgonovo and E. Cadorin, "HR^4-Net: A hierarchial random-routing reliable and reconfigurable network for metropolitan area," in *Proc. INFOCOM*, 1987, pp. 320–326.

[2] C. Chou and D. H. C. Du, "Hierarchial uni-directional hypercubes," in *Proc. 1991 Int. Conf. Parallel Processing*, vol. I, pp. 530–533.

[3] T. Y. Chung and D. P. Agrawal, "On network characterization and optimal broadcasting for the MSN," in *Proc. IEEE INFOCOM 1990*, San Francisco, CA, June 5–7, 1990, pp. 465–472.

[4] T. Y. Chung, S. Rai, and D. P. Agrawal, "A routing scheme for datagram and virtue circuit services in MSN," presented at the Int. Phoenix Conf. Comput. and Commun., Phoenix, AZ, Mar. 1989, pp. 214–218.

[5] W. J. Dally and C. L. Seitz, "Deadlock-free message routing in multiprocessor interconnection networks," *IEEE Trans. Comput.*, vol. C-36, pp. 547–553, May 1987.

[6] S. P. Dandamudi and D. L. Eager, "Hierarchial interconnection networks for multicomputer systems," *IEEE Trans. Comput.*, vol. C-39, pp. 786–797, June 1990.

[7] P. W. Dowd and K. Jabbour, "A unified approach to local area network interconnections," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 1418–1424, Dec. 1987.

[8] N. Kini, A. Kumar, and D. P. Agrawal, "Quantative reliability analysis of redundant multistage interconnection networks," in Amer. Math. Soc./ACM DIMACS series, *Reliability of Computer and Communication Networks*, vol. 5, 1991, pp. 153–170.

[9] N. F. Maxemchuk, "Routing in the Manhattan Street Network," *IEEE Trans. Commun.*, vol. COM-35, pp. 503–512, May 1987.

[10] D. A. Reed and D. C. Grunwald, "The performance of multicomputer interconnection networks," *IEEE Comput. Mag.*, pp. 63–73, June 1987.

Finite Buffer Analysis of Multistage Interconnection Networks

Jianxun Ding and Laxmi N. Bhuyan

***Abstract*—We propose an analysis technique for a class of Multistage Interconnection Networks (MIN's) that have finite buffers at their switch inputs and operate in a synchronous packet-switched mode. We examine the issue of clock period in design and analysis of synchronous MIN's and propose a model based on small clock periods. Then we analyze our "small cycle" design and compare the results with those obtained from the standard "big cycle" model that is currently used. The significant performance improvement of our model is shown based on various clock width, data width, and buffer length.**

***Index Terms*— Multistage interconnection networks (MIN's), finite buffers, packet-switching performance.**

Manuscript received April 21, 1992; revised December 11, 1992. This work was supported in part by the NSF under grant MIP-9002353.

The authors are with the Department of Computer Science, Texas A & M University, College Station, TX 77843-3112.

IEEE Log Number 9212767.

I. INTRODUCTION

The Multistage Interconnection Networks (MIN's) are extremely suitable for building large scale shared memory multiprocessors and high performance broad-band communication switching networks [1]–[5]. A MIN can be operated synchronously or asynchronously; it also can be circuit-switched or packet-switched [4]. In packet switching, the packets pass through the network stages in a pipelined way, resulting in high throughput. In case of a conflict, a blocked packet is stored in an intermediate switch without occupying the whole path. In this brief contribution, we present analytical techniques to evaluate the performance of a synchronous packet-switched MIN with finite buffers at each switch.

There are some analytical models for infinite-buffered MIN's [6], [7]. Dias and Jump analyzed single-buffered MIN's by using Timed Petri Net [8]. Jenq used a probabilistic model and by iteratively evaluating a set of probabilistic formulas, he was able to get the latency and throughput for single-buffered MIN's [9]. Yoon, Lee, and Liu extended Jenq's model to cases where buffer length at a MIN switch is more than one [10]. Kim and Garcia extended Jenq's analysis to nonuniform traffic patterns [11]. Theimer, Rathgeb, and Huber introduced "blocked" state in Jenq's model to get more accurate results [12]. Thus, Jenq's model [9] has become somewhat of a standard for the analysis of synchronous packet-switched MIN's. However, it is assumed in Jenq's model that the basic synchronous clock periods are big enough to let the control signals pass from the last stage to the first stage. This assumption points to the major drawback of Jenq's model [9]. Rather, in practical designs, the basic clock periods should be small based on the control signals passing one stage to ensure the flow of packets in a pipelined way.

In this brief contribution, we present an analysis that considers designs based on small clock cycles and show that the throughput of the MIN is increased and the delay is reduced. We also study the MIN performance with various clock width, data width, and buffer length, etc. The brief contribution is organized as follows. In Section II, we examine the design problems of the Jenq's model and propose a "small cycle" clock design. In Section III-A, a probabilistic MIN analysis considering the "small cycle" is presented. In Section III-B, we provide the simulation technique that is used to verify the analytical model. In Section IV, our model is compared with Jenq's model based on switch parameters such as data width, switch size, and buffer length. Finally, Section V concludes the brief contribution.

II. CLOCK DESIGN

In this section, we first try to explain the importance of the clock period in a synchronous network analysis. An 8×8 MIN with multiple buffers at the input of a switch is illustrated in Fig. 1. The conflict situation, shown in the figure, will be explained later. The analysis of such synchronous packet-switched and finite-buffered MIN's have been reported in [9]–[12]. However, the clock synchronization mechanism used in these models is not practical to MIN implementations. They all assume that each cycle τ has two phases, $\tau = \tau_1 + \tau_2$, as shown in Fig. 2(a). In the first phase τ_1 that consists of a number of small phases, the control signals are passed from the last stage of the MIN toward the first stage so that every packet knows whether it can go to the next stage's input buffer in phase τ_2. In the second phase τ_2, the selected packets may move forward by one stage. During τ_2, a buffer may be emptied to the output and filled with a new packet from the input simultaneously. We call this model as Jenq's model [9].

In a MIN, a packet can move forward only if it is selected among the competing packets by the routing logic of the switch. In Jenq's model, a packet can move when *either* the buffer of the switch to which it is destined in the next stage is not full *or* a packet in the next stage buffer will move forward, creating a space in the buffer. The control mechanisms for this kind of synchronization

0018–9340/94$04.00 © 1994 IEEE