PERFORMANCE, FAULT TOLERANCE AND RELIABILITY OF MULTISTAGE INTERCONNECTION NETWORKS FOR BROADBAND PACKET SWITCH ARCHITECTURES

SIVAKUMAR HARINATH

Performance, Fault Tolerance and Reliability of Multistage Interconnection Networks for Broadband Packet Switch Architectures

by

© Harinath Sivakumar, B. E.

A thesis submitted to the School of Graduate
Studies in partial fulfillment of the
requirements for the degree of
Master of Engineering

Faculty of Engineering and Applied Science
Memorial University of Newfoundland
December 1995

St. John's Newfoundland Canada

ISBN 0-612-17599-5

Canada

Abstract

Multistage Interconnection Networks (MINs) are being considered for use in switch fabrics of broadband packet switch architectures. Most of the MINs considered are based on 2×2 switching elements (SEs). Throughput performance is degraded mainly due to the basic size of the SEs. In this work, the Balanced Gamma (BG) network, a multipath MIN, which features 4×4 SEs has been studied in detail. Architecture, fault tolerance, reliability and hardware complexity of the BG network are compared with the 2-replicated 2-dilated Banyan (R2D2) networks and the Batcher Banyan (BB) networks. A new simple routing algorithm has been proposed for the BG network to enhance its fault tolerance capability. Switching performance is enhanced by increasing the size of the basic SEs and thereby providing multiple paths between each input-output pair of the BG network. The throughput performance of the BG network is studied under certain idealistic uniform and non-uniform traffic conditions. These performance results are compared with the R2D2 and the BB networks under the same traffic conditions. Performance analysis and simulation results have shown high throughput performance of the BG network even in the presence of an SE fault. It has been found that the fault tolerance properties, reliability and throughput performance of the BG network are much superior to those of the R2D2 and the BB networks. Due to increased throughput performance of the BG network it may be considered as a potential candidate for use in switch fabrics of broadband packet switch architectures.

Acknowledgements

I would like to express my deep gratitude and thanks to my supervisor Dr. R. Venkatesan for his invaluable guidance, useful discussions, criticisms and constant encouragement, and help in preparing this manuscript. I sincerely thank Dr. R. Venkatesan, the Faculty of Engineering and Applied Science and the School of Graduate Studies for the financial support provided to me during my M. Eng. program.

I would like to thank Dr. C. A. Sharpe and Dr. J. Malpas, former and present Dean of the School of Graduate Studies, Dr. R. Seshadri, Dean of the Faculty of Engineering and Applied Science and to Dr. J. J. Sharp, Associate Dean for Graduate Studies for their support during my study in Canada.

I sincerely acknowledge Ms. J. Withers, a graduate of Electrical Engineering, Faculty of Engineering, Memorial University of Newfoundland, for explaining the simulation program developed by her during her final year project. Special thanks go to Mr. J. Swamidas, fellow graduate student, for useful discussions on the simulation program. I would also wish to express my thanks to Dr. W. Raman-Nair, Dr. L. Lye and Dr. C. R. Moloney. I would like to thank all my fellow graduate students for their friendship and constant encouragements.

I would also like to acknowledge fellow graduate students Mr. S. P. Mangalaramanan, Mr. P. Mehrotra and Ms. K. Subramaniam for their help in preparing this manuscript. Finally, I would like to express my profound gratitude to my parents for their constant encouragement, understanding and support during the course of my study.

Contents

List of Figures

List of Tables

List of Abbreviations

ACN	: Augmented C-Network
ATM	: Asynchronous Transfer Mode
BG	: Balanced Gamma
BB	: Batcher Banyan
B-ISDN	: Broadband Integrated Services Digital Network
BP	: Broadcast Path
BR	: Broadcast Reliability
ESC	: Extra Stage Cube
FT	: Fault Tolerance
IU	: Input Upper
IUM	: Input Upper Middle
ILM	: Input Lower Middle
MDN	: Merged Delta Network
MTTF	: Mean Time to Failure
MIN	: Multistage Interconnection Network
NP	: Network Path
NR	: Network Reliability
OU	: Output Upper

OLM	: Output Lower Middle
OUM	: Output Upper Middle
HC	: Hardware Complexity
R2D2	: 2-Replicated 2-Dilated
SE	: Switching Element
TP	: Terminal Path
TR	: Terminal Reliability
VLSI	: Very Large Scale Integrated Circuits

List of Symbols

n	: Number of stages in a MIN.
p	: Reliability of a switching element of a MIN.
$BR(t)$	: Broadcast reliability of a network after time t.
F_i	: Total possible combinations of failure of i SEs causing the BG network lose full access property.
$HC(i)$	: Hardware Complexity of network i.
INF_i	: Different possible combinations of i SE failures in the intermediate stages of the BG network which do not make the BG network lose full access.
LNF_i	: Different possible combinations of i SE failures in the last stage of the BG network which do not make the BG network lose full access.
$LSR(t)$	: Reliability of the last stage of a BG network.
$MTTF_N$	: Network MTTF of a network.
$MTTF_{N-LOW}$	: Lower bound network MTTF for the BG network.
$MTTF_T$	: Input-output MTTF of a network.
N	: Size of a MIN/ number of inputs/outputs of a MIN.
$NR(t)$	: Network reliability of a network after time t.
$NR_{LOW}(t)$	: Network reliability lower bound for a BG network after time t.
$NR_1(t)$	: Network reliability of the BG network after time t.

$NR_{UP}(t)$	: Network reliability upper bound for a BG network after time t.
$SE_{i,j}$	: Switching element i in jth stage of a MIN.
$SR(t)$	: Reliability of any intermediate stage in the BG network.
TN	: Total SEs of the BG network excluding the SEs of the first stage.
$TR(t)$	: Terminal reliability of a network after time t.

Chapter 1

Introduction and Literature Review

1.1 Introduction

Rapid advancements in computer and communication technologies have resulted in a virtual merger of these two fields. The advances in communication technology are in the areas of transmission and switching devices while those in the field of computers can be attributed to high speeds, multiple processors, developments in software, effective interconnection of these processors etc. There has also been rapid developments in the field of Very Large Scale Integrated Circuits (VLSI). Due to increasing needs from users and greater demand of various services such as voice, full motion video, data etc., researchers started looking into one transmission medium which could effectively handle all these services. ITU-T (originally called CCITT) has defined Broadband Integrated Services Digital Network (B-ISDN) as "a service requiring transmission channels capable of supporting rates greater than the primary rate offered by the existing circuit and packet switched networks" [1]. This integrated network will be capable of supplying services such as voice, data, full-motion video etc. individually or combined as in the case of multi-media to

various places at varying speeds. This is illustrated in Figure 1.1 which shows a B-ISDN network carrying different payloads. Asynchronous Transfer Mode (ATM) has been identified by ITU-T as a switching system capable of meeting the requirements of B-ISDN such as very high throughput, low switching delay, a low probability of packet loss, expandability, testability, fault tolerance (FT), low cost, and ability to achieve broadcasting as well as multicasting.

1.2 Asynchronous Transfer Mode

ATM is a packet-oriented transfer mode that uses asynchronous time division multiplexing techniques with the multiplexed information flow being organized into blocks of fixed size, called cells [3]-[6]. ATM is also known by names such as Asynchronous Time Division and Fast Packet Switching [4]. The packets in the ATM networks are called cells. Each cell consists of 5 bytes of header and 48 bytes of information field. This is shown in Figure 1.2. The information field is transported transparently by an ATM network without any processing. Cell sequencing is preserved in an ATM network.

Virtual circuits are established between the users before information is exchanged in an ATM network. This is done by a connection set-up procedure. Similar to other switching systems, B-ISDN protocol reference model for ATM consists of several layers. This model has been explained in detail in references [4]-[6]. An ATM public switched network is shown in Figure 1.3. It can be seen that the public ATM network connects the local ATM networks. These local ATM networks are in turn are made up of a number of ATM switches.

ATM switching networks are currently expected to operate at 155.52 or 622.08

Mega bits per second [1]. Future networks are expected to operate at 1 Giga bits per second or more. Research is being carried out in order to find an ATM switch having very high throughput, low cell loss ratio of the order of 10^{-6} and lower.

1.3 Asynchronous Transfer Mode Switch Fabrics

Various ATM switches architectures have been proposed [4], [7]. Some of these broadband packet switch architectures include the Knockout, the Roxanne, the Coprin, the Athena, the St. Louis, the Starlite, the Moonshine, Turner's ISPN switches and the Banyan based switches. These switches can be classified into three categories: the shared medium, the shared memory and the space division architectures [8]. In this work, the ATM switch fabric which uses space division architecture is considered.

An ATM switch consists of N input ports and N output ports. Each input port is connected to every output port by an interconnection network. Cells arriving at the input ports are switched to the requested output port by the interconnection network. This is shown in Figure 1.4.

There are several interconnection networks reported in references [9]and [10]. Of all these interconnection networks, the crossbar type is the simplest one. Figure 1.5 shows a crossbar interconnection network connecting input links to the output links. There are several factors involved during the evaluation of the cost of an ATM network [11], [12]. The cost of an ATM switch fabric and hence the cost of an ATM network is mainly dependent on the crosspoint complexity of the ATM switch fabric. The cost of a switch fabric is directly dependent on its crosspoint complexity [11].

The crosspoint complexity of an $N \times N$ crossbar is N^2 [8]. Due to the fact that the crosspoint complexity of a crossbar switch is greater, multistage interconnection networks (MINs) are being considered for use in ATM switch fabrics. The crosspoint complexity of an $N \times N$ MIN is of the order $N \times \log_2 N$. Furthermore multi-packet acceptance is made possible at the output ports in certain MINs to increase their throughput performances.

1.4 Multistage Interconnection Networks

Most of the ATM switch fabrics are composed of a large number of identical basic building blocks or switching elements (SEs) which form a MIN. Complex SEs have been used only in certain cases such as the Athena and the Roxanne switches. Most of the MINs employ very simple self-routing SEs. Most of the MINs considered for use in ATM switch fabrics employ Banyan networks [13] and enhanced versions of the Banyan networks [14]-[17]. The Banyan network is a MIN consisting of $\log_2 N$ stages of 2×2 SEs, with $\frac{N}{2}$ SEs in each stage and follow destination tag algorithm for routing the packets. In the destination tag algorithm, the binary representation of the destination of a packet is used to route the packet through the network. If the routing bit for a particular stage is 0, then the packet is routed through the upward link. Otherwise, the packet is routed through the downward link of the SE. The Banyan networks are unipath MINs since there exists exactly one path between an input port and an output port. The Banyan networks suffer throughput limitations because of the low throughput of the 2×2 SE. Due to this reason, enhanced versions of the Banyan networks are being considered for use in the ATM switch fabrics. These MINs along with certain other MINs which are used

in connecting multiprocessor systems are explained in the following sections.

1.4.1 Batcher Banyan Network

The main drawback of the Banyan networks is that they are internally blocking in the sense that two packets destined for two different outputs may collide in one of the intermediate nodes. It has been shown that packets will not block within the Banyan networks if the incoming packets are compacted and their destination addresses are monotonic and non-repeated [11]. This is the basic idea behind the BB network.

The BB network consists of a Batcher network followed by the actual Banyan network. In an $N \times N$ BB network, the Batcher network is based on bitonic sorting elements, which are arranged in $\log_2 N \times \frac{(\log_2 N - 1)}{2}$ stages with $\frac{N}{2}$ such elements per stage. An 8 × 8 BB network is shown in Figure 1.5. The total number of SEs in a BB network is equal to $\frac{N}{4} \times ((\log_2 N)^2 + \log_2 N)$. A destination tag algorithm is employed to route the packets within the Banyan networks.

1.4.2 2-Replicated 2-Dilated Banyan Network

The R2D2 networks are formed by dilation of each link in the Banyan network followed by a replication of this dilated Banyan network. The duplicated links are connected to the same SEs as those of the original links. The first stage SEs of the 2-replicated 2-dilated Banyan networks are connected to multiplexers. The two 2-dilated Banyan networks are called as subnetworks A and B in this thesis. An SE in an R2D2 network is referred as $SE_{i,j,k}$ where i indicates the layer number, j indicates the stage number and k indicates the row number. An 8 × 8 R2D2 network is shown in Figure 1.6.

The incoming packets are fed to the multiplexers. The multiplexers then route the packets to their respective destination through one of the fault-free 2-dilated Banyan network. Faults in either of the 2-dilated Banyan networks are notified to the multiplexers. If a fault is located in one of the subnetworks, then the multiplexers do not send any packets to that subnetwork. All packets are routed through the fault-free network. There are exactly two paths between each input and output port in the R2D2 network.

1.4.3 Extra Stage Cube Network

The extra stage cube (ESC) network [18], [19] is derived from the generalized cube MIN [20] by adding an extra stage to the input side of the network along with multiplexers and demultiplexers at the input and output stages, respectively as shown in Figure 1.7. The last stage and the first stage have a similar connection pattern. Each of these stages can be enabled or disabled by the use of the available multiplexers and demultiplexers.

Normally, the network will be set so that stage n is disabled and stage 0 is enabled. If a fault is found after running fault detection and location tests, the network is reconfigured. A fault in a stage n switch requires no change in network configuration; stage n remains disabled. If the fault occurs at stage 0, then stage n is enabled and stage 0 is disabled. For a fault in a link or in a switch in stages $n-1$ to 1, both stages n and 0 are enabled.

1.4.4 Augmented C-Network

The Augmented C-network (ACN) [21] derives from the C-network, an $N \times N$ MIN having an arbitrary number of stages of $\frac{N}{2}$ 2×2 crossbar switches each.

Stages are numbered 0 to $n-1$ from input to output. C-networks are networks which satisfy the property that the SEs in every stage, except the last one, can be grouped into pairs such that each pair is connected to a common pair of SEs in the next stage. Such SEs are called as conjugates.

The ACN provides 2^n distinct paths between any source and destination; but most of these paths are not disjoint. Routing in the ACN is predicated on a routing tag scheme existing for the particular base C-network. When there are no faults or when certain SEs are busy, the tag is determined and interpreted by the ACN in the same manner as it would have been done in the base C-network. Otherwise, the two proposed routing strategies utilize both the standard path and conjugate path switches [21]. An 8×8 ACN constructed from an Omega network [22] of the same size is shown in Figure 1.8.

1.4.5 Merged Delta Network

An $N \times N$ merged delta network (MDN) results from cross-linking the corresponding stages of C copies of an $\frac{N}{C} \times \frac{N}{C}$ delta network. The delta networks are a class of networks that include the generalized cube network. An MDN denoted by C-MDN indicates explicitly the number of copies. The basic switch for the MDN is a $2C \times 2C$ crossbar, with $\log_2 N/C$ stages. An 8×8 MDN with C=2 is shown in Figure 1.9.

Routing in MDN is based on the delta network routing procedure. Routing information is initially passed to a non-faulty switch in stage 0. From there, each switch in stage j forwards the information by choosing with equal probability one of the copies. If the selected switch is faulty, another copy is chosen. At stage $n-1$ the switch output is chosen to reach the intended destination.

1.4.6 F-Network

The F-network [23] is a $2^n \times 2^n$ with $n+1$ stages of $N = 2^n$ switches each, that are in general 4×4 selectors. An 8×8 F-network is shown in Figure 1.10. The F-network can emulate the structure of the generalized cube network. At each stage except the output stage, two different switches can be selected while maintaining the same destination. A faulty or a busy switch is avoided by taking the alternate path.

1.5 Motivation for this Thesis

The Banyan networks suffer throughput limitation due to their basic SEs which are 2×2 switches. Even though the BB network is a nonblocking switch for permutation traffic it is blocking under realistic traffic conditions. The throughput of the BB is severely degraded under realistic traffic conditions. Moreover, neither the Banyan nor the BB networks possess any FT properties. All the MINs mention in the previous section also have low throughput because they accept only one packet at their output ports in each cycle. In order to be used as a switch fabric for broadband packet switch architectures, MINs should possess high throughput rate. The throughput of the R2D2 and the BG networks is increased by increasing the size of their SEs. This thesis work aims in evaluating the throughput performance of the BB and the R2D2 MINs and the BG network [24]. Since it is difficult to simulate the real time traffic conditions expected in Broadband Packet Switch Architectures, the performance of MINs under certain well known traffic conditions are studied in order to estimate their performance under realistic traffic conditions. If the switches are to be used in a remote environment, MINs considered for use

should be highly reliable. Apart from throughput performance, the MINs studied in this thesis are compared with respect to hardware complexity, FT and reliability, in this thesis.

1.6 Thesis Organization

This thesis consists of six chapters. An introduction about ATM and the objective of this thesis work were provided in previous sections. The BG network is discussed in Chapter Two. The hardware complexity, fault tolerance and reliability analysis, of MINs is presented in Chapter Three. Performance of MINs under certain uniform and non- uniform traffic patterns is presented in Chapter Four. Performance analysis results of the BG network in the presence of SE fault is discussed in Chapter Five. Contributions of this thesis along with some recommendations for future work in this area are presented in Chapter six.

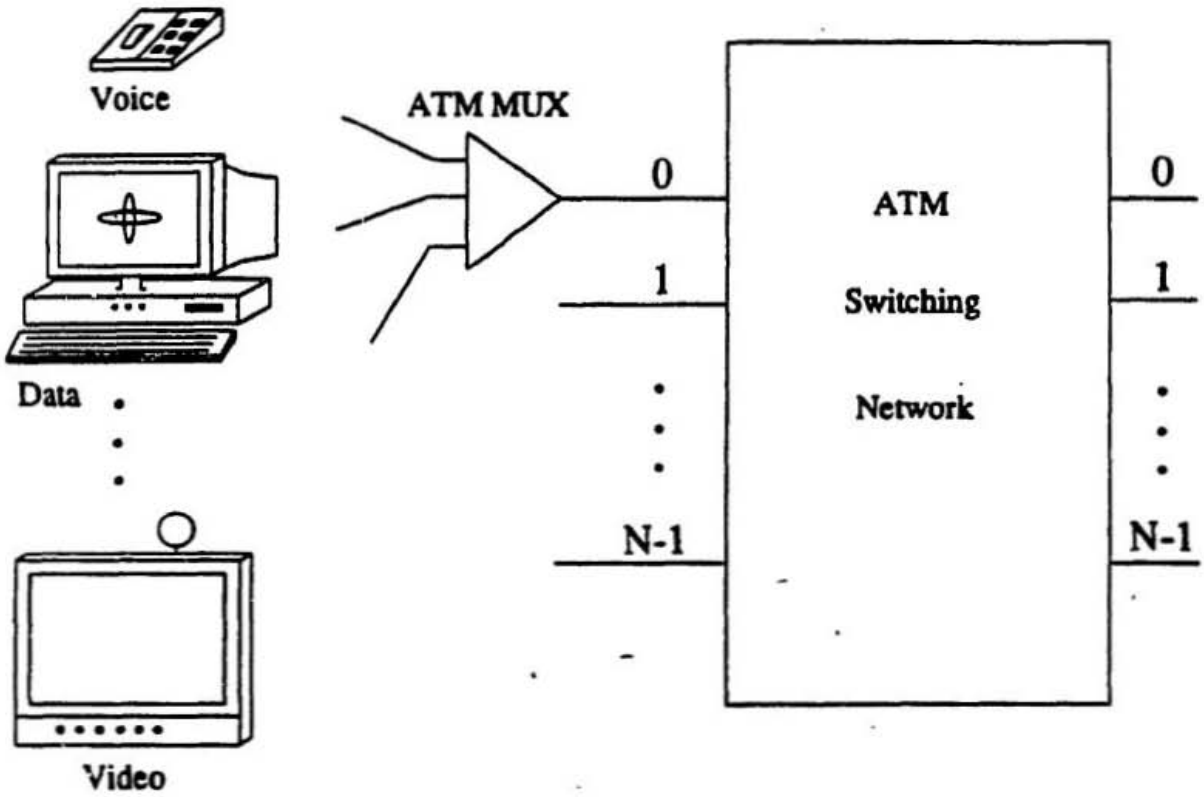

Figure 1.1: A B-ISDN Network [2].

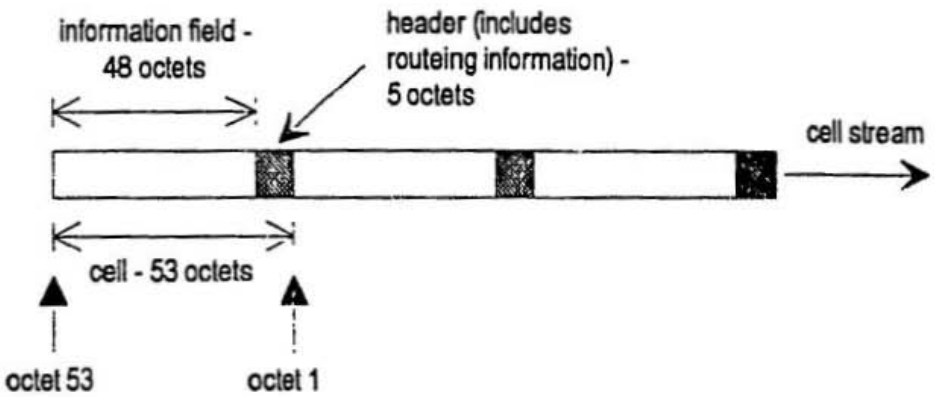

Figure 1.2: A basic ATM cell format [6].

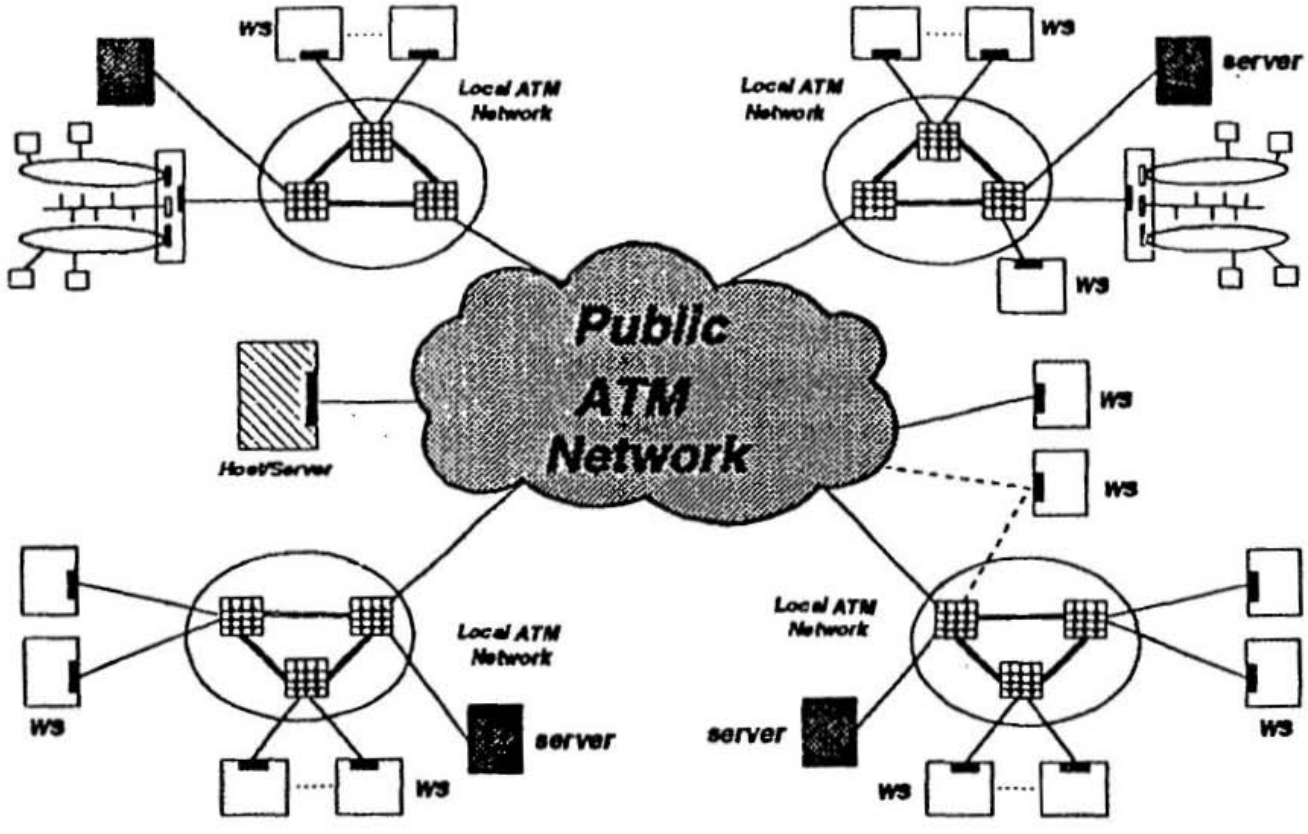

Figure 1.3: An ATM Public Switched Network.

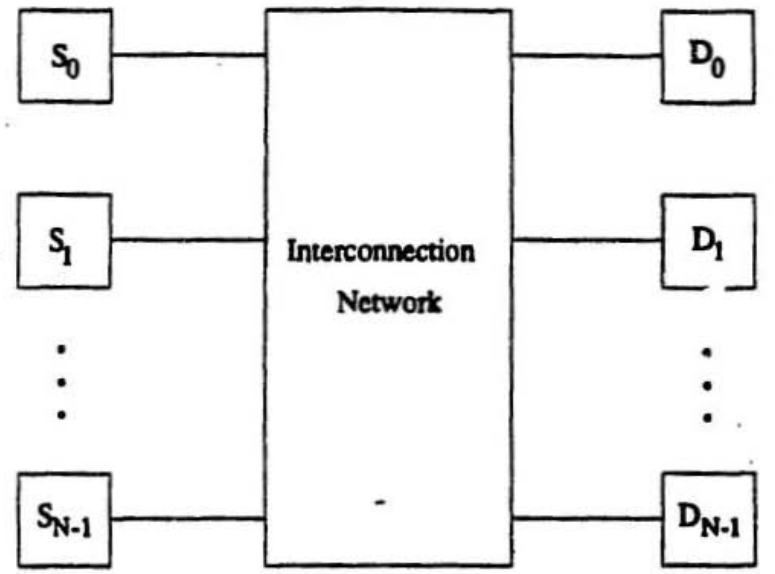

Figure 1.4: An ATM switch fabric [2].

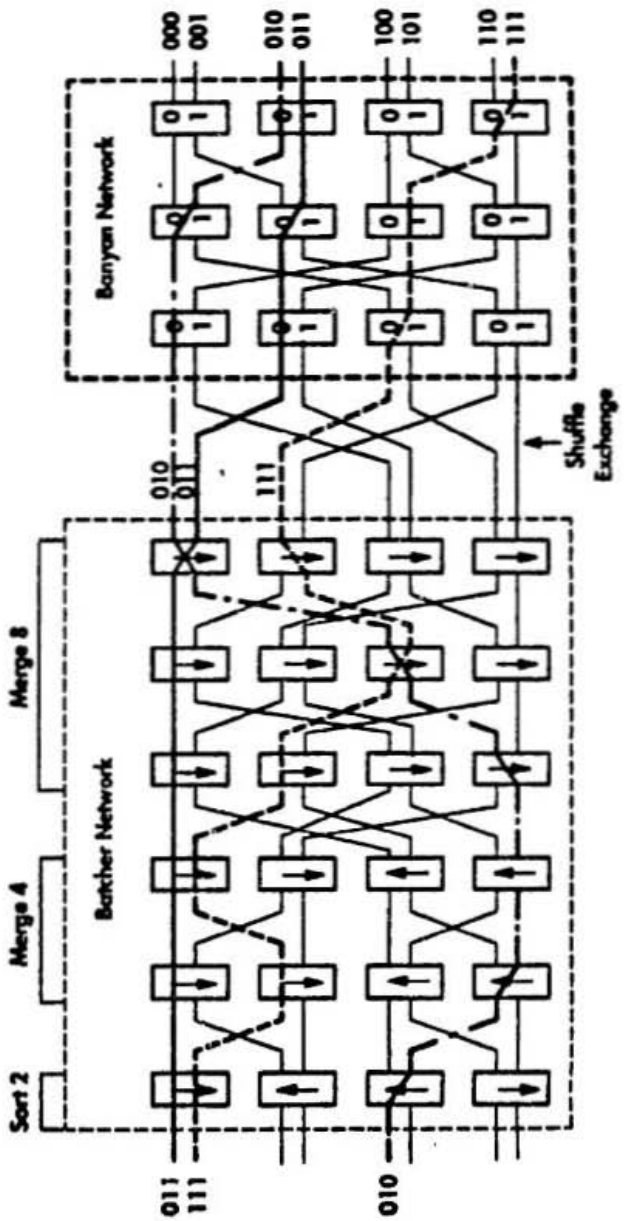

Figure 1.5: An 8×8 Batcher Banyan Network [4].

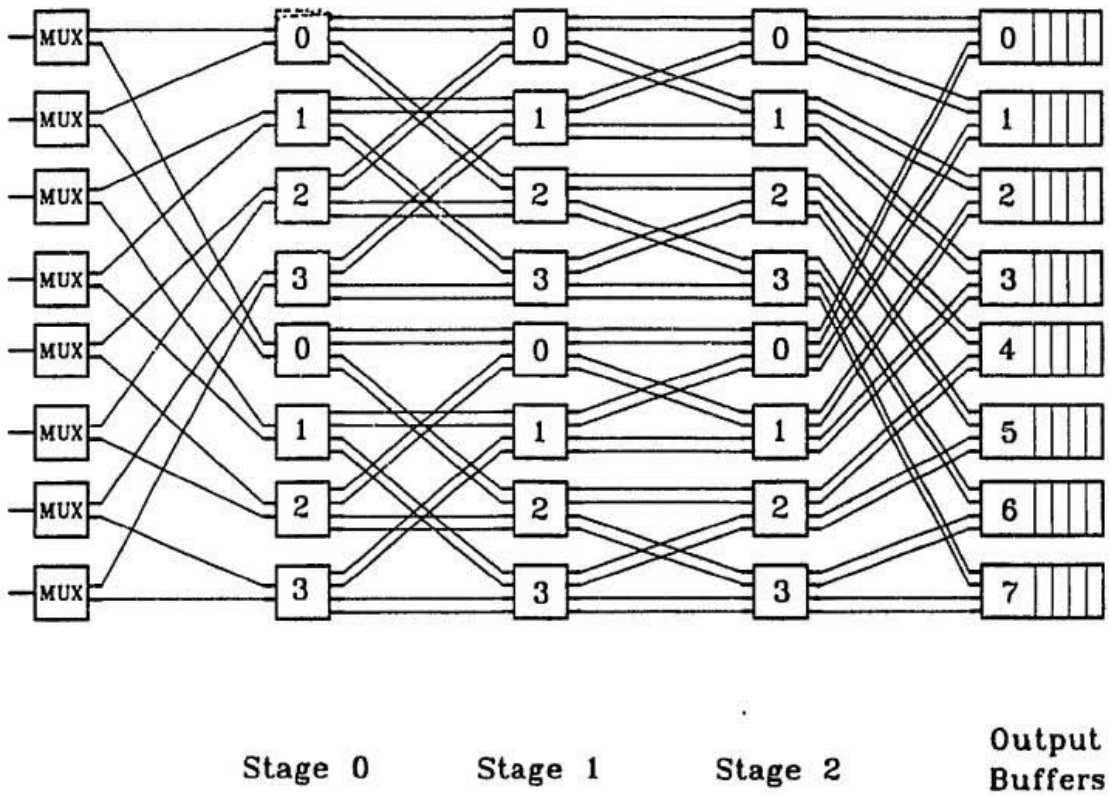

Figure 1.6: An 8×8 2-Replicated 2-Dilated Banyan Network.

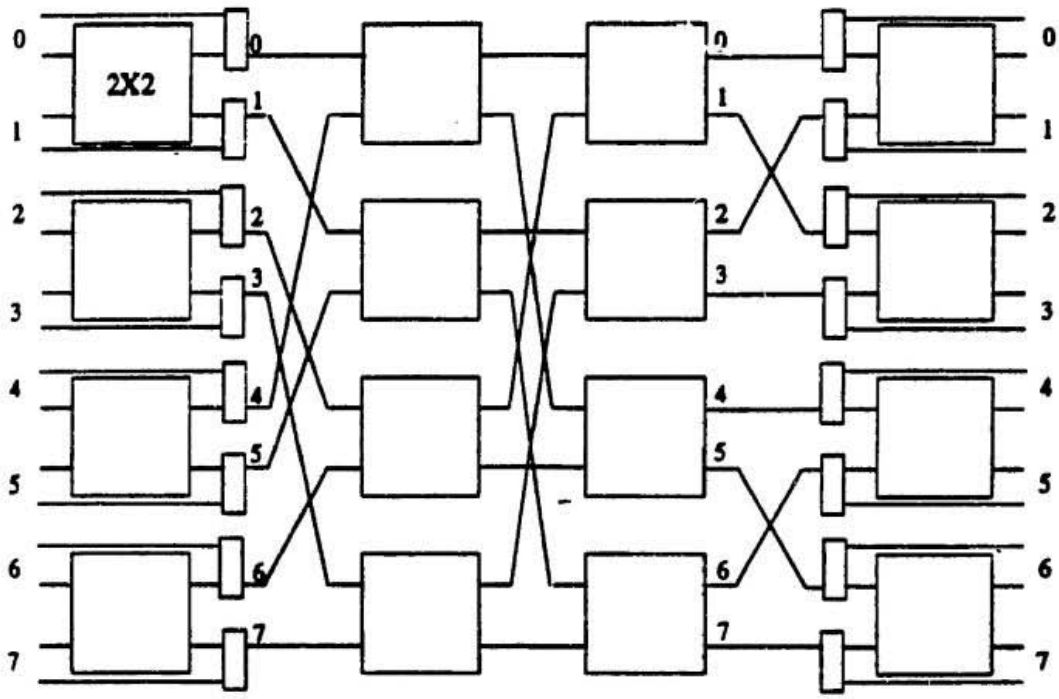

Figure 1.7: An Extra Stage Cube Network for $N = 8$ [2].

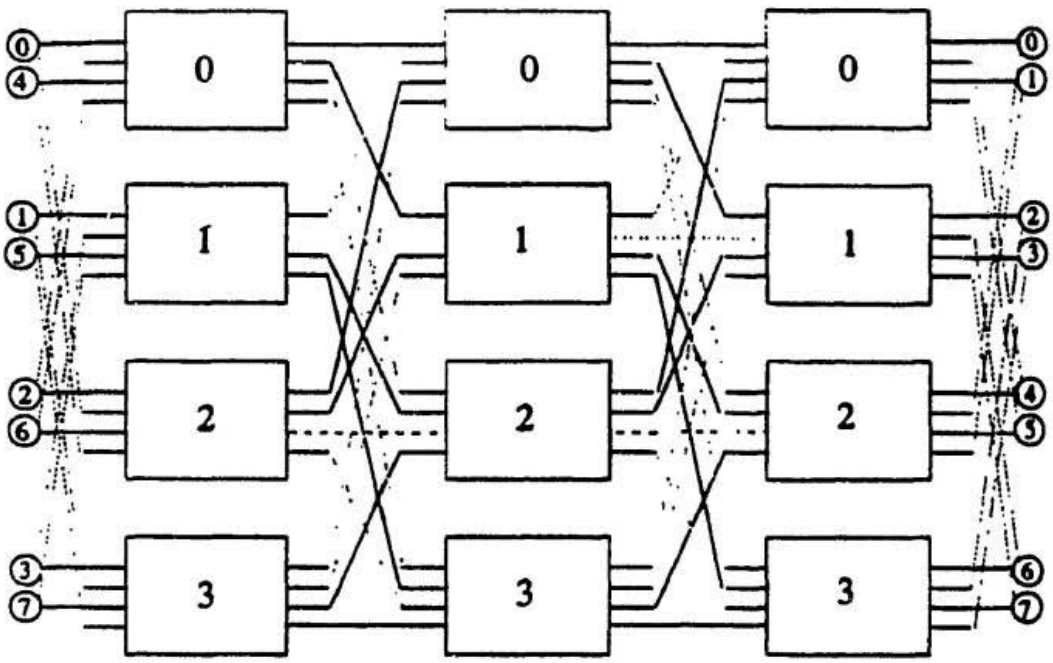

Figure 1.8: An 8×8 Augmented C-Network [2].

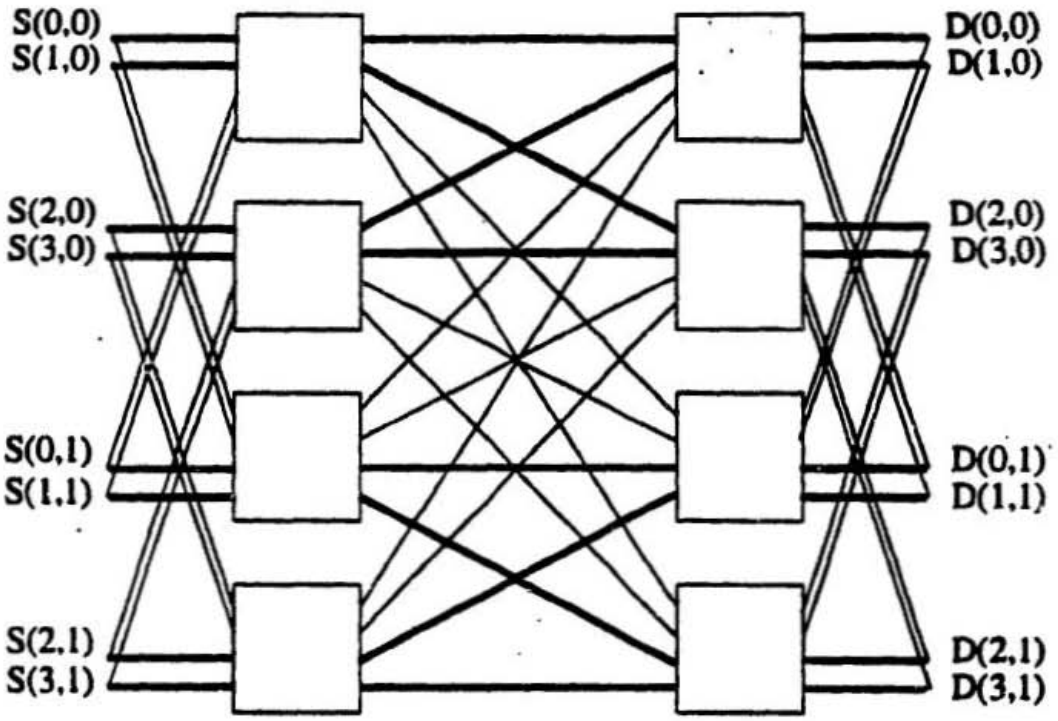

Figure 1.9: An 8×8 Merged Delta Network [2].

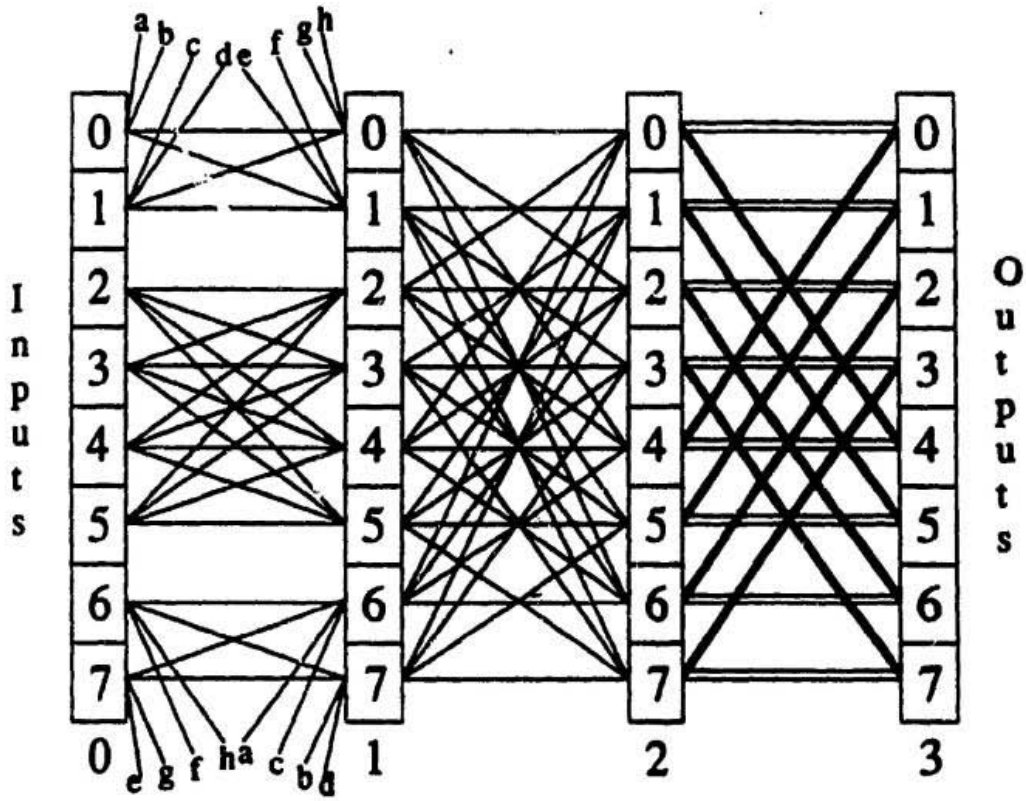

Figure 1.10: An 8×8 F-Network [2].

Chapter 2

The Balanced Gamma Network

2.1 Introduction

Unipath and multipath MINs discussed in the previous chapter are being experimented for use in the switch fabric of broadband packet switch architectures. In this chapter, yet another multipath MIN called the BG network is introduced.

The BG network has multiple paths to route a packet from a source to a destination. Due to multiple paths in the network, packets can be routed through the network even in presence of failures of some of the switching elements (SEs) in the network. Moreover the BG network is capable of accepting up to four packets at each destination during each cycle.

Section 2.2 discusses the structure of the BG network. The routing algorithm followed in the BG network is presented in Section 2.3 followed by a summary in Section 2.4.

2.2 Structure

The BG network [24] features 4 × 4 SEs and is derived by the enhancement of the gamma network. An 8 × 8 gamma network [25] is illustrated in Figure 2.1. An

$N \times N$ gamma network consists of one stage of 1×3 SEs, followed by $log_2 N - 1$ stages of 3×3 SEs and finally one stage of 3×1 SEs. Each stage of the gamma network consists of N SEs. The ith $(0 \leq i \; leq \; N)$SE in the jth $(0 \leq j \leq log_2 N - 2)$ stage is connected to SEs i, $(i + 2^j) \; mod \; N$ and $(i - 2^j) \; mod \; N$, in the $(j + 1)$th stage. The gamma network consists of twice the number of SEs in each stage as that of the Banyan network for the same N.

In case of failure of a SE in the gamma network, one of the non-straight links in the gamma network can be considered as an alternative for the other non-straight link and a re-routing algorithm can be developed to exploit this redundancy. However, the straight link does not have an alternative link and becomes the most critical component. Hence the gamma network can be viewed as an unbalanced network.

The gamma network is balanced by adding an extra link to each SE. This additional link in an ith SE of jth stage will be connected to SE $(i + 2^{j+1}) \; mod \; N$ in the $(j+1)$th stage. The SEs in the last stage of the gamma network are replaced by buffers to form the BG network. These buffers are designed to accept up to four packets during each cycle. These buffers collect the outcoming data from the network and feed them to the respective destinations. Similar to the gamma network, the BG network consists of N SEs numbered 0 to $N - 1$ in each stage and $n = log_2 N$ stages numbered 0 to $n - 1$. This network consists of 1×4 SEs in the first stage and 4×4 SEs in the subsequent stages. A 16×16 BG network is shown in Figure 2.2. Figure 2.3 shows a reconfiguration of the BG network that explicitly shows its block structure and modularity. If the output buffers in Figure 2.3 are replaced by concentrators, then this network is identical to the Kappa network [26].

Each SE in the BG network is addressed as $SE_{i,j}(0 \leq i < N, 0 \leq j < n)$ where i indicates the SE number within a stage and j the stage number. Within the 4×4 SEs the input ports are named as input upper (IU), input upper middle (IUM), input lower middle (ILM) and input lower (IL). The output ports are named as output upper (OU), output upper middle (OUM), output lower middle (OLM) and output lower (OL). Each 4×4 SE, $SE_{i,j}$, is connected to SEs $SE_{i-2^j,j-1}$, $SE_{i-2^{j-1},j-1}$, $SE_{i,j-1}$, and $SE_{i+2^{j-1},j-1}$ from the previous stage and to SEs $SE_{i-2^j,j+1}$, $SE_{i,j+1}$, $SE_{i+2^j,j+1}$, $SE_{i+2^{j+1},j+1}$ in the next stage as shown in Figure 2.4. Each SE $SE_{i,n-1}$ in the last stage is connected to the output buffers $(i-2^j, j+1)$, $(i, j+1)$, $(i+2^j, j+1)$, $(i+2^{j+1}, j+1)$. The links connected to OUM and OLM are called *normal links* and the links connected to OU and OL are called *alternate links*. Hereafter the links connected to output ports OU, OUM, OLM and OL will be referred to as OU link, OUM link, OLM link and OL link respectively.

2.3 Routing Algorithm

The BG network originally used a distance tag algorithm [24]. To enhance the fault tolerance properties of the BG network a reverse destination tag routing algorithm has been proposed in this thesis [27]. In the reverse destination tag routing algorithm, the routing tag for a cell to be routed from source S to destination D is the binary representation of D, viz. $d_{n-1}d_{n-2}d_{n-3}...d_0$. The SEs interpret the tag in the reverse order, i.e., the SE in *stage* 0 switches a packet based on bit d_0, SE in *stage* 1 switches based on bit d_1, and so on until the SE in *stage* $n-1$ which switches based on bit d_{n-1}. Each SE is associated with a value α given by the

formula

$$a = \lfloor \frac{i}{2^j} \rfloor \ mod\ 2. \tag{2.1}$$

This value a is used in switching a packet at an input port of an SE in one stage to the next stage. An SE $SE_{i,j}$ routes a packet through the OUM link when a is 0 and the tag bit is 0, or when a is 1 and tag bit is 1, and the OLM link when a is 0 and the tag bit is 1, or when a is 1 and the tag bit is 0. Packets arriving at the input port of an SE are routed through the alternate links only if more than one packet is to be routed through with the same tag bit or when the SEs connected to normal links in the next stage are faulty. As there are four inputs coming into each SE, it is quite likely that, in a given cycle, more than one packet may arrive at an SE with a 0 (or a 1) as the tag bit. It is easy to see that up to 2 packets with the same tag bit can be routed without any loss. If three or more inputs require the same output link, any two are arbitrarily chosen and the rest are dropped. If the cells from inputs are to be routed through the OUM link, then one cell is routed through the OUM link and the other is routed through the OL link. Similarly, if two of the incoming cells are to be routed through the OLM link, then one is routed through the OLM link and the other is routed through the OU link. The detailed routing procedure is given in Appendix A.

An example of this routing algorithm is depicted in Figure 2.5 which shows the path taken by a packet from source 7 to destination 12 in a 16 × 16 BG network. The destination tag used for routing the packet is 12 in binary which is 1100.

Figure 2.6 shows the routing of a packet from source 7 to destination 12 in the reconfigured 16 x 16 BG network. It can be noted that in the reconfigured BG network, the upper two links of each SE are used for switching a packet with a 0

tag bit while the lower two links are used for switching a packet with a 1 tag bit.

2.4 Simulation Program

Research on evaluation of MINs for use in broadband packet switch architectures has been done for the past four years under the guidance of Dr. Venkatesan in Faculty of Engineering and Applied Science at Memorial University of Newfoundland. As a result of this endeavor, a software package, written in C++, was developed to simulate several MINs and study their performance. This simulation program was capable of simulating the performance of the Banyan, the R2D2 and the BG networks under certain traffic conditions. The current work included modification and enhancement of this software.

The simulation program consists of four main sections - network definition, traffic generation, switching and results sections. The network definition section defines parameters used by the program to simulate the network. The expected traffic pattern is generated by the traffic generation part and then switched through the network. Finally the results are calculated in the results section.

The new routing algorithm proposed for the BG networks has been implemented in the simulation program. The program is now capable of simulating the performance of the BB networks. Performances of the Banyan, the R2D2, the BB and the BG networks under the the permutation, the hotspot, the community of interest and the bursty traffic conditions, to be discussed in Chapter 4, have been either added to the existing program, or the existing software has been modified to include these. The failure of an SE in the BG network has also been simulated so that the performance under failed components can be studied.

2.5 Summary

A multipath MIN, called the BG network featuring 4×4 switching elements has been presented in this chapter. Evolution, structure and properties of the BG network are explained in detail. Symmetry and modularity of the BG network are explicit in the reconfigured BG network. The routing algorithm followed in the BG network is illustrated with the help of an example.

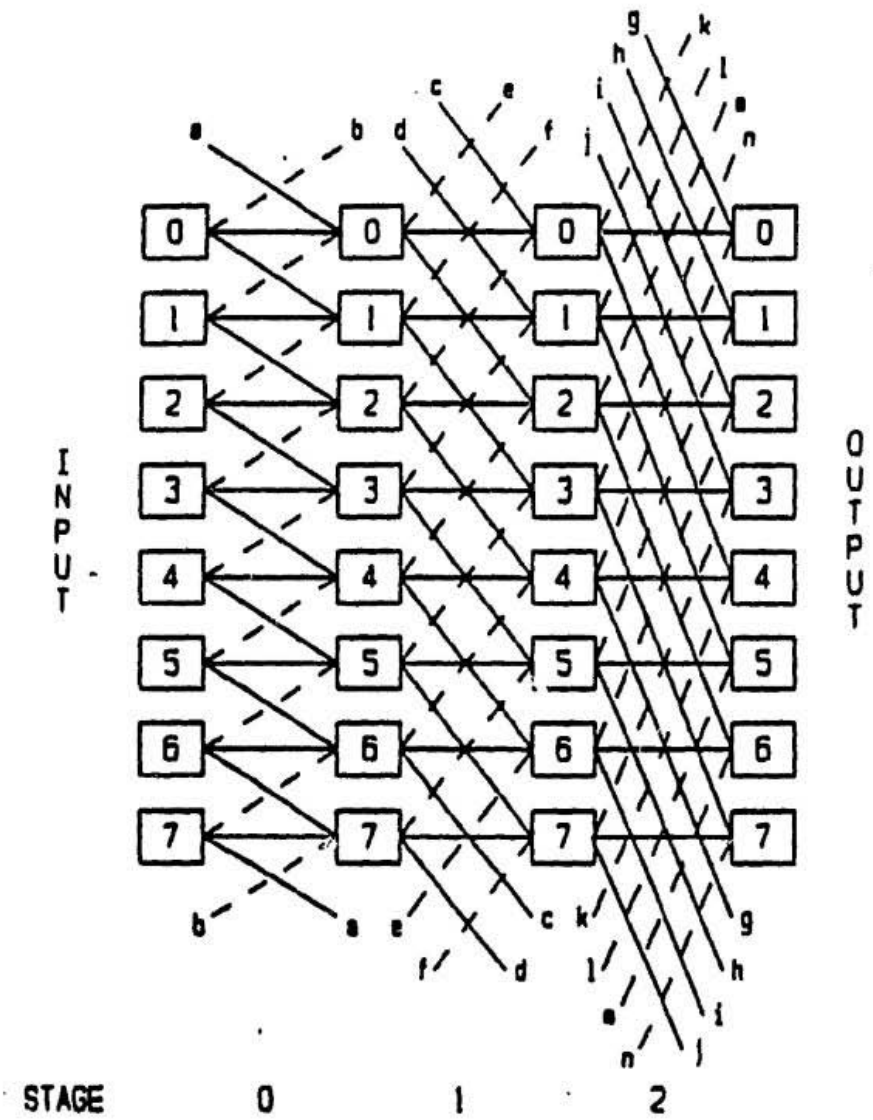

Figure 2.1: An 8 × 8 Gamma Network [24].

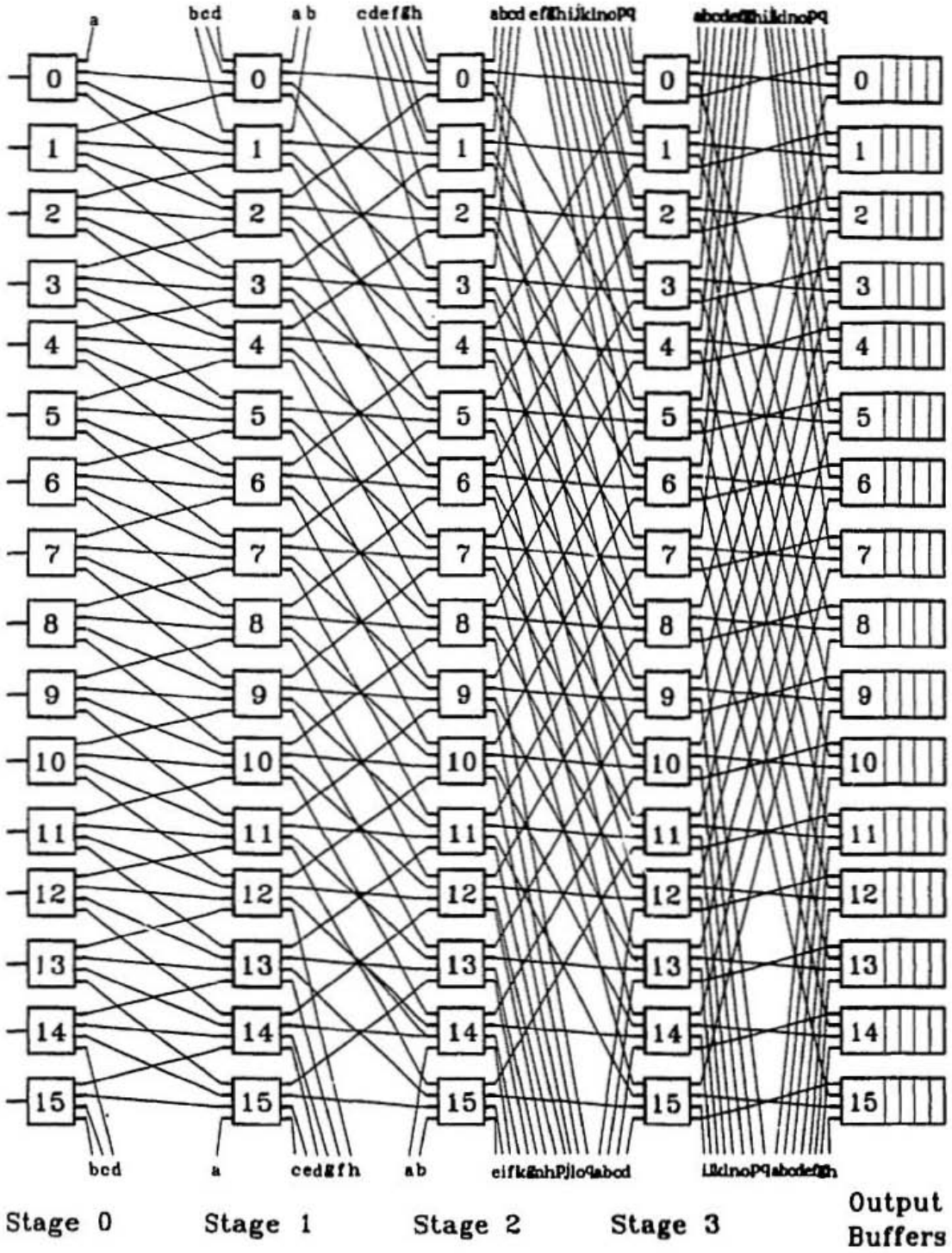

Figure 2.2: A 16 × 16 Balanced Gamma Network.

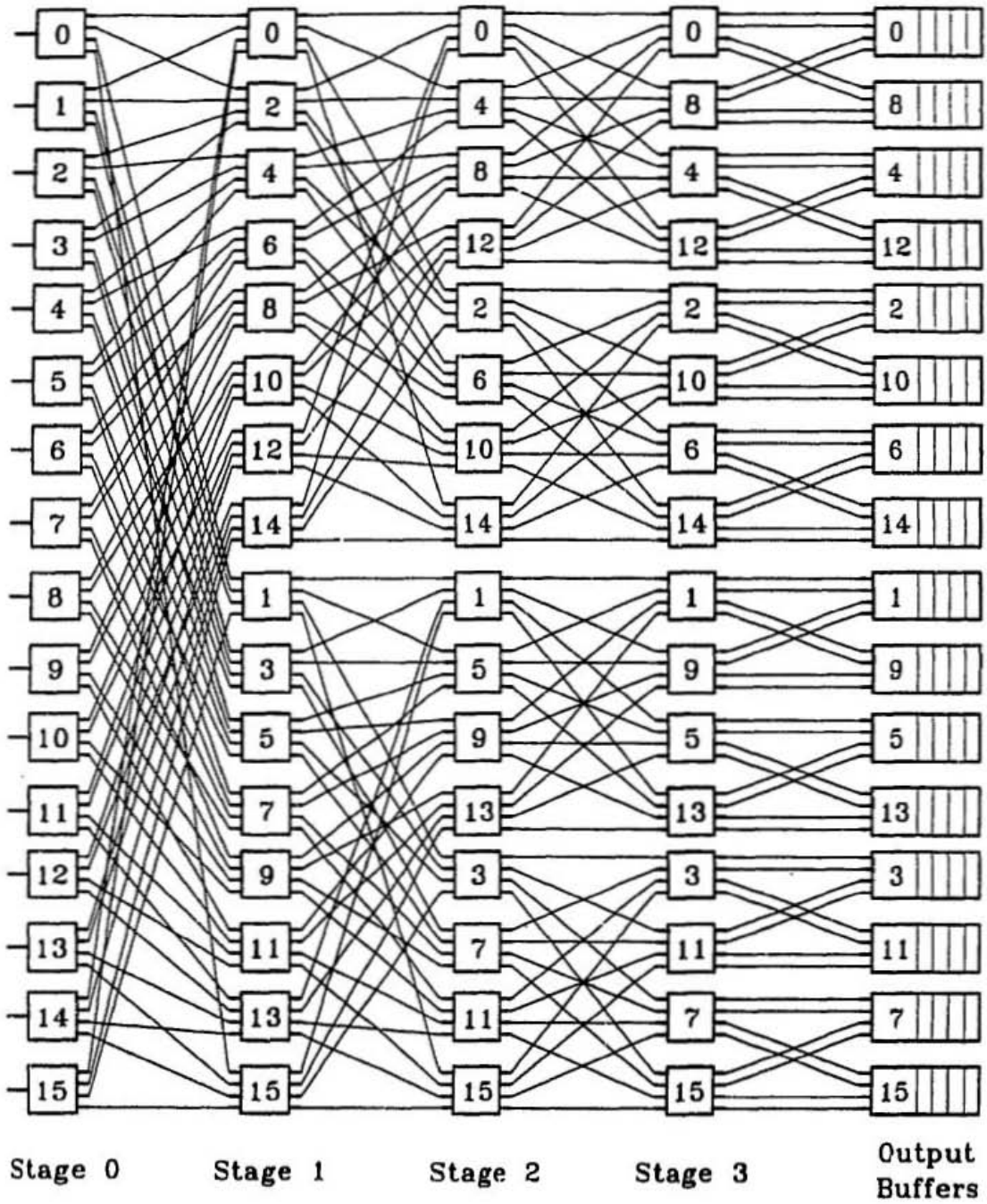

Figure 2.3: A Reconfigured 16 × 16 Balanced Gamma Network.

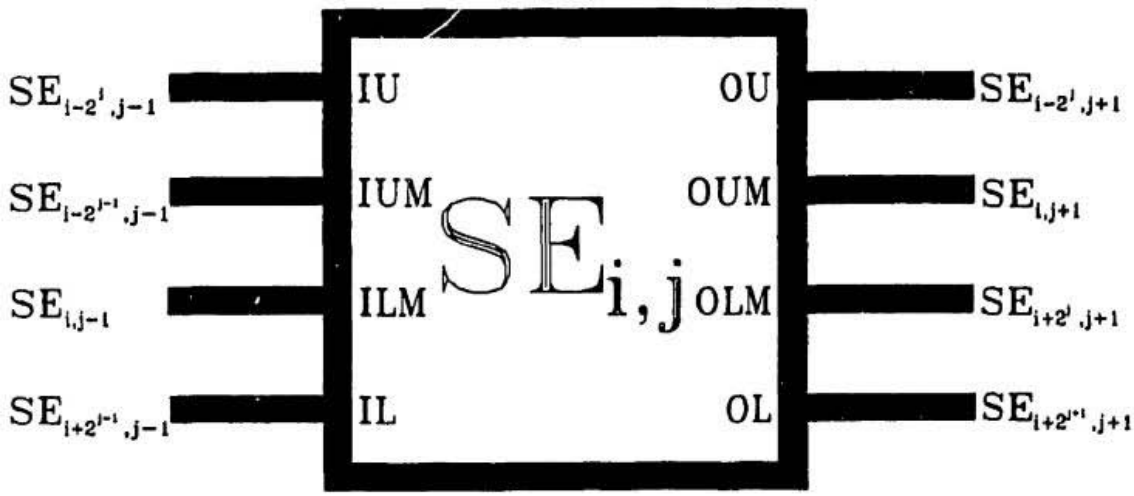

Figure 2.4: Connections of Switching Element in Balanced Gamma Network.

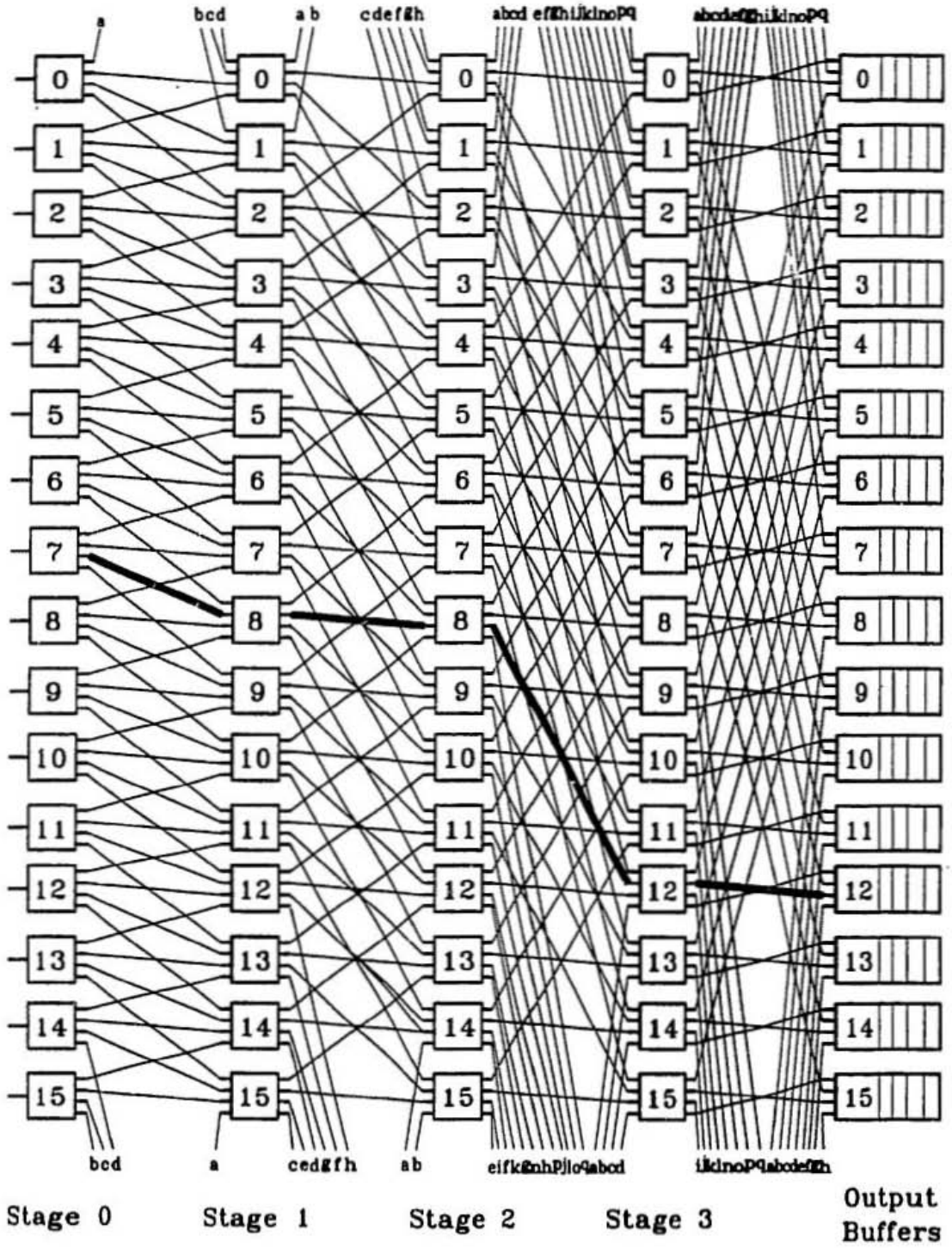

Figure 2.5: Routing in 16 × 16 Balance Gamma Network.

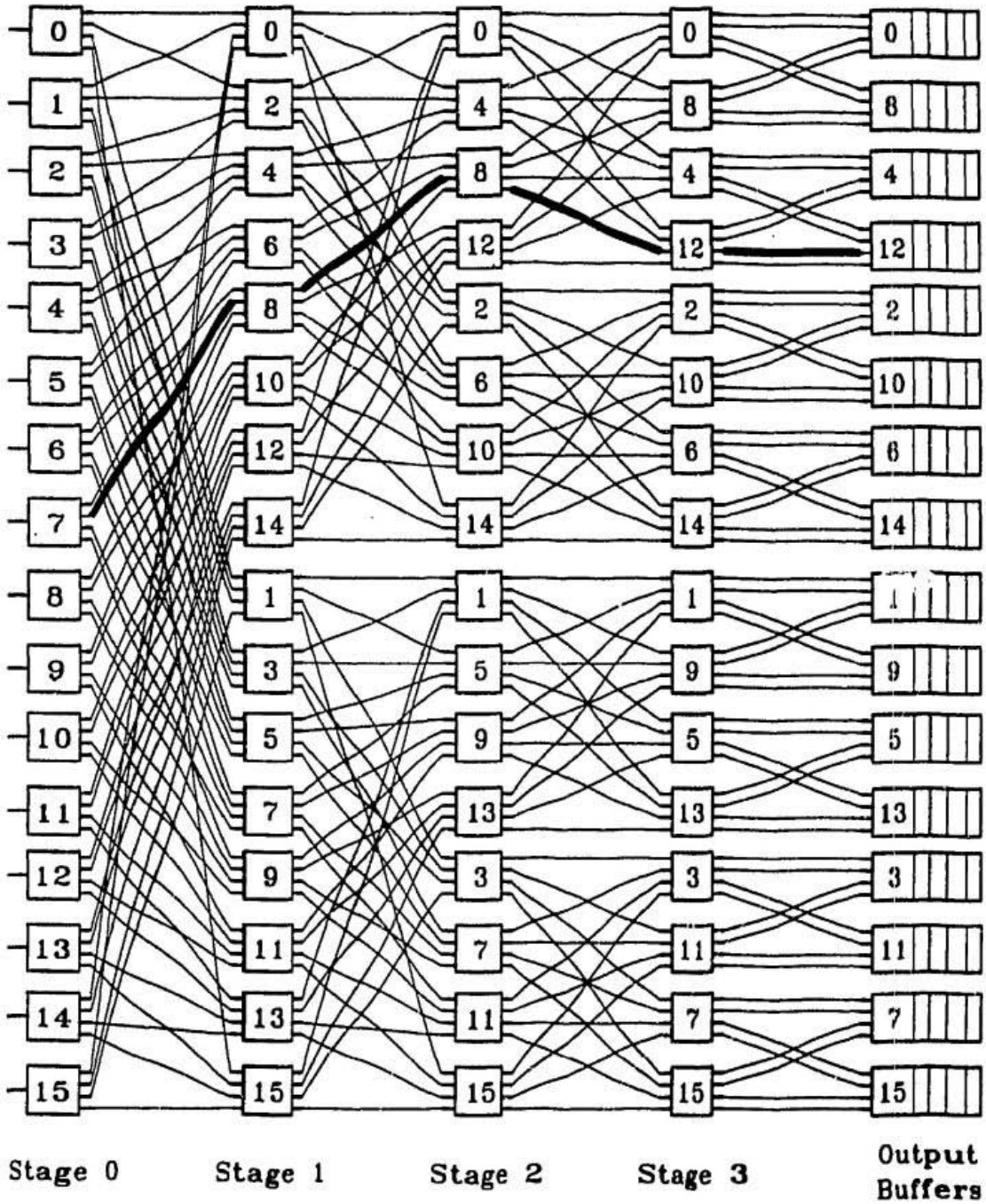

Figure 2.6: Routing in Reconfigured 16 x 16 Balanced Gamma Network.

Chapter 3

Properties of MINs

3.1 Introduction

MINs are being experimented with for use in broadband packet switch architectures. Some MINs along with their routing algorithm have been explained in Chapters 1 and 2. There are certain properties of MINs that are to be considered for use in switch fabrics of the broadband packet switch architectures. These properties of MINs are detailed in this chapter.

The hardware complexity (HC) of the MINs are explained in Section 3.2. The FT properties of MINs are explained in Section 3.3 followed by their reliability analysis in Section 3.4. Finally a summary is provided in Section 3.5.

3.2 Hardware Complexity

The MINs being experimented for use in switch fabrics of broadband packet switch architectures have to be finally implemented on hardware. Most of the studies indicate the number of crosspoints required to implement the switching network as proportional to the hardware cost [11]. So the

With rapid advancements in the VLSI technology, a great deal of HC can be implemented on a single chip. Here, the main focus is on the crosspoint complexity of the entire MIN instead of taking the number of chips required for hardware implementation of the network.

The HC of a network can be equated to the sum of the hardware complexities of the SEs of the MIN. The HC of an SE is also referred to as the crosspoint complexity because it depends on the total number of connections between the input ports and the output ports within an SE. The crosspoint complexity of the SEs can be found by different methods [11], [33]. One such method states the HC to be equal to the product of number of the input ports and output ports. This method has been chosen for calculation of HCs of MINs in this thesis. Primarily the MINs discussed in the previous chapters are made up of either 2×2 SEs or 4×4 SEs.

The crosspoint complexity of an 2×2 SE is equal to 4. This has been illustrated in Figure 3.1. Each input port is connected to two other output ports within the SE. Hence there are 4 connections within the SE and hence the crosspoint complexity or the HC of an 2×2 SE is equal to 4. Similarly the crosspoint complexity of a 4×4 SE is equal to 16 and is shown in Figure 3.2. Here, there are 4 input ports connected 4 output ports and hence a total of 16 connections exist within the SE.

Although several researchers have proposed more complex measures for computing the HC of an SE (for example, a scheme which give different weights to the number of crosspoints, the total number of input and output ports, and the fabrication technology), the most popular approach is to employ the crosspoint complexity. Hence the choice here.

3.2.1 Banyan Network

An $N \times N$ Banyan network consists of $\log_2 N$ stages and $\frac{N}{2}$ SEs in each stage. The HC of an $N \times N$ Banyan network is calculated to be

$$HC(Banyan) = 4 \times \frac{N}{2} \times \log_2 N. \tag{3.1}$$

3.2.2 Batcher Banyan Network

The BB network consists of a Batcher network to sort the incoming packets followed by a Banyan network to route the packets. An $N \times N$ Batcher network consists of $\frac{N}{2} \times log_2 N$ SEs in the Banyan network and $\frac{N}{2} \times \sum_{i=1}^{\log_2 N} i$ SEs in the Batcher network.

As discussed in Section 1, the BB is made up of 2×2 SEs. The sorting of the packets by SEs of the Batcher network has to be done as fast as possible, while the SEs of the Banyan network just check one bit to route a packet. In order to match the speed of the Banyan network, the SEs in the Batcher network must be at least $\log_2 N$ times faster, which will be the time taken to compare $\log_2 N$ bits. Hence, unlike the HC of SEs in the Banyan network, the 2×2 SEs used in the Batcher network have a slightly higher complexity than 4. The HC of the SEs of the Batcher network is denoted by 4^+ indicating it to be slightly greater than that of a normal 2×2 SE. These SEs must have extra hardware in order to compare all the bits used for routing, in order to sort the packets.

The overall HC of the BB network is calculated to be

$$HC(BB) = 4^+ \times \frac{N}{2} \times \sum_{i=1}^{\log_2 N} i + 4 \times \frac{N}{2} \times \log_2 N. \tag{3.2}$$

3.2.3 2-Replicated 2-Dilated Banyan Network

An $N \times N$ R2D2 network consists of two $N \times N$ Banyan subnetworks. It also contains N 1×2 multiplexers.

Since the multiplexers in the R2D2 network have one input and two outputs, the HC of these can be equated to 2. Since the R2D2 network consists of two Banyan networks, the overall HC of the R2D2 network is calculated to be

$$\begin{aligned} HC(R2D2) &= N \times 2 + 2 \times 4 \times \frac{N}{2} \times \log_2 N \\ &= 2 \times N \times (1 + 2 \times \log_2 N) \end{aligned} \tag{3.3}$$

3.2.4 Balanced Gamma Network

An $N \times N$ BG network consists of 1×4 SEs in the first stage and 4×4 in $\log_2 N - 1$ stages: there are N SEs in each stage. The HC of the 1×4 SEs is equal to 4. As calculated earlier, the HC of a 4×4 SE is equal to 16. Therefore the HC of the BG network is calculated to be

$$HC(BG) = 4 \times N + 16 \times N \times (\log_2 N - 1). \tag{3.4}$$

The HCs for different sizes of the Banyan, the BB, the R2D2 and the BG networks are given in Table 3.1. It can be seen from Table 3.1 that the BB, the R2D2 and the BG networks have comparable HCs. Therefore these three networks have been taken into consideration for comparison of their performances under different kinds of traffic which can be expected in broadband packet switch architectures. FT calculations and reliability analysis are also done mainly for these networks. Comparison results of other existing networks which have been provided at certain places are taken from available references.

3.3 Fault Tolerance Properties

The MINs which are to be used in switch fabric of broadband packet switch architectures are quite complex systems which must be highly reliable. The performance of the network should not be affected by the failure of any of the component(s) within the MIN which forms a part of the switch fabrics. FT is an essential metric for comparison of the MINs [28].

The SEs and the links between the SEs can be collectively termed as network components. A $fault-tolerant\ MIN$ is one that is able to route packets from input ports to the requested output ports, in at least some cases, even when some of its network components are faulty. A fault can be either permanent or transient. Here, it is assumed that faults are always permanent.

FT is defined only with respect to a chosen *fault tolerance model* which has two parts. The fault model characterizes all faults assumed to occur, stating the failure modes (if any) for each network component. The *fault tolerance criterion* is the condition that must be met for the network to be said to have tolerated a given fault or faults. Most of the MINs have the *fault tolerance criterion* satisfying the full access property. A network is said to have full access property if a packet at one of its input ports can be routed through the network to any of its output ports. The *fault tolerance model*, the *fault tolerance criterion* and the *fault tolerance method* used under component failures to be used for the MINs discussed in earlier chapters are given below. Any additional *fault tolerance model* specific for a MIN is given under the section explaining the FT properties of that MIN.

Fault tolerance model

1. SE-faults are random and independent.

2. SE-faults are permanent.

3. Each faulty SE is totally unusable.

4. The network is considered to have failed when the number and locations of the faulty SEs prevent the connection between at least one input and output pair of the network.

5. All faults are detected 100%.

Fault tolerance criterion

The network has to satisfy the full access property.

Fault tolerance method

There exists alternate route(s) between any input and output pair in case of SE-faults.

A network is called *single fault-tolerant* if it can function as specified by its FT criterion despite any single fault conforming to its fault model. In general, if a network can tolerate any set of i faults, then it is said to be *i-fault tolerant*. A network that can tolerate some instances of i faults is said to be *robust* although it is not *i-fault tolerant*.

3.3.1 Banyan and Batcher Banyan Networks

An $N \times N$ BB network consists of an $N \times N$ Banyan network for routing the packets. The Banyan network which is made up of 2×2 SEs is a unipath MIN. Hence any component failure within the Banyan network will disconnect a path between one

or more input-output pairs. Hence both the BB network and the Banyan network are not fault-tolerant MINs. Similarly, all sorting elements are also critical.

3.3.2 2-Replicated 2-Dilated Banyan Network

Modifications to Fault tolerance model

1. The output buffers and multiplexers are highly reliable.

2. The links connecting last stage SEs with buffers are highly reliable.

An $N \times N$ R2D2 network is a multipath MIN. Let the replicated Banyan networks be termed as subnetworks A and B. In the case of link failures packets can still be routed through the same network A or B. In case of SE failures, the multiplexers have to be notified as to which subnetwork has failed. Therefore, depending upon this, the multiplexer will route an incoming packet to the fault-free subnetwork.

All link faults in the R2D2 network are reported to the SEs to which they are connected. Therefore in the case of a link fault, the SE will route a packet through the alternate link when the regular link has failed. When m $(1 < m \leq 4)$ packets are to be routed for a particular tag bit value (0 or 1), and if one of the links corresponding to that tag bit has failed, then arbitrarily $m-1$ packets are dropped and only one packet is routed to the next SE.

Figure 3.3 shows routing in an R2D2 network in case of link fault. The link fault is indicated by a broken line. In this figure a packet from input port 0 is to be routed to output port 1. Instead of taking the regular link at Stage 0, the packet takes the alternate link because the regular link connecting SEs $SE_{0,0,0}$ and $SE_{0,1,0}$ is faulty. The failure of both the regular and alternate links connecting two SEs

$SE_{i,j,k}$ and $SE_{i,j+1,k}$ will be considered as SE failure and will be reported to the multiplexers.

R2D2 networks can tolerate up to $N \times \log_2 N$ SE failures provided there exists at least one path between any input-output pair. This happens in the case when all the SEs of one of the subnetworks fail. This is shown in Figure 3.4. All SEs within the block denote SE faults.

If a complex routing is followed in the R2D2 networks then it can tolerate certain SE faults in both the subnetworks and still satisfy the FT criterion. In this complex routing algorithm, each faulty SE has to report its failure to the multiplexers. The multiplexers will route an incoming packet only through a fault-free path, if one such path exists. Checking for a fault-free path involves checking faulty SEs in the path taken by the packet. Since this routing algorithm is complex, and requires more hardware, this has not been proposed for the R2D2 network. This modification also introduces some degree of centralized control which may not be desirable. With developments in VLSI technology this could be accomplished. If such a routing algorithm is followed for routing packets, the R2D2 network can tolerate up to $(\frac{N}{2} \times (\log_2 N - 1))$ SE failures in each of the subnetworks. In such a case, all the even SEs of stages 0 to $\log_2 N - 1$ in one subnetwork and all the odd SEs of stages 0 to $\log_2 N - 1$ of the other subnetwork must have failed or vice versa. There should be no SE faults in the last stage. This is indicated in Figure 3.5. All SEs within blocks indicate SE faults. The failure of SEs, $SE_{i,j,0}$ and $SE_{i,k,1}$, (j and k need not necessarily be equal) makes the R2D2 network lose its FT criterion. This is shown in Figure 3.6 where SEs, $SE_{0,0,0}$ and $SE_{0,0,1}$, of an 8×8 R2D2 network are faulty. Hence the network is single fault-tolerant.

Routing of a packet in an R2D2 network under SE fault is shown in Figure 3.7. The failure of the SE, $SE_{0,1,0}$, in the R2D2 network is notified to the multiplexers which have to route packets to that subnetwc 'ch will route the packet through the network. The R2D2 network is single fault-tolerant and robust under multiple faults.

3.3.3 Balanced Gamma Network

Modifications to Fault tolerance model

1. Faults can occur in any switching element (SE) except the SEs in the first stage.
2. The output buffers are highly reliable.
3. The links connecting the last stage SEs to the output buffers are highly reliable.

An $N \times N$ BG network is also a multipath MIN. In each stage of the BG network, packets can be routed through two different paths. Similar to the R2D2 network, link faults are notified to the corresponding SEs to which the links are connected. If both the regular and alternate links for a specific switching bit from a SE $SE_{i,j}$ are faulty then this SE will inform this to the SEs in stage $(j-1)$ to which it is connected. The connection pattern of a 4×4 SE was given in Figure 2.4. Failure of an SE $SE_{i,j}$ is notified to SEs $SE_{i-2^j,j-1}$, $SE_{i-2^{j-1},j-1}$, $SE_{i,j-1}$, and $SE_{i+2^{j-1},j-1}$ of stage $(j-1)$.

There are certain pairs of SEs in each stage which are called as critical pairs. Each SE forms a critical pair with two other SEs in the same stage. In the BG

network each SE, $SE_{i,j}$ forms critical pair with exactly two other SEs, $SE_{i+2^j,j}$ and $SE_{i-2^j,j}$, for $0 < j < n-1$ and with $SE_{i+2^j,j}$ for $j = n$ (If $i + 2^j > N$ then $(i + 2^j) = (i + 2^j) \bmod N$).

Lemma 1 *The failure of any of the critical pair SEs in the BG network will result in the loss of full access property.*

Proof : If the critical pair of SEs, $SE_{i,j}$ and $SE_{i+2^j,j}$, fail in the BG network, then it loses the full access property because, a packet originating at input port i cannot to be routed to output port i through the BG network. This can be seen from Figure 3.9 that the failure of certain pairs of SEs will result in loss of full access property for the BG network. The faulty SEs are enclosed within a box. A packet originating at input port i will be routed to either $SE_{i,1}$ or $SE_{i+2,1}$ from $SE\,SE_{i,0}$. At stage 1, this packet will be routed to either $SE_{i,2}$ or $SE_{i+4,2}$. Similarly the packet will be routed till the last stage where it is routed to its final destination i either from $SE_{i,n-1}$ or $SE_{i+\frac{N}{2},n-1}$. This is depicted in Figure 3.10 for i=0, which shows all the paths that can be taken by a packet from the input port 0 to the output port 0 in a 16 × 16 reconfigured BG network.

It can be noticed that in each stage a packet is either routed to $SE_{i,j}$ ($0 < j \leq n-1$) or to another SE which forms a critical pair with $SE_{i,j}$ at that stage. So if any of the critical pair SEs fail, then a connection between one of input-output pairs cannot be established. If the critical pair of SEs $SE_{i,j}$ and $SE_{i+2^j,j}$ fail, then a packet cannot be routed from the input port i to the output port i if $i < (i + 2^j)$ and a packet from the input port $i + 2^j$ cannot be routed to the output port $i + 2^j$ if $i < (i + 2^j)$.

□

Lemma 2 *There are no other pair(s) of failures that can cause loss of full access property for the BG network other than the critical pair failures.*

Proof : Lemma 1 states that the failure of any critical pair of SEs in the BG network will result in loss of full access property for the BG network. For a packet to be routed through the BG network, the packet has to be routed through each stage of the BG network. A packet will be routed from stage j to stage $(j+1)$ $(0 \leq j < n-1)$ if at least one of the two SEs to which it has to be routed in stage $j+1$ is fault-free. The two SEs to which a packet has to be routed from stage j to stage $(j+1)$ always form a critical pair. Packets arriving at the last stage will always be delivered to their respective output ports because the output buffers and links of the last stage are highly reliable according to the chosen FT model.

□

Theorem 1 *The BG network will lose full access property if and only if any of the critical pair of SEs fail.*

Proof : Lemma 1 states that, the failure of any critical pair of SEs will result in loss of full access property for the BG network. From Lemma 2, it can be observed that the BG network will not lose its full access property due to failure of any pair(s) of SEs except the critical pair SEs.

□

Similar to the R2D2 network, all link faults in the BG network are reported to the SEs to which they are connected. Therefore, in the case of a link fault, the SE will route a packet through the alternate link when the regular link has failed. When m $(1 < m \leq 4)$ packets are to be routed for a particular (0 or 1) tag bit, and

if one of the links corresponding to that tag bit has failed, then arbitrarily $m-1$ packets are dropped and only one packet is routed to the next SE. Figure 3.11 shows routing in an BG network in the case of a link fault. The link fault is indicated by a broken line. In this figure a packet from input port 0 is routed to output port 1. Instead of taking the normal link from $SE_{0,1}$ to $SE_{0,2}$ the packet takes the alternate path and goes to $SE_{1,2}$.

Routing of a packet in a 16 × 16 BG network under SE fault is shown in Figure 3.12. In this figure $SE_{0,2}$ is faulty. This is notified to $SE_{0,1}$, $SE_{2,1}$, $SE_{12,1}$ and $SE_{14,1}$. Hence as shown in Figure 3.12 a packet arriving at $SE_{0,1}$ is routed to $SE_{1,2}$ because of SE fault at $SE_{0,2}$. Each SE of the BG network in stage j forms critical pair with two other SEs in the same stage j. The BG network can tolerate up to $\frac{N}{2}$ SE faults in each stage. In total, the BG network can tolerate $\frac{N}{2} \times (n-2)$ SE faults in the whole network. Figure 3.13 shows an 16 × 16 BG network having tolerated maximum SE faults. There are two such combination of maximum SE failures in the BG network : boxed and unboxed SEs in which the network possesses the full access property even after multiple SE failures. The failure of one or more critical pairs in the BG network causes loss of full access property. It has a simple routing algorithm as SE faults in stage j are checked only in SEs in stage $(j-1)$ compared to the multiplexers in the R2D2 network. The BG network is single fault-tolerant MIN and robust under multiple faults.

Having studied the FT properties of certain MINs, a comparison of these properties with the hypothetical network and other networks described in [28] would give a better understanding of the discussed MINs. This hypothetical fault-tolerant MIN is assumed to have the following ideal engineering characteristics:

1. Fault model - any network component can fail, and failed components are unusable.

2. FT criterion - full access.

3. Routing complexity - low as network discussed in [28].

4. HC - low as any network discussed [28].

5. FT capability - single fault-tolerant and robust with respect to multiple faults.

Table 3.2 shows a summary of FT information for the networks discussed in [28] along with the networks discussed in this section.

3.4 Reliability Analysis

MINs being considered for use in broadband communication switch fabrics, apart from having high throughput performance, must also be highly reliable as they would be used in real-time communication networks. For real-time systems the principal concerns are time-dependent reliability and the mean time to failure (MTTF) [29].

The reliability of a component, or a system, is the conditional probability that the component operates correctly throughout the interval $[t_0, t]$ given that it was operating correctly at time t_0 [30]. Exact reliability analysis of a complex system is complicated even under simplifying assumptions such as perfect coverage and no repair [31]. The reliability analyses of several MINs have been reported in [31]-[41]. There are three basic forms of connections, and consequently three main reliability measures are computed for MINs [37]. These are the terminal reliability, the

broadcast reliability and the network reliability which are explained in subsequent sections. The MTTF which specifies the quality of a system is the expected time that a system will operate before the first failure occurs. The two figures which are usually calculated for MINs are the input-output MTTF and the network MTTF. For the reliability analysis, the following assumptions are made.

1. The SEs in the first stage in case of the BG network and the multiplexers in the case of the R2D2 network are highly reliable.

2. The output buffers are highly reliable.

3. All SEs which can fail have an identical and constant failure rate of λ, and SE failures are statistically independent.

4. SE faults are permanent and each faulty SE is totally unusable.

5. Without loss of generality, we consider only SE failures. Failures of a link connecting two SEs $SE_{i,j}$ and $SE_{i,j+1}$, can be considered as SE failure of $SE_{i,j+1}$.

As mentioned in the earlier section both the Banyan and the BB networks are non-fault tolerant network and therefore the reliability analysis has been done only for the R2D2 and the BG network.

3.4.1 Terminal Reliability

Terminal reliability (TR) is the probability that there exists at least one fault-free path from a particular input port to a particular output port. TR is always associated with a terminal path (TP) which is a one-to-one connection between an input port (the source) and an output port (the destination). A network is

considered failed if it is not able to establish a connection from a given source to a given destination. TR is normally used as a measure of the robustness of a communication network [33]. The set of paths in a network between a given input-output pair is represented as a directed graph, sometimes referred to as the redundancy graph (or R-graph) [42], with its vertices representing the SEs and the edges representing the connecting links. This R-Graph is used to determine the TR of a network. Here the reliability of each SE is assumed to be p where p is given by

$$p = e^{-\lambda t}. \tag{3.5}$$

3.4.1.1 2-Replicated 2-Dilated Banyan Network

In the R2D2 network, there are exactly two distinct paths between each input-output pair. Decisions are made at the multiplexer stage to route the incoming packet through one of the paths. After the multiplexer stage each SE in the TP has to be fault-free. The R-graph for an $N \times N$ R2D2 network is shown in Figure 3.14. From the R-graph it can be noted that there are only two paths from a given input port to a particular output port. There the TR of the R2D2 network is given by

$$TR(t) = 1 - (1 - p^n)^2. \tag{3.6}$$

3.4.1.2 Balanced Gamma Network

In the BG network, a packet has an option of being routed through either the regular or the alternate link at each stage. The R-graph of the BG network is not the same for each input-output pair and is dependent on the destination. Certain input-output pairs use only a pair of SEs in each stage of their TPs and hence have a lower TR than the other pairs which use more than two SEs at certain stages.

The TR of an input-output pair is the lowest when the least significant $\lceil \frac{n}{2} \rceil$ tag bits are identical. Therefore, the number of former input-output pairs decreases with an increase in the size of the BG network. While Figure 3.15 shows the R-graph corresponding to a destination in a 16 × 16 BG network having the best TR, Figure 3.16 shows an example of the worst case. Figure 3.17 shows the R-graph of a general $N \times N$ BG network having the worst case TR. At each stage one of the critical pair of SEs has to be fault free and there are $n-2$ such stages for a TP. Therefore, the worst case TR of the BG network is given by

$$TR(t) = (1-(1-p)^2)^{n-2}, \tag{3.7}$$

where p is the reliability of each SE in stages 1 to n-1 which is equal to $e^{-\lambda t}$.

Due to rapid advancements in VLSI technology, component (SE) failure rates of 10^{-6} failures/hour (or better) are quite realistic. The worst case terminal reliabilities of the R2D2 and the BG network for 10 years, assuming a failure rate of $\lambda = 10^{-6}$, are given in Table 3.3. The exact terminal reliability of the R2D2 and the BG networks are obtained by multiplying the figures in the table by the reliabilities of the components which are assumed to be fault-free, viz., the SEs in the first stage, the destination buffers and the links connecting the last stage SEs to the buffers in case of the BG network and the multiplexers, destination buffers and the links connecting the last stage SEs to the buffers in case of the R2D2 network. From Table 3.3 it can be concluded that the BG network has better TR than the R2D2 network. It is also to be noted that the TR of both these networks decreases with an increase in the size of the network. This is due to the fact that larger networks have more SEs and hence the probability of failure of SEs in these network is more.

3.4.2 Broadcast Reliability

Broadcast reliability (BR) is the probability that at least one path exists from a particular input port to all the output ports. BR is always associated with a broadcast path (BP) which is a connection from one source to all destinations in the network. BR is usually referred to as the source-to-multiple terminal (SMT) reliability [33]. Under this criterion, the network is considered failed when a connection cannot be made from a given input port to at least one of the output ports. For both the R2D2 and the BG networks, broadcast capability can be obtained by having a copy network [11] in order to make N copies of an incoming packet having routing tags corresponding to each of the outputs. These N copies of the incoming packet have to be buffered in a queue and supplied to the corresponding input port.

3.4.2.1 2-Replicated 2-Dilated Banyan Network

The BR for each input port is identical in the case of the R2D2 network and therefore the R2D2 network has a uniform BR. The BR R-Graph for an 8×8 R2D2 network is shown in Figure 3.18. In the R2D2 network there are exactly two BP for each input port due to the replication. All SEs in the BP in each of the subnetworks of the R2D2 networks have to be reliable in order to establish broadcast connection. The BR R-Graph of the R2D2 network shown in Figure 3.18. It can be seen that all the SEs in the BP have to be fault free and there are 2 such BPs. Therefore, the BR of an $N \times N$ R2D2 network can be expressed as

$$BR(t) = 1 - \left(1 - \prod_{i=0}^{n-1} p^{2^i}\right)^2 . \tag{3.8}$$

3.4.2.2 Balanced Gamma Network

The BR for each input port of the BG network is also identical as in the case of the R2D2 network and therefore the BG network has a uniform broadcast reliability. The broadcast reliability R-graph for an 8×8 BG network is shown in Figure 3.19. The equivalent reduced broadcast reliability R- graph for an 8×8 BG network is given in Figure 3.20. Each of the composite nodes in Figure 3.20, shown by a double circle, represents a pair of SEs in the network; the edge between two composite nodes indicates the set of edges between any one of the SEs in the first composite node and any one of the SEs in the second. From the broadcast reliability redundancy graph shown in Figure 3.20, the broadcast reliability of an $N \times N$ BG network is

$$BR(t) = \prod_{i=1}^{n-1}(1-(1-p)^2)^{2^i}. \tag{3.9}$$

The BRs of the R2D2 and the BG network for 10 years assuming a failure rate of $\lambda = 10^{-6}$ are given in Table 3.4

As in case of the TR, the exact BRs of the R2D2 and the BG networks are obtained by multiplying the figures in Table 3.4 by the reliabilities of the components which are assumed to be fault-free, viz., the SEs in the first stage, the destination buffers and the links connecting the last stage SEs to the buffers in case of the BG network and the multiplexers, destination buffers and the links connecting the last stage SEs to the buffers in case of the R2D2 network. It can be seen from Table 3.4 that the BR of the R2D2 is very low when compared to that of the BG network. BRs of the R2D2 for network sizes of 512 and 1024 have not been provided as they are of very low values.

3.4.3 Network Reliability

Network reliability (NR) is the probability of maintaining full access capability throughout the network. Full access property is defined as the property of network to establish connection between each input-output pair. This measure considers the tolerable average number of switch failures [35]. NR is associated with network path (NP) which is a many-to-many connection, linking sources to many destinations.

3.4.3.1 2-Replicated 2-Dilated Banyan Network

According to the routing algorithm of the R2D2 network discussed in Section 1, one of the Banyan subnetworks of the R2D2 network has to be completely fault-free in order to establish connections between every input-output pair in the network. Therefore all the SEs in one of the subnetworks have to be fault-free for the R2D2 network to have full access property at a particular time t. Therefore the NR of the R2D2 network is given by

$$NR(t) = 1 - (1 - p^{\frac{N}{2} \times \log_2 N})^2. \qquad (3.10)$$

3.4.3.2 Balanced Gamma network

The BG network loses full access property only when one or more critical pairs fail. Considering all the possible combinations of critical pair failures, we arrive at the NR of the BG network as

$$NR(t) = 1 - \sum_{i=2}^{TN} F_i \times (1-p)^i \times p^{TN-i}, \qquad (3.11)$$

where,

$T.N = N \times (n-1)$- Total SEs in the BG network excluding the SEs of the first stage.

F_i - Total possible combinations of failure of i SEs causing the BG network to lose full access property where $2 \leq i \leq T.N$.

As N increases the calculation of F_i becomes a very tedious process. The evaluation of NR of a network is known to be an NP-hard problem [36]. Using Equation 3.11, with a failure rate of $\lambda = 10^{-6}$, the NR for an 8×8 BG network is given by

$$NR(t) = 1 - \sum_{i=2}^{16} F_i \times (1-p)^i \times p^{16-i} = 0.922639 \tag{3.12}$$

One alternative approach is to develop approximate methods as the one given in [36].

For the BG network to retain full access, none of the critical pairs can be faulty. Both the SEs in any critical pair are always located within a stage of the network. Therefore, each stage of the BG network has to be reliable for the BG network to be reliable. Stages 1 to $(n-2)$ are identical as they have N critical pairs and so each of them has the same reliability; stage $(n-1)$ has only $\frac{N}{2}$ critical pairs and so has a different reliability. Since the reliability of each stage of the BG network is independent of other stages, the NR of the BG network is the product of reliabilities of the each stage in the BG network. The stage reliability (SR) of stage i, $(i \epsilon \{1, 2, ..., n-2\})$ is given by

$$SR(t) = \sum_{i=0}^{N} INF_i \times (1-p)^i \times p^{N-i}. \tag{3.13}$$

and the reliability of the last stage (LSR) is given by

$$LSR(t) = \sum_{i=0}^{N} LNF_i \times (1-p)^i \times p^{N-i}. \tag{3.14}$$

where INF_i and LNF_i indicate all possible combinations of i SE failures in the intermediate stages and in the last stage respectively, which do not make the BG network lose full access. The NR of the BG network is given by

$$NR_1(t) = SR(t)^{n-2} \times LSR(t). \tag{3.15}$$

The NR of the BG network given by $NR_1(t)$ is slightly less than the actual NR of the BG network because Equation 3.15 does not take into account the different possible combinations of SE failures in the overall network. The difference in NR given by Equation 3.11 and Equation 3.15 is 8.13574×10^{-4} for an 8×8 BG network. The computation of INF_i and LNF_i is cumbersome for larger values of N. Moreover, the values of INF_i and LNF_i for i=0 and 1 have a major contribution to the values of SR and LSR respectivelv. Therefore we truncate the values of LR and SR at $i = 4$ and get the NR lower bound of the BG network as

$$\begin{aligned} NR_{LOW}(t) \;=\; & \left(\sum_{i=0}^{3} INF_i \times p^{N-i} \times (1-p)^i\right)^{n-2} \times \\ & \left(\sum_{j=0}^{3} LNF_j \times p^{N-j} \times (1-p)^j\right) \end{aligned} \tag{3.16}$$

The computation of F_i is even more complex than INF_i and LNF_i. The values of F_2 and F_3 constitute the major factor in Equation 3.10. Therefore, the upper bound of NR obtained by approximating NR given in Equation 3.10 is

$$NR_{UP}(t) = 1 - \sum_{i=2}^{3} F_i \times (1-p)^i \times p^{TN-i}. \tag{3.17}$$

The general expressions for INF_i, LNF_i and F_i used in $NR_{LOW}(t)$ and $NR_{UP}(t)$ are

$$
\begin{aligned}
INF_0 &= 1 \\
INF_1 &= N \\
INF_2 &= {}^{N}C_2 - N \\
INF_3 &= {}^{N}C_3 - N \times (N-3) \\
LNF_0 &= 1 \\
LNF_1 &= N \\
LNF_2 &= {}^{N}C_2 - \frac{N}{2} \\
LNF_3 &= {}^{N}C_3 - \frac{N}{2} \times (N-2) \\
F_2 &= N \times (n-2) + \frac{N}{2} \\
F_3 &= N \times (N-3) \times (n-2) + \frac{N}{2} \times (N-2) + \\
&\quad N^2 \times (n-2)^2 + \frac{N}{2} \times N \times (n-2)^2
\end{aligned}
$$

For an 8 × 8 BG network the values of NR(t), $NR_{LOW}(t)$, and $NR_{UP}(t)$ are

$$
\begin{aligned}
NR(t) &= 0.92263944 \\
NR_{LOW}(t) &= 0.92182587 \\
NR_{UP}(t) &= 0.94496154
\end{aligned}
$$

It can be clearly seen that the NR of an 8 × 8 BG network is closer to the lower bound than to the upper bound.

Assuming a failure rate of $\lambda = 10^{-6}$, the NR for different sizes of the R2D2 network is given in Table 3.5 and the lower and upper bounds of NR for the BG networks of different sizes are given in Table 3.6.

It can be seen from Tables 3.5 and 3.6 that the NRs of the R2D2 and the BG networks drop down with an increase in the network size similar to their terminal and broadcast reliabilities. Since NR has been shown to be closer to the lower bound, it can be concluded that the NR of the BG network reduces to quite low values with an increase in the network size.

Since the values of NR of the R2D2 networks of sizes greater than 64, the lower bound of the BG networks are nearly zero and those of the upper bound of the BG networks are one for $N > 128$, these values have not been included in the above table.

The lower bound $NR_{LOW}(t)$ values of the BG network suggest the the BG network to be more more reliable than the R2D2 network. Larger size BG networks exhibit higher reliability than those of similar size R2D2 networks. It should also be noted that the actual value of NR would be much lower if we take into account the network components which have been considered to be fault-free during this analysis. However, NR gives the probability that full access is never lost over a period of ten years assuming no repair is possible.

3.4.4 Input-Output MTTF

The input-output MTTF is the expected time a network will be functional before the failure of at least one of its terminal paths. It is given by the formula

$$MTTF_T = \int_0^\infty TR(t)dt. \tag{3.18}$$

3.4.4.1 2-Replicated 2-Dilated Banyan Network

Substituting the TR(t) of the R2D2 network given in Equation 3.6, the input-output MTTF for the R2D2 network is calculated to be

$$MTTF_T = \int_0^\infty 1 - (1 - p^N)^2 dt. \tag{3.19}$$

3.4.4.2 Balanced Gamma Network

Similar to the input-output MTTF of the R2D2 network, the input-output MTTF for the BG network is calculated as

$$MTTF_T = \int_0^\infty (1 - (1 - e^{-\lambda t})^2)^{n-1} dt. \tag{3.20}$$

The input-output MTTF for different sizes of the R2D2 and the BG networks are given in Table 3.7.

It can be noted from Table 3.7 that the input-output MTTF of the BG network is nearly twice than that for the R2D2 network. It can be concluded that, if t is the expected time that the BG network will be successful in establishing any terminal connection before it is unsuccessful in one or more such terminal connections, then $\frac{t}{2}$ is the corresponding expected time for the R2D2 network.

3.4.5 Network MTTF

The network MTTF is the expected time a network will be functional before it loses the full access property. The network MTTF for the BG network is given by

$$MTTF_N = \int_0^\infty NR(t)dt. \tag{3.21}$$

3.4.5.1 2-Replicated 2-Dilated Banyan Network

Substituting the NR(t) of the R2D2 network given in Equation 3.10 in Equation 3.21, the network MTTF for the R2D2 network is calculated to be

$$MTTF_N = \int_0^\infty 1 - (1 - p^{\frac{N}{2} \times \log_2 N})^2 dt. \tag{3.22}$$

3.4.5.2 Balanced Gamma Network

Since the exact NR of the BG network is difficult to compute we use the lower bound given in Equation 3.16 as this is almost certainly closer to the exact NR than the upper bound given in Equation 3.17. Substituting this lower bound expression in Equation 3.21 we have

$$MTTF_{N-LOW} = \int_0^\infty \left(\sum_{i=0}^{3} INF_i \times p^{N-i} \times (1-p)^i \right)^{n-2} \times \left(\sum_{j=0}^{3} LNF_j \times p^{N-j} \times (1-p)^j \right) dt \tag{3.23}$$

$$\tag{3.24}$$

The network MTTF for different sizes of the R2D2 network is given by Table 3.8 and the network MTTF lower bound for various sizes of the BG network are shown in Table 3.9. A failure rate of $\lambda = 10^{-6}$ has been assumed in calculating the values given in Tables 3.8 and 3.9. It can be seen from Tables 3.8 and 3.9, that for $N=$

1024, the BG and the R2D2 networks lose their full access property in around 2 months and 12 days respectively. Therefore, if repair is not possible, each SE of the large sized R2D2 and BG networks should have very low failure rates, say of the order of 10^{-10}.

3.5 Summary

The HC, FT and reliability of the MINs have been studied in this chapter. The three MINs – the BB, the R2D2 and the BG networks were found to be of comparable HC. The Banyan and the BB networks are unipath networks and therefore do not possess any FT properties. The R2D2 and the BG networks were found to be single fault-tolerant and robust under multiple faults. Reliability analysis of the R2D2 and the BG networks showed that the BG network had better terminal, broadcast and network reliabilities due to the existence of multiple paths between each input-output pair within the network. The reliability of these networks reduced considerably with increase in size. Even though the R2D2 and the BG network are more reliable than the Banyan and the BB networks, they have to have higher reliability measures in order to be considered for use in broadband packet switch architectures. In order to have higher reliability, either the reliability of the basic network components has to be increased or repair has to be possible. Since the BB, the R2D2 and the BG networks are of comparable hardware complexities, performance study of these networks under different traffic patterns will give a better idea of how well the hardware has been utilized. Next chapter studies the performance of these networks under certain traffic types.

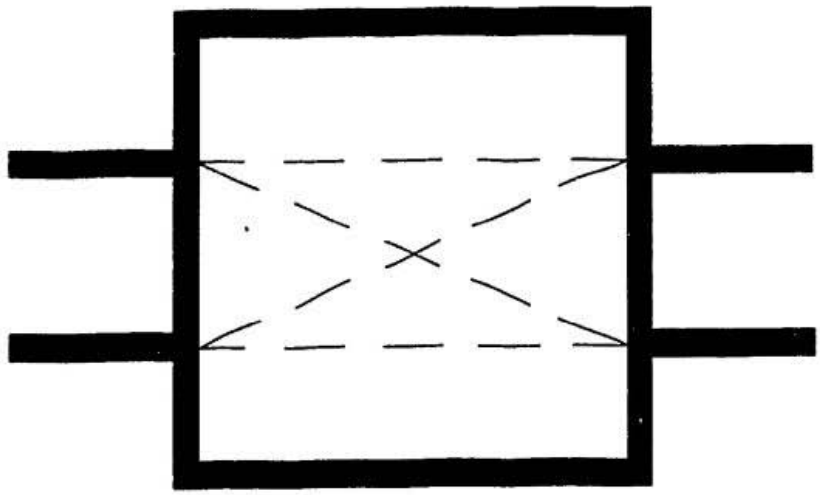

Figure 3.1: Crosspoint complexity of an 2×2 SE.

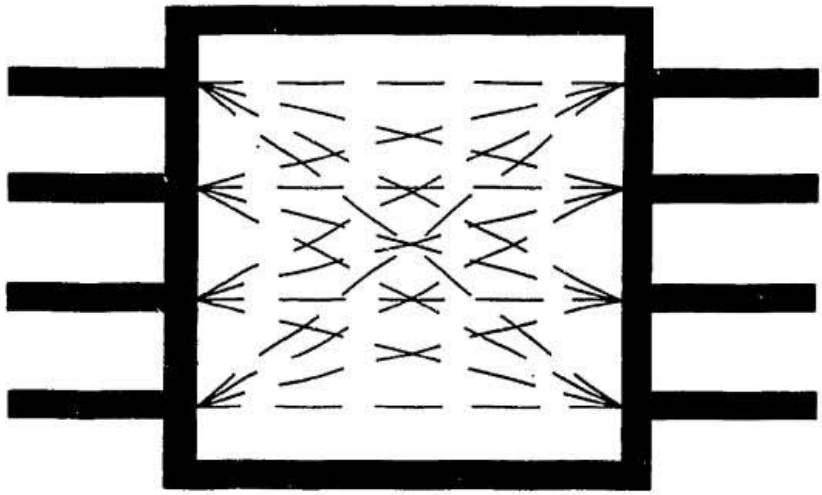

Figure 3.2: Crosspoint complexity of an 4×4 SE.

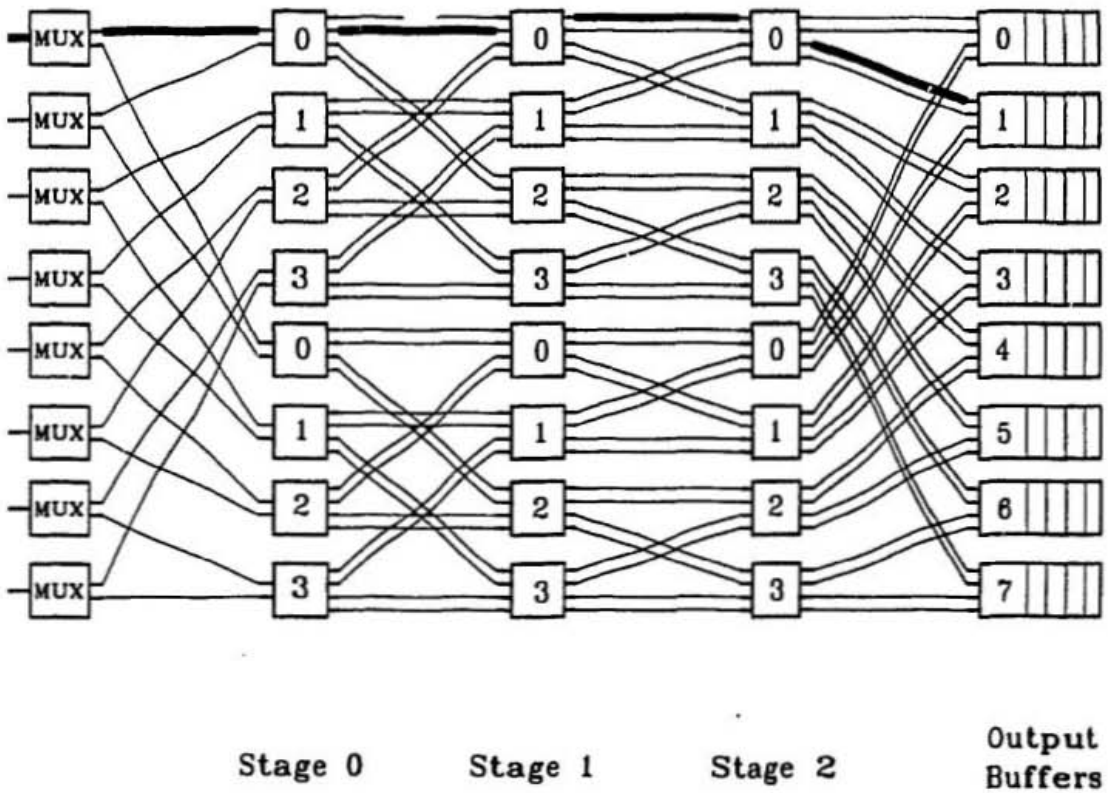

Figure 3.3: Routing in R2D2 network in case of link fault.

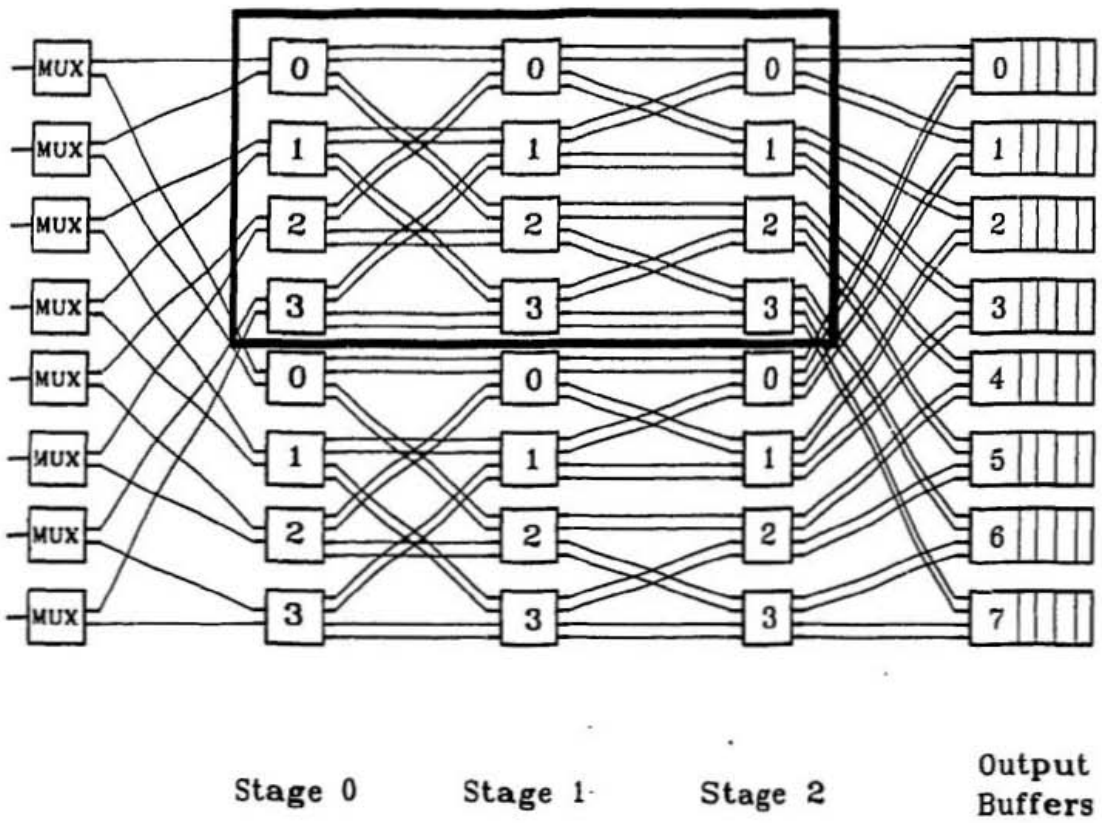

Figure 3.4: An 8×8 single-fault tolerant R2D2 network with maximum SE faults.

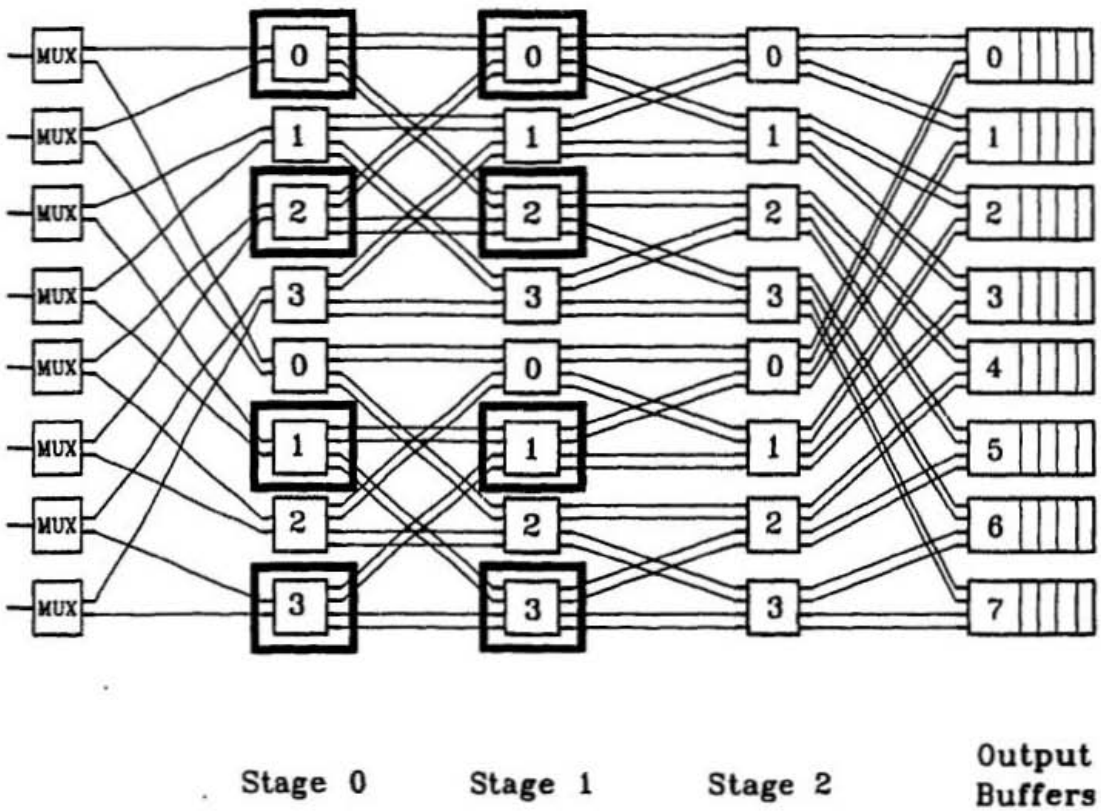

Figure 3.5: An 8×8 single-fault tolerant R2D2 network with maximum SE faults in each subnetwork.

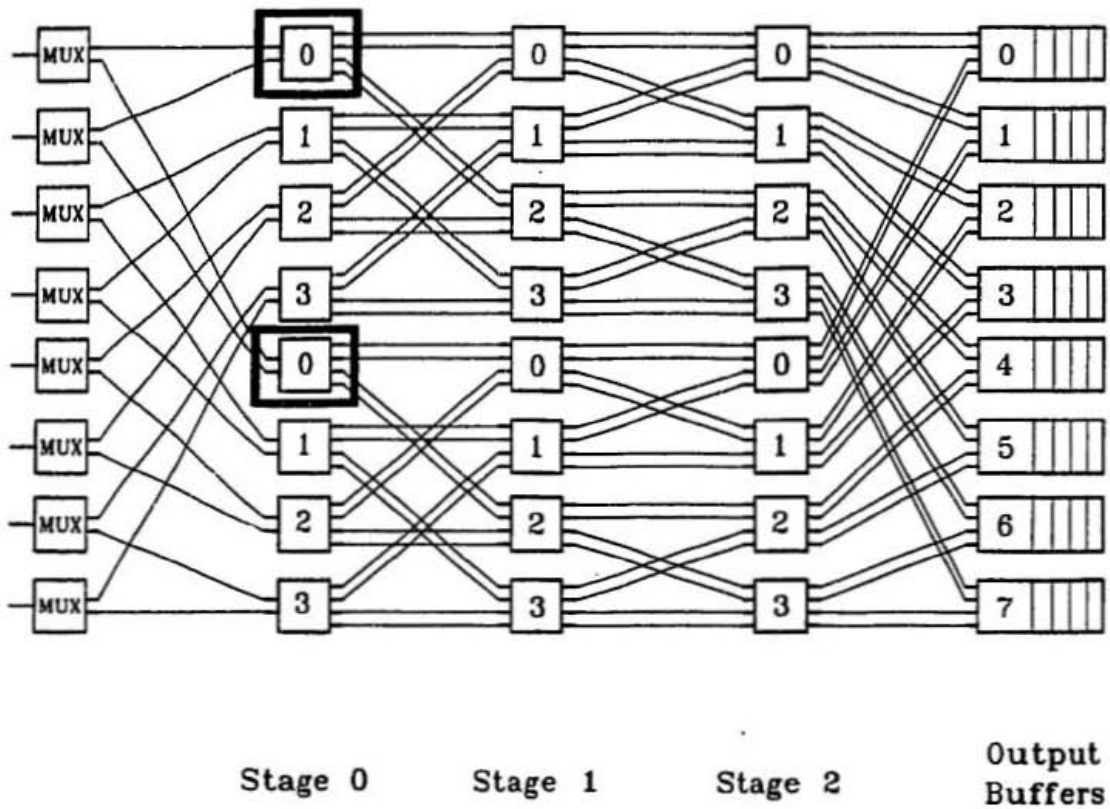

Figure 3.6: An 8 × 8 R2D2 network losing full access property due to SE faults.

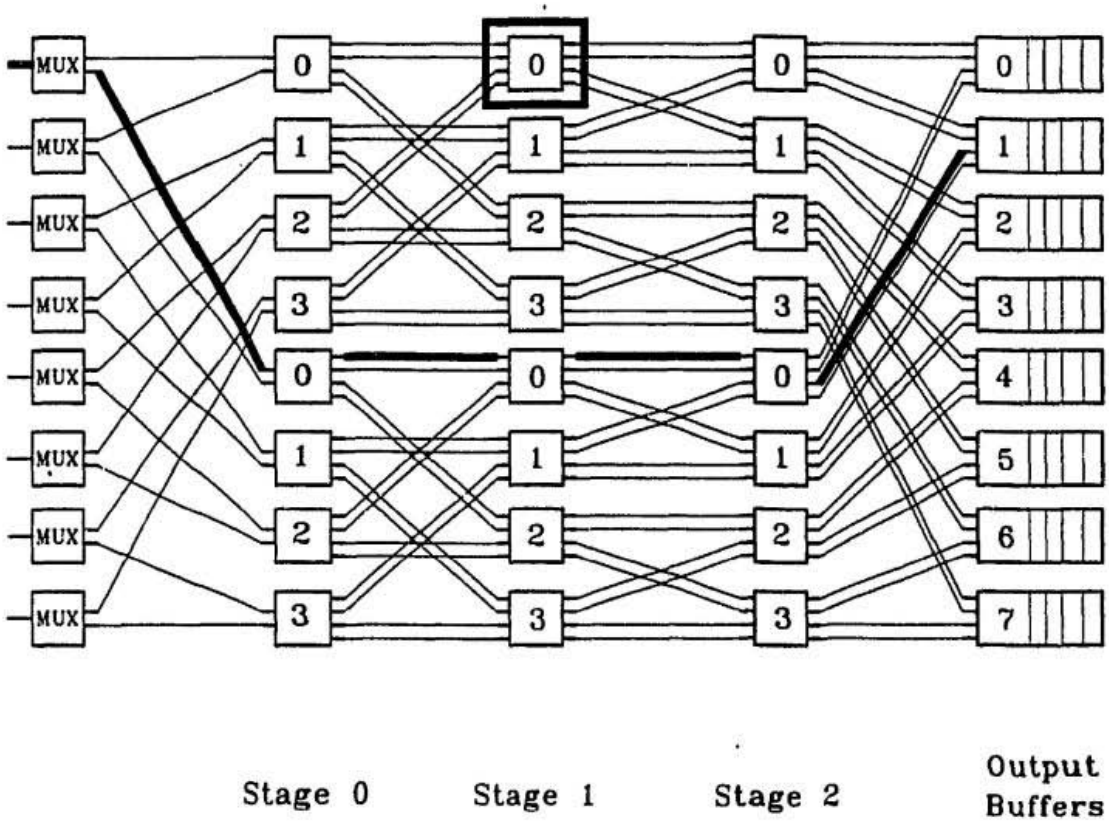

Figure 3.7: Routing in R2D2 network in case of SE fault.

Figure 3.8: Connection pattern for an SE in the BG network.

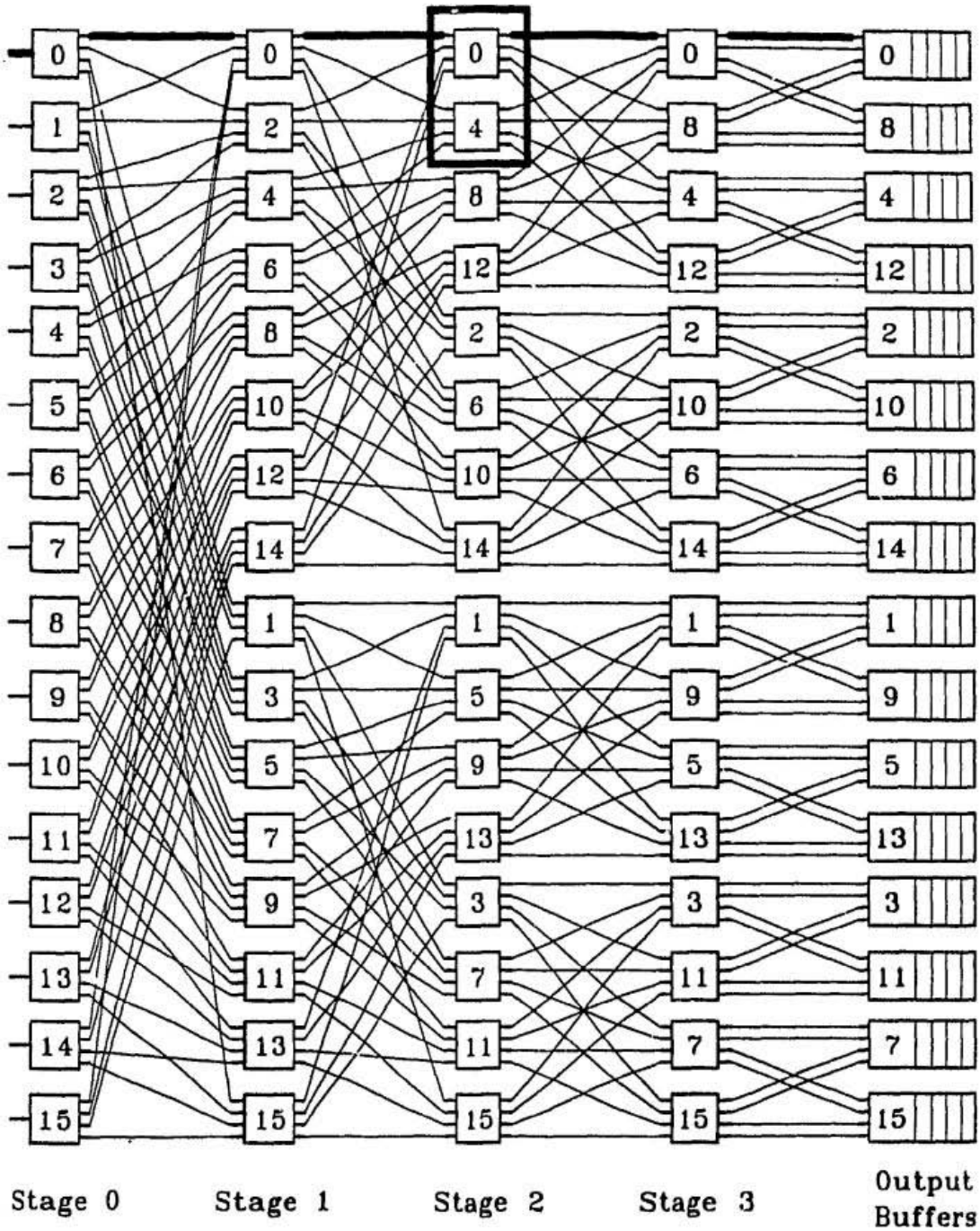

Figure 3.9: Loss of full access property by the BG network due to SE faults.

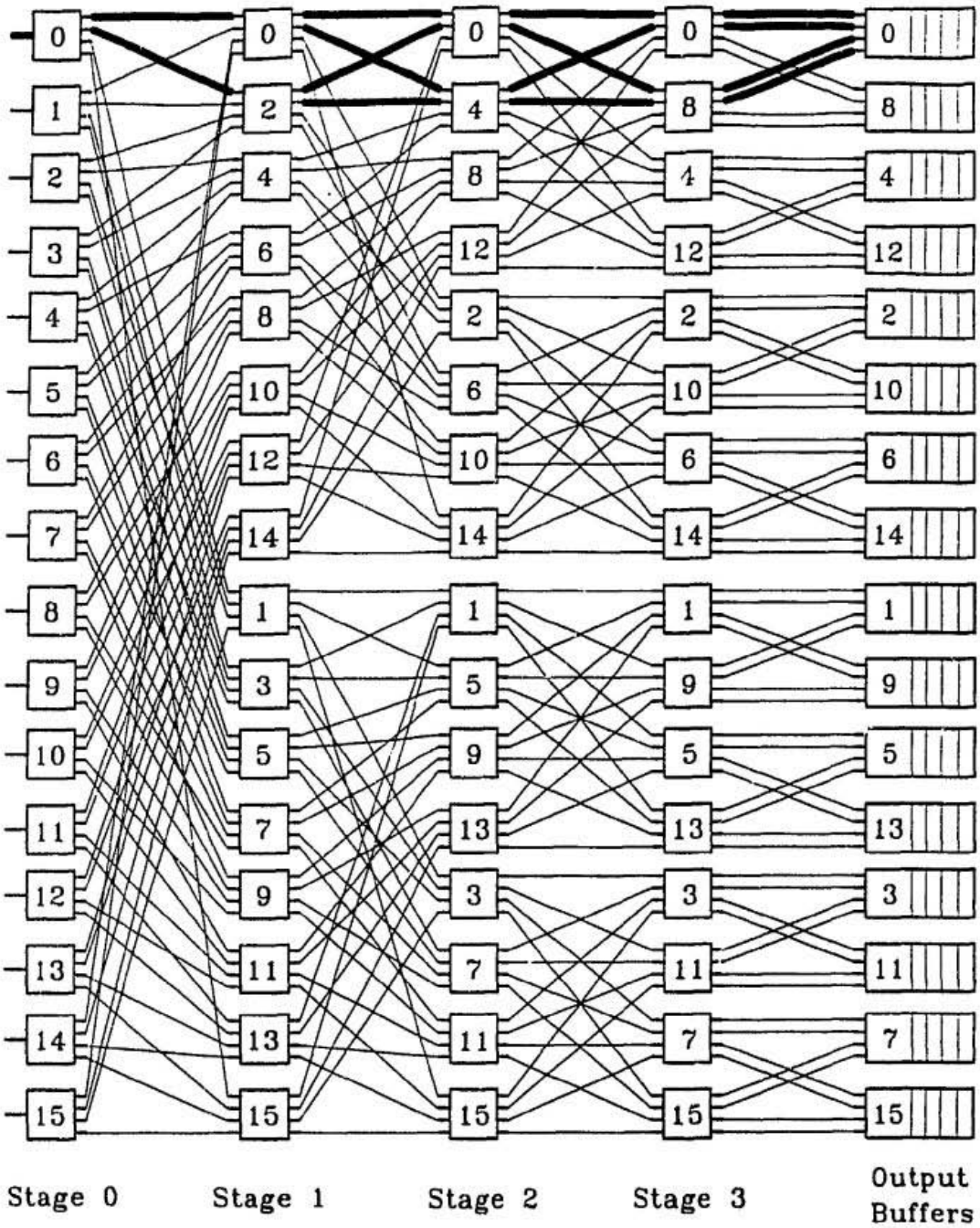

Figure 3.10: Different paths available during routing of a packet in a 16 × 16 BG network.

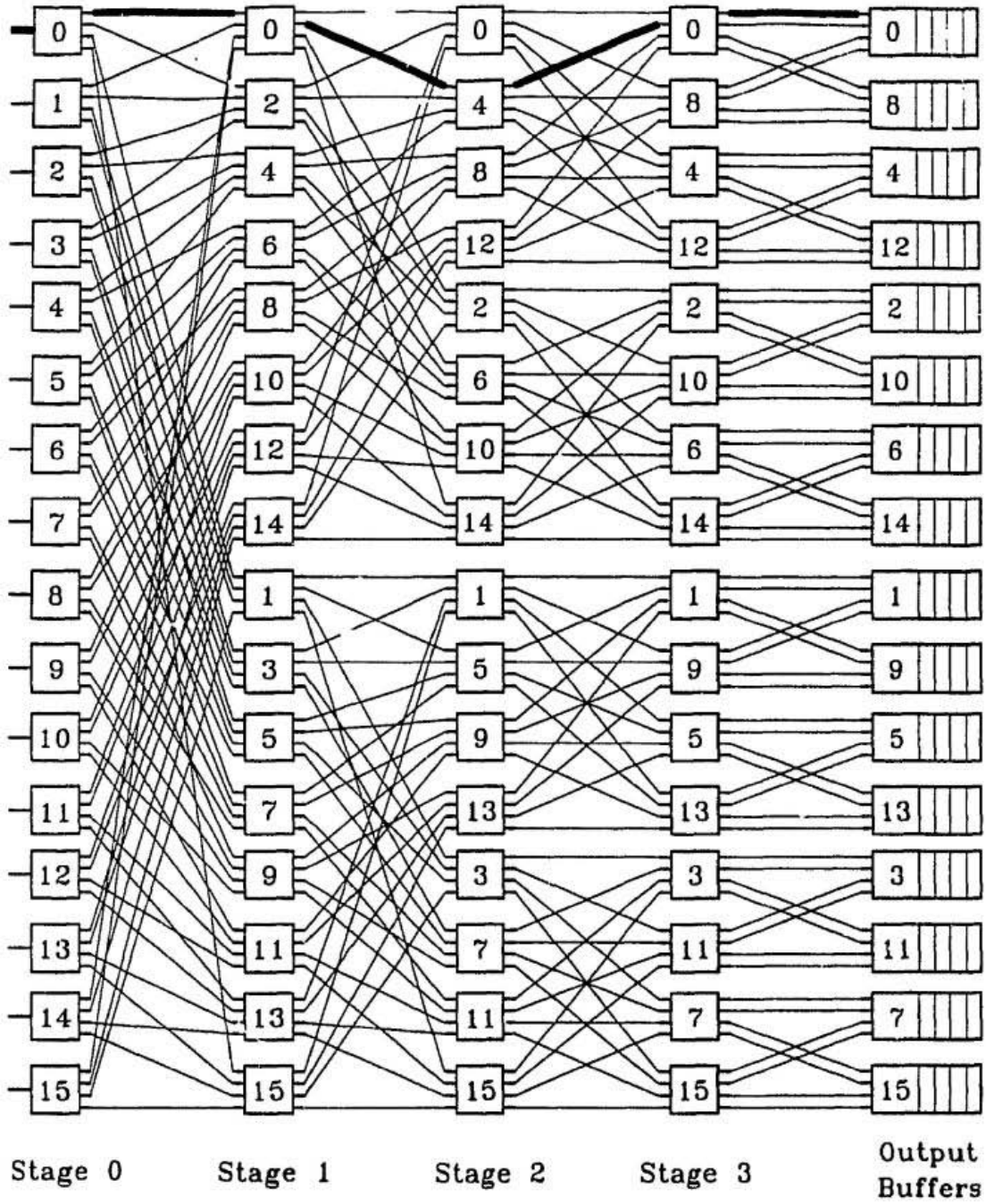

Figure 3.11: Dynamic rerouting in a 16 × 16 BG network in case of link fault.

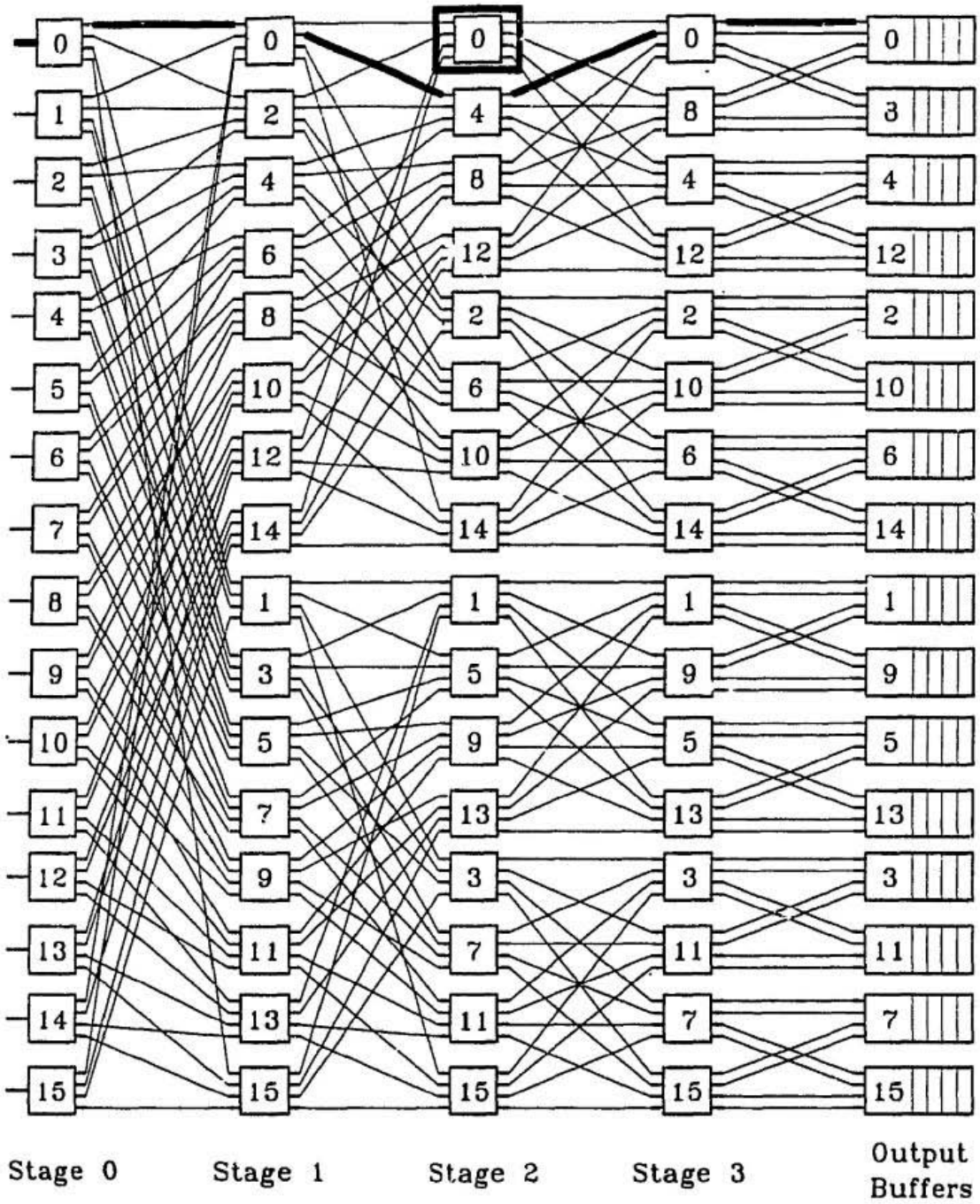

Figure 3.12: Dynamic rerouting in an 16 × 16 BG network in case of SE fault.

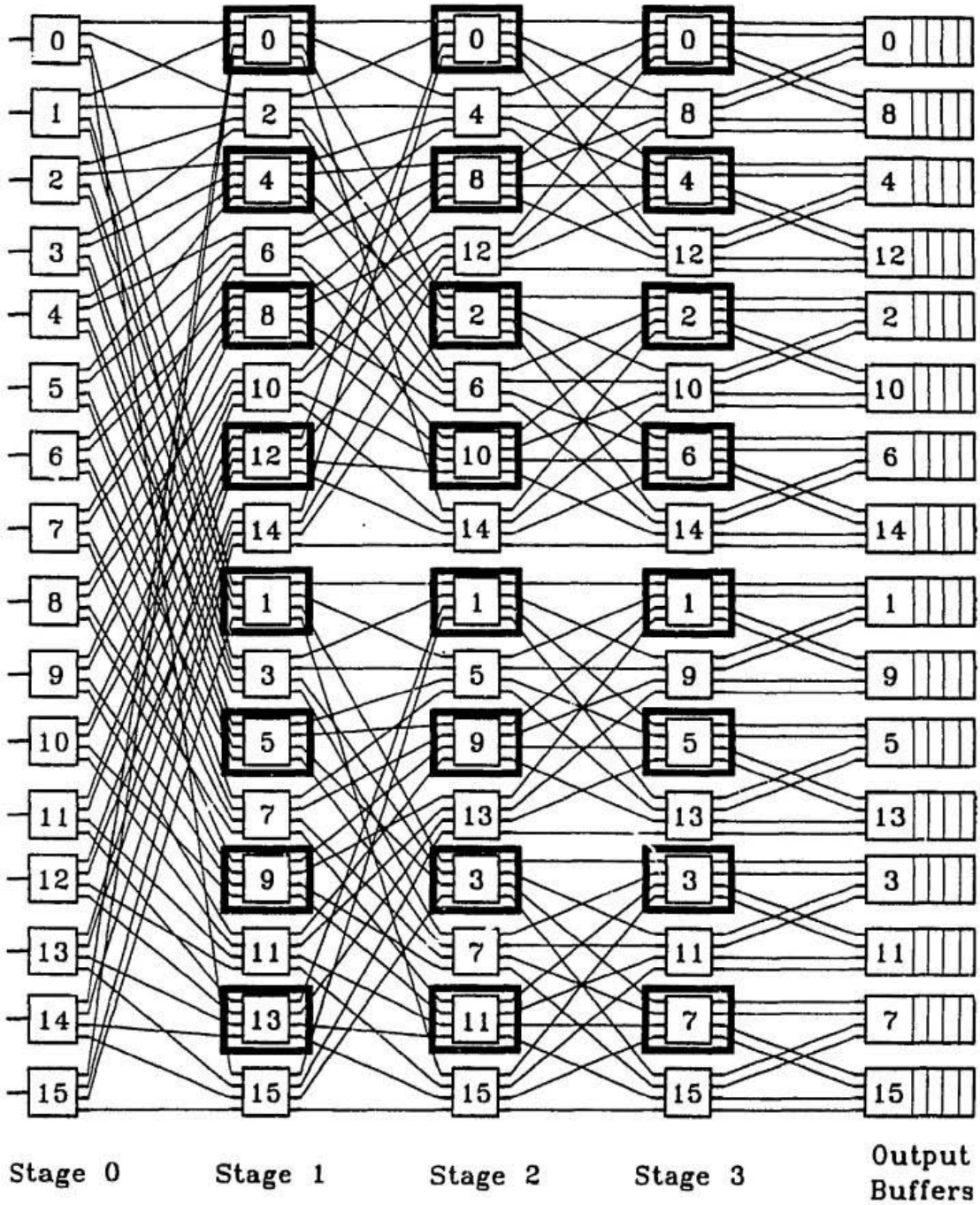

Figure 3.13: Single-fault tolerant 16 × 16 BG network with maximum SE faults.

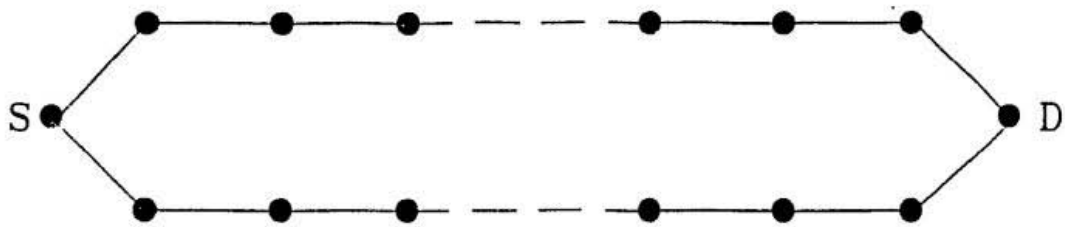

Figure 3.14: TR R-Graph of an $N \times N$ R2D2 network.

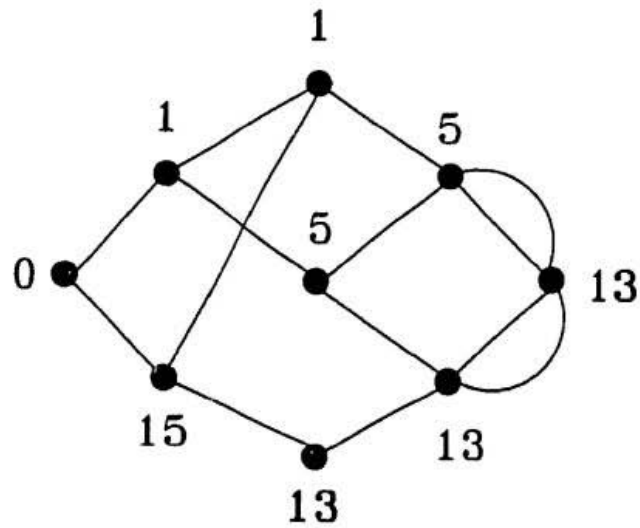

Figure 3.15: R-Graph of a 16 × 16 BG network having best case TR.

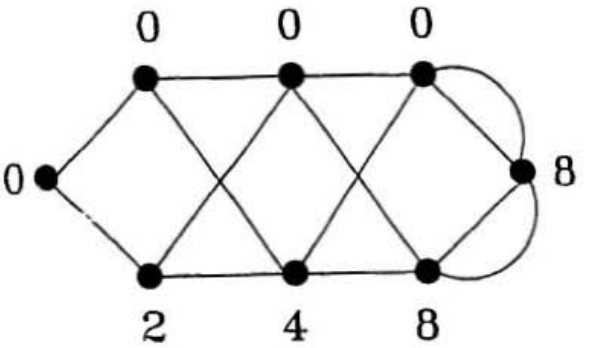

Figure 3.16: R-Graph of a 16 × 16 BG network having worst case TR.

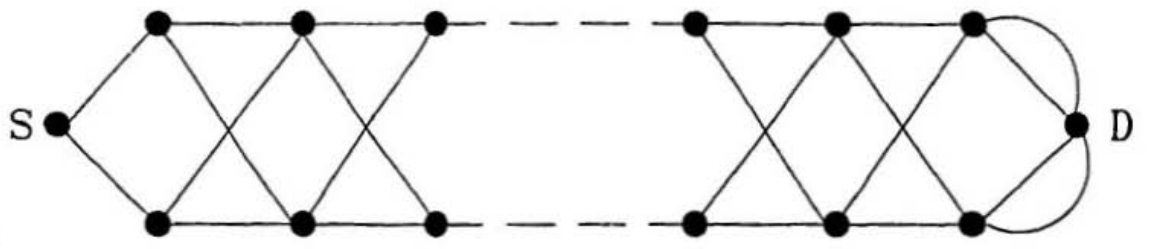

Figure 3.17: R-Graph of an $N \times N$ BG network having worst case TR.

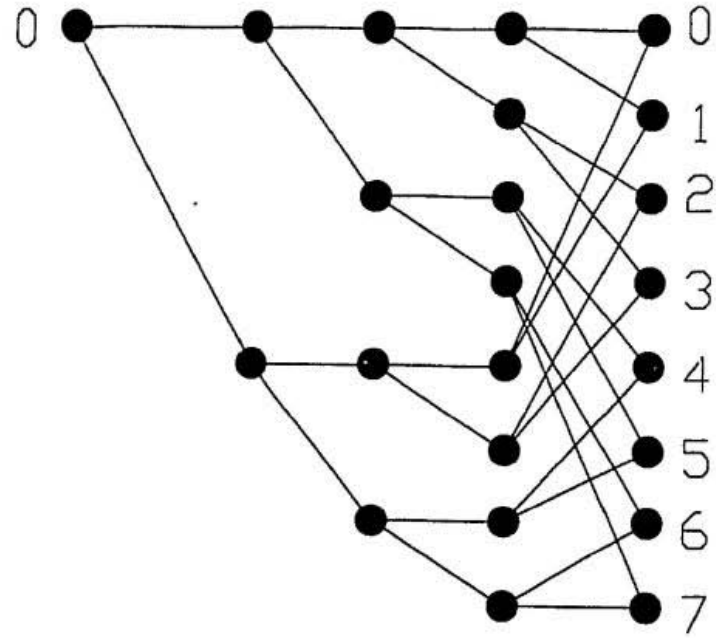

Figure 3.18: Broadcast Reliability R-Graph of an 8×8 R2D2 network.

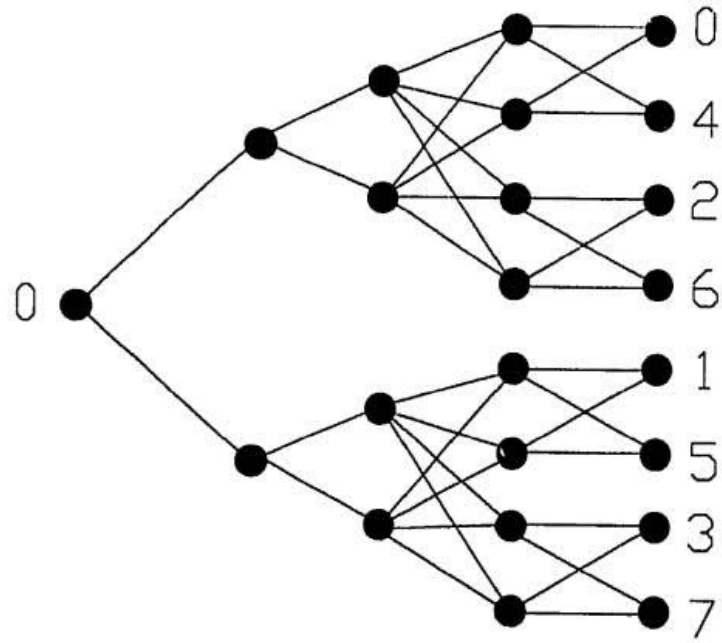

Figure 3.19: Broadcast Reliability R-Graph of an 8×8 BG network.

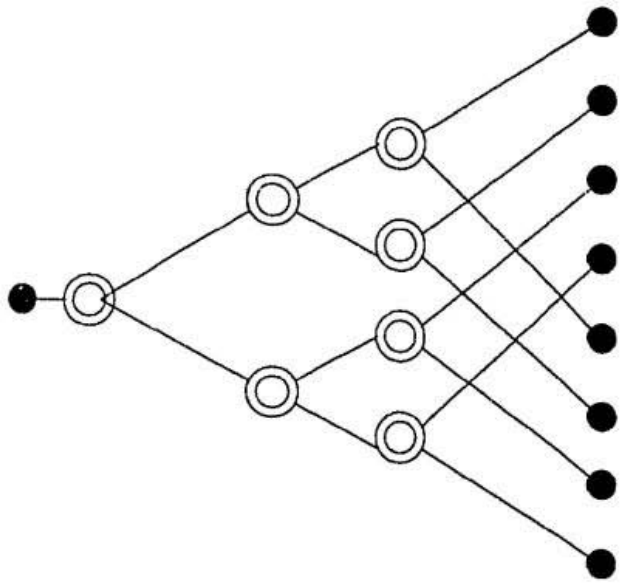

Figure 3.20: Reduced Broadcast Reliability R-Graph of an 8×8 BG network.

Network Size	Banyan Network	BB Network	R2D2 Network	BG Network
2	4	$4^{+}+4$	34	8
4	16	$24^{+}+16$	130	80
8	48	$96^{+}+48$	388	288
16	128	$320^{+}+128$	1032	832
32	320	$960^{+}+320$	2576	2176
64	768	$2688^{+}+768$	6176	5376
128	1792	$7168^{+}+1792$	14400	12800
256	4096	$18432^{+}+4096$	32896	29696
512	9216	$46000^{+}+9216$	73984	67584
1024	20480	$112640^{+}+20480$	164352	151552

Table 3.1: HC of MINs for varying values of N

Network	Fault Model	Fault-Tolerance Criterion	Routing Complexity	Hardware Complexity	Fault-Tolerance Capability
Extra Stage Cube	As strict	Same	Comparable	Slightly greater	Comparable
Augmented C-network	As strict	Same	Comparable: complexity hidden in switch	Much greater	Comparable
Merged Delta	Less strict	Same	Comparable: complexity hidden in switch	Much greater	Greater
F-network	Less strict	Same	Comparable	Slightly greater	Comparable
BG network	Slightly less strict	Same	Comparable, complexity hidden in switch	Greater	Comparable
R2D2 network	Less strict	Same	Comparable, complexity hidden in switch	Greater	Comparable

Table 3.2: Comparison of properties of MINs with the hypothetical network and other networks discussed in [28].

Network Size	R2D2 Network	BG Network
4	0.974140	0.992956
8	0.946527	0.985962
16	0.912524	0.979017
32	0.874069	0.972121
64	0.832709	0.965273
128	0.789675	0.958474
256	0.745934	0.951723
512	0.702240	0.945019
1024	0.659172	0.938362

Table 3.3: Worst case terminal reliabilities of different sizes of the R2D2 and the BG network.

Network Size	R2D2 Network	BG Network
4	0.946527	0.985962
8	0.789675	0.958474
16	0.464910	0.905776
32	0.127723	0.808913
64	0.007975	0.645155
128	2.925E-05	0.410382
256	3.9E-10	0.166049
512	-	0.027185
1024	-	0.000729

Table 3.4: Broadcast reliability for different sizes of the BG network

Network Size	$NetworkReliability$
4	0.912526
8	0.576546
16	0.117339
32	1.8 E-03
64	9.8E-08

Table 3.5: Network reliability for different sizes of the R2D2 network

Network Size	$NR_{LOW}(t)$	$NR_{UP}(t)$
4	0.985962	0.986012
8	0.921826	0.944962
16	0.727123	0.971385
32	0.199360	0.999811
64	1.637E-06	1.000000
128	7.27E-19	1.000000

Table 3.6: Network reliability bounds for different sizes of the BG network

Network Size	R2D2 Network	BG Network
4	0.750000	1.500000
8	0.500000	0.916667
16	0.375000	0.700000
32	0.300000	0.582143
64	0.250000	0.506349
128	0.214286	0.452742
256	0.187500	0.412421
512	0.166667	0.380760
1024	0.150000	0.355094

Table 3.7: Input-Output MTTF (in 10^6 hours) for different sizes of the R2D2 and the BG Networks.

Network Size	$MTTF_N$ (in 10^6 hours)
4	0.375000
8	0.125000
16	0.046875
32	0.018750
64	0.007813
128	0.003348
256	0.001465
512	0.000651
1024	0.000293

Table 3.8: Network MTTF for different sizes of the R2D2 network.

Network Size	$MTTF_{N-LOW}$ (in 10^6 hours)
4	0.916667
8	0.315664
16	0.136992
32	0.063382
64	0.029908
128	0.014223
256	0.006795
512	0.003258
1024	0.001568

Table 3.9: Lower Bound Network MTTF for different sizes of the BG network.

Chapter 4

Performance of Multistage Interconnection Networks

4.1 Introduction

Throughput performance of MINs is one of the major aspects to be researched for their use in broadband communication switch fabrics. This chapter discusses the performance of MINs discussed in earlier chapters, under uniform and non-uniform traffic patterns. All packets arriving at an input are assumed to be switched during the same cycle. It is also assumed that there are no network component failures within the MINs.

Performance of MINs is mainly based on the extent of blocking within MINs. Throughput and cell loss rate are the main parameters of importance in performance analysis of MINs for use in switch fabrics of broadband packet switch architectures. In this thesis work performance analysis of the MINs has been done without buffers.

Throughput of a MIN is defined as the ratio of total number of delivered packets to the total number of incoming packets over a given period of time. If *total_input* packets are the packets which are fed to the MIN and *total_output* cells are the packets delivered to the destinations during a given period of time t, then, throughput

is given by

$$Throughput = total\_output/total\_input. \tag{4.1}$$

Cell loss rate is defined as the ratio of total number of lost packets to the total number of incoming packets over a given period of time. If *total_lost* packets denotes the total packets lost within the MIN and *total_input* packets are the packets which are fed to the MIN over a given period of time t, then the cell loss rate is given by

$$Cell\ loss\ rate = total\_lost/total\_input. \tag{4.2}$$

In case of MINs without buffers the cell loss rate is also given by

$$Cell\ loss\ rate = 1 - Throughput. \tag{4.3}$$

Section 4.2 discusses blocking in MINs. Section 4.3 explains about the expected traffic patterns in B-ISDN. Performance analysis of MINs under uniform traffic patterns – permutation traffic and uniform random traffic patterns are explained in Section 4.4. The performances of MINs under non-uniform traffic types hotspot traffic, community of interest traffic and bursty traffic are explained in Section 4.5. Finally, a summary is provided in Section 4.6.

4.2 Blocking in MINs

Blocki..g in MINs is defined as failure to satisfy a connection requirement. Based on this property networks can be classified as strict sense nonblocking, wide sense nonblocking, rearrangeably nonblocking and blocking networks [43].

A network is said to be *nonblocking in the strict sense* or *strict − sense nonblocking* if any desired connection between unused ports can be established immediately without interference from any arbitrary existing connections.

A network is said to be *nonblocking in the wide sense* or *wide−sense nonblocking* if any desired connection between two unused ports can be established immediately without interference from arbitrary existing connections, provided that the existing connections have been inserted using some routing algorithm. If the algorithm has not been followed, some attempted connections may be blocked.

A network is said to be *rearrangeable* or *rearrangeably nonblocking* when a desired connection between unused ports may be temporarily blocked, but can be established if one or more existing connections are rerouted or rearranged.

A network is said to be *blocking* if there exist connection sets that will prevent some additional desired connections from being established between unused ports, even with rearrangement of the existing connections.

It should be noted that the above terminology deals with the ability of the networks to make permutation connections which is explained in Section 4.2.1. The BB network is a wide-sense non-blocking network, while all other MINs explained in previous chapters are blocking networks.

The throughput of a MIN is mainly degraded due to two kinds of blocking within a MIN. They are internal blocking and output contention blocking.

The internal blocking can occur within an SE at any stage other than the input stage and the output stage of a MIN. For example in a Banyan network, when two incoming packets to an SE, $SE_{i,j}$ (j is any stage other than first or last stage), are to have the same tag bit for switching, then one of the packets is dropped arbitrarily.

This constitutes loss of packets and hence degradation of the throughput of the MIN. In general, if p incoming packets to an SE have the same routing tag, and the SE has q output links for that routing tag ($p > q$), then $p - q$ packets will be dropped. An example of internal blocking in an 8×8 Banyan network is shown in Figure 4.1. In this figure two packets from input ports 0 and 6 need to be routed to output ports 6 and 7 respectively. Internal blocking occurs at SE numbered 2 at stage 1.

Packets are also lost due to output contention blocking in a MIN. When p packets are at the output of a SE $SE_{i,laststage}$ and if the destination can accept only q packets ($q < p$) during a cycle, then arbitrarily $p - q$ are dropped of the p packets at the output of SEs of the last stage. This is known as output contention blocking. An example of output contention blocking in an 8×8 Banyan network is shown in Figure 4.2. In this figure two packets from input ports 0 and 3 have to be routed to output port 7. Since only one packet can be accepted in one cycle one of the packets is dropped.

4.3 Traffic Patterns in B-ISDN

When voice, video, data and other payloads are carried in an integrated network, different types of traffic can be expected. Although it may not be possible to exactly simulate realistic traffic patterns, study of MINs under certain traffic patterns may help in understanding their performances under realistic traffic patterns. Performance analysis of ATM networks has been reported in references [43] and [44].

The traffic patterns expected in broadband packet switch architectures can be

classified as uniform traffic patterns and non-uniform traffic patterns. The following two sections will explain about these traffic patterns in detail.

4.4 Uniform Traffic Patterns

The two main uniform traffic patterns which are being studied are *permutation traffic* and *uniform random traffic*. In the previous chapter it has been shown that the BG network, the BB network and the R2D2 network have comparable hardware complexities. A comparison of the performances of these MINs will enable us to understand how well the hardware has been utilized. The following sections give a comparison of the performances of the BB network, the R2D2 network and the BG network.

4.4.1 Permutation Traffic

The *permutation traffic* pattern refers to the case where the output ports requested by packets arriving to the switch in the same time slot are distinct from one another. It is referred to as permutation traffic because, at full load, the list of requested destinations ordered by input lines forms a permutation of the set (0, 1, ..., N-1).

The amount of internal blocking in the MINs is distinctly visible under permutation traffic pattern. In the case of permutation traffic, at full load, only one packet is destined to each destination. So packets will be lost only due to internal blocking in the network and not due to output contention at a particular destination.

4.4.1.1 Banyan and R2D2 networks

Both the Banyan and the R2D2 networks do not have 100% throughput under

permutation traffic patterns as they do not have non-conflicting paths for certain combinations of permutation traffic. Hence they are called blocking networks.

4.4.1.2 Batcher Banyan network

In the BB network, the incoming packets are sorted by the Batcher network, in either ascending or descending order based on their destinations. Due to this, packets are compacted and they are monotonically increasing or decreasing depending on the sorting employed. Then a perfect shuffle of these packets is done and given to the Banyan network. It has been shown that non-conflicting paths exist between each source and destination and hence the throughput of the BB network is always 1 under permutation traffic pattern [11] .

4.4.1.3 Balanced Gamma network

Due to the connection pattern of BG network, stage 0 and stage 1 of the BG network are nonblocking for any kind of traffic. Under permutation traffic, stage $(n-1)$ and stage $(n-2)$ are also non blocking. Hence an $N \times N$ BG network is nonblocking under permutation traffic, if and only if $N \leq 16$. The BG networks suffers cell loss for networks of size $N > 16$. These properties are in the following Lemmas and Theorem 1.

Lemma 1 *Stage 0 of an $N \times N$ BG network is non-blocking.*

Proof : Stage 0 of the BG network consists of 1×4 SEs. Each SE can receives only one packet per cycle. This packet can be routed through one of the regular links based on the switching bit. The alternate links at stage 0 are provided for FT and modularity.

□

Lemma 2 *Stage 1 of an $N \times N$ BG network ($N > 2$) is non-blocking.*

Proof : All SEs at stage 1 have four input links and four output links. The packets arriving at stage 0 are routed through the regular link to stage 1. Therefore any SE at stage 1 can at most receive two packets from stage 0 through the regular links. At any stage of BG network, packets are lost if and only if more than three packets have the same tag bit for routing to proceed to the next stage. Since the maximum number of incoming packets at stage 1 is only two there can be no loss of packets at this stage. When both the incoming packets to a SE at stage 1 have the same tag bit, one of them will be routed onto the regular link and the other onto the alternate link.

□

Similarly it can be shown that an $N \times N$ network, having similar connection patterns as that of the BG network, constructed of $k \times k$ SEs(N and k are powers of 2), will have its first $log_2 k$ stages non-blocking for any kind of traffic.

Lemma 3 *Stage $n-1$ of an $N \times N$ BG network is non-blocking under permutation traffic.*

Proof : Due to the connection pattern of the BG network, any SE in the $(n-1)th$ stage is connected to exactly two outputs. Moreover any two SEs separated by distance $N/2$ are connected to the same pair of outputs. Suppose we have two SEs SE1 and SE2 separated by a distance $N/2$, and if SE1 is connected to a particular output D for tag bit 0, then SE2 will be connected to the same output D for tag bit 1 and vice versa. Under permutation traffic each output receives exactly one

packet in each cycle. Therefore, any SE in stage n-1 will not receive more than two packets.

□

Lemma 4 *Stage $n-2$ of an $N \times N$ BG network($N > 2$) is non-blocking under permutation traffic.*

Sketch of proof : Due to the connection pattern of the BG network, any SE in stage $n-1$ is connected to only two outputs. Any two SEs at stage $n-2$ separated by distance $N/2$ are connected to the same pair of SEs in stage $n-1$. Let us assume that we have two SEs SE1 and SE2 separated by a distance $N/2$ in stage $n-1$. Let these be referred as $SE_{i,n-1}$ and $SE_{i+N/2,n-1}$. Due to the connection pattern of the BG network, two incoming links for $SE_{i,n-1}$ from stage $n-2$ are from SEs $SE_{i,n-2}$ and $SE_{i+N/2,n-2}$ which are separated by a distance $N/2$ at stage $n-2$. Similarly two incoming links for $SE_{i+N/2,n-1}$ from stage $n-2$ are also from the same SEs $SE_{i,n-2}$ and $SE_{i+N/2,n-2}$. This is shown clearly using a 16 × 16 BG network in Figure 4.3.

Here the two SEs selected are 7 and 15. These two SEs are chosen arbitrarily, but without loss of generality. Any two SEs separated by a distance 16/2 = 8 can be chosen. The SEs 7 and 15 of stage 3 are connected to outputs 7 and 15. Two links of SE 7 of stage 2 are connected to SEs 7 and 15 of stage 3. These two links correspond to the regular and alternate links for tag bit 1 for stage 2. Similarly SE 15 is also connected to SEs 15 and 7 of stage 3 for the same tag bit 1. This is shown by thick lines in Figure 4.3. So, any packet with tag bit 1 for stage 2, and arriving at one of the SEs 7 or 15 at stage 2, has to go to one of the outputs 7 or 15 due to connection pattern of BG network. During permutation traffic not more

than one packet is destined to a particular output at any given cycle. Therefore, at most two packets can arrive at one of the SEs 7 or 15, with tag bit 1. One packet will be routed onto the regular link and the other onto the alternate link. Hence there is no chance of packet loss at this stage which corresponds to stage n-1.

□

Lemma 5 *In an $N \times N$ BG network, any stage i ($i \notin$ 0, 1, n-1, n-2) is blocking.*
Sketch of proof : An $N \times N$ BG network contains a stage i, ($i \notin 0, 1, n-1, n-2$) only if $N > 16$. So, for proving this a 32×32 BG network is considered. Figure 4.4 shows a 32×32 BG network with connections showing the routing of packets from inputs 13, 15, 16 and 17 to outputs 15, 31, 19 and 23 respectively.

The routing tags for the above packets are 01111, 11111, 10011 and 10111 respectively. According to the routing tag the packets from SE's $SE_{13,0}, SE_{15,0}, SE_{16,0}$ and $SE_{17,0}$ are routed to SE's $SE_{13,1}, SE_{15,1}, SE_{17,1}$ and $SE_{17,1}$ respectively. SE $SE_{17,1}$ has two packets having the same tag bit 1. Let us assume that the packet destined to output 19 from input 16 takes the regular link and the packet destined to output 23 from input 17 takes the alternate link. Then at stage 2 we have three packets arriving at SE $SE_{15,2}$ and one packet to SE $SE_{19,2}$. Now at stage 2, all the three packets going out of $SE_{15,2}$ have routing tag bit 1, but since we have only one regular link and one alternate link, one of the packets is lost. Hence packets can be lost at stage 2 in a 32×32 BG network. For each SE in stage 2 we can find a permutation traffic, which will result in packet dropped. Since packets can be lost at stage 2 of a 32×32 BG network, stage 2 of the BG network is blocking. For an $N \times N$ BG network where $N > 16$, similar permutation traffic combinations can be found to show that any stage i ($i \notin \{0, 1, n-1, n-2\}$) blocking.

□

Theorem 1 *An $N \times N$ BG network is non - blocking under permutation traffic, if and only if $N \leq 16$.*

Proof : Any $N \times N$ BG network has $log_2 N$ stages where $log_2 N$ is an integer. When $N \leq 16$ we have four different possible values of N, namely 2, 4, 8 and 16. A network is non-blocking if and only if one or more stages of the network are blocking. A 2×2 BG network is non-blocking as it contains only one stage, stage 0 and from lemma 1 this stage is non-blocking and hence this network is non-blocking. A 4×4 BG network contains 2 stages , stage 0 and stage 1. From lemmas 1 and 2 this network is also non-blocking. An 8×8 BG network contains three stages 0, 1, and 2. From lemmas 1, 2 and 3, this network is also non-blocking. Finally a 16×16 BG network contains 4 stages, stage 0, 1, 2 and 3 and from lemmas 1, 2, 3 and 4 this network is also non-blocking. From lemma 5 we have, any stage i, $(i \notin \{0, 1, n-1, n-2\})$ of an $N \times N$ network $(N > 16)$, is blocking. If stage i, $(i \notin \{0, 1, n-1, n-2\})$ of an $N \times N$ network $(N > 16)$ is blocking, some packets are lost at that stage. Therefore packets are lost in an $N \times N$ BG network where $N > 16$, which means that the network is blocking.

□

Table 4.1 shows the performance of the BB, the BG and the R2D2 networks under permutation traffic at full load, where load indicates the amount of incoming traffic rate.

It can be seen from the above results that the BB network is completely non-blocking under permutation traffic.

4.4.2 Uniform Random Traffic

The uniform random traffic pattern refers to the case where the each output port has equal probability of being requested by packets arriving at the input ports. As destination addresses can be repeated, it is referred as uniform random traffic.

Even though the uniform random traffic does not properly describe the real traffic, it is a much more realistic traffic pattern when compared to the permutation traffic pattern explained in Section 4.6. Performance of MINs used in multiprocessor interconnections and in switch fabrics of broadband packet switch architectures is usually studied under uniform random traffic. It is a much simpler traffic pattern to be analyzed as compared to the non-uniform traffic types.

Table 4.2, 4.3 and 4.4 show the performance of the Banyan, the BB, the R2D2 and the BG networks under uniform random traffic under 50% load, 70% load and full load respectively.

It can be clearly seen from the above tables that the BG network has higher throughput than those of the R2D2 and the BB networks under varying loads of uniform random traffic types. A severe degradation on the performance of the BB network was noted.

4.5 Non-Uniform Traffic Patterns

The type of traffic expected in B-ISDN may not be of the uniform traffic type explained in Section 4.1. Much research is going on in determining a traffic type more close to the expected real traffic. Since it is quite difficult to simulate and analyze the exact traffic expected in B-ISDN, researchers have come up with certain tractable non-uniform traffic types. Studying the performance of MINs to be used

in switch fabrics of broadband packet switch architectures, under the non-uniform traffic patterns could give a better understanding of performance of the MINs under real traffic.

Non-uniform traffic patterns [45]-[50] are being studied in evaluating the performance of MINs. Three most widely studied non-uniform traffic types are the hotspot traffic, community of interest traffic and bursty traffic. The following sections will explain about these traffic models and provide simulation results of the performance of MINs under the above mentioned traffic types.

4.5.1 Hotspot Traffic

Hotspot traffic [47], [46] is defined as the traffic type in which one or more nodes of the switch fabric, or of the destinations receive a given percentage of the incoming packets. The rest of the incoming traffic is of the type uniform random. These nodes or destinations are referred to as hotspots. Hotspot traffic is also referred to as output-concentration traffic [48].

It is difficult to model hotspots at nodes within the switch fabrics and hence performance study of MINs under hotspot destinations was carried out.

10% of the incoming traffic is destined to one output; in this example, it is destination 0. Performance of the BG network is much better than the BB network and the R2D2 network even in the case of the hotspot traffic.

Table 4.6 gives the performance of the BG network under varying degrees of hotspots. Performance of the BG network under uniform random traffic is also provided for comparison. It is clearly seen that the throughput of the BG network drops down considerably with an increase in the percentage of hotspot.

4.5.2 Community of Interest Traffic

In the case of Community of Interest traffic, the traffic pattern is in such a way that a certain percentage of the traffic arriving at a certain input(s) is always directed to certain output(s). The remaining traffic originating at these inputs are of the uniform random traffic type. Uniform random traffic type is expected at all other inputs.

In this traffic type, certain output request patterns cause degradation of throughput primarily due to contention on internal links as opposed to output conflicts[48]. This occurs when the number of community of interest input-output pairs increases. Due to this more path conflicts occur in the intermediate stages of the MIN. The throughput degradation is more pronounced in case of large size networks.

Table 4.7 gives the performance of MINs under 100% community of interest traffic with one input-output pair. The input load offered to the MINs is 1. This indicates that during each cycle any packet arriving at a community of interest input will always be destined to that output which forms the community of interest. The number of input-output pairs was not increased with increase in network size. Due to this, performance of MINs under community of interest traffic is more or less equivalent to the performance under uniform random traffic type. Hence the throughput performance of the MINs is nearly similar to that shown in Table 4.4.

By increasing the number of input-output pairs in this traffic type, the degradation of the throughput can be seen clearly because there will be more path conflicts in the intermediate stages of the MINs. Still the throughput performance of the BG network is much superior to that of the BB network and of the R2D2 network.

4.5.3 Bursty Traffic

Bursty traffic is a traffic type in which the inputs of the switch fabric receive sudden bursts of packets [47], [49]. The traffic source at each input port alternates between active and idle periods [49]which are geometrically distributed with mean $m(A)$ and $m(S)$. The active period has packets for m continuous cycles called the burst length. The idle period is also referred to as burst gap as it occurs between two active periods which have bursty traffic. Packets arriving at an input port within the same burst are always directed to the same output port and are separated by fixed or random spacing [49]. It has been assumed that the active periods have a probability p and the idle periods a probability of $1 - p$. The burst length for an input port is not constant because the burstiness is assumed to be caused due to different payloads that are to be integrated in broadband communications.

For the MINs studied here, the gap between bursts under full load bursty traffic has been assumed to be zero. It is also assumed that the distribution of burst length is the same for all burst arriving on any given input port, and burst lengths and gap between bursts are drawn independently from a geometric distributions [51] with mean L packets/burst. The output port requested by a burst is assumed to be uniformly distributed over all the output ports independent of all other bursts entering the switch [48].

Simulation results of the performance of the BG networks for different probabilities of bursty traffic p, and having a burst length of mean equal to 5 packets/burst is shown in Table 4.8. The performance of the BG network under uniform random traffic pattern is also provided in Table 4.8. It can be seen that the throughput of the BG network increases with a decrease in probability of the bursty traffic

but does not vary much between probabilities of 0.5 and 0.8. The bursty traffic throughput is slightly less than that of the uniform random traffic type. The uniform random traffic is a special case of bursty traffic under full load where the bursty length is equal to one. The performance of the BG networks under bursty traffic at full load and with a burst length of mean equal to 5 packets/burst is compared with the performances of the BB and R2D2 networks in Table 4.9. It can be seen that the BG network features a better performance as compared to the BB and the R2D2 networks which are of comparable hardware complexities. This has been reported in [50].

The throughput performance of a $32 \times 32 BG$ network under full load for varying mean burst lengths is shown in Table 4.10. The performance of the BG network decreases with increase in the burst lengths. It can be inferred that the throughput of a network decreases with an increase in burst lengths in case of bursty traffic. In order to meet the needs of ATM standards of very high throughput, buffers are required. The BG network requires 136 buffers in order to achieve a packet loss rate of 10^{-6} under full load bursty traffic with mean equal to 15 packets per burst. This buffer size required by the BG network is much less when compared to 256 required by R2D2 network with identical traffic conditions and 1300 required by the tandem Banyan network even under 0.9 probability bursty traffic [48].

4.6 Summary

This chapter introduced the classification of networks based on the extent of blocking the network offers in establishing permutation connections. Later sections provided the performance of MINs under uniform and non-uniform traffic patterns.

While the BB network had better performance under permutation traffic type,

the performance of the BG network was found to be much better in all other traffic patterns. The exact traffic pattern(s) that can be expected during the normal function of an ATM network is not clearly known; it is only logical to expect that the traffic will be bursty and non-uniform with possibly repeated destination addresses at any given time – that is permutation traffic is far from reality. Furthermore, as the ATM network typically consists of interconnected clusters of smaller switch fabrics (each of which could be a MIN), the traffic arriving at the input ports of any switch is a more difficult problem to predict. However, it is reasonable to expect that this traffic pattern is, if anything, farther removed from being a permutation traffic. Therefore, the BG network which performs better under bursty traffic conditions is a wiser choice than networks, such as the BB combination, which are non-blocking under permutation traffic but perform poorly under other traffic conditions. It was noticed that the throughput of all the MINs decreased with an increase in switch size because of increase in number of stages and more conflicting paths between the packets to be routed in these stages.

Table 4.5 gives the performance of MINs under 10% hotspot traffic with probability of packet arrival at inputs equal to one. This indicates that

Network Size	BB Network	R2D2 Network	BG Network
2 × 2	1.000000	1.000000	1.000000
4 × 4	1.000000	1.000000	1.000000
8 × 8	1.000000	1.000000	1.000000
16 × 16	1.000000	0.993625	1.000000
32 × 32	1.000000	0.981562	0.997500
64 × 64	1.000000	0.966156	0.993203
128 × 128	1.000000	0.949203	0.986844
256 × 256	1.000000	0.933164	0.979062
512 × 512	1.000000	0.916697	0.970928
1024 × 1024	1.000000	0.901682	0.963329

Table 4.1: Throughput performance of MINs under Permutation Traffic

Network Size	BB Network	R2D2 Network	BG Network
2 × 2	0.889432	1.000000	1.000000
4 × 4	0.810005	0.995978	1.000000
8 × 8	0.780773	0.991580	0.999751
16 × 16	0.761559	0.987135	0.997996
32 × 32	0.753663	0.978900	0.996650
64 × 64	0.743854	0.974748	0.994755
128 × 128	0.746248	0.968893	0.993120
256 × 256	0.741558	0.963810	0.991517
512 × 512	0.743932	0.957198	0.990516
1024 × 1024	0.743313	0.951463	0.988663

Table 4.2: Throughput performance of MINs under Uniform Random Traffic - 50% Load

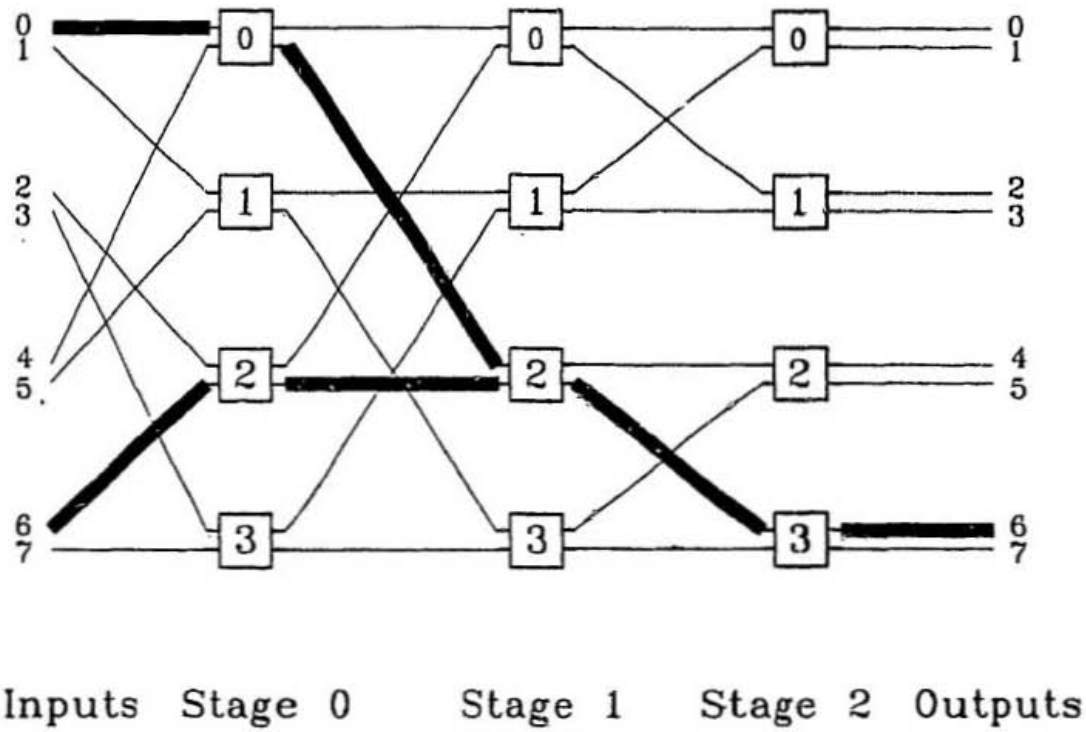

Figure 4.1: Internal Blocking in an 8 × 8 Banyan Network

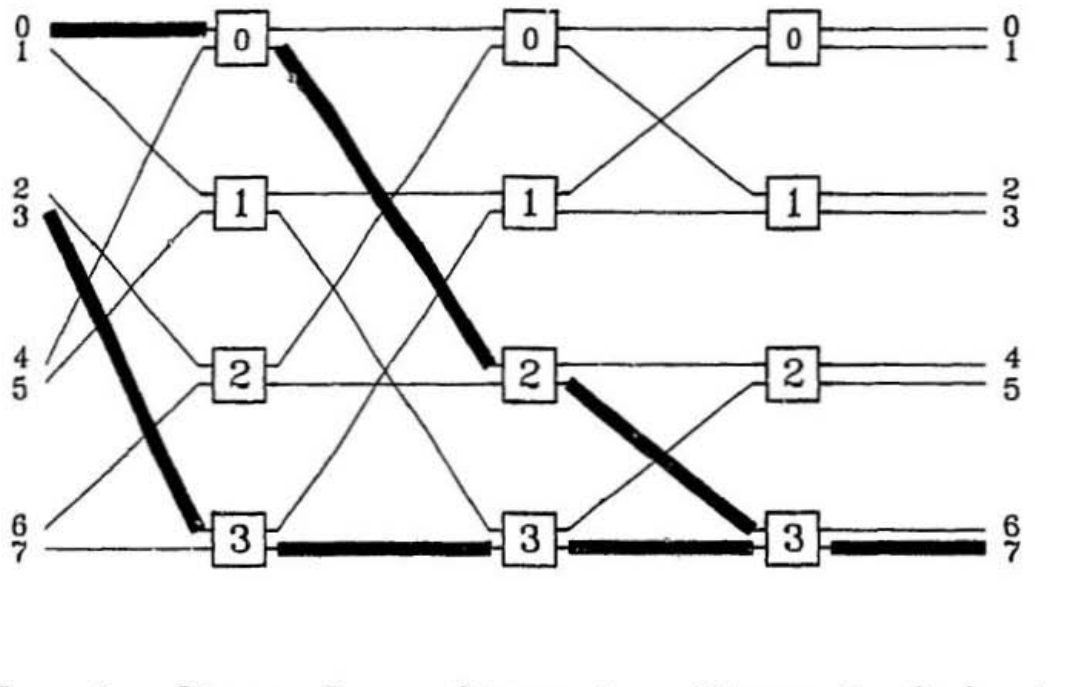

Figure 4.2: Output Contention Blocking in an 8 × 8 Banyan Network

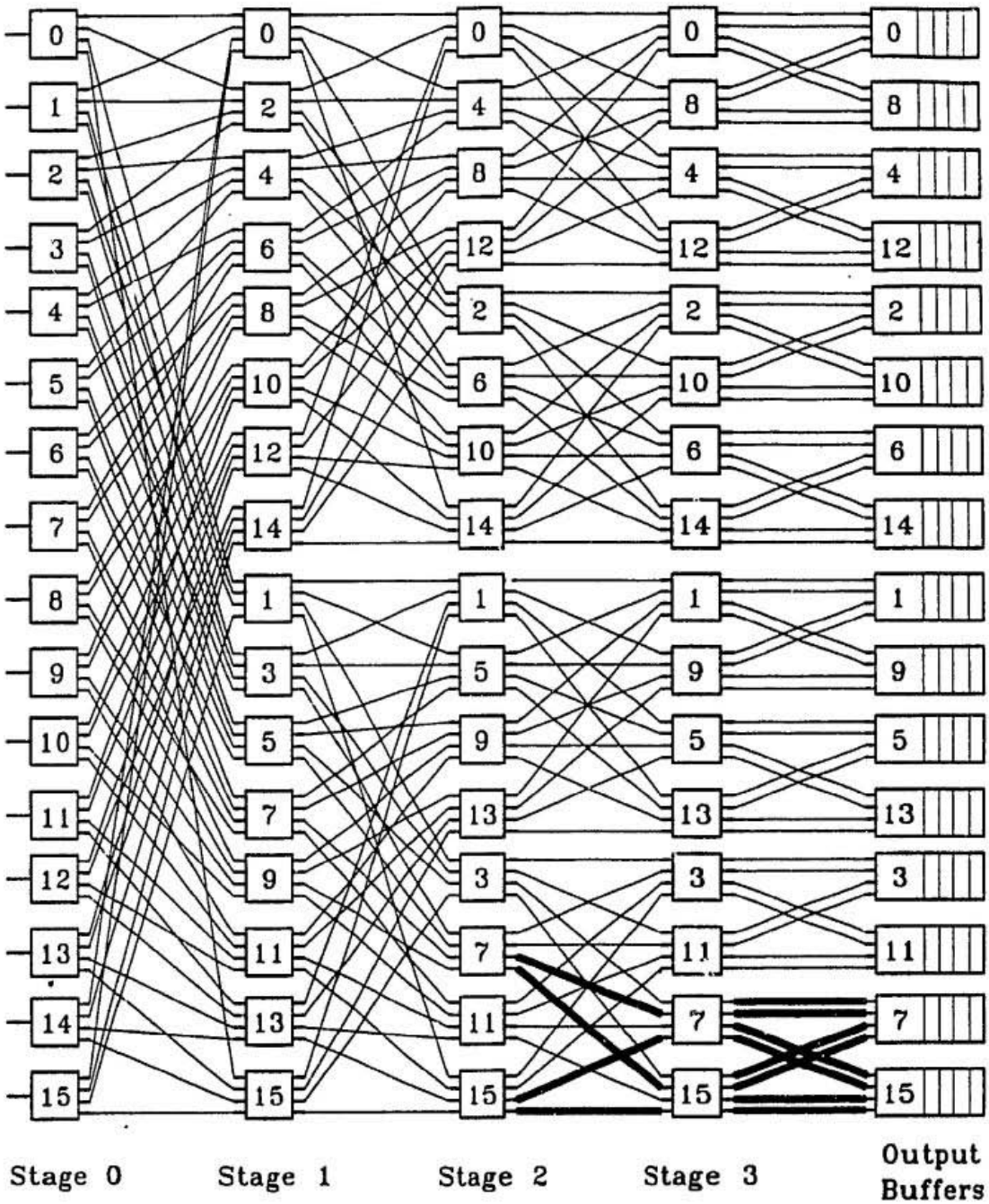

Figure 4.3: A 16 x 16 Balanced Gamma Network showing connection pattern between SEs in Stage2, Stage 3 and the output buffers.

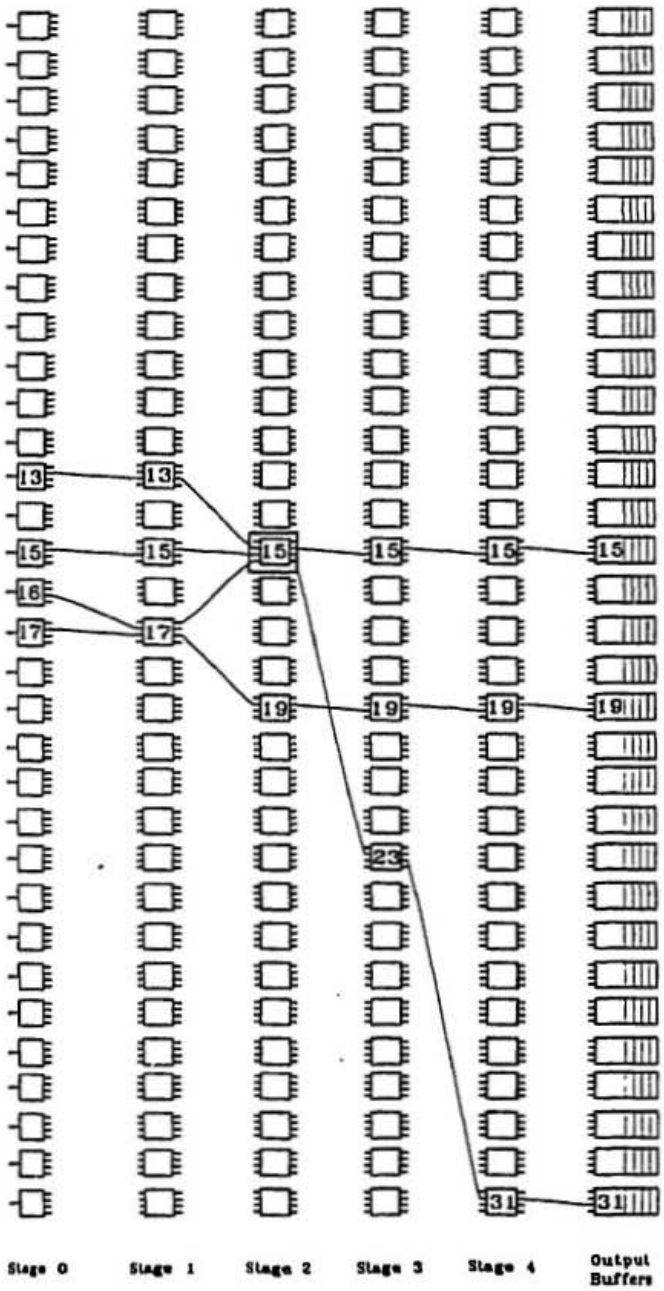

Figure 4.4: Blocking in a 32 x 32 Balanced Gamma Network

Network Size	BB Network	R2D2 Network	BG Network
2×2	0.812282	1.000000	1.000000
4×4	0.731993	0.990845	1.000000
8×8	0.690316	0.974937	0.998401
16×16	0.647365	0.975640	0.994952
32×32	0.647592	0.962567	0.991767
64×64	0.643381	0.954872	0.986901
128×128	0.640545	0.944439	0.983319
256×256	0.636841	0.934695	0.979464
512×512	0.637367	0.925141	0.976234
1024×1024	0.636125	0.916000	0.972447

Table 4.3: Throughput performance of MINs under Uniform Random Traffic - 70% Load

Network Size	BB Network	R2D2 Network	BG Network
2×2	0.774000	1.000000	1.000000
4×4	0.617250	0.988750	1.000000
8×8	0.552000	0.966500	0.993250
16×16	0.517563	0.953000	0.985625
32×32	0.490219	0.933625	0.976250
64×64	0.468250	0.916438	0.968750
128×128	0.460781	0.898008	0.959656
256×256	0.456086	0.883840	0.951191
512×512	0.449947	0.869176	0.942209
1024×1024	0.446400	0.854897	0.934344

Table 4.4: Throughput performance of MINs under Uniform Random Traffic - Full load

Network Size	BB Network	R2D2 Network	BG Network
2 × 2	0.747000	1.000000	1.000000
4 × 4	0.614250	0.983000	1.000000
8 × 8	0.535250	0.962000	0.989250
16 × 16	0.490813	0.937375	0.975125
32 × 32	0.474000	0.903406	0.950688
64 × 64	0.458906	0.867922	0.917203
128 × 128	0.451234	0.836727	0.887461
256 × 256	0.441719	0.810633	0.865406
512 × 512	0.441041	0.792773	0.853939
1024 × 1024	0.438835	0.777368	0.845227

Table 4.5: Throughput performance of MINs under 10% Hotspot Traffic

Switch Size	Throughput					
	Uniform	10%	30%	50%	70%	90%
2 × 2	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
4 × 4	0.988750	0.984000	0.982250	0.965500	0.924000	0.869500
8 × 8	0.966500	0.968125	0.948375	0.871500	0.747625	0.580625
16 × 16	0.953000	0.946063	0.874750	0.716250	0.531687	0.346812
32 × 32	0.933625	0.917594	0.783813	0.595312	0.403031	0.217125
64 × 64	0.916438	0.879094	0.707563	0.522172	0.337750	0.154047
128 × 128	0.898008	0.842023	0.666406	0.485594	0.304469	0.121211
256 × 256	0.883840	0.813082	0.642500	0.467578	0.288367	0.105074
512 × 512	0.869176	0.792910	0.627281	0.455584	0.278625	0.097406
1024 × 1024	0.854897	0.778297	0.617014	0.448571	0.272758	0.093805

Table 4.6: Throughput performance of the BG network for varying hotspot traffic.

Network Size	BB Network	R2D2 Network	BG Network
2 × 2	0.774000	1.000000	1.000000
4 × 4	0.617250	0.985250	1.000000
8 × 8	0.552250	0.966250	0.992250
16 × 16	0.520375	0.949000	0.984062
32 × 32	0.487094	0.933375	0.975688
64 × 64	0.472953	0.914594	0.967719
128 × 128	0.460523	0.900297	0.958164
256 × 256	0.455781	0.884215	0.950824
512 × 512	0.450291	0.869408	0.942762
1024 × 1024	0.446037	0.854766	0.934229

Table 4.7: Throughput performance of MINs under 100% Community of Interest Traffic

Network Size	BB Network	R2D2 Network	BG Network
2 × 2	0.73667	1.000000	1.000000
4 × 4	0.61925	0.986917	1.000000
8 × 8	0.543042	0.968458	0.993667
16 × 16	0.512812	0.951187	0.982500
32 × 32	0.490271	0.932469	0.976083
64 × 64	0.47175	0.914969	0.967443
128 × 128	0.462409	0.898536	0.959146
256 × 256	0.454299	0.884061	0.950008
512 × 512	0.449662	0.868557	0.943021
1024 × 1024	0.44724	0.855262	0.934652

Table 4.8: Throughput performance of the MINs under full load Uniform Random and Bursty Traffic

Network Size/ Network Type	Uniform Random Traffic 100%	Bursty Traffic 100%	Bursty Traffic 80%	Bursty Traffic 50%
2 × 2	1	1	1	1
4 × 4	1	1	1	1
8 × 8	0.993667	0.998628	0.999136	0.993375
16 × 16	0.9825	0.998183	0.998365	0.985687
32 × 32	0.976083	0.99666	0.996415	0.975781
64 × 64	0.967443	0.994402	0.995217	0.968266
128 × 128	0.959146	0.993538	0.993882	0.958625
256 × 256	0.950008	0.991575	0.991466	0.950297
512 × 512	0.943021	0.990036	0.990153	0.94209
1024 × 1024	0.934652	0.988639	0.988654	0.934768

Table 4.9: Throughput performance of the BG network under varying loads of Bursty Traffic

Mean Burst Length	Throughput
5	0.976005
10	0.975844
15	0.975058

Table 4.10: Throughput performance of a 32 × 32 BG network under varying burst lengths of Bursty Traffic

Chapter 5

Performance of the BG network in the presence of SE faults

5.1 Introduction

The various reliability measures explained in Chapter 3 do not provide a clear indication of the dependability of the network. On the other hand, the performance of a network in the presence of faulty components gives us a better insight in this regard.

Performance of most of the networks is studied under the uniform random traffic pattern. Theoretical performance analysis of the BG network under uniform random traffic pattern has been done in [24] and the simulation results of the performance of the BG network under uniform random traffic without the presence of faults have been given in the previous chapter.

This chapter presents the performance analysis of the BG network in the presence of faults. Similar to the assumptions made during reliability analysis, here we consider only SE failures. Failures of links connecting two SEs, $SE_{i,j}$ and $SE_{i,j+1}$ are considered as SE failures of $SE_{i,j+1}$. The performance analysis of the BG network in the presence of an SE [41] has been dealt with in detail in Section 5.2

followed by the summary in Section 5.3.

5.2 Analysis and Simulation

The analysis and simulation of the BG network under faults has been done under uniform random traffic. Let T(N) be the throughput performance of an $N \times N$ BG network without faults under uniform random traffic. Let the failed SE be $SE_{i,j}$ ($j \neq 0$). The performance of the BG network with the failure of $SE_{i,j}$ is lower than that of the BG network without the failure of $SE_{i,j}$ because the traffic which normally goes through $SE_{i,j}$ would be routed through the SEs which form critical pairs with $SE_{i,j}$. Due to increased contention at the SEs which form critical pairs with $SE_{i,j}$ the throughput of the BG network drops down.

The performance of an $N \times N$ BG network varies depending upon the stage in which the SE fault occurs. In this thesis the stage 0 SEs have been assumed to be fault free. The performance analysis of the BG network in the presence of faults is divided into two parts – the SE fault located at stage 1 and the SE fault located at stage $j (2 \leq j \leq n-1)$.

5.2.1 SE fault at stage 1

There is no cell loss at stage 1 of an $N \times N$ BG network if there are no SE faults at stage 1 because each SE in stage 1 receives no more than two packets during each cycle. Due to the presence of an SE fault at stage 1 some cells could be lost at stage 1. Without loss of generality let us consider a SE fault at $SE_{4,1}$ in an $N \times N$ BG network where $N > 4$. Due to SE fault at $SE_{4,1}$, cells which would have normally been routed through it (under no fault condition) from $SE_{4,0}$ and $SE_{3,0}$ would now

be routed through $SE_{6,1}$ and $SE_{2,1}$. SEs $SE_{6,1}$ and $SE_{2,1}$ form critical pairs with the faulty SE $SE_{4,1}$. Now SEs $SE_{6,1}$ and $SE_{2,1}$ have a possibility of receiving three packets each from stage 0. If all these three packets have the same switching bit for stage 1, then one packet would be dropped. There are four possible combinations in which packets can be lost - each of the SEs $SE_{6,1}$ and $SE_{2,1}$ receiving three packets, where all these three packets have the same switching bit (0 or 1) in each of the SEs. In the worst case, 4 out of N incoming packets are lost. The probability $P_{3-stage1}$, that each of the SEs $SE_{6,1}$ and $SE_{2,1}$ receive three packets and all these three packets have the same switching bit for stage 1 is given by the expression

$$P_{3-stage1} = \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{2}\right)^3. \tag{5.1}$$

Therefore the probability of cell loss at stage 1, $P_{cell-loss-stage1}$, will be

$$P_{cell-loss-stage1} = \frac{4}{N} \times \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{2}\right)^3. \tag{5.2}$$

Hence the throughput of an $N \times N$ BG network in the presence of an SE fault can be approximated as

$$Throughput = T(N) - P_{cell-loss-stage}. \tag{5.3}$$

Table 5.1 gives the throughput of an $N \times N$ BG network using Equation 5.3. The throughput of the BG network without fault, T(N), obtained by analysis is taken from [24]. The throughput of the BG network without fault, T(N), obtained by simulation is taken from Table 4.4.

The simulation results of throughput of BG network under single SE fault at $SE_{1,1}$ has also been given in Table 5.1.

The throughput analysis of the BG network carried out in [24] was approximate. Therefore using the throughput T(N), from [24] in column 1 has given rise to further

approximations. In column 2 of Table 5.1, we have taken T(N), from the simulation results presented in Table 4.4. It can be seen clearly that the throughput analysis of the BG network in the presence of a fault at stage 1 shown by column 2 of Table 5.1 is comparable to the throughput results of the BG network in the presence fault at stage 1 provided by simulation. It is to be noted that the presence of a single SE fault at stage 1, does not affect the throughput performance of the BG network under uniform random traffic for larger sizes. This is due to the fact that the probability of cell loss is inversely proportional to the size of the network, as given by Equation 5.3.

There is a large variation between the throughput performances of a 4×4 BG network obtained by analysis and simulation. This is because stage 1 in a 4×4 BG network is also the last stage and we have not accounted for the cell loss due to this effect. The throughput of a fault-free 4×4 BG network under uniform random traffic is 1. When there is a SE fault in stage 1, which also forms the last stage, cell loss occurs mainly due to multiple packets being destined to a particular output port. Apart from the cell loss explained in the earlier analysis, 2 out of 4 cells are dropped whenever 4 cells are destined to 2 out of 4 output ports, during an SE fault in stage 1 of a 4×4 BG network. The probability of having 4 cells being destined to 2 out of 4 output ports, in which 2 cells are dropped is equal to $\left(\frac{1}{2}\right)^4 \times \left(\frac{1}{2}\right)^4$. The SE fault could be any one of the 4 SEs of the 4×4 BG network. Therefore, the probability of cell loss due to the effect of all cells generated going to 2 out of 4 output ports is

$$P_{cell-loss-4} = 4 \times \frac{2}{4} \times \left(\frac{1}{2}\right)^4 \times \left(\frac{1}{2}\right)^4. \qquad (5.4)$$

Therefore the overall throughput of a 4 × 4 BG network in the presence of a single SE fault at stage 1, by analysis is

$$\text{Throughput} = 0.984375 - P_{cell-loss-1} = 0.980469.$$

The above throughput analysis for a 4 × 4 BG network produces a result which is comparable to that obtained by simulation.

5.2.2 SE fault at stage i ($i \in \{2,3, \ldots,n\text{-}1\}$)

The throughput analysis of an $N \times N$ BG network ($N > 4$) in the presence of single SE fault in stage j ($j > 1$) is presented in this section. For this analysis, $SE_{i,j}$ has be assumed to be faulty. Since $SE_{i,j}$ has failed, the packets which are normally routed through $SE_{i,j}$ will now be routed through SEs which form critical pairs with $SE_{i,j}$. $SE_{i,j}$ receives packets from four SEs in stage $j-1$. Of these four SEs, the regular links of two SEs are connected to $SE_{i,j}$ and alternate links of the other two SEs are connected to $SE_{i,j}$. Let these SEs be SE1, SE2, SE3 and SE4 respectively where the regular links of SE1 and SE2 are connected to $SE_{i,j}$ and the alternate links of SE3 and SE4 are connected to $SE_{i,j}$. SE1 and SE2 can reroute the packets to be switched to $SE_{i,j}$ through their alternate links. Out of the k packets which have arrived at SE3 or SE4 having the same switching bit as that of a packet to be switched to $SE_{i,j}$, $k-1$ packets are dropped at these respective SEs causing packet loss. This is due to the failure of $SE_{i,j}$. Similarly if SE3 and SE4 have packets to be switched to $SE_{i,j}$ then these packets are dropped at SE3 and SE4 respectively. The SEs SE1, SE2, SE3 and SE4 can each drop a cell due to the failure of $SE_{i,j}$. Therefore, the probability of packet loss at stage $j-1$ due to failure of $SE_{i,j}$ is

$$P_{CL-STG-j-1} = \frac{4}{N} \times al_{j-1}. \tag{5.5}$$

where al_{j-1} is the probability that the alternate links of stage $j-1$ carry a packet. Due to the failure of $SE_{i,j}$, packets are lost even at stage j. If SE1 or SE2 has a packet p_k, which is to be routed to $SE_{i,j}$ then this packet will be routed through the alternate links of SE1 or SE2 making p_k to go to SEs which form critical pair with $SE_{i,j}$. Let these SEs which form critical pairs with $SE_{i,j}$ be called SEc1, SEc2. If SEc1 and SEc2 already have a minimum of two packets, having the same switching bit as that of p_k, then the packet p_k is dropped at stage j. The probability that packet loss occurs at stage j due to the failure of $SE_{i,j}$ is

$$P_{CL-STG-j} = \frac{4}{N} \times al_{j-1} \times al_j. \tag{5.6}$$

Therefore the overall probability of packet loss due to a failure of $SE_{i,j}$ is the sum of Equation 5.4 and Equation 5.5. Hence the throughput of the BG network under single SE fault at a stage other than its first two stages is

$$Throughput = T(N) - P_{CL-STG-j-1} - P_{CL-STG-j}. \tag{5.7}$$

The throughput performance of the BG network without SE failure, under uniform random traffic T(N) is taken from Table 4.4. The values of al_j and al_{j-1} are calculated by the recursive expression given in [24]. The throughput performance of an $N \times N$ BG network using Equation 5.6 is presented in Table 5.2, while the simulation results are presented in Table 5.3. The throughput performance of the BG network obtained by analysis and simulation are found to be nearly equal.

It can be noted that the throughput of large size BG networks is very minimally affected by the presence of single SE fault.

5.3 Summary

The throughput performance of the BG network in the presence of a single SE fault under uniform random traffic has been obtained by analysis and also by simulation. Results from analysis and simulation exhibit a close match. Performance of the BG network is not significantly degraded due to occurrence of a single SE fault. Smaller size BG networks have more throughput degradation in the presence of an SE fault and this degradation decreases with an increase in size of the network. The performance of the BG network in the presence of faults gives a better understanding of the dependability of the network than the reliability metrics which have been reported in literature for most of the MINs studied so far.

Network Size (N)	Throughput by Analysis T(N) from [24]	Throughput by Analysis T(N) from Table 4.4	Throughput from Simulation
4 × 4	0.984375	0.984375	0.978250
8 × 8	0.985188	0.985438	0.987125
16 × 16	0.982094	0.981718	0.983000
32 × 32	0.977047	0.974297	0.974875
64 × 64	0.971023	0.967773	0.966391
128 × 128	0.965512	0.9591677	0.959031
256 × 256	0.958756	0.950946	0.950348
512 × 512	0.952878	0.942087	0.942387
1024 × 1024	0.946939	0.934283	0.934273

Table 5.1: Throughput performance of BG network in the presence of a single SE fault at stage 1 under Uniform Random Traffic

Network Size	Throughput							
	Stage in which SE has failed							
N	2	3	4	5	6	7	8	9
8	0.9598	NA	NA	NA	NA	NA	NA	NA
16	0.9689	0.9671	NA	NA	NA	NA	NA	NA
32	0.9679	0.9669	0.9669	NA	NA	NA	NA	NA
64	0.9646	0.9641	0.9641	0.9642	NA	NA	NA	NA
128	0.9576	0.9573	0.9573	0.9574	0.9574	NA	NA	NA
256	0.9502	0.9500	0.9500	0.9500	0.9500	0.9506	NA	NA
512	0.9417	0.9416	0.9416	0.9416	0.9416	0.9417	0.9417	NA
1024	0.9341	0.9341	0.9341	0.9341	0.9341	0.9334	0.9341	0.9341

Table 5.2: Throughput Analysis of the BG network in the presence of a single SE fault at different stages under uniform random traffic.

Network Size	Throughput							
	Stage in which SE has failed							
N	2	3	4	5	6	7	8	9
8	0.9579	NA	NA	NA	NA	NA	NA	NA
16	0.9671	0.9605	NA	NA	NA	NA	NA	NA
32	0.9666	0.9644	0.9616	NA	NA	NA	NA	NA
64	0.9633	0.9636	0.9610	0.9607	NA	NA	NA	NA
128	0.9571	0.9567	0.9568	0.9554	0.9558	NA	NA	NA
256	0.9496	0.9489	0.9483	0.9491	0.9493	0.9487	NA	NA
512	0.9418	0.9424	0.9420	0.9417	0.9418	0.9413	0.9414	NA
1024	0.9246	0.9343	0.9345	0.9341	0.9342	0.9341	0.9344	0.9348

NA - Not applicable

Table 5.3: Simulation results of the performance of the BG network in the presence of a single SE fault at different stages under uniform random traffic.

Chapter 6

Conclusions

6.1 Contributions in the Thesis

There has been rapid advancements in the field of communication systems. ATM has been identified as the potential transfer mode for B-ISDN. Researchers are experimenting with switch fabrics for ATM networks. One of the main components of an ATM network is the switch fabric. Most of the switch fabrics which are being considered have a MIN for transporting data from an input port to an output port within the switch fabric. The performance of most of the switch fabrics experimented is not sufficient for meeting the needs of B-ISDN. This thesis has focused on the evaluation of a MIN for use in switch fabrics of broadband packet switch architectures. Most of the MINs which have been considered for use in ATM networks use the Banyan networks or its enhanced versions. Three MINs - the BB, the R2D2 and the BG networks have been studied in this thesis.

Performance of a MIN has been the main aspect considered so far, for employing a MIN in ATM networks. Apart from the throughput performance, hardware complexity, fault tolerance, reliability and performance under faults are certain other important aspects which are to be considered while choosing MINs for use in broad-

band communication switch fabrics. The cost involved in manufacturing an ATM switch fabric is mainly dependent on the hardware complexity of the switch fabric. The HC of a network is directly proportional to its manufacturing cost. The BB, the R2D2 and the BG are found to have comparable hardware complexities.

The BB network is a unipath network and therefore it does not possess any FT. A detailed study has been done on the FT properties of the R2D2 and the BG network. A new simple routing algorithm has been proposed for the BG network in order to enhance its FT capability. It has been shown that a complex routing algorithm would increase the FT capability of the R2D2 network, but at the same time, also increases the hardware cost. The BG network has better FT properties than the R2D2 networks since there are multiple paths between each input-output pair as compared to two paths between each input-output pair in the R2D2 networks.

Reliability analysis has shown the BG networks to be more reliable than the R2D2 networks. The network reliability of a network is of greater importance than the network's terminal and broadcast reliabilities because, its denotes the working of the overall MIN at a particular instance of time. Since the actual computation of the NR for the BG network is complex, lower and upper NR bounds were derived for the BG networks. The lower bound NR of the BG network was found to be closer to its NR. The NR of large size BG network was found to be very low. Considering a failure rate of 10^{-6} for each SE, the network MTTF of a 1024×1024 BG network was found to be around 2 months. This is not practical if the MIN is to be used in a remote environment. Therefore, SEs of lower failure rate have to be used for constructing the BG networks. In environments where repair is possible, we can still make use of SEs with the above mentioned failure rates.

Different payloads are to be integrated in B-ISDN. Therefore, it is difficult to exactly predict the realistic traffic patterns in an ATM network. The performance of an ATM network has to be analyzed before its use in B-ISDN. Since MINs are being considered for use in switch fabrics of an ATM network, study of these MINs, under the realistic traffic patterns expected in B-ISDN, will help in their evaluation. Simulation of real time traffic patterns expected in B-ISDN is quite tedious. Therefore, researchers have studied the performance of MINs under certain traffic patterns. This helps in understanding the performance of the network under real time traffic conditions. Networks have been classified as blocking or non-blocking networks based on their ability to establish permutation connections. The Batcher Banyan network is a non-blocking network for permutation traffic [11]. Such a classification does not reveal the performance of the networks under real time traffic conditions. Therefore, the performances of networks are studied under certain uniform and non-uniform traffic patterns. The performances of the BB, the R2D2 and the BG networks under uniform random, hotspot, community of interest and bursty traffic conditions have been studied in this thesis. Even though these networks have comparable HCs, simulation results have shown the BG networks to possess superior performance in all the above mentioned traffic patterns. Therefore it can be concluded that hardware has been effectively utilized in the BG networks. It was found that the performances of all these MINs decrease with an increase in size of the MIN. The bursty traffic pattern can be considered as a traffic pattern nearly equal to those of the real-time traffic conditions. It was observed that the performance of the BG networks under bursty traffic patterns dropped down with an increase in the burst length.

Since the exact traffic pattern(s) entering an ATM network is still not clearly known, it is only logical to expect that the traffic would be non-uniform and mainly bursty rather than the permutation traffic. Even though the BB network performs better than the BG network under permutation traffic, the BG network performs much better under the bursty traffic conditions. So it may be a wiser choice to choose the BG network for use in switch fabrics of broadband packet switch architectures.

The reliability metrics just indicate the working of a network after a particular period of time. This does not show the dependability of the network. It has been shown that the performance of a MIN under fault gives a better understanding of the dependability of the network. The performance of the BG network in the presence of single SE fault, under uniform random traffic pattern, has been studied in this thesis. Both performance analysis and simulation results for the same have been provided. Results from analysis and simulation have shown a close agreement. It was found that the performance of large sized BG networks are not significantly degraded due to the occurrence of a single SE fault.

Even though the BG network was found to be have better throughput performances than the BB and the R2D2 networks, this performance is not high enough to meet the high throughput required in B-ISDN. The high throughput requirements of B-ISDN could be attained by either employing different buffering schemes in the BG networks [52] or by the method of dilation or replication or enlargement [53]. Enhanced BG networks by buffering, dilation, replication and enlargement were found to have very low cell loss rate (of the order 10^{-6} and lower). If buffers were to be employed for increasing the throughput capability of the BG networks,

the cells arriving at their final destinations may be out of sequence. In order to avoid this, one has to formulate some ordering mechanism at the output ports. Enhanced BG networks, apart from having high throughput, also have better FT and reliability. Therefore the BG networks may be considered as a potential candidate for use in switch fabrics of broadband packet switch architectures.

6.2 Recommendations for Future Research

The work reported in this thesis can be extended to the following areas :

1. In an environment where there is a faster repair rate of the faulty components, the failure rate of the SEs of the BG network can still be chosen to be 10^{-6}. This will considerably decrease the cost of the BG network as compared to the one with lower failure rate SEs. In such a case availability analysis [30] of the BG network should be done. This would give a better idea on the availability of the network at a particular instance of time.

2. Since the BG network may be considered as a potential candidate for use as switch fabrics in broadband packet switch architectures, an in depth study of the hardware implementation of the BG network to form a switch fabric may be a worthwhile task. This involves VLSI design for the switch fabric followed by the actual fabrication process. After the fabrication process the performance of the switch fabric can be tested in real time traffic conditions. Modifications catering to the needs of the requirements of B-ISDN can be done after a thorough testing of the switch fabric.

Bibliography

[1] W. Stallings, "ISDN and Broadband ISDN", Second Edition, Maxwell Macmillan Canada, 1992.

[2] M. Belkadi, "Towards High Performance Highly Reliable Interconnection Networks for ATM Switch Architectures", Ph.D. Dissertation, Queen's University, 1995.

[3] M. Listani and A. Roveri, "Switching structures for ATM", Computer Communications, Vol. 12, no. 6, pp. 349-358, December 1989.

[4] M. de Prycker, "Asynchronous Transfer Mode : solution for broadband ISDN", Ellis Horwood Ltd., Second Edition, 1993.

[5] R. O. Onvural, "ATM networks : Performance Issues", Artech House, 1994.

[6] L. G. Cuthbert and J-C. Sapanel, "ATM the Broadband Telecommunications Solution", The Institution of Electrical Engineers, 1993.

[7] H. Ahmadi and W. E. Denzel, "Survey of Modern High-Performance Switching Techniques", Journal on Selected Areas in Communications, Vol. 7, no. 7, pp. 1091-1103, September 1989.

[8] R. Venkatesan, "Research on Switch Fabrics for Integrated Broadband Communication Networks", Proceeding of NECEC 95, St. John's, May 1995.

[9] H. J. Siegel, "Interconnection networks for Large Scale Parallel Processing", Second Edition, McGraw Hill, 1990.

[10] S. Sundaram and R. Venkatesan, "Reconfigurable Interconnection Networks", East Report No: 89-001, Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, Canada, August 1989.

[11] H. Y. Hui, "Switching and traffic theory for integrated broadband networks", Kluwer Academic Publishers, 1990.

[12] D. Minoli, "Broadband Network Analysis and Design", Artech House, 1993.

[13] L. R. Goke and G. P. Lipovski, "Banyan Networks for partitioning Multiprocessor Systems", in Proceedings of 1st Annual Symposium on Computer Architecture, pp. 21-28., December 1973.

[14] C-L. Wu and T-Y. Feng, "On a Class of Multistage Interconnection Networks", IEEE Transactions on Computers, Vol. C-29, no. 8, pp. 694-702, August 1980.

[15] D. M. Dias and J. R. Jump, "Analysis and Simulation of Buffered Delta Networks", IEEE Transactions on Computers, Vol. C-30, no. 4, pp. 273-282, April 1981.

[16] Y. C. Jenq, "Performance Analysis of a Packet Switch based on a Single-buffered Banyan Network", IEEE Journal of Selected Areas in Communications, Vol. SAC-3, no. 6, pp. 1014-1021, December 1983.

[17] H. F. Badran and H. T. Mouftah, "Performance of output-buffered Broadband Switch Architectures", Proceedings of Canadian Conference on Electrical and Computer Engineering, paper no. 39.2, Ottawa, September 1990.

[18] G. B. Adams III and H. J. Siegel, "The Extra Stage Cube: A Fault-Tolerant Interconnection Network for Supersystems", IEEE Transactions on Computers, pp. 443-454, May 1982.

[19] G. B. Adams III and H. J. Siegel, "Modifications to Improve the Fault Tolerance of the Extras Stage Cube Interconnection Network", 1984 International Conference on Parallel Processing, Computer Society Press, Silver Spring, Md., pp. 169-173, 1984.

[20] H. J. Siegel and R. J. McMillen, "The Multistage Cube: A Versatile Interconnection Network", Computer, pp. 65-76, December 1981.

[21] S. M. Reddy and V. P. Kumar, "On Fault-Tolerant Multistage Interconnection Networks", 1984 International Conference on Parallel Processing, Computer Society Press, Silver Spring, Md., pp. 155-164, 1984.

[22] D. H. Lawrie, "Access and Alignment of Data in an Array Processor", IEEE Transactions on Computers, Vol. C-24, No. 12, pp. 1145-1155, December 1975.

[23] L. Ciminiera and A. Serra, "A Connecting Network with Fault Tolerance Capabilities", IEEE Transactions on Computers, pp. 578-580, June 1986.

[24] R. Venkatesan and H. T. Mouftah, "Balanced Gamma Network - A new candidate for Broadband Packet Switch Architectures", IEEE INFOCOM '92, Florence, May 1992.

[25] D. S. Parker and C. S. Ragavendra, "The Gamma Network : A Multiprocessor Interconnection Network with Redundant Paths", Proceedings of 9th Annual Symposium on Computer Architecture, pp. 73-80, April 1982.

[26] S. C. Kothari, G. M. Prabhu and R. Roberts, "The Kappa Network with Fault-tolerant Destination Tag Algorithm", IEEE Transactions on Computers, Vol. C-37, pp. 612-617, May 1988.

[27] H. Sivakumar and R. Venkatesan, "Blocking in Multistage Interconnection Networks for Broadband Packet Switch Architectures", NECEC '95, St. John's, May 1995.

[28] G. B. Adams, D. P. Agrawal and H. J. Siegel, "Fault-tolerant Multistage Interconnection Networks", IEEE Computer, pp. 14-27, June 1987.

[29] A. Goyal, S. Lavenberg and K. Trivedi, "Probabilistic Modeling of Computer System Availability", Annals of Operations Research, Vol. 8, pp 285-306, March 1987.

[30] B. W. Johnson, "Design and Analysis of Fault Tolerant Digital Systems", Addison Wesley Publishing Company, 1989.

[31] J. T. Blake and K. S. Trivedi, "Reliability Analysis of Interconnections Networks using Hierarchical Composition", IEEE Transactions on Reliability, Vol. 38, no. 1, pp. 111-119, April 1989.

[32] J. T. Blake and K. S. Trivedi, "Multistage Interconnection Reliability", IEEE Transactions on Computers, Vol. 38, no. 11, November 1989.

[33] A. Varma and C. S. Ragavendra, "Reliability Analysis of Redundant-Path Interconnection Networks", IEEE Transactions on Reliability, Vol. 38, no.1, pp. 130-137, April 1989.

[34] C. S. Ragavendra and D. S. Parker, "Reliability Analysis of an Interconnection Network", Proceedings of the 4th International Conference on Distributed Computing Systems, pp. 461-471, May 1984.

[35] C. Booting, S. Rai and D. P. Agrawal, "Reliability Computation of Multistage Interconnection Networks", IEEE Transactions on Reliability, Vol. 38, no.1, pp. 138-145, April 1989.

[36] J. S. Provan, "Bounds on the Reliability of Networks", IEEE Transactions on Reliability, Vol.R-35, no.3, pp. 260-268, August 1986.

[37] S. Rai and D. P. Agrawal, "Reliability of Program Execution in a Distributed Environment", IEEE Transactions on Reliability, October 1987.

[38] K. Padmanabhan and D. H. Lawrie, "Fault-tolerance Schemes in Shuffle/Exchange type Interconnection Networks", Proceedings of International Conference on Parallel Processing, pp. 71-75, August 1983.

[39] X. Cheng and O. C. Ibe, "Reliability of a Class of Multistage Interconnection Networks", IEEE Transactions on Parallel and Distributed Systems, Vol. 3, No. 2, pp. 241-246, March 1992.

[40] V. Cherkassky and M. Malek, "Reliability and Fail-Softness Analysis of Multistage Interconnection Networks", IEEE Transactions on Reliability, Vol. R-34, No. 5, pp. 524 527, December 1985.

[41] H. Sivakumar and R. Venkatesan, "Reliability and Performance Analyses of Balanced Gamma Network for use in Broadband Communications Switch Fabric", submitted to Broadband Communications-96, to be held in Montreal, Canada, April, 1996.

[42] D. P. Agrawal and J. S. Leu, "Dynamic Accessibility Testing and Path Length Optimization of Multistage Interconnection Networks", IEEE Transactions on Computers, Vol. C-34, pp. 255-266, March 1985.

[43] C. Dhas, V. K. Konangi and M. Sreetharan, "Edited: Broadband Switching : Architectures, Protocols, Design and Analysis, Chapter 4 : Switch Fabric Design and Analysis", IEEE Computer Society Press, 1991.

[44] Y. Viniotis and R. O. Onvural, "Ed: Asynchronous Transfer Mode Networks", Plenum Press, 1993.

[45] S. Urushidani, "Rerouting Network : A High-Performance Self-Routing Switch for B-ISDN", IEEE Journal on Selected Areas in Communications, Vol. 9, No. 8, pp. 1194-1204, October 1991.

[46] S. L. Scott and G. S. Sohi, "The Use of Feedback in Multiprocessors and Its Applications to Tree Saturation Control", IEEE Transactions on Parallel and Distributed Systems, Vol. 1, No. 4, pp. 385-398, October 1990.

[47] H. F. Badran and H. T. Mouftah, "Performance of Broadband Integrated Switch Architectures with Input-Output-Buffering under Backpressure Mechanisms", Research Report, no. 90-7, Queen's University, Kingston, Ontario, Canada, 14 p., 1990.

[48] F. A. Tobagi, Timothy Kwok and Fabio M. Chiussui, "Architecture, Performance and Implementation of the Tandem Banyan Fast Packet Switch, IEEE Journal on Selected Areas in Communications", Vol. 9, no. 8, October 1991.

[49] H. Ahmadi, "Tree Switch: A Modular Fast Packet Switching Fabric for Synchronous and Asynchronous Traffic", Telecommunications Research Institute of Ontario, November 1990.

[50] H. Sivakumar and R. Venkatesan, "Performance of Balanced Gamma Network under Bursty Traffic", submitted to etaCOM-96, to be held at Portland, Oregon, U.S.A., May 1996.

[51] P. Bratley, B. L. Fox, L. E. Scharge, "A Guide to Simulation", Second Edition, Springer-Verlag New York Inc., 1987.

[52] K. P. Singh and R. Venkatesan, "Performance Evaluation of Different Buffering Schemes for Balanced Gamma Networks", Proceeding of CCECE '93, Vancouver, September 1993.

[53] R. Venkatesan, "Performance Analysis of Kappa Networks and Enlarged Kappa Networks under uniform and non-uniform traffic", Proceedings of CCECE '94, Halifax, September 1994.

Appendix A

Routing Algorithm for the Balanced Gamma Network

```
Route_packets_BGNetwork()
{
  for j varying from 0 to n-1 do
  {
    for i varying from 0 to N-1 do
    {
      alpha = int(i/2j)mod2: //calculate alpha for SE
      count = total inputs for SE_{i,j}: //calculate number of Input packets
      if(count > 0)
      {
        case (count)
        {
          1: switch_one_packet(SE_{i,j} , alpha);
             break:
          2: switch_two_packets(SE_{i,j} , alpha):
             break:
          3: switch_three_packets(SE_{i,j} , alpha):
             break;
          4: switch_bit0 = number of packets at input ports having bit d_j=0;

             case(switch_bit0)
             {
               0: delete two of the four packets at random;
                  switch_two_packets( SE_{i,j} , alpha);
                  break;
               1: deleted one of the three packets having d_j = 1;
                  switch_three_packets(SE_{i,j} , alpha);
                  break;
               2: switch_packet0 = 0;
                  switch_packet1 = 0;
                  for k varying from 1 to 4 do
                  {
                    if( SE_{i,j} has a packet at input port k and if d_j == 1)
                    case (alpha)
                    {
                      0: if(packet_switched1 == 0)
```

```
                {
                  switch packet at SE_{i,j} through OLM link:
                  packet_switched1 =1:
                }
                else
                  switch packet at SE_{i,j} through OU link:
                break:
              1: if(packet_switched1 == 0)
                {
                  switch packet at SE_{i,j} through OUM link:
                  packet_switched1 =1:
                }
                else
                  switch packet at SE_{i,j} through OL link:
                break:
            }
          if( SE_{i,j} has a packet at input port k and if d_j == 0)
          case (alpha)
          {
            0: if(packet_switched0 == 0)
              {
                switch packet at SE_{i,j} through OUM link:
                packet_switched0 =1:
              }
              else
                switch packet at SE_{i,j} through OL link:
              break:
          1: if(packet_switched0 == 0)
            {
              switch packet at SE_{i,j} through OLM link;
              packet_switched0 =1;
            }
            else
              switch packet at SE_{i,j} through OU link:
            break;
          }
        }
      break:
    3: deleted one of the three packets having d_j = 0;
      switch_three_packets(SE_{i,j} , alpha);
```

```
                break;
              4: delete two of the four packets at random;
                switch_two_packets(SE_{i,j} , alpha);
                break;
              }
            }
          }
        }
      }
}

switch_one_packet (SE_{i,j} , alpha)
{
  for k varying from 1 to 4
  {
    if (packet at input port k of SE_{i,j})
      switch_bit = d_j of packet;
  }
  case(switch_bit)
  {
    0: case (alpha)
       {
         0: switch packet at SE_{i,j} through OUM link;
            break;
         1: switch packet at SE_{i,j} through OLM link;
            break;
       }
       break;
    1: case (alpha)
    {
      0: switch packet at SE_{i,j} through OLM link;
         break;
      1: switch packet at SE_{i,j} through OUM link;
         break;
    }
  }
}

switch_two_packets(SE_{i,j} , alpha)
{
```

```
switch_bit0 = number of packets at input ports having bit d_j =0:
case(switch_bit0)
{
  0: packet_switched = 0;
     for k varying from 1 to 4 do
     {
       if(SE_{i,j} has a packet at input port k)
       case (alpha)
       {
         0: if( packet_switched == 0)
            {
              switch packet at SE_{i,j} through OLM link:
              packet_switched =1:
            }
            else
              switch packet at SE_{i,j} through OU link:
            break:
         1: if(packet_switched == 0)
            {
              switch packet at SE_{i,j} through OUM link:
              packet_switched =1:
            }
            else
              switch packet at SE_{i,j} through OL link:
            break:
       }
     }
     break:
  2: packet_switched = 0:
     for k varying from 1 to 4 do
     {
       if(SE_{i,j} has a packet at input port k)
       case (alpha)
       {
         0: if(packet_switched == 0)
            {
              switch packet at SE_{i,j} through OUM link:
              packet_switched =1;
            }
            else
```

```
            switch packet at SE_{i,j} through OL link;
          break;
        1: if(packet_switched == 0)
          {
            switch packet at SE_{i,j} through OLM link;
            packet_switched =1;
          }
          else
            switch packet at SE_{i,j} through OU link;
          break;
      }
    }
    break;
1: for k varying from 1 to 4 do
    {
      if(SE_{i,j} has a packet at input port k)
      {
        switch_bit = d_j of packet;
        case (alpha)
        {
          1: case (switch_bit)
            {
              0: switch packet at SE_{i,j} through OLM link;
                 break;
              1: switch packet at SE_{i,j} through OUM link;
                 break;
            }
            break;
          0: case (switch_bit)
            {
              1: switch packet at SE_{i,j} through OLM link;
                 break;
              0: switch packet at SE_{i,j} through OUM link;
                 break;
            }
            break;
        }
      }
    }
    break;
```

```
    }
}

switch_three_packets(SE_{i,j} , alpha)
{
  switch_bit0 = number of packets at input ports having bit d_j =0:
  case(switch_bit0)
  {
    0: delete one of the packets in input port at random:
       switch_two_packets(SE_{i,j} , alpha):
       break:
    1: switch_two_and_one(SE_{i,j}, 0 , alpha):
       break:
    2: switch_two_and_one(SE_{i,j}, 1, alpha):
       break:
    3: delete one of the packets in input port at random:
       switch_two_packets(SE_{i,j} , alpha):
       break:
  }
}

switch_two_and_one(SE_{i,j} , bit, alpha)
{
  packet_switched = 0:
  for k varying from 1 to 4
  {
    if(packet at input port k of SE_{i,j} and d_j ==bit)
    {
      case (alpha)
      {
        0: case (bit)
           {
             0: switch packet at input port k of SE_{i,j} through OUM link:
                break;
             1: switch packet at input port k of SE_{i,j} through OLM link;
                break;
           }
           break;
        1: case (bit)
           {
```

```
            0: switch packet at input port k of SE_{i,j} through OLM link:
               break:
            1: switch packet at input port k of SE_{i,j} through OLM link:
               break:
          }
          break:
    }
}
else
{
    if(SE_{i,j} has a packet at input port k)
    case (alpha)
    {
       0: case (bit)
          {
            0:if(packet_switched == 0)
              {
                switch packet at SE_{i,j} through OUM link:
                packet_switched =1:
              }
              else
                switch packet at SE_{i,j} through OL link:
              break:
            1: if(packet_switched == 0)
              {
                switch packet at SE_{i,j} through OLM link:
                packet_switched =1;
              }
              else
                switch packet at SE_{i,j} through OU link:
              break;
          }
          break:
       1: case (bit)
          {
            0:if(packet_switched == 0)
              {
                switch packet at SE_{i,j} through OLM link:
                packet_switched =1;
              }
```

```
                else
                  switch packet at SE_{i,j} through OU link;
                break;
              1: if(packet_switched == 0)
                {
                  switch packet at SE_{i,j} through OUM link;
                  packet_switched =1;
                }
                else
                  switch packet at SE_{i,j} through OL link;
                break;
            }
            break;
        }
    }
  }
}
```