Journal of the Korean
OR/ MS Society
Vol. 21, No. 2, August 1996 

# MIN (Multistage Interconnection Network) System Performance Analysis under the Presence of Failure*

Kang Won Lee**

Abstract

The purpose of this paper is to critically review and compare the evaluation methods of fault tolerant MIN system and to suggest future study direction. The reliability analysis of MIN system is discussed. The system performance modeling under the faulty components is another main focus of this study.

## I. Introduction

In recent years, design of efficient ATM switching systems for broadband ISDN (B-ISDN) has received considerable attention. Among the desirable goals of the design are switches that have low delays, distributed routing, high throughput, and low hardware complexity. Multistage Interconnection Networks (MIN's) are strong candidate for implementation of ATM switching fabrics in B-ISDN.

They are also used to provide high speed communication paths between the processors and memories in multiprocessor system. The reasons lie in their suitability to VLSI implementation, simple hardware complexity, and their self-routing capability. Compared to the time shared bus or the cross bar switch, it is shown that they provide cost-effective and high bandwidth communication [8, 18,32].

* 본 연구는 연암문화 재단의 지원을 받아 1995년 1월부터 1996년도 1월까지 미국 Georgia Institute of Technology 에서 수행하였습니다.
** Department of Industrial Engr. Seoul National Polytechnic University.

MIN's are usually designed for $N=m^n$ inputs and N outputs, and contain $\log_m^n$ stages of m× n crossbar switches. The typical example is shown in Fig. 1 with N=8, M=2, i. e., 8x8 shuffle exchange MIN.

The switches in adjacent stage are connected by an interconnection pattern in such a way that only one path is available between any inlet and outlet. This class of network is known as Banyan type networks, as originally defined in [21,22], that include Omega network and Delta network. The permutation capability of MIN is determined by the interconnection pattern of the network. The combinatorial power of an MIN is the fraction of the n! one-to one mappings that can be realized by compatible sets of routes. Since each of the $(n/m)\ \log_m^n$ switches can realize m! distinct permutations and since each switch setting yields a different permutation for the network as a whole, the number of distinct permutations realized by the Delta network is exactly $(m!)^{(N/M)\log_m^n}$ (For the rearrangeable non-blocking and stritcly non-blocking networks such as Benes and clos-networks respectively, the combinatorial power becomes 1.)

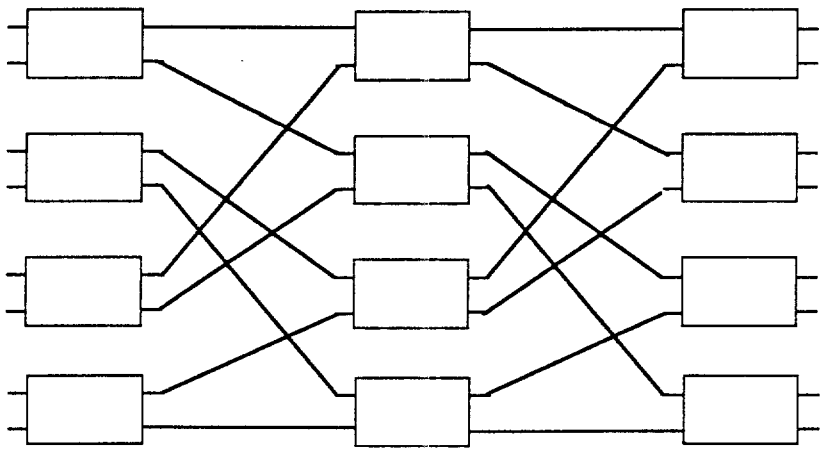


Fig. 1 8 X 8 Shuffle-Exchange MIN

A basic property of Banyan type networks is that there is a unique path between any source-destination pair and distinctive input-output path may have common links. This leads to two serious disadvantages of MIN.

1. 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.

2. As shown in Fig. 2 input connection may be blocked by either internal blocking (different input-output path may share common links) or output contention (same output address), thereby causing poor performance in random access environment.

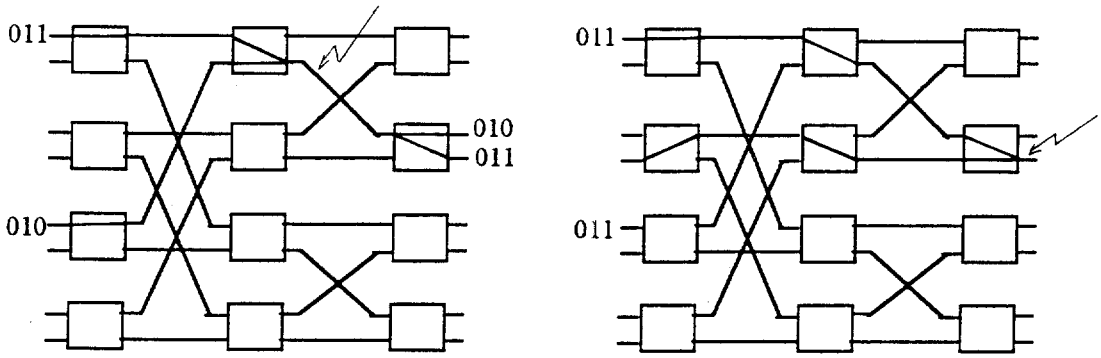



Fig. 2 Blocking Phenomena in 8x8 Delta Network

These lack of fault-tolerant capability, low reliability, and poor performance has received considerable attention, and many ways of providing fault tolerance and increasing performance of MIN have been proposed [38, 63]. The basic idea for fault tolerance is to provide multiple paths or passes for a source-destination pair so that alternate paths can be used in case of faults in a path, thereby leading to high reliability, and good performance even in the presence of faults.

Now, to evaluate and compare the proposed fault tolerant MIN system, two system evaluation measures have been suggested.

1. System reliability

The probability of maintaining full access capability (network reliability) is derived under the presence of failure. This measure considers the tolerable number of switch failure. MTTF of MIN is obtained as the result of analysis. And also probability that at least one path exists from a particular input to a particular set of output is calculated.

2. System performance

The fault tolerance makes the MIN to maintain full access property even under the several faulty components, but to function with some degraded performance. That is, some faulty components do not affect the system reliability measure, but the performance such as blocking probability, delay, and throughput. Therefore, to evaluate the fault tolerant MIN system accurately system performance measure as well as reliability should be taken into consideration.

The purpose of this paper is to critically review and compare the evaluation methods of fault-tolerant MIN system and to suggest future study direction. In section 2, the reliability of MIN system is discussed. The system performance modeling under the faulty components is the main focus of section 3. Based on the discussion of section 2 and 3, the further research topics are proposed in section 4.

# Ⅱ. MIN System Reliability

Several important problems arising in reliability modeling of MIN system are discussed. Based on this discussion system reliability models are analyzed.

## 2. 1 Issues in Modeling of MIN System Reliability

Several important issues are discussed, which will be used for analyzing criteria of MIN system models.

### 1) System under Consideration

Several systems have been studied for the reliabilty analysis of MIN.

① unique path system

There is a unique path between any input-output pair, and distinct input-output path may have same common link. Therefore, the failure of even a single link or switch in the network disconnects several input-output pairs. They also experience both internal blocking and output contention. Banyan type network is typical example of this system.

② space redundancy fault tolerant network

Multiple paths are provided from each input to each output so that alternate path can be

used in case of faults. The methods include to increase the number of stages, to use multiple links between switches (intrastage or interstage link), to partition a unique path network into several subnetworks, and to incorporate multiple copies of a basic network. There are many examples of such fault-tolerant MINs : the Extra-Stage Cube [63], the Augmented Data Manipulator and its variations [64, 65, 66], the Gamma Network and its variations [67,68], the F- network [69], the Modified Omega Network [70,71], the Augmented C-Network[69], the Modified Omega Network [70,71], the Augmented C-Network and the Merged Delta Network [72], the Augmented Shuffle-Exchange Network [73,74], the Chained Baseline Network [75], and the INDRA network [76].

③ time redundancy fault tolerant network

Suppose that a MIN is used to connect processing elements, with end-around connections as shown in Fig. 3. In this case, multiple passes through the network can be allowed to route data from an input to an output. For instance, if PE(Processing Element) X can not route directly to PE Y in one pass through the network, it might be possible to route from PE X to PE Z in one pass, and then PE Z to PE Y in the second pass. The ability to route from any input to any output of the network in a finite number of passes is called the dynamic full access(DFA).

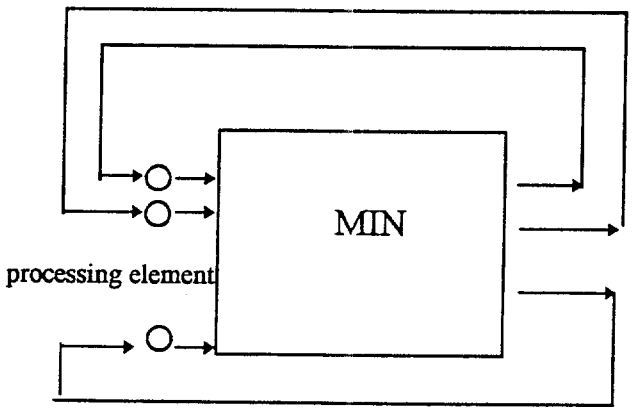



Fig. 3 MIN with End-Around Connections for Multiple Pass Routing

④ reconfigurable fault tolerant network

When a switch (or a module) fails, a replacement switch (or a module) assumes all operational functions that were performed by faulty one. And the reconfigurable links should be

properly assigned to support all the replacement structure. With this replacement model, the design of reconfigurable fault tolerant is transformed into the two subproblems :

a. replacement design : design a feasible replacement topology to achieve a certain performance indices of interest, e. g. maximum fault tolerance and reliability, for a given hardware overhead.
b. link assignment : properly design the interconnections based on feasible technology such that the replacement topology is achieved.

2) Reliability Measure

Recently there have been much analysis and comparisons of MIN's from the following aspects of reliability.

① network reliability

It is the probability of maintaining full access capability throughout the network. That is, it can be defined as the probability that any output port can be reached from any input port of the network in a cycle pass. This measure considers the tolerable number of switch (or link) failure.

② broadcast reliability

It is defined as the probability of maintaining a connection from one source to many destinations

③ terminal reliability

It is the probability that at least one path exists from a particular input port to a particular output port.

In this paper, network reliability is mainly focused.

3) Fault Mode

There can be three types of fault modes adopted to the reliability analysis of the MIN.

① link fault

It is the failure of an individual link connecting two switches. Interconnection switches are decomposed into links and if one link is faulty, the other link can still operate properly.

② switch fault

It is the failure of an individual switching element. Switch faults can be used to pessimistically approximate link faults. That is, all the links connected to the faulty switch are considered to be totally unusuable. In some cases, the reliability of an switching element cmopo-

nent may be considered as an independent failure/success entity. An switching element having any one of its components failed still works in degraded mode, thus a switching element behaves like a multistate element.

③ module fault

In some fault tolerant systems, module (loop, or group) can be a replaceable unit. When a switch is identified to be faulty, the module containing the faulty switch is replaced by a fault-free module. During repair or replacement all the switching elements in a module are considered to be faulty. Generally, in reliability model, switch and module fault modes are adopted to conservatively estimate system reliability.

For the purpose of network fault tolerance and reliability analysis, we should also distinguish between network interface faults and inner network faults.

① network interface fault

faults on the link connecting resources (e. g. computer, processor, memory) to the network

② inner network fault

faults on all the inner links and switches of the network

4) Repair/Replacement

In real systems utilizing fault tolerant networks, it is anticipated that the detection of a fault in a switching element initiates the repair of fault to guard against the occurrence of a second, potentially catastrophic fault.

① simple Markov chain reliability model with switch repair

Let the constant failure rate of individual switch be $\lambda$ and the constant repair rate $\mu$. Suppose we have a MIN with M switches. Even if the MIN can tolerate more than one faulty switch in many cases, we assume that more than one faulty switch lead to the system failure. Therefore, the whole system can be conservatively represented by the Markov chain model shown in Figure 4, where there are three state: state 0 is the no fault state, state 1 is the single fault state, while state 2 is the two fault state.

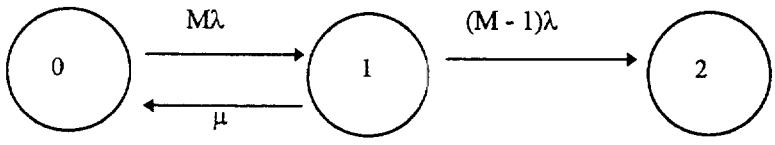



Fig. 4 Markov Chain Model for System with Switch Repair

② Markov chain reliability model with imperfect coverage

Often when a component fails, the detection, isolation, and reconfiguration procedures of the system are less than perfect. This notion of imperfection is called imperfect coverage, and is defined as the probability, c, that the system successfully reconfigures given that a component fault occurs. Imperfect coverage is important in considering the reliability of MINs since as their size increases, the number of components increases, and the potential for an uncovered fault to occur increases, as well. This system also can be represented by the Markov chain model as shown in Fig. 5.

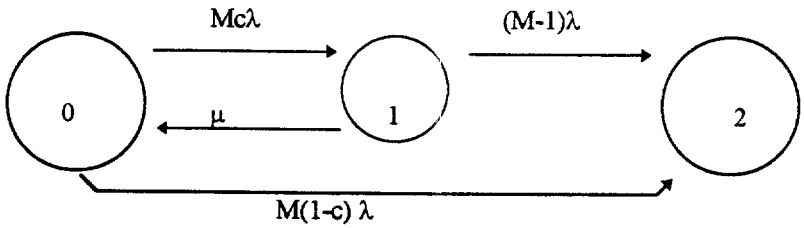



Fig. 5 Markov Chain Model for System with Imperfect Coverage

③Markov chain availability model with loop(or module) replacement

In some cases, module can be a replaceable unit. When a switch is identified to be faulty, the module containing the faulty switch is replaced by a fault-free module. Simple Markovian model is shown in Fig. 6. Here the model is simplified by ignoring the possibility of a second switch fault.

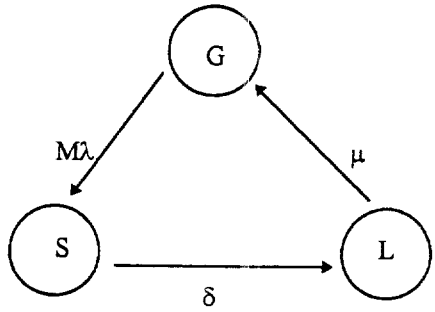



Fig. 6 Markov Chain Availability Model with Module Replacement

Initially the system is fully operational (in state G). If a fault occurs and a switch ceases to operate the system moves to S. State S represents all possible equally likely types of switch faults. Faults are detected at rate $\delta$. Once the fault is detected, and the repair begins at a rate of $\mu$, the entire module containing the faulty switch is out of service (state L)

Following Fig. 7 is the availability model of system with imperfect coverage as shown is Fig. 5.

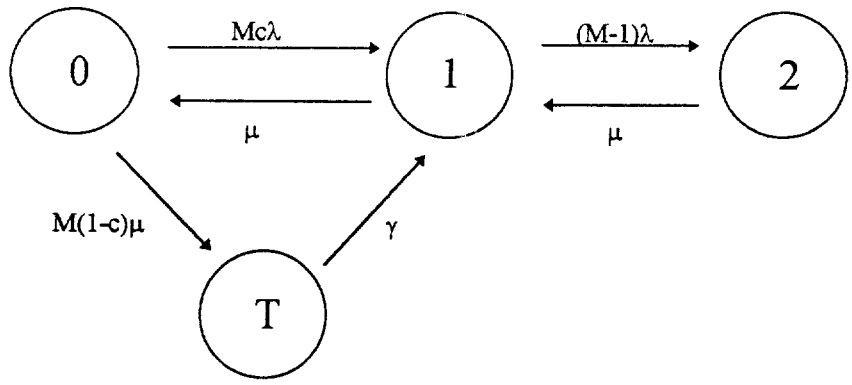



Fig. 7 Markov Chain Availability Model with Imperfect Coverage

State 2 of Fig. 5 is split into state 2 and T. Then, a repair transition (with rate $\gamma$) from the coverage failure state T to state 1 and a repair transition from exhaustion failure state 0 to state 1 are added.

5) Cost-Effectiveness

Compared to the unique path network, the multiple path networks certainly have high reliability, but with increased hardware complexity, which not only increase cost but also give some wrinkle on the claim of enhanced reliability.

To estimate the cost of a network, one common method is to calculate the switch complexity with an assumption that the cost of a switch is proportional to the number of gates involved, which is roughly proportional to the number of cross points within a switch. For example, a 2x2 switch has 4 units of hardware cost, whereas 4x4 switch has 16 units. For a kx1 multiplexer or 1xk de-multiplexer, we roughly assume that each has a k units of cost. Fol-

lowing Table 1 shows cost functions of networks with 2x2 switches.

Table 1 Cost Functions of Networks with 2x2 Switches

| Name | Cost Function |
|---|---|
| Omega | $2\ N\ \log_2 N$ |
| ESC | $2\ N\ (\log_2 N + 3)$ |
| 3-Rep | $6N\ \log_2 N$ |
| INDRA | $4\ N\ (\log_2 N + 1)$ |
| EGN-m | $2(N+N/M)(\log_2(N/M)+m)$ |

( N : network size)

Now a simple measure of the cost effectiveness for reliability can be given by comparing MTTFs and the costs of these networks. The cost-effectiveness of fault tolerant MIN can defined as the ratio of its MTTF to its cost.

## 2. 2 Reliability Modeling

The analysis and modeling of MIN reliability are summarized in the following Table 2 according to systems under consideration discussed in section 2.1. Under each system following factors are taken into consideration.

- reliability definition
- fault tolerance measure
- fault mode
- modeling method
- repair / replacement policy

Table 2 Reliability Modeling of MIN

| System type | Description and characteristics | Reliability definition (network reliability) | Fault tolerant measure | Failure mode |
|---|---|---|---|---|
| Unique path MIN | • simple and distributed Banyan type networks that include OMEGA, and DELTA networks | • the probability that all the switching elements in the MIN are not faulty | • reliability function, MTTF | • switch fault |
| Space redundancy (multiple path MIN) | • several fault tolerant methods (extra stages of switches, extra switches in each stage,extra link in each switch, etc.) proposed to provide alternate paths in the presence of faults | • the probability that number of component failure is less than the maximal number of fault tolerance the system can guarantee | • the maximal number of fault tolerance<br>• reliability function, MTTF<br>• cost-effectiveness measure (MTTF/cost) | • switch fault |
| Time redundancy (multiple pass MIN) | • dynamic full access property of multiple pass MIN is investigated | • the probability that dynamic full access exists | • reliability function, MTTF | • switch fault<br>• loop fault |
| Reconfigur-able network | • to give unique path MIN maximum fault tolerance and reliability, it focuses on the replacement design and reconfigurable link assignment | • the probability that number of component failure is less than the maximum number of fault tolerance the reconfigurable network can provide | • the maximal fault tolerance and number of links required for a node of switch<br>• reliability function, MTTF | • switch fault |

Table 2 (Continued)

| Modeling method | Repair / replacement | Reference |
|---|---|---|
| $R(t)=[R_{SE}(t)]^N$, $R_{SE}(t)$ : reliability function of each switching element<br>N : number of switching element in MIN | switch repair | 53 |
| ① continuons time Markov chain : It can give the exact reliability function, but the number of states grows enormously as the MIN size increases. Therefore, this method might not be suitable for practical use. In some cases,state lumping technique is used.<br>② reliability block diagram : Based on the number of fault tolerance and MIN structure, both the reliability upper and lower bound can be approximated from seires-parallel block diagram. It can give a first insight of MIN reliability with little effort.<br>③ connection matrix : First enumerate the requisite connections of MIN by using the concept of inter-stage and intra-stage connectivities. Then use the exclusive-operator to obtain disjoint terms and hence to generate the reliabillity expression.<br>④ graph theoretic : Based on the bipartite graph of MIN and edge failure mode, reliability function is derived.<br>⑤ combinatorial approach : The upper and lower bound of reliability polynomial coefficient $C_k$ are estimated. The $k^{th}$ term in the polynomial is the probability of an operational network with exactly k faulty switches and $C_k$ is the number of ways in which k switches in the subnets may fail without destroying full access | • switch repair (imperfect / perfect coverage)<br>• several availability functions are determind according to the switch repair policy | 40, 42, 47, 48, 53, 54, 55, 56, 61, 62 |
| • Using set and graph theoretic approach, the necessary and/or sufficient conditions for DFA existence are derived<br>-Based on these conditions, the upper and/or lower bound of reliability are abtained<br>-Comparisons are made with simulation result. | • switch repair policy<br>• module repair policy | 38, 60 |
| • Using graph theorectic approach, the maximum fault tolerance and its associated link requirement are derived.<br>-Based on these derivations, the lower bound of reliability is calculated | | 58, 59 |

# Ⅲ. MIN Performance Modeling under the Presence of Failure

Several important factors and assumptions affecting the MIN performance are discussed in section 3. 1. Based on this discussion system performance models considering faulty componts are analyzed in section 3. 2. In section 3. 3, the performability models are discussed, which show ways to combine both reliability and performance measures.

## 3. 1 Issues in MIN System Performance Modeling

### 1) System Under Consideration

① unique and multiple path MIN

Several architectural designs for fast packet switches have emerged in recent years. These can be classified into the following three categories.

- space-division type
- shard-memory type
- shard-medium type

Space-division type, which is considered to be most promising candidate for ATM switching fabric, can be further subdivided into

- Banyan type
- buffered Banyan based fabrics
- Batcher (or sort) Banyan based fabrics
- switch fabrics with disjoint path topology
- cross-bar based fabrics

Each fabric was developed to reduce the blocking problem and to achieve high throughput at lower hardware cost and complexity. Extensive study has been taken to evaluate the performance of each switch fabric mentioned above. Table 3 summarizes the key issues in performance analysis of each system.

In this study, we just focus on the Banyan type switches of unique and multiple path MIN's with faulty components. The definitions of unique and multiple path MIN's are given in section 2. 1. The mulitiple path MIN can be thought as the space redundancy fault tolerant network : the performance analysis of the other systems under the faulty components is very complicated tasks and has not been taken yet. It will be a challenging and mportant future research topic.

Table 3 Key Issues in Performance Analysis of Space Division Type Switches

| Switch type | Proposed system architecture | Issues in performance analysis | References |
|---|---|---|---|
| Banyan type | • unique path.<br>• multiple path | fault tolerant structure | 6, 20, 21, 22 |
| Buffered Banyan type | • buffers at each switching element<br>-input<br>-output<br>-shard | • buffer location and size<br>• speed up factor for output buffer<br>• flow control mechanism | 1, 2, 34, 5, 7, 19, 23, 24, 25 34, 35, 36, 37 |
| Non-internal blocking type(Batcher-Banyan, cross-bar) | buffer at each I/O of network<br>-input<br>-output<br>-input/output | • buffer location and size<br>• speed up problem for output queue<br>• flow control mechanism | 26, 28, 29, 30 33 |
| Switch fabric with disjoint path topology | • knockout switch<br>• MIN based on knockout switch | • concentration ratio<br>• output(shard) buffer size | 27, 31 |

② synchronous vs 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. 8(sec. Thus, the switch fabric has to be designed in such a way so that it can handle approximately 350,000 packets per second per input port.

• asynchronous system : Packets can arrive in the input trunk according to some independent stochastc process at a given rate per unit of time.

③ flow control

Several types of flow control mechanism are suggested to regulate the flow of packets between stage to prevent packet loss due to blocking or buffer overflow. These can be categorized as shown in the following Table 4.

Table 4 Flow Control Mechanism

| block and lost (or queue loss) | | |
|---|---|---|
| backpressure | global | ack |
| | | grant |
| | local | ack |
| | | grant |

Depending upon the flow control mechanism employed in MIN, different performance models can be constructed. Since we do not consider any internal buffer in Banyan type switch fabric with unique or multiple paths, there is no packet loss due to buffer overflow. In our study we can assume "block and lost" flow control mechanism. Also, for the block call to be resubmitted in the next cycle, some kind of backpressure with "ack" can be assumed.

④ clock cycle

In synchronous system clock cycle is very much related to the flow control mechanism. For the "block and lost" protocol, any request for output access is issued at the beginning of a clock cycle and all successsful connections terminate at the end of the cycle. The length of a clock cycle is equal to the output access time plus the network delay. For the backpressure model each cycle has two phases basically : the control signal and packet movement phase. The constituting components and size of each phase vary acccording to the various flow control mechanism used in the backpressure model.

2) General Assumptions

① steady state condition

For analysis purpose it is assumed that the mean time between component failures is very large compared to the length of clock cycle, implying that once a component fails, the next failure will occur after a very long period of time. This enables us to study the system's behavior under a statistical steady-state, resulting from its having a fixed combination of faulty and non-faulty components for a relatively long time.

② failure mode

As in the failure mode of reliability modeling link, switch, and loop failure modes are assumed. Moreover, input and output links of a network are considered to be non-decomposable

components. Therefore, for example, any fault in a input link results in the removal of that input from the system.

③ blocked call

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

• random model : In next trial, it selects the next switch at random. Since it allows a blocked packet to be routed around a congested switch, the results obtained from this model might be too optimistic.

• memorized model : A blocked packet always hunts for the same switch during consequtive cycles, and does not obey static routing probability.

The "blocked call" problem imposes a lot of difficulties in analyzing the performance of the MIN. For example,in case of buffered Banyan network, it introduces the interdependency between the interstage buffer state since buffered packets are correlated as a result of having contended for output. For our study we just consider two cases : ignored blocked call and random model. Also, it is assumed that the contention in switch is resolved randomly. In some cases such as switches with intrastage link, it can be assumed that requests from intrastage are with lower priority than interstage request. For the analysis of memorized model, additional state called "blocked state" is required. Under the faulty component assumption, it can be a topic of future study.

3) Traffic Assumptions

Incoming traffic has been modeled as a independent Poisson process or Bernoulli process, which will be called homogeneous input traffic. However, to represent traffic patterns of ATM network where a wide range of bandwidths need to be accomodated some bursty traffic model is required. Also, the output 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 characteristic can be classified into 4 types as shown in the following Table 5.

Table 5 Offered Traffic Characteristic

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

In this study, homogeneous-balanced traffic is assumed. Considering that component failures make the traffic unbalanced, however, the other traffic assumptions need to be adapted to see the failure effect on the network performance more realistically. The work under bursty and/or unbalanced traffic considering component failure has not been performed yet. It will be a good topic of future study. For a homogeneous-balanced traffic, it is assumed that

- Each input i generates random and independent request in a given cycle according to Poisson process with rate (i or Bernoulli process with probability Pi.
- Packets arriving at input link at stage 1 are destined uniformly for all output link at stage n

The homogeneous-balanced traffic and "block and lost" (or random block call) assumptions make it easy to analyze the switch network. That is, these assumptions lead to

- Independence between interstage switch element.
- All the switching elements in a stage are identical.

Therefore, the state of a system can be characterized by that of a switching element in a stage.

4) Performance Measure

Some performance measures of the MIN are observed during each network cycle. Based on the above discussions, the performance affecting factors can be summarized as shown in Table 6. Under the some assumptions and specifications related to factors mentioned in Table 6, several performance measures have been proposed.

① largest realizable system : It is defined as maximum number of inputs and outputs that can be completely connected when the network is in some fault state. This measure is just good for unique path system, but not for fault tolerant system. The multiple path MIN can still maintain full access property under some faulty components.

② bandwidth : It is defined as the total number of requests that can be routed through the network in a cycle. The bandwidth measures the effect of blocking which results from either the internal blocking, the output contention, or the presence of some faulty components. It is a good measure which can estimate the average performance of the entire network under the symmetric traffic assumption. However, in some cases this measure is not adequate. When a fault occurs, the symmetry of traffic passing through the network is lost. Or as in case of multiservice environment of ISDN, the incoming traffic can be totally unsymmetric. So different network locations may experience significantly different

Table 6 Performance Affecting Factors

| | |
|---|---|
| Failure mode | -number of faulty switches or links<br>-fault tolerant scheme |
| Flow control | -block and lost<br>-back pressure (global/ack, global/grant, local/ack, local/grant) |
| Incoming traffic | -stochastic nature of incoming traffic request (homogeneous or bursty)<br>-input request destination distribution (uniform or non-uniform)<br>-interdependency of input traffic request within a cycle<br>-interdependency of input traffic request between cycles (blocked call problem)<br>-contention resolving policy (priority) |
| Blocked call | -ignored<br>-random retry<br>-memorized retry |
| Path selection algorithm | In case of multiple path<br>-random<br>-preferred |

loads. Because the loads in different parts of the network can vary greatly, one can't simply average performance over each input in the network. For example, a telecommunication network must provide uniformly high quality service. Even when the average service quality is high, an individual customer who receives degraded service may be dissatisfied.

③ performance of individual input and output : To remedy the problem of bandwidth measure, we predict the probability that a particular memory is referred in a cycle. That is, when faulty components are present or incoming traffic is unsymmetric, we can estimate the output reference probabilities individually. And it can be shown that these probabilities vary widely from output to output.

④ normalized 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

⑤ blocking probability : If the blocked call is lost, this is defined as the probability that

the packet is lost due to internal blocking, output port contention, or full buffer in case of buffered switch. In case of backpressure with some ack, this can be defined as the probability that the IBC buffer is full.

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

⑦ maximum throughput : maximum allowable throughput at an output port which has the worst congestion among all the output ports.

⑧ maximum delay : delay of the input to output path which has largest delay of all the paths

⑨ maximum blocking probability : blocking probability of a input link which has the largest blocking probability of all input links

### 3. 2 Performance Modeling under the Presence of Failure

To derive the performance measures mentioned above, several performance models are proposed with the proper assumptions and specific MIN structure. In the research [78] previous to this study extensive investigation was taken on the performance modeling of the fast packet switch fabrics, which does not consider component failure. Each model was analyzed according to

- type of switch fabric
- modeling method
- assumptions related to the buffer state
- switching element type and size
- flow control and clock cycle
- buffer location and size
- blocked call
- incoming traffic assumption
- measure

In this study our attention is limited to Banyan type switch with unique or multiple paths considering the faulty component in the switch. Compared to the performance modeling with "no fautly component", very few works have been done to see the effect of faulty component on performance and to combine the measure of reliability and system performance. Table 7 summarizes the modeling under two general categories : unique MIN and multiple path MIN.

Table 7 Failure Dependent Performance Modeling

| Switch network type | System characteristics | Incoming traffic | Flow control | Blocked call | Performance measure |
|---|---|---|---|---|---|
| Unique path MIN | • With single faults it loses the full access property. Its main concern is to calculate the maximum number of connected I/O pair and average network performance such as bandwidth under some faulty condition. | • independent homogeneous/ uniform<br>• random contention resolving policy | • block and lost<br>• back-pressure with ack | • lost<br>• randomly resubmit in next cycle | • maximum realizable system interconnection<br>• average bandwidth |
| Multiple path MIN | • It can maintain full access property even with some faulty components. Since the faulty components make the network load unsymmetric, the individual component performance vary widely. | • independent homogeneous/ uniform<br>• random contention resolving policy (In case of fault tolerance using intrastage loop, the priority scheme is required between interstage and intrastage input request)<br>• some path selection assumption: random or preferred | • block and lost | • lost | • performance of individual input and output link<br>• average bandwidth |

Table 7 (Continued)

| Fault mode | Modeling method | References | Discussion |
|---|---|---|---|
| • link fault mode<br>• I/O component is also subject to failure | • Calculation of exact number of largest realizable system is exponential. Using elementary set theory, simple approximation is used.<br>• To derive average bandwidth, Markov chain is used. It is expressed as an input request probability, link and I/O component failure probabilities. | 49<br>50<br>51 | • These measure already combine the component reliability and system performance measure in themselves. |
| • switch and loop fault mode | • It considers the traffic at every link in the network individually. Analyzing the entire network, stage by stage, gives the busy probabilities for all links in the network, and for the destination.<br>• Based on the combinatorial analysis of the MIN, two steps of analysis are performed : single switch and intrastage link. Intrastage link analysis invloves the probability calculation of contention in a switch and intrastage request propagation. | 52 | • To derive some performability measure, Markov reward model is used. |

### 3. 3 Performability Model

To determine how faults degrade the overall MIN performance, we need to combine the results from the reliability model with the results from the MIN performance model. Generally two approaches have been suggested : Markov reward model and combinatorial model.

#### 1) Markov Reward Model

The evolution of a degradable system through various configurations with different sets of operational components can be represented by a discrete-state, continuous-time Markov chain, $\{z(t),\ t \geq 0\}$. Let $r_i$ denote the reward rate associated with state i, for example, network bandwidth at failure state i. The reward rate of the system at time t is given by the process. $x(t)=r_{z(t)}$. The expected reward rate at time t is

$$E[x(t)]=\sum r_i \cdot P_r[Z(t)=i]$$

If we let Y(t) be the amount of reward accumlated by a system during the interval (0,t), then

$$Y(t)=\int_0^t x(u)du$$

Then, the expected accumulated reward is

$$E[Y(t)]=\sum r_i \cdot \int_0^t P_r[Z(u)=i]du$$

The construction and solution of an overall Markov model for Pr [Z(t)=i] become intractable for fault-tolerant MINs of size greater than 16×16. State lumping reduces the state space of the continuous time Markov chain. To extend the conditions for lumpability to Markov reward model, it is required that every pair of state u and v in the "lumped state" must have indentical reward rate (i. e., $r_u=r_v$)

2) Combinatorial Methods

The critical problem in the Markov reward model is that the number of states in Markov model of realistic system tends to be extremely large, creating difficulties in generation, storage, and solution of such models. Combinatorial reliability model takes advantage of the system structure and avoids generation and solution of the underlying Markov model. The limitation of this combinatorial model is due to inherent assumptions of stochastic independence, two-state behavior of components, and restrictive repair. Trivedi et. al. [41] proposed a combinatorial algorithm for the combined performance and reliability analysis of coherent repairable system with multistate components. To model multistate system with multistate components, this algorithm transforms the system to a binary one with binary components. This transformation is done by assigning a binary variable to each state of each multistate component and for each state of the system. This transformation introduces dependencies among the binary variable. Then the combinatorial binary network model is converted to "Event Based Reliability Model" to account for stochastic dependencies between components, if any. Now Satyanarayana and Prabhakar [77] algorithm is used to obtain the probability of the system being in state k at time t. From these system state probabilities, several performability measure can be computed as needed.

# Ⅳ. Future Works

The reliability and performance analysis of MIN are still in infant stage, and require much more refinement and enhancement in following areas.

### 1) MIN Architecture

Until now, the performability analysis is limited to the area of pure Banyan type network. It needs to be extended to

- tandem and parallel Banyan
- buffered Banyan
- Batcher-Banyan with I/O buffer
- switch fabrics with disjoint path topology

### 2) Traffic Characteristics

In most cases, traffic is assumed to be homogeneous and balanced. When fault occurs, however, the symmetry of traffic passing through the network is lost. Or, as in case of multiservices environment of ISDN, the incoming traffic can be totally unsymmetric. Therefore, to see the effect of faulty components on system performance some realistic traffic assumptions are required such as

- bursty type traffic
- unbalanced traffic

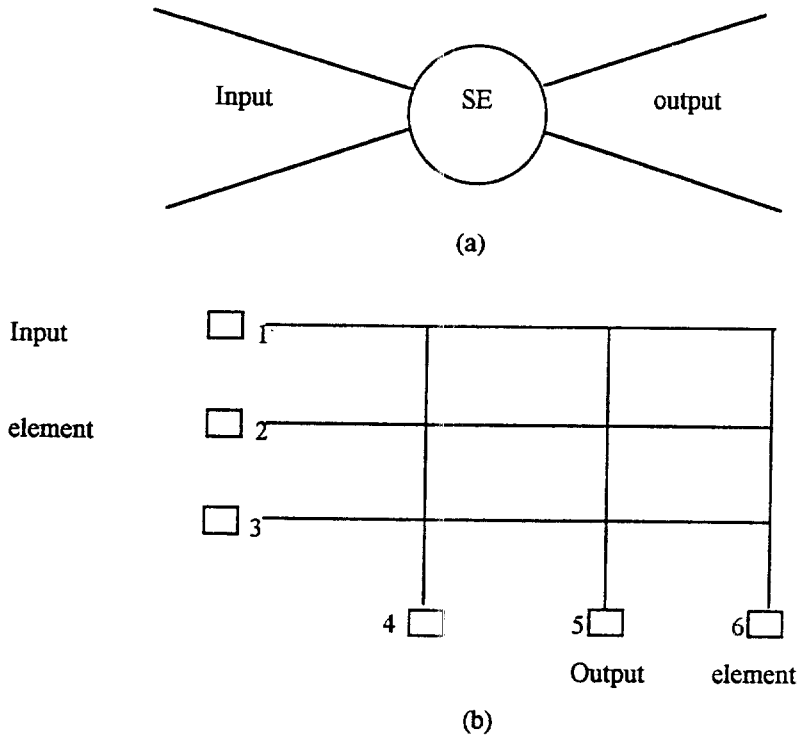
### 3) Flow control

For flow control "block and lost" is generally assumed. To model the system realistically backpressure models with specific protocol need to be considered such as global/ack, global/grant, local/ack, or local/grant. The memory model for blocked call also needs to be considered to enhance the modeling accuracy.

### 4) Fault Model for Switching Element(SE)

For the switch fault model, the two-state fault model is assumed that an SE is in either up

or down. It is represented as a single node as shown is Figure 8. a. A failure in the node causes all of its links to be deleted from the corresponding MIN. In some cases, however, SE is not looked upon as a single node, but as several nodes along with their connection links (intrastage link, input and output). Figure 8. b represents the connection pattern. The reliability of SE may be considered as an independent failure/success entity. An SE having one of its elements failed still works in degraded mode, thus an SE behaves like a multistate component.




8 Node Representation for
a. Two-State SE
b. Multi-State SE

5) Dependent Failure

Algorithms for the computation of reliability generally assume that component failures are

statistically independent. For most practical applications, this assumption is not valid. Spragins [43] outlined examples of the strongly co-related failures of different communication lines at a single geographical location. Several models have been proposed to introduce the dependencies in communication network such as q-$\psi$ model [45, 46], $\varepsilon$-model [44], CNM[45], and EBRM[46]. Based on these models, analysis of failure dependency in MIN's can be considered.

# References


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

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

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

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

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

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

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

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

[13] Eng, K. Y., M. G. Hluchyj, Y. S. Yeh, "Multicast and broadcast services in a knockout

packet switch," IEEE INFOCOM, 1988, pp 29-34

[14] Nojima, S., 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] Killat, U., "Asynchrone Zeitvielfachübermittlung für Breitbandnetze", Nachrichtentech. Z., Vol. 40, No. 8, 1987, pp 572-577

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

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

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

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

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

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

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

[23] Zeng, T., "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] Kim, H. S., 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] Bianchi, G., J. S. Turner, "Improved queueing analyis of share buffered switching networks," IEEE INFOCOM, 1993, pp 1392-1399

[26] Iliadis, I., 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] Yoon, H., et al., "The knockout switch under nonuniform traffic", IEEE Transaction on Comm., Vol. 43, No. 6, 1995 June, pp2149-2156

[28] Karol, M., M. G. Hluchyj, S. P. Morgan, "Input versus output queueing on a space-division packet switch", IEEE Transaction on Comm., Vol. 35, 1987 Dec., pp 1347-1356

[29] Hui, J. Y., 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] Chang, C. Y., A. J. Paulraj, T. KaiLath, "A broadband packet switch architecture with

input and output queueing", GLOBECOM, 1994, pp 448-452

[31] Kim, Y. M., 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., pp 2391-2398

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

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

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

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

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

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

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

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

[42] Botting, C., et al., "Reliability computation of multistage interconnection networks", 1989 Apr., pp 138-145

[43] Spragins, J., J. Assiri, "Communication network reliability calculations with dependent failures", Proc. National Telecomm. Conf., 1980. pp 25. 2. 1-25. 2. 5

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

[45] Lam, Y., V. Li, "On reliability calculations of network with dependent failure", GLOBECOM,

1983, pp 1499-1503

[46] Lam, Y., V. Li, "Reliability modeling and analysis of communication networks with dependent failure", IEEE Transactions on Comm., Vol. 34, 1986 Jan, pp82-84

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

[48] Cherkassky, V., 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] Koren, I., Z. Koren, "An alyzing the connectivity and bandwidth of multi-processors with multi-stage interconnection networks", in Concurrent Computations : Algorithms, Architecture, and Technology, NY-Plenum, 1988, pp 525-540

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

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

[54] Blake, J. T., 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] Blake, J. T., 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] Bansal, P. K., 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] Rai, S., 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] Yang, S. C., J. A. Silvester, "Reconfiguable fault tolerant networks for fast packet switching", IEEE Transtions on Reliability, Vol. 4, 1991 Oct., pp 474-486

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

[60] Shen, J. P., 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] Menezes, B. L., 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] Menezes, B. L., 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] Adams, G. B., 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] McMiller, R. J., 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] Siegel, H. J., R. J. McMillen, "Dynamic rerouting tag schemes for the augmented data manipulator network", Proc. 8th International Symposium Computer Architecture, 1981 May, pp 505-516

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

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

[70] Padmanabhan, K., D. H. Lewis, "Fault-tolerant schemes in shuffle exchange type interconnection network", Proc. International Conf. Parallel Processing, 1983 Aug., pp 71-75

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

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

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

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

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

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

[78] Lee, K. W., "Performance Analysis of MIN Switching System", Journal of the Korean Institute of Industrial Engineer, Vol. 22, No. 2, 1996 June, pp 277~302