Throughput Analysis of a Fault-Tolerant Optical Switch

A. M. Memon, M. Guizani and M. S. Benten
College of Computer Sciences and Engineering
King Fahd University of Petroleum & Minerals
Dhahran - 31261, Saudi Arabia
email: {atif,mohsen,benten}@ccse.kfupm.edu.sa

Abstract

The throughput and reliability analysis of a fault-tolerant high-performance optical switch is presented. Comparison with two major fault-tolerant networks is provided. The results show that without using redundant switches, better network survivability is achieved. The results of the analysis show that the analyzed switch [1] has a performance comparable to other fault-tolerant networks whereas it outperforms them in other aspects such as number of extra switches/links and therefore provides better cost-effectiveness.

1 Introduction

Massively parallel processing applications executing on multiprocessor systems require a robust interconnection network [2]. This is due to their need to exchange large volumes of data. These could either be in the form of frequent short bursts or large continuous streams.

Several implementations of fault tolerant interconnection networks can be found in the literature [3, 4, 5, 6]. Almost all of these introduce redundancy in the network in terms of adding extra links and switches. These solutions are expensive since in most cases they increase the number of links [3].

Optical interconnections offer a high communication bandwidth. Using conventional fault tolerant schemes, extra links and switches must be introduced in the network. Under normal operating conditions, these are hardly ever used. However, once there is a fault in the network (e.g. a faulty switch), these links/switches completely take over the operations. Once a switch is detected as faulty, its incoming and outgoing links become useless. This results in wastage of bandwidth. In massively parallel processing applications, this is highly undesirable.

The analyzed 2×2 switch architecture [1] is an extension of existing switches that achieve straight and exchange connections. Fault-tolerance is increased by introducing buffers in the switch that act as queues and store the packets that cause contention, hence preventing their loss. Additional fault tolerance is provided by augmenting the switch with two circuits. These are responsible for detecting faults in the network and correcting them. They can also route data to other switches in case the main module of the switch fails.

The network is compared with existing fault-tolerant networks – the Benes' network and Itoh's network. The Benes' network introduces fault-tolerance in the network by adding additional links/switches. Itoh's self-routing fault-tolerant ATM switching network [3] provides multiple paths by adding sub-switches between switching stages of a Banyan network. It has a large number of redundant paths. It maintains high throughput with acceptable switching delay even when element failures occur. However, substantial amount of redundant SEs/links are required and the fault-tolerance decreases drastically on switch failure in the first and last stages of the network.

2 Performance Analysis

In order to have meaningful discussion about the throughput analysis, we should first discuss the reliability analysis.

2.1 Reliability Analysis

The assumptions made in carrying out this analysis are: First, the failure of one switch can in no way affect the reliability of any other switch in the network. Second, the network is said to have failed if at least one connection between input and output ports cannot be realized. Most of other reliability analyses given in the literature [3] assume the first and last stages of the network are fully operational under all conditions. This is not a realistic assumption since all the switches in the network are equally likely to fail. Therefore, in this analysis, all switches have equal probability of failure, that is including all switches of the first and last stages.

Figures 1 and 2 show the results of both Itoh's and Benes' networks using the above assumptions. These results show that the survival probability, $Q(k)$, is very low (in the orders of ≤ 0.2). Figure 3 shows the survival probability, $Q(k)$, of the analyzed network. It is clear that it performs better for the same range of faults in addition to having lower slope of the corresponding curves. Note that in the analyzed network, no additional switches/links are required. The total number of switches in the network is $n2^{(n-1)}$. Any of these switches can fail with equal probability.

If the number of failures is k, where $0 \leq k \leq n \times 2^{(n-1)}$, then the number of configurations in which k failures can occur is $\binom{n \times 2^{(n-1)}}{k}$.

0-7803-3255-5/96 $5.00 © 1996 IEEE

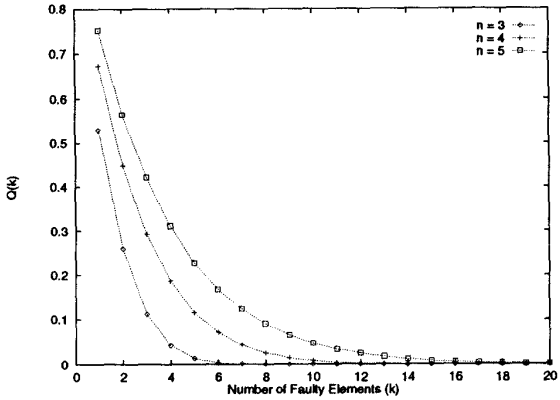

Figure 1: Network Survivability of Itoh's Network for Different Number of Faults.

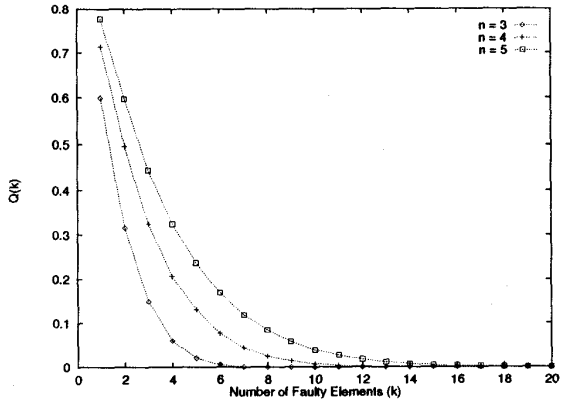

Figure 2: Network Survivability of Benes' Network for Different Number of Faults.

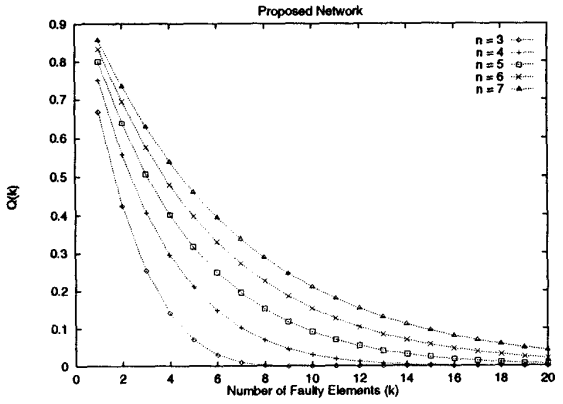

Figure 3: Network Survivability of Analyzed Network for Different Number of Faults.

Switch Size n	Benes	Itoh	Analyzed
2	6	5	4
4	56	49	32
6	352	321	192
8	1920	1793	1024
10	9728	9217	5120
12	47104	45057	24576
14	221184	212993	114688

Table 1: Total Number of Switches.

Therefore the survival probability $Q(k)$ is given by:

$$Q(k) = \frac{\binom{n2^{(n-1)}}{k}}{\binom{(n-1)2^{(n-1)}}{k}} \quad (1)$$

The total number of switches in all the three networks, Benes', Itoh's and the analyzed network are given by the following formulae:

$$Benes' \ Network = 2^{-1+n} \times (-1 + 2 \times n) \quad (2)$$

$$Itoh's \ Network = \sum_{i}^{n-1} 2^i \times 2^{n-i} - 1 \quad (3)$$

$$Analyzed \ Network = n \times 2^{-1+n} \quad (4)$$

Table 1 compares the total number of switches needed to achieve the reliability for all three cases. It is clear that this number is doubled in the case of Itoh's and Benes' networks. This reduction of the number of switches in the analyzed network is desirable since it reduces the cost and complexity.

Table 2 shows that the number of redundant paths in the analyzed network is similar to that of the Benes'

Switch Size n	Benes	Itoh	Analyzed
2	2	2	2
3	4	5	4
4	8	14	8
5	16	42	16
6	32	132	32
7	64	429	64
8	128	1430	128
9	256	4862	256
10	512	16796	512

Table 2: Total Number of Redundant Paths.

network and much less than Itoh's network. This is also a desirable feature since the number of links is reduced. The only advantage of introducing redundant paths in a network is to increase its fault-tolerant. In this case the analyzed network increases the fault-tolerance without increasing the physical links in the network.

The definition of $Q(k)$ can be used to find $P(k)$, the probability that entire network will fail after k faults. $P(k)$ can be computed as

$$P(k) = 1 - Q(k). \tag{5}$$

If $p(i)$ is the probability that the i^{th} fault will cause the entire network to fail, then:

$$P(k) = \sum_{i=2}^{k} p(i). \tag{6}$$

Let $\mathcal{E}$ be the expected number of faulty switches that will cause the entire network to fail. Then $\mathcal{E}$ can be obtained as:

$$\mathcal{E} = \sum_{i=2}^{L} ip(i). \tag{7}$$

The parameter L in Eq. 7 represents the total number of switches that can fail. This parameter is a cause for concern when performing the analysis. The assumption used in [3, 4] does not take into consideration the first and last stages. This assumption weakens the analyses as the switches in these stages are equally likely to fail. A more realistic analysis would consider all the switches of the network. This results in two different kinds of analyses. The first one with L as the number of switches in the intermediate stages (same as [3, 4]), and the second with L as the total number of switches in the network.

Using the first assumption i.e., L is the number of switches that can fail in the intermediate stages, and using exactly the same values of n, $\mathcal{E}$ and L as [3], the results can be summarized in Table 3. This table shows that the failure of all the switches in the intermediate stages will not affect the analyzed network in any way. Such a failure will result in decrease of throughput but the network will continue to operate.

Network	n	$\mathcal{E}$	L	$\mathcal{E}/L(\%)$
Benes				
	3	5.1	32	16.1
	4	8.3	96	8.7
	5	12.1	256	4.7
Itoh				
	3	7.1	36	19.6
	4	15.1	116	13.0
	5	29.8	324	9.2
Analyzed				
	3	4	4	100
	4	8	8	100
	5	16	16	100

Table 3: $\mathcal{E}$, L, and Cost-Effectiveness of the three networks ignoring the first and last stages.

Network	n	$\mathcal{E}$	L	$\mathcal{E}/L(\%)$
Benes				
	3	5.1	40	12.75
	4	8.3	112	7.41
	5	12.1	256	4.20
Itoh				
	3	7.1	44	16.13
	4	15.1	132	11.43
	5	29.8	356	2.58
Analyzed				
	3	4	12	33.33
	4	8	32	25.00
	5	16	80	20.00

Table 4: $\mathcal{E}$, L, and Cost-Effectiveness of the three networks including the first and last stages.

Using the second definition of L, where all the switches of the network are considered, it is seen (in Table 2.1) that the cost-effectiveness of all the networks reduces. However, even here, the analyzed network performs much better than the others.

2.2 Throughput Analysis

Let's consider the analyzed switch with at least one preceding stage and one succeeding stage. This situation behavior can be considered as an open queueing model with three queues in tandem (see Figure 4). Consider also a single M/M/1 queueing system. Let the input Poisson process have mean arrival rate λ. Then, as is shown in [7, 8], the output process is also Poisson with the same rate λ. This property of an M/M/1 system leads to *Kleinrock's independence assumption* for open queueing models of computer networks. When this assumption is feasible, a tractable analysis is possible.

Consider now the analyzed system represented by the three-stage open queueing model shown in Figure 4. All external arrival data packets are assumed to be Poisson, and it is assumed that message lengths can be approximated by an exponential distribution.

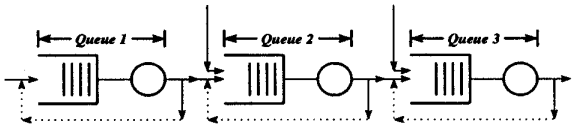

Figure 4: An Open Queuing Model with Three Queues in Tandem.

By Burke's Theorem [8], the output of the first queue is Poisson. However, the length of a particular message, after it is chosen from a particular distribution, is fixed. This means that the service times at the three nodes are not independent, so the three-stage network can not be modeled exactly as three separate M/M/1 queues. However, if it is assumed that a new choice of message length is made at each length, then the average delay for a message passing through the system is the sum of the average delays given by the three independent M/M/1 submodels. For the independence assumption to be valid, the message lengths at the input to the second and third queues must be selected from independent exponential distributions. This, of course is not exactly the case, but at least it serves the purpose of this analysis. Under the assumption that the length of the message entering the first queue in Figure 4 is chosen from an exponential distribution, the first queue is M/M/1 since its arrivals are from a Poisson process, its message lengths are exponentially, and arrivals and message lengths are independent random variables. As the message proceeds to queue 2, there is clearly a correlation between arrival time and message length since large messages take longer to be processed out of the first queue. Thus, we can conclude that message lengths and arrival times can not be independent, so the second and subsequent queues can not be M/M/1. Kleinrock [7] has shown that the independence assumption gives accurate results if the traffic intensity is small (which is not the case at hand) or if there is sufficient mixing at each node in the sense that messages join the queue from several input lines. The breakdown of complicated multistage queueing models, representing any multistage interconnection network, into a collection of independent M/M/1 or M/G/1 submodels makes the analysis of such networks, such as the analyzed switching architecture, relatively straightforward. When this is not possible, such as this case, a simulation technique must be used.

The throughput of a network is a measure of the speed of the network, i.e., the time it takes the input data to reach an output port. The throughput can be influenced in a variety of ways.

- The number of collisions in the network.
- The number of recirculations that the data has to undergo.
- The number of faulty elements in the network.

One switch of the network was modelled as an object and the whole network was assembled as a collection of instances of this object. Simulation involved creating three different networks of different sizes. Input data was generated and faults were introduced in the network. The throughput was measured as the data packets reached the output ports.

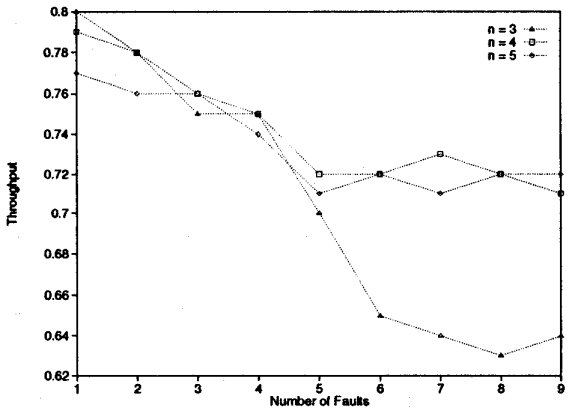

Figure 5: Throughput of the Analyzed Network for Different Network Sizes.

Figure 5 shows a graph of the throughput for different network sizes. Although the throughput is not high for $n = 3$, i.e., small network size, it stabilizes for larger networks. The main reason for this stability is the duplication of data in the network and its removal upon detection. As the faults occur randomly in the network, it is rare to have consecutive faults. Hence, the replicated data is not allowed to propagate through many stages.

3 Conclusion

In this paper, the reliability and throughput analyses of a fault-tolerant switch were presented. The results of the performance and reliability analyses showed that without using any additional hardware, better network survivability is achieved. Cost-effectiveness analysis showed that the analyzed switch is much more cost-effective than other fault-tolerant switches, while maintaining the same throughput.

Acknowledgement

The authors would like to acknowledge the support of the King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia.

References

[1] A. M. Memon, M. Guizani, "Design of a Self-Routing Optical Switch for Fault-Tolerant Interconnection Networks," Accepted for publication in the *International Journal of Communication Systems.*

[2] T. Y. Feng, "A survey of interconnection networks," *IEEE Computer Mag.*, Vol. 4, Dec. 1981, pp. 12-27.

[3] A. Itoh, "A fault-tolerant switching network for B-ISDN," *IEEE J. Select. Areas Commun.*, Vol. 9, No. 8, Oct. 1991, pp. 1218-1226.

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

[5] S. M. Reddy and V. P. Kumar, "On fault-tolerant multistage interconnection networks," *Proc. of the International Conference on Parallel Processing*, Aug. 1984, pp. 155-164.

[6] G. B. Adams and H. J. Siegel, "Modifications to improve the fault-tolerance of the extra stage cube interconnection network," *Proc. of the International Conference on Parallel Processing*, Aug. 1984, pp. 169-173.

[7] L. Kleinrock, it Queueing Systems: Computer Applications, Vol. 2, New York: Wiley-Interscience, 1976.

[8] B. J. Burke, "The Output of a Queueing System", *Operations Research*, Vol. 4, (1966), pp. 699-706.