Journal of the Korean Institute
of Industrial Engineers
Vol.22, No.2, June, 1996

Performance Analysis of Fast Packet Switch*

Kang Won Lee**

Abstract

The overall performance of BISDN depends significantly on the interconnection network or the switch fabric. Hence, it is extremely important to evaluate the performance of the network or the fabric. The well developed performance models also provide insight into the factors that determine design tradeoffs as well as quantitative estimates of their importance. The objective of this paper is to investigate and critically compare all the developed performance analysis models of FPS according to internal switch fabric structure, traffic assumptions, performance measures, methodologies, etc. FPSs are described according to their internal fabric structure. Brief history of FPS performance analysis is mentioned and performance analysis modeling is discussed.

1. Introduction

Asynchronous transfer mode (ATM) technology is one of the most promising solutions for broadband integrated services digital network (BISDN). To meet the demand for an expanding diversity of services in the backbone network, efficient switching techniques are becoming increasingly desirable. Advances in the field of VLSI technology, while reducing the cost of circuit fabrication, have brought completely new principles in the design and architecture of high performance switching fabrics which can accomodate a wide range of bandwidths. Many switch architectures have been proposed in the literature[8,18, 32].

Packet switching evolved primarily as an efficient way to carry data communication traffic. The switching capacity of current conventional packet switches ranges from 1 to 4 thousand packets per second with average delays of 20-50ms[8]. The capabilities of these packet switches are very attractive for applications requiring low throughput and low delay

* This paper was supported (in part) by 1994 Non Directed Research Fund, Korea Research Foundation

** Department of Industrial Engineering, Seoul National Polytechnic University

such as inquiry/response type traffic and those requiring high throughput but which can tolerate high delays such as file transfer. However, these capabilities are not sufficient for real-time types of traffic with following diverse traffic characteristics.

- Both narrowband and broadband rate
- Both stream traffic (such as voice and video) and bursty traffic (such as interactive data)
- Unforeseen demands
- Delay and/or loss sensitive quality requirements

The rapid pace of technological change has brought about new switching concepts in order to satisfy the high performance requirement for future systems. Among the desirable goals of the design are switches that have low delays, distributed routing, high throughput and low hardware complexity.

Many different switch fabric designs have been proposed and developed at various research orgnizations around the world over the past few years. All current approaches of high performance switch fabrics employ a high degree of parallelism and distributed control, and the routing function is performed at the hardware level. The focus in this paper is on those switch fabrics which incorporate the fast packet switching (FPS) concepts as their underlying switching technique to support a wide range of services with different bit rates.

The overall performance of BISDN depends significantly on the interconnection network or the switch fabric. Hence, it is extremely important to evaluate the performance of the network orthe fabric. The well developed performance models also provide insight into the factors that determine design tradeoffs as well as quantitative estimates of their importance.

The objective of this paper is to investigate and critically compare all the developed performance analysis models of FPS according to internal switch fabric structure, traffic assumptions, performance measures, methodologies, etc. In section 2, FPSs are described according to their internal fabric structure. Before discussing performance analysis modeling in section 4, brief history of FPS performance analysis is mentioned in section 3.

2. Fast Packet Switch

Packets are assumed to be of a constant length; these packets are called "cells"in an ATM environment. ATM specifies fixed-length packets comprising 5 octets for the header and 48 octets for the information payload. If we assume that all input links to the network are slotted and synchronized with bit rates equal to 150 Mb/s, the resulting packet slot times is approximately 2.8μsec. Thus, the switch fabric has to be designed in such a way that it can handle approximately 350,000 packets per second per input port. Several architectural designs for fast packet switch have emerged in recent years. These can be classified into the

Table 1. Description and Brief Summary of Fast Packet Switch

Type of Switch	Description	Blocking	Design Issue	Application Switch
Banyan Type Network	· It belongs to a class of MIN with the property that there is exactly one path from any input to any output · Banyan networks are subdivided into serveral classes such as Omega, Delta, etc. Although these networks have different interconnection patterns and permutation capabilities, they have the same performance in a packet switching environment.	· internal link blocking · output port blocking	· fault tolerant MIN pattern · multiple parallel paths · tandem Banyan switches	
Buffered Banyan Network	· To reduce the blocking in the switching network and increase the throughput of Banyan type switches, buffers are placed in every switch node. · Depending on the location of buffer, three types have been suggested -input -output -shared	· no blocking in the switch network except input buffer controller(IBC) -input buffer : HOL blocking -output buffer : hardware complexity and required speed up for output queue -shared buffer : compromise	· location of buffer · buffer size · in case of output buffering, required speed up for the output queue	· wideband packet technology network [9,10] · broadcast packet switching network [11]
Sort Banyan Network with Input and/or Output Queue	· If packets are first sorted based on their destination addresses and then routed through the Banyan network, the internal blocking problem can be avoided. The first segment consists of a sorting network, followed by a shuffle exchange or Banyan network which routes the packet. · Since this fabric still has output port blocking, queues can be placed at the input and/or output port to prevent output port contention.	· no internal link blocking · output port blocking -input queue : HOL blocking - output queue : H/W complexity and output queue speed up -input/output queue : some compromise	· speedup factor for output queue	· Starlite switch [12]

Type of Switch	Description	Blocking	Design Issue	Application Switch
Switch Fabric with Disjoint Path Topology	· The switch fabric is based on a fully interconnected topology in a sense that every input has a nonoverlapping direct path to every output so that no blocking or contention may occur internally · $N \times N^2$ crossbar matrix	· output port blocking -output (shared) buffer · packet loss due to knock-out concentration	· ratio of knockout concentration · output buffer size · no speed up required for output queue · It can be used as a building block for larger MIN	· Knockout switch[13]
Crossbar Based Switch Fabrics	· They have always been attractive to switch designers since they are internally nonblocking and simple. · They have the property of square growth and are not economical for large switches · To overcome the output port contention problem, a queueing function has to be added to the pure crosspoint matrix -input queue -output queue -queue with crosspoint	· no internal blocking · output port blocking	· location of queue · input queue -control mechanism for output port contention (backpressure) -HOL blocking · output queue -buffer size -speedup problem · crosspoint queue -self routing concept · It is either designed for relatively small application or just for a building block which is used in larger switch	· crossbar switch with input queue[15] · bus matrix switch[14]
Shared Memory Fast Packet Switches	· They are very similar in design concept to today's conventional packet switch. · The switching memory is shared by all input and output links	· call blocking in case of full memory buffer	· the switch memory bandwidth access time, and size · maximum acceptable packet loss probability	· Prelude switch[16]
Shared Medium Fast Packet Switch	· They employ a bus or ring topology as the switching medium. · They provide flexibility in terms of the access protocol and distribution of traffic · Their bandwidth capacity and throughput are limited compared to the multipath switch architecture.	· output port contention blocking	· how to implement the high speed bus or ring · allocation of the medium	· PARIS Switch[17]

following three categories[32].

- Space division type
- Shared memory type
- Shared medium type

Space division type can be further subdivided into

- Banyan type
- Buffered Banyan based fabrics
- Batcher(Sorted)-Banyan based fabrics
- Switch fabrics with disjoint path topology
- Crossbar based fabrics

Table 1 gives brief summary and description on the fast packet switches according to their internal fabric structure discussed above.

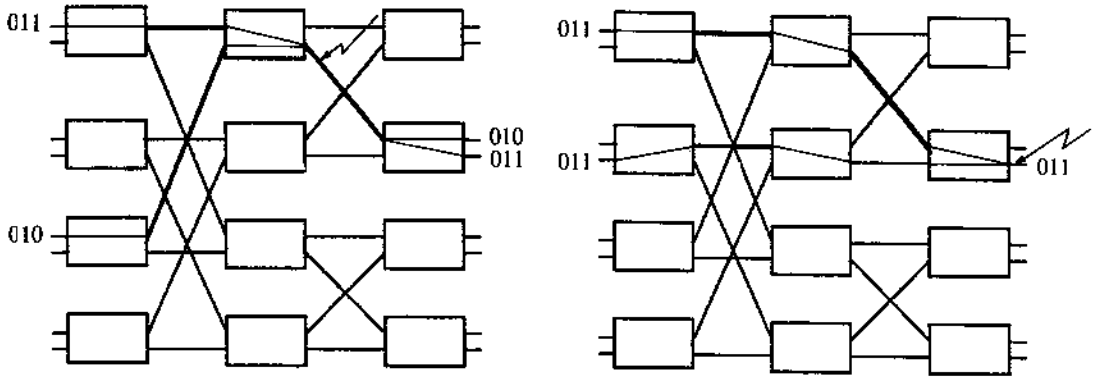
3. Brief History of Fast Packet Switching Performance Analysis

Banyan type networks have shown to be a very promising candidate as the switching fabrics for the broadband ISDN. The reasons lie in their suitability to VLSI implementation and to their self-routing capability[21,22]. For the brief history of fast packet switching performance analysis, the Banyan type networks including buffered and Batcher-Banyan are mainly focused. Switch fabrics with disjoint path topology such as knockout switch are also considered.

Pure Banyan network with log N stages have two important properties: a unique path exists from any input to any output, and distinct input/output paths may have common links. This leads to serious disadvantages. An input/output connection may be blocked by a previously established connection (even if outputs involved are distinct), thereby causing poor performance in a random access environment. That is, the network can experience two forms of blocking: internal blocking and output port blocking. The internal link blocking refers to the case where packets are lost due to contention for a particular link inside the network. The output port blocking, however, occurs where two or more packets are contending for the same output port. Figure 1 shows their effects in an example of an 8*8 delta-2 network. These two effects result in the reduction of the maximum throughput of the switch. Therefore, several ways have been proposed to reduce the blocking or to increase the throughput of Banyan type switches

- increasing the internal link speeds relative to the external speeds,
- placing buffers in every switching node,
- using a handshaking mechanism between stages or a back-pressure mechanism to delay the transfer of blocked packets,
- using multiple networks in parallel to provide multiple paths from any input to any output or multiple links for each switch connection,
- using a distribution network in front of the Banyan network to distribute the load evenly.

Naturally previous performance analysis works are mainly focused on the above issues and have tried to develop models with realistic



a) Internal Blocking b) Output Contention

Figure 1. Blocking Phenomena in 8×8 Delta Network

traffic assumptions. Their concern has been directed to two goals:

1) To propose a switch fabric with higher throughput (and less blocking)

· Buffering problem

—buffer location (input, output, and shared)

In case of input buffering, HOL blocking is a main concern. Bypass buffering algorithm (serial or parallel) is developed. In case of output buffering, speed up for output queue is a main concern

—buffer size (single or multiple buffer)

· Parallel (or tandem) network or multiple link

—place additional Banyan network in parallel or tandem

—multiple links for each switch connection

· Internal protocol

Generally, for flow control in the switch network, following two protocols are suggested. (In the following chapter it will be further classified)

—backpressure (delay is a main concern)

—block and lost (blocking probability is a main concern)

2) To develop a more realistic model and to reduce computational complexity

· Approach

—simple probabilistic calculation

—Markov chain

—fluid flow model

· Offered traffic model

—uniform

—general traffic pattern (bursty and/or non-uniform)

· Some general assumptions

—independent assumption between input links of a stage, or addresses of packets stored within a switch buffer.

—blocked call (memory, or random)

Followings are the list of some important works.

- Performance analysis of Delta network [20]
- Performance analysis of a packet switch based on a single input buffered Banyan network [23, 37]
- Performance analysis of buffered Banyan network under uniform traffic

—multiple input buffer size [34]
—output buffer with arbitrary size [19, 36]
—shared buffer with arbitrary size [5]
—finite input buffer with memorized blocked call [4]

- Performance analysis of a buffered Banyan network under non-uniform traffic

—single input buffered [24]
—shared buffered [2]
—output multibuffered and memorized blocked call [7]
—output multibuffered under bursty and non-uniform traffic [1]

- Performance analysis of internally non-blocking switches with input/output queue [28, 29, 30, 36]
- Performance analysis of knockout switch under non-uniform traffic [27]
- Performance analysis of tandem Banyan switch fabric [6]

4. Performance Analysis Modeling

4.1 Several Issues in FPS Performance Analysis Modeling

1) System Description and Characteristics

① System under Consideration

In a space division type switch, multiple paths are provided between the inlets and outlets.

Then, the total capacity of a space-division type switch is the product of the number of paths and the bandwidth of each path. Therefore, the capacity of a space division type switch is generally orders of magnitude higher than that of a time-division type switch, which makes space-division type designs preferable for realizing large dimension broadband packet switch. For the performance modeling, therefore, the switch fabrics of space-division type are only considered as shown in Table 2.

② Flow Control (or Internal Protocol) of Switch Network

Two type of flow control mechanism are suggested to regulate the flow of packets between stages to prevent packet loss due to blocking or buffer overflow

a. block and lost : There is no exchange of signaling information within the network, so that packets are always transmitted downstream by each switching element, independent of the current buffer status of the destination switching element. Packet storage in the switching element takes place as long as there are enough idle buffer positions, whereas packets are lost when the buffer is full.

b. backpressure : Request/acknowledgement signals are exchanged between switching elements in adjacent stages so that each switching element can grant the transmission of a packet to its upstream switching elements only within

Table 2. System under Consideration

Switch Type	Proposed System Architecture
Banyan	· unique path topology : Delta, Omega, or Shuffle · multiple path topology : ASEN, MD-Omega, or indra · parallel or tandem
Buffered Banyan -Based Network	· Buffers are placed at each switching element -input -output -shared
Buffered Internally Non-Blocking Switch Network (Batcher-Banyan or Crossbar)	· Buffers can be placed at each input and/or output of the network -input -output -input and/or output
Switch Fabrics with Disjoint Path Topology	· knockout switch -concentration -output shared buffer · MIN employing knockout switch as a switching element

the current idle buffer capacity. Consequently, we can think of the system as operating in two phases. In the first phase, flow control information passes through the network from right to left. In the second phase, packets flow from left to right, in accordance with the flow control mechanism. Two different cases can be distinguished based on the number of packet transmissions that is granted to the upstream switching elements:

· global backpressure : A switch will allow its predecessor to send it a packet if it has an empty buffer slot currently or if one of the packets in its buffer will leave during the second phase of the current cycle.

· local backpressure : A switch allows its predecessor to send a packet only if its buffer has an empty slot.

While local flow control doesn't make as effective use of switch buffers, it is more straightforward to implement, particularly in high speed systems where the propagation time required for global flow control can lead to unacceptable overheads. There can be two additional possibilities for implementing back-pressure flow control. Assume that each switching element has d input and output links.

- grant method : A switch with x empty ok slots grants permission to send a packet to min{x, d} of its upstream neighbors at the start of an operation cycle of the switch. If $x < d$, we assume that x predecessors are chosen at random.
- acknowledgement method : It allows all predecessors with packets to send them. The receiving switch stores as many as it can in its buffer and acknowledges their receipt by means of a control signal. Unacknowledged packets are retransmitted during a subsequent cycle.

Local/grant control is the easiest to implement and appears to have most practical interest while global/acknowledgement control provides best performance.

Following Table 3 shows the effect of internal protocols on the switch performance measures for the output buffered network.

Table 3. Effect of Flow Control Model on the System Performance Measure

Flow Control	Blocking Probability	Delay	Throughput
Block and Lost	· It experiences blocking in the switch network · Blocking probability can be very serious under the non-uniform bursty input traffic.	· At the low throughput level, there might be no difference between two controls · At the high throughput level, however, it can give lower delay at the expense of higher blocking probability.	At high offered traffic, it gives more carried traffic, i.e., higher throughput
Backpressure	· No internal blocking in the switch network · Blocking can occur only in IBC, and it can't be an issue.	· In highspeed system, the propagation delay for flow control can lead to unacceptable overhead. · For high throughput, the delay increases very rapidly	

③ Synchronous or Asynchronous system

· synchronous system : This reflects the situation in an ATM environment where all packets have a fixed length, called cell. All input links to the network are slotted and synchronized with a specific bit rate, for example 150 Mb/s. The resulting packet slot time is approximately 2.8s. Thus the switch fabric has to be designed in such a way that it can handle approximately 350,000 packets per second per input port.

· asynchronous system : Packets are assumed to arrive in the input trunk according to independent Poisson process at a rate of packets per unit of time.

④ Clock Cycle

It is considered the switch operates in a clocked format. How the clock cycle is composed depends on the internal protocol being used. For the grant method, each cycle has two phases, $\tau=\tau_1+\tau_2$, as shown in Figure 2. 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 in phase τ_2. During τ_2, a buffer may be emptied to the output and filled with a new packet from the input simultaneously. As mentioned above, in global method, 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 main problem with this approach is that the control signals in one stage can't be decided until the next stage's control signals arrive. If we consider that control signal passing accross one stage takes ϕtime units, then in a 2^n*2^n

network using 2*2 crossbar switch, τ_1 equals to $n^* \phi$ times as shown in Figure 2. This τ is called "big cycle". For the local method, control information is exchanged only between two successive stages that are required for the data transfer. Every cycle also consists of two phases, τ_1 and τ_2, as shows in Figure 3, called small cycle.

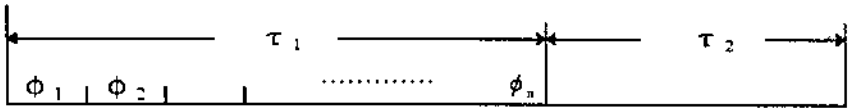


Figure 2. Big Clock Cycle

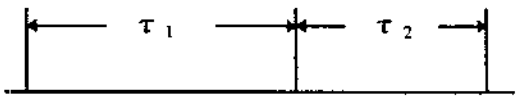


Figure 3. Small cycle

During τ_1 every input port in a switch informs its connected preceding switch about the buffer status, full or not full. If the succeeding switch's input port has a "not full" buffer, a packet destined to this output is selected and it can be passed in the next phase τ_2. Otherwise, no packet can go through the output port. Reference[3] shows the performance comparison between "big" and "small" clock cycles according to buffer length, switch size, and data width.

2) General Assumptions

Some assumptions for the performance modeling are discussed in this subsection (Assumptions related to the offered traffic is dealt with in the following subsection).

① Buffer Size

It can be assumed to be finite on infinite. Buffer size plays an important role in determining the maximum throughput achieved. For the buffered Banyan network, buffers can be placed at input or output of the switching element. Table 4 shows the different throughput value for various buffer sizes and locations.[5, 3, 19, 23].

Table 4. Throughput of the Buffered Banyan Network

Buffer Size	Throughput	
	Input Buffer	Output Buffer
1	.45	.45
2	·	.55
16	·	.90
:		
∞	.75	1

For the internally non-blocking switch fabrics such as cross-bar or Batcher-Banyan, buffers can be placed at input or output of the switch network. Table 5 shows the various throughput values depending on the buffer locations and sizes.[5, 3, 19, 23].

It is obvious that the output buffer gives

Table 5. Throughput of the Buffered Non-Blocking Switch Fabrics

Buffer Size	Throughput	
	Input Buffer	Output Buffer
1	.586	.586
2		.886
4		.976
:		:
∞		1

higher throughput compared to the input buffer due to the HOL blocking, which is intrinsic to input queue. However, the output queue requires to accept all packets arriving simultaneously at the input of a specific output queue. This can be realized by operating a simple queue at appropriate higher speed or by highly paralleled implementation, which leads to much more hardware complexity and cost.

② Blocked Call

In a block and lost protocol, blocked calls are completely ignored. In a backpressure protocol, however, rejected packets stay in a buffer and would have to retry to make request again in the next time slot. Two ways can be assumed

- random model : In next trial, it selects the next queue in random. Since it allows a blocked packet to be routed around a congested queue, the result obtained from this model might be optimistic.
- memory model : A blocked packet always hunts for the same queue during consecutive cycle, and does not obey static routing probability

The "blocked call problem" does not seem to allow the independence assumption employed in some analysis [5,25] that addresses of the packets stored within a switch buffer are independent. Buffered packets are correlated as a result of having contented for output. The correlations are strongest when buffer size is large and number of input/output ports are small. This independence assumption is used to calculate the interstage traffic flow in the switch network.

③ System State

In order to accurately model the behavior of MIN, it is essential to consider the correlation between the states of the buffers of two adjacent stages. Figure 4 shows the packet flow relationship between stages.

For example, incase of the input buffered MIN, the state of buffers at the $(k+1)$th stage depends on the state of buffers at the k-th stage at previous clock cycle. However, in order to simplify the analysis most works assume statistical independence between the states of buffers at the $(k+1)$th stage from states of buffers at the kth stage. In other words, the event that buffer space is available at the $(k+1)$th stage and the event that a packet does not encounter a conflict are assumed to be independent. Some works have been done to introduce dependency between buffers in different stages as follow.

- memorized blocked call
- the address of the packets stored within a

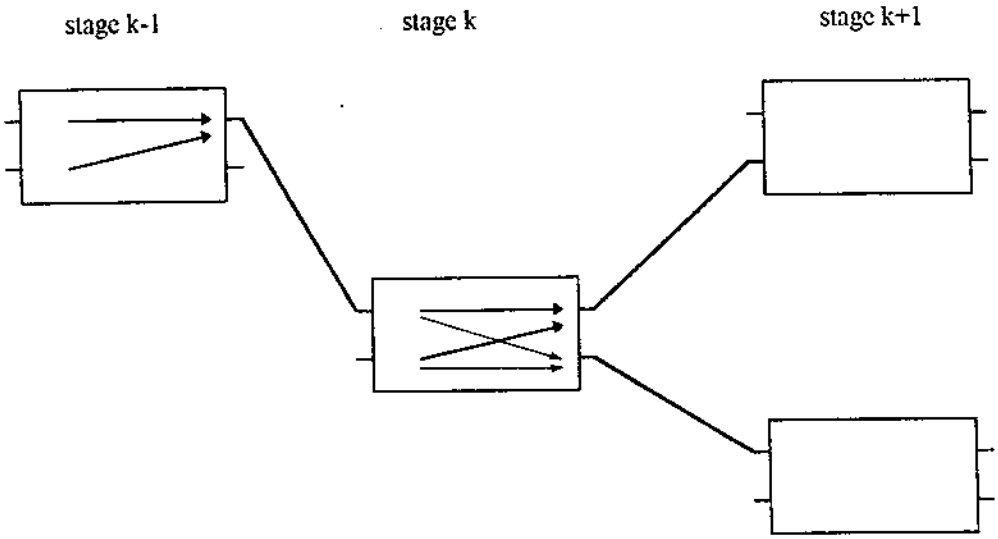


Figure 4. Relationship between Stages

switch buffer

(while it is true that the addresses of packets arriving at a switch are independent, buffered packets are correlated as a result of having contented for output)

The buffers in the same switching element can be assumed to be independent from each other under the homogeneous and uniform traffic pattern. Under the independence assumption, the state of a stage can be reduced further to that of a single buffer. However, packets in the buffers of a switching element do interfere with each other, and then it is not clear this independence assumption is reasonable.

For the single input buffered MIN, the buffer state can be represented as the two state Markov chain as shown below.

In order to accurately model the behavior of a MIN, it is essential to consider the

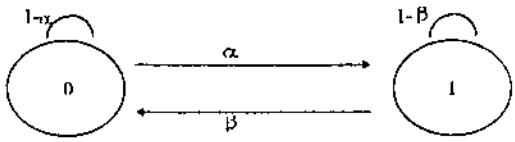


Figure 5. The State Transition Diagram of the Buffer State

correlation of packet movement between two subsequent network cycles. Also, the correlation between the states of the buffers of two adjacent stages needs to be considered. For example, in the case of memorized blocked call the "blocked" state must be included to correctly reflect these correlations into a model. The behavior of a buffer in the blocked state should be modeled differently from the one in the normal state. The destination of the packet in the normal state can be presumed to be uniformly distributed to any one of the output link, while that of a packet in the blocked state

is predetermined. The state transition diagram for memorized blocked call model is shown in Figure 6.

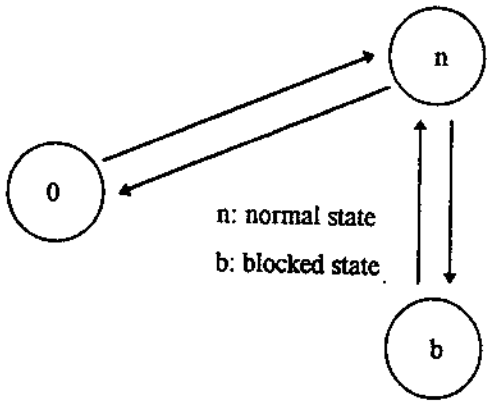


Figure 6. The State Transition Diagram for Memorized Blocked Call Model

Table 6 summarizes several assumptions related to the buffer state. It can be used to properly determine the system state variable

3) Offered Traffic Assumptions

Incoming traffic has been modeled as a Poisson process or Bernoulli process, called homogeneous input traffic. And output port destination has been assumed to be uniformly distributed or balanced. However, to represent traffic patterns of ATM network where a wide range of bandwidths needs tobe accommodated, some bursty traffic model is required. Also the output port destination might be unbalanced. The traffic pattern which reflects the network with one dedicated video channel and many voice or data channels is really unbalanced. The offered traffic characteristics can be classified into 4 different types as shown in the Table

Table 6. Assumption Related to the Buffer State

Modeling Element	Assumption
Interstage (or Inter-network cycle) Buffer Interdependency	· buffered packet address: -independent -various dependency assumptions · blocked call -random model -memorized model
Switching Elements in the Stage	· under balanced load the state of a switching element is statistically not distinguishable from that of another switching element of the same stage · under unbalanced load, it should be distinguished
Buffers in the Switching Element	· statistically independent · statistically dependent

7.

① Homogeneous-Balanced Traffic

· Each input i generates random and independent requests in a given cycle according to Poisson process with rate λ_1 or Bernoulli process with probability P_1.

· Packets arriving at input link at stage 1 are destined uniformly (balanced) for all output links at stage n.

These two assumptions make the load balanced in the whole switch network. Therefore, the state of a "stage" can be characterized by that of a "switching element".

Table 7. Offered Traffic Characteristics

Output port Destination / Incoming Traffic	Balanced (Uniform)	Unbalanced (Non-Uniform)
Homogeneous	homogeneous-balanced	homogeneous-unbalanced
Bursty	bursty-balanced	bursty-unbalanced

② Bursty-Balanced Traffic

· Each network input is assumed to be loaded by a bursty source, which can be represented by a two state Markov chain. A source in the ON-state (or bursty state) transmits a packet at the beginning of each time slot with probability P.

- Packet arrival events at different network inputs in the same or different time slots are mutually independent.
- All those packets belong to the same burst address the same network output
- Each burst is given a network output address that uniformly loads all the outputs.

③ Homogeneous-Unbalanced

- During a given time slot (or cycle) each input generates packets from source i to destination j by Poisson process with arrival rate λ_{ij} specified in a load matrix.
- Packets arrival events at different network inputs in the same or different time slots are mutually independent.

One particular unbalanced traffic pattern is the single-source-to-single destination (SSSD) traffic pattern in which each input source sends all its packets to a single output destination. Banyan networks require fewer switching elements because their structures allow links between stages to be shared by different traffic paths. Thus, the SSSD traffic may place extreme loads on common links shared by other traffics. Obviously, conflict due to packet collision occurs frequently as the traffic intensifies in the shared links. Wideband traffic such as large data or image can be modeled by SSSD traffic patterns. Another significant traffic pattern involves one SSSD path imbedded in a uniform traffic pattern as shown below.

$$\begin{vmatrix} \lambda & 0 & \lambda & \lambda\cdots & \lambda \\ 0 & \lambda_{sssd} & 0 & 0\cdots & \lambda \\ \lambda & 0 & \lambda & \lambda\cdots & \lambda \\ : & : & : & : & : \\ \lambda & 0 & \lambda & \lambda\cdots & \lambda \end{vmatrix}$$

This traffic pattern may reflect the network with one dedicated video channel and many voice or data channels .

There can be a alternative way to represent the homogeneous-unbalanced traffic.

- Each input link i receives a packet with probability P_i during a given time slot.
- The traffic pattern is unbalanced on the

network outputs in such a way that the outputs can be divided into a number of group such as "hot output group" and "cold output group".

④ Bursty-Unbalanced

The traffic arriving to the switch is completely characterized by the load matrix λ_{ij} and the burst matrix B.

- λ_{ij} represents the load (average number of cells per time slot) appearing on input port i destined for output port j.
- B_{ij} is the average number of cells per burst when input i generates a burst destined for output j.

Figure 7 shows a diagram of the Markov chain governing Source i. States are labeled such that when in state 0, the source is off (not generating any cell), and when in state k, $1 \leq k \leq N$, the source generates cells at a rate of 1 per slot destined to output port k.

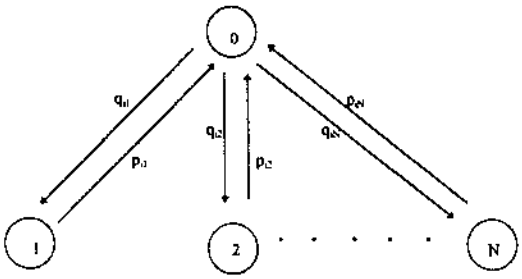


Figure 7. Bursty-Unbalanced Traffic Model for source i

In above figure it can be found

$$P_{ij}=\frac{1}{B_{ij}} \text{ and } q_{ij}=p_{ij}\cdot\lambda_{ij}/(1-\sum_{k=1}^{n}\lambda_{ik})$$

Additionally following assumptions are required.

- All cells within a burst at the input to the switch are assumed to be destined to the same output port.
- Packet arrival events at different network input in the same or different time slots are mutually independent.

4) Measures

Even though the switch networks have the same architecture, their performance might vary according to their buffer sizes, offered traffic characteristics, and internal protocol being used. Also, the performance measures of concern would be different according to internal protocols being used and incoming traffic characteristics. That is, for block and lost model, the blocking probability is of main concern. For backpressure model which does not have any internal blocking, however, delay is important measure. Under uniform traffic pattern, all throughputs of destination ports and the average delays of all possible source-destination paths are equal. Therefore, the throughput and the delay of the network can be obtained from a single destination port and a single source-destination path, respectively. And "average" delay and throughput are sufficiently enough to represent the network performances. However, the performance measures of the network under a non-uniform traffic pattern require different considerations. Obviously, some source-to-destination paths may have better or worse performance than

other pairs for non-uniform traffic patterns. Therefore, we are interested in the maximum allowable throughput at a destination port and maximum delay.

In this section, formal definitions of performance measures are given and the performance affecting factors are identified.

① Definitions

- Bandwidth : number of output requests (packets) accepted per clock cycle or time slot.
- Throughput : number of output requests per output link per clock cycle
- Delay : number of clock cycles a packet takes to reach the destination port from the source port
- Blocking probability :

—probability that the packet is lost due to internal blocking, output port contention, or full buffer in case of buffered switch (block and lost model)

—probability that IBC(Input Buffer Controller) buffer is full (backpressure model)

Under the nonuniform traffic pattern, following measures might be more suitable

- maximum throughput : maximum allowable throughput at an output port which has the worst congestion among all the output ports.
- maximum delay : delay of the input to output path which has the longest delay of all the paths.
- maximum blocking probability : blocking probability of an input link which has the largest blocking probability of all the input links.

② Performance Affecting Factors

As discussed above briefly, the performance of a given switch network can vary widely according to the factors shown in the Table 8. The performance of a switch network can be determined by the interactions of following factors.

4.2 Performance Analysis Model

Based on the discussion of section 4.1, the literatures on the performance analysis modeling were surveyed and analyzed in the Table 9. Only the literatures are selected, which are considered to have some meaningful progresses

Table 8. Performance Affecting Factors

Switching Element	Size	2×2 or d×d
	Type	crossbar, or knockout, etc.
Buffer	Location	input, output, input/output, or shared
	Size	finite, or infinite
Flow Control		-block and lost -backpressure : global/grant global/ack local/grant local/ack
Incoming Traffic		-homogeneous/balanced -homogeneous/unbalanced -bursty/balanced -bursty/unbalanced
Blocked Call		-random -memory

Table 9. Performance Analysis Modelings

Modeled Systems	Modeling Methods	SE Size	Flow Control	Clock Cycle	Buffer Location	Buffer Size	Blocked Call	Offered Traffic Assumption	Measure	Reference
Banyan Network	· simple probabilistic iterative	2×2 crossbar	block and lost		no buffer	no buffer	ignored	homogeneous-balanced	· bandwidth · probability that arbitrary request will be accepted	20
Tandem Banyan Network	· combinatorials	2×2 crossbar	block and lost		no buffer	no buffer	ignored	homogeneous-balanced	· blocking probability	6
Buffered Banyan	· Markov chain model on the input buffer state of switching element	2×2 crossbar	backpressure	big	input ink at SE	1	random	homogeneous-balanced	· throughput · delay · blocking probability for finite IBC buffer	23
					IBC	finite	ignored			
						infinite				
	· Markov chain model on the input buffer · describe the relationships among different SE through average flow constraint	2×2 crossbar	backpressure	big	input link at SE	1	random	homogeneous-unbalanced	· maximum throughput · maximum delay	24
					IBC	infinite				
	· simulation analysis for backpressure and block and lost · —Markov chain model on the output buffer in each SE Modeling Methods —describe the relationship among different SE's through probabilistic means	k×k knockout switch	backpressure	big	output link at SE	arbitrary	random	homogeneous-balanced	· throughput · delay · blocking probability for IBC	19
					IBC	finite	ignored			
			block and lost	big	output link at SE	arbitrary	ignored	homogeneous-balanced	· throught · delay · blocking probability	

Table 9. Performance Analysis Modeligs (계속)

Model System	Modeling Methods	SE Size	Flow Control	Clock Cycle	Buffer Location	Buffer Size	Blocked Call1	Offered Traffic Assumption	Measure	Reference
	· Markov chain model on the shared buffer state of the SE · derive the performance measure by iterative computational method	k×k crossbar	backpressure (local/grant)	small	shared between input and output links of SE	arbitrary	random	homogeneous-balanced	· throughput · delay	5
	· Markov chain model on the input buffer	2×2 crossbar	backpressure	small	input	arbitrary	memory	homogeneous-balanced	· throughput · delay	4
	· Markov chain model on the input buffer	k×k crossbar	backpressure	small	input	arbitrary	random	homogeneous-balanced	· throughput · delay	3
	· Markov chain model on the shared buffer · By using recursive method, steady-state Probability is derived	k×k crossbar	block and lost	big	shared between input and output links of SE	arbitrary	ignored	bursty unbalanced	· blocking probability · throughput · delay	2
			backpressure	big	shared between input and ouput links of SE	arbitrary	random	bursty unbalanced	· blocking probability (IBC) · throughput · delay	
					IBC	finite	ignored			
	· Markov chain model on the output buffer	2×2 crossbar	backpressure	small	output link of SE	arbitrary	random memory	homogeneous-unbalanced	· throughput · delay	7

Table 9. Performance Analysis Modelings (계속)

Modeled System	Modeling Methods	SE Size	Flow Control	Clock Cycle	Buffer Location	Buffer Size	Blocked Call	Offered Traffic Assumption	Measure	Reference
	· fluid-flow approach · It is very attractive because its complexity depends only on the number of states in the chain, but is independent of buffer size.	k×k crossbar	block and lost	big	output link of SE	arbitrary	ignored	bursty-unbalanced	· delay(max) · blocking Probability · throughput(max)	1
			backpressure	big	· output link of SE · IBC	arbitrary	random	bursty-unbalanced	· delay(max) · blocking probability · throughput(max)	
Buffered Internally Non-Blocking Switch Network	· equivalent discrete time queueing system with bulk arrival	N×N crossbar(for crossbar switch) 2×2 crossbar (for Batcher Banyan Switch)	backpressure	small	input and/or output of the switch network	· output-arbitrary · input-large enough to avoid blocking probability	random	homogeneous-unbalanced	· delay	26,30
Switch Fabrics with Disjoint Path Topology	Knock-out Switch · Markov chain analysis on the output(shared) buffer state		block and lost		output of switch network	arbitrary	ignored	homogeneous-unbalanced	· delay · blocking probability · throughput	27
	MIN based on Knockout Switch · Markov chain analysis on the output buffer size	k×k knockout switch	block and lost	small	output at each SE (knockout switch)	arbitrary	ignored	homogenrous-balanced	· delay · blocking probability · throughput	31

Table 10. Assumption Related to the SE and Buffer State

Reference	Assumptions Related to the SE and Buffer State		
	Interstage Buffer Interdependency	SE's in the Stage	Buffer's Interdependency in SE
20	independent	identical	independent
23	independent	identical	i) independent ii) dependent
24	independent	distinguish	dependent
19	independent	identical	independent
5	(buffered packet address) i) independent ii) dependent	identical	shared buffer
2	independent	i) identical for bursty traffic ii) distinguish for unbalanced traffic	shared buffer
7	(blocked call) i) independent(random) ii) dependent(memory)	distinguish	dependent
1	independent	distinguish	dependent
26,30	independent	identical	independent
31	independent	identical	independent
6	independent	identical	independent
4	(blocked call) dependent (memory)	identical	independent
3	independent	identical	independent

toward more realistic and accurate modeling.

Table 10 shows various assumptions about buffer state and SE, which is very basic to the model analysis. They are very closely related to the "blocked call problem", "offered traffic", and "modeling method".

5. Conclusions

For the performance analysis investigations of the fast packet switch, the switch fabrics of space division type are only considered with the emphasis on the buffered Banyan type network. Among the numerous performance analyses on the space division switches, those

modelings are selected and compared, which are considered to have some meaningful progresses toward more realistic and accurate modelings, or computational simplicity.

In this investigation, all the components in the switch network are assumed to be fault free. However, the components or devices are surely subject to failure. A failure of any switch or link can disconnect several source - destination pairs, because several pairs usually share a common link in the network, leading to a lack of fault tolerance and low reliability. To provide fault tolerance and improve system reliability and performance, several types of fault - tolerant switch network have been proposed. There have been a lot of work in this field [38 - 77]. How these component failures affect the switch network performance is the main topic of the study [78] sequel to this one.

References

[1] B. R. Collier, H. S. Kim, " Performance of multistage ATM switch architecture under nonuniform bursty traffic" , IEEE INFOCOM, 1995, pp 667-674

[2] S. Gianatti, A. Pattavina, "Performance analysis of shared - buffered Banyan networks under arbitrary traffic pattern," IEEE INFOCOM, 1993, PP 943-952

[3] J. Ding, L. N. Bhuyan, "Finite buffer analysis of multistage interconnection networks" , IEEE Transactions on Computer, Vol. 43, No.2, 1994 Feb., PP 243-247

[4] Y. Mun, H.Y.Youn, "Performance analysis of finite buffered multistage interconnection network", IEEE Transactions on computers, Vol. 43, No.2, 1994 Feb., PP 153-162

[5] J. S. Turner, " Queueing analysis of buffered switching networks", IEEE Transactions on Comm., Vol. 41, No.2, 1993 Feb. , PP 412-420

[6] S. Sibal, J. Zhang, " On a class of Banyan networks and tandem banyan switching fabrics", IEEE Transactions on Comm., Vol 43, No.7, 1995 July , PP 2231-2240

[7] B. Zhou, M. Atiquzzaman, "Performance of output-multibuffered multistage interconnection networks under general traffic pattern", IEEE INFOCOM, 1994, PP 1448-1455

[8] H.Ahmad: W.E. Denzel, "A survey of modern high performance switching techniques", IEEE Journal on Selected Areas in Comm., Vol. 7, No.7, 1989 Sep., PP 1091-1103

[9] R. W. Muise, T. J. Shonfeld, G. H. Zimmerman, "Digital communictions experiments in wideband packet technology", in Proc. Int. Zurich Seminar Digital Comm., Zurich, Switzerland, 1986 Mar., PP. 135-140

[10] G. W. R. Luderer et. al., "Wideband packet technology for switching systems", in Proc. ISS '87, Phoenix, AZ, 1987 Mar., PP. 448-454

[11] J. S. Turner, "Design of a broadcast switch network", IEEE INFOCOM, 1986, PP 667-675

[12] A. Huang, S. Knauer, "Starlite : A wideband digital switch", GLOBECOM, 1984, PP 121-125

[13] K. Y. Eng , M. G. Hluchyj, Y.S. Yeh, "Multicast and broadcast services in a knockout packet switch," . IEEE INFOCOM, 1988, PP 29-34

[14] S. Nojima, et al., "Integrated services packet network using bus matrix switch" IEEE Journal on Selected Areas in Comm., Vol. SAC-5, 1987 Oct., PP 1248-1292

[15] U. Killat, "Asynchrone Zeitvielfach(bermittlung f(r Breitbandnetze", Nachrichtentech. Z., Vol. 40, No.8, 1987, PP 572-577

[16] J. P. Coudreuse, M. Servel, "Prelude : An asynchronous time division switched network", ICC, 1987, PP 769-773

[17] I. S. Gopal, H. Meleis "Paris : An approach to integrated private networks" , ICC, 1987, PP 764-768

[18] T.Y. Feng, "A survey of interconnection networks", Computer, Vol 14, 1981 Dec., PP 12-27

[19] H.S. Kim, I. Widjaja, A. L. Garcia, "Performance of output buffered Banyan networks with arbitrary buffer sizes", IEEE INFOCOM, 1991,, PP 701-710

[20] J. H. Patel, "Performance of processor-memory interconnections for multiprocessor", IEEE Transactions on Computers, Vol. C-30, No 10, 1981 Oct., PP 771-780

[21] L. R. Goke, G. J. Lipovski, "Banyan networks for partitioning multiprocessing systems", Proc. First Annual Computer Architecture Conference, 1973 Dec., PP 21-28

[22] D. H. Lawrie, "Access and alignment of data in an array processor", IEEE Transactions on Computers, Vol. 24, 1975 Dec., PP 99-109

[23] T.Zeng, "Performance analysis of a packet switch based on single-buffered Banyan network", IEEE Journal on Selected Areas in Comm., Vol. SAC-1, No.6 1983 Dec., PP 1014-1021

[24] H. S. kim, A. L. Garcia, "Performance of buffered Banyan networks under nonuniform traffic patterns, IEEE Transactions on Communications, Vol. 38, No.5. 1990 May, PP 648-658

[25] G. Bianchi, J. S. Turner, "Improved queueing analyis of share buffered switching networks, IEEE INFOCOM, 1993, PP 1392-1399

[26] I. Iliadis, W. E. Denzel, "Analysis of packet switches with input and output queueing", IEEE Transaction on Communications, Vol. 41, No.5, 1993 May, PP 731-740

[27] H.Yoon et al., "The knockout switch under nonuniform traffic", IEEE Transaction on Comm., Vol.43, No.6, 1995 June,

PP2149-2156

[28] M.J.Karol, M.G. Hluchyj, S.P. Morgan, "Input versus output queueing on a space-division packet switch", IEEE Transaction on Comm., Vol.35, 1987 Dec., PP1347-1356

[29] J.Y.Hui, E Arthurs, "A broadband packet switch for integrated transport," IEEE Journal on selected Areas in Comm., Vol SAC-5, 1987 Oct., PP 1264-1273

[30] C. Y. Chang, A.J. Paulraj, T. KaiLath, "A broadband packet switch architecture with input and output queueing", GLOBECOM, 1994, PP 448-452

[31] Y.M. Kim, K.Y. Lee, "KSMIN's : knockout switch based multistage interconnection networks for highspeed packet switching", IEEE Transactions on Comm., Vol. 43, No.8, 1995 Aug., PP2391-2398

[32] F.A. Tobag: "Fast packet switch architecture for broadband integrated services digital networks," Proceeding of IEEE, Vol. 78, No.1 1990 Jan, PP 133-167

[33] M. Narasimha, "The Batcher-Banyan self routing network : universality and simplification", Vol. 36, No. 10, 1988 Oct., PP 1175-1178

[34] T. Szymanski, S. Shaikh, "Markov chain analysis of packet switched Banyans with arbitrary switch sizes, queue sizes, link multiplicities and speedups, IEEE INFOCOM, 1989, PP 960-971

[35] M. Atiquzzaman, M.S. Akhtar, "Performance of buffered multistage interconnection networks in non-uniform traffic environment", 7th International Parallel Processing Symposium, 1993, PP 762-767

[36] C.P. Kruskal, M.Snir, "The performance of multistage interconnection networks for multiprocessors", IEEE Transaction on Computers, Vol. 32, 1983 Dec, PP 1091-1098

[37] D.M. Dias, J.R. Jump, "Analysis and simulation of buffered delta networks", IEEE Transactions on computers, Vol. 30, 1981 Apr., PP 273-282

[38] V.P. Kumar, S.J. Wang, "Reliability enhancement by time and space redundancy in multistage interconnection networks", IEEE Transactions on Reliability, Vol. 40, No.4, 1991 Oct., PP 461-472

[39] A. Varma, C.S. Raghavendra, "Fault-tolerant routing in multistage interconnection networks", IEEE Transactions on computers, Vol. 38, 1989 MR., PP 385-393

[40] S.Wei, G. Lee, "Extra group network : a cost-effective fault-tolerant multistage interconnection network", GLOBECOM, 1988 , PP 108-115

[41] M. Veeraraghavan, K.S. Trivedi, "A combinatorial algorithm for performance and reliability analysis using multistate models", IEEE Transactions on Computer, 1994 Feb., PP 229-234

[42] C.Botting et al., "Reliability computation of multistage interconnection networks",

1989 Apr., pp 138-145

[43] J. Spragins, J.Assiri, "Communication network reliability calculations with dependent failures", Proc. National Telecomm. Conf., 1980. PP 25.2.1-25.2.5

[44] S. Pan, J. Spragins, "Dependent failure reliability models for tactical communication networks", Proc. International Conf. Communication, 1983, PP 765-771

[45] Y.Lam, V. Li, "On reliability calculations of network with dependent failure", GLOBECOM, 1983 , PP 1499-1503

[46] Y. Lam, V. Li, "Reliability modeling and analysis of communication networks with dependent failure", IEEE Transactions on Comm., Vol. 34, 1986 Jan, PP82-84

[47] V.Cherkassky, M. Malek, "Reliability and fail-softness analysis of multistage interconnection network", IEEE Transactions on Reliability, Vol. R-34, No.5, 1985 Dec., PP 524-527

[48] V. Cherkassky, E. Opper, M. Malek, "Reliability and fault diagnosis analysis of fault tolerant multistage interconnection networks", 14th International Symposium Fault Tolerant Computing, 1984 June, PP 246-251

[49] I. Koren, Z.Koren, "Analyzing the connectivity and bandwidth of multiprocessors with multi-stage interconnection networks", in Concurrent Computations : Algorithms, Architecture, and Technology, NY- Plenum, 1988, PP 525 - 540

[50] I. Koren, Z.Koren, "On the bandwidth of a multistage networks in the presence of faulty components", in Proc. 8th International Conference Distributed Computing System, 1988, PP 26-31

[51] I. Koren, Z. Koren, "On Gracefully degrading multiprocessors with multistage interconnection networks", IEEE Transactions on Reliability, Vol. 38, No. 1, 1989 Apr., PP 82-88

[52] V.Kumar, A. L. Reibman, "Failure dependent performance analysis of a fault tolerant multistage interconnection network", IEEE Transactions on Computers, Vol.38, No.12 1989 Dec., PP 1703-1713

[53] J.T.Blake, K.S. Trivedi, "Multistage interconnection network reliability", IEEE Transactions on computers, Vol. 38, No.11, 1989 Nov., PP 1600-1604

[54] J.T.Blake, K.S. Trivedi, "Reliability analysis of interconnection networks using hierarchical composition", IEEE Transactions on Reliability, Vol.38, No. 1, 1989 Apr, PP 111-119

[55] J. T. Blake, A. Reibman, K.S. Trivedi:, "Sensitivity analysis of reliability and performability measures for multiprocessor system," in Proc. ACM SIGMETRICS Conf., Measurment Modeling Computing System, 1988, PP 177-186

[56] P.K. Bansal, K.Singhi, R.C. Josh, "Reliability and performance analysis of a modular multistage interconnection network", Microelectronics and Reliability,

Vol.33, No. 4, 1993, PP 529-534

[57] S. Rai, K. K. Aggarwal, "An efficient method for reliability evaluation of a general network", IEEE Transactions on Reliability, Vol. 27, No. 3, 1978 Aug., PP 206-211

[58] S.C. Yang, J.A. Silvester, "Reconfiguable fault tolerant networks for fast packet switching", IEEE Transtions on Reliability, Vol. 4, 1991 Oct., "PP 474-486

[59] S.C. Yang,. J.A. Silvester, "A fault-tolerant reconfiguable ATM switch fabric", IEEE INFOCOM, 1991, PP 1237-1244

[60] J.P. Shen, J. P. Hayes , "Fault tolerance of dynamic full access interconnection networks", IEEE Transactions on Computers, Vol. 33, No.3, 1984 Mar., PP 241-248

[61] B.L. Menezes, U. Bakhru, R. Sergent, "New bounds on the reliability of two augmented shuffle exchange networks", in Proc. International Conf. On Parallel Processing, 1991 Aug., PP 1318-1322

[62] B.L. Menezes, U. Bakhru, "New bands on the reliability of augmented shuffle exchange networks" IEEE Transtions on Computers, vol.44, No.1, 1995 Jan., PP 123-129

[63] G.B. Adams, H.J. Siegal, "The Extra stage cube : A fault-tolerant interconnection network for supersystems", IEEE Transactions on Computer, Vol. 31, 1982 May, PP 443-454

[64] R.J. McMiller, H. J. Siegel, "Routing schemes for the augmented data manipulator network in an MIND system", IEEE Transactions on Computers, vol. 31, 1982 Dec., PP 1302-1214

[65] H.J. Siegel, R. J. McMillen, "Dynamic rerouting tag schemes for the auagmented data manipulator network", Proc. 8th International Symposium Computer Architecture, 1981 May, PP 505-516

[66] R.J. McMillen, H.J. Siegel, "Performance and fault-tolerance improvements in the inverse augmented data manipulator network", Proc. 9th Annual Symposium Computer Architecture, 1982 June, PP 63-72

[67] D.S. Parker, C.S. Raghavendra, "The gamma network : A multiprocessor interconnection network with redundant paths", Proc. 9th Annual Symposium Computer Architecture, 1982 June, PP 72-80

[68] C.S. Raghavendra, D. S. Parker, "Reliability analysis of an interconnection network", Proc 4th International Conference Distributed Computing Systems, 1984 May, PP 461-471

[69] L. Ciminiera, A. Serra, "A fault tolerant connecting network for multiprocessor system", Proc. International Conf. Parallel Processing, 1982 Aug., 99 113-122

[70] K. Padmanabhan, D. H. Lewis, "Fault-tolerant schemes in shuffle exchange type interconnection network", Proc. International Conf. Parallel Processing, 1983

Aug., PP 71-75

[71] K. Padmanabhan D.H. Lewis, "A class of redundant path multistage interconnection networks", IEEE Transactions on computer, Vol. 32, 1983 Dec., PP 1099-1108

[72] S.M. Reddy, V.P. Kumar, "On fault tolerant multistage interconnection networks", Proc. International Conf. Parallel Processing, 1984 Aug., PP 155-164

[73] V. P. Kumar, S. M. Reddy, "Design and analysis of fault-tolerant multistage interconnection networks with low link complexity", Proc. 12th International Symposium Computer Architecture, 1985 June, PP 376-386

[74] V. P. Kumar, S. M. Reddy, "Angmented shuffle exchange multistage interconnection network", Computer, 1987 June, PP 30-40

[75] N. Tzeng, P.Yew, C. Zhu, "A fault tolerant scheme for multistage interconnection network", Proc. 12th International Symposium Computer Architecture, 1985 June, PP 368-375

[76] C.S. Raghavendra, A. Varma, "INDRA : a class of interconnection networks with redundant paths", Proc. 1984 Real Time Systems Symposium, 1984 Dec.

[77] A. Satyanarayana, A. Prabhakar, "New topological formular and rapid algorithm for reliability analysis of complex networks", IEEE Transactions on Reliability, Vol. 27, 1978 June, PP 82-100

[78] K. W. Lee, "Performance Analysis of MIN Switching System under the Presence of Failure", submitted to Korea Management Science, 1996

96년 2월 최초 접수, 96년 4월 최종 수정