고 신뢰성 큐브 네트웍☆

Enhanced Cube Network for the High Reliability

문 영 성*
Youngsong Mun

요 약

고성능 컴퓨팅 및 통신을 위하여 상호연결네트워크는 효율적이며 신뢰성이 높아야 한다. 상호연결네트워크의 웰트 톨러런스를 높이기 위해 제안되었던 이전 방안들은 효율성이 떨어지거나 추가적인 오버헤드가 너무 많이 필요하였다. 따라서 본 논문에서는 비교적 간단한 구조로 높은 신뢰성을 제공하면서도 효율적인 상호연결네트워크를 제안한다. 제안된 구조의 신뢰도를 이전의 제안되었던 구조와 비교하여 우수함을 입증한다.

Abstract

Multistage Interconnection networks (MIN) for the high performance computing and communications must be efficient and reliable. While a number of fault tolerance schemes have been developed, some of them are not efficient enough with respect to all evaluation measures or overheads of others are too significant. In this paper we develop a new efficient fault tolerant MIN which displays high reliability and fault tolerance capability using a simple structure. Structure and reliabilities of Enhanced Cube Network are evaluated and compared with previous designs to show the effectiveness of new design.

⌐ Keyword : MINs, Fault Tolerance, Reliability

1. Introduction

One of major problems suffered in parallel and distributed processing is the degradation of performance due to the communication overhead. Therefore, interconnection networks by which efficient communication can be achieved are crucial to this system. The simplest types of interconnection networks are crossbar and single shared bus which have problems in cost and performance, respectively. To solve these problems, various multistage interconnection networks(MINs) have been proposed[1-5].

To improve the reliability and performance of MINs, fault tolerance capabilities should be implemented for MINs. A number of different fault tolerance designs for MINs, which involve modification of original topology, have been proposed in the literature[6-15]. In providing path redundancy several typical schemes employed such as adding extra stages or rows of switches[6-8], varying switch complexities[8-9], extra links[10], and duplicating networks[14]. These schemes are used alone or some of them are combined together to obtain the desired fault tolerance capability. Effectiveness of these fault tolerance designs are evaluated usually with respect to terminal and network reliability, Mean Time To Failure(MTTF), and network survivability.

The proposed Enhanced Cube Network(ECN) is developed based on the fact that the network reliability will be very high and a large number of fault can be tolerated if original connectivity of MIN is enhanced efficiently. This brought a design which allows the network to be alive as far as each switch in the conjugate pairs (and attached

* 종신회원 : 숭실대학교 컴퓨터학부 부교수
mun@computing.ssu.ac.kr(제1 저자)
☆ 본 연구는 숭실대학교 교내연구비 지원으로 이루어졌음.

link) in the path is good. Additional hardware required to support this feature is as simple as two multiplexers and demultiplexers (mux/demux) in each switch node and the doubled links between stages. The network reliability is much higher than other previously proposed designs with a much more significant or comparable hardware overhead. Due to the structural simplicity and regularity of the design, routing is also very simple and straightforward.

2. Reliability Measures

• Terminal Reliability

Terminal reliability(TR) is defined to be the probability of existence of at least one fault-free path between a given pair of input and output terminal. Redundancy graph[17] is a directed graph showing the possible paths between a given input and out terminal.

• Network Reliability

Network Reliability(NR), also called all-terminal reliability, addresses the probability that at least one path exists between every source and destination pair.

• Mean Time To Failure(MTTF)

If $NR(t)$ is the time dependent network reliability, MTTF can be obtained by

$$MTTF = \int_0^\infty NR(t)dt$$

• Network Survivability

Network survivability is the probability that the network is alive in the presence of a fixed number of faults. Therefore, it is a good measure of the detrimental effect of an additional fault on the network reliability.

3. The Proposed Enhanced Cube Network

3.1 Motivation

Even though the approach employed in ESC is efficient, it has several shortcomings which requires some improvement. First of all, ESC is effective for only a small number of faults (so requires high component reliability) due to its somewhat rigid utilization of extra stages[16,19]. For example, for a 16x16 ESC, the network survivability rapidly drops down to 0.089 from 1.0, when the number of switch faults increases to three from one[19]. Moreover, this shortcoming becomes more important as the network size grows. This is because the role of the extra stage for providing redundant paths becomes less significant.

3.2 Structure of ECN

An 8x8 ECN is shown in Figure 1. As explained, one additional set of links(what I call conjugate links) is augmented between every stage to realize the connections of each conjugate switch. For example, $SW_{0,2}$ has additional connections to $SW_{1,1}$ and $SW_{3,1}$, to which $SW_{1,2}$(conjugate of $SW_{0,2}$) has connections. All other additional links are made by the same way.

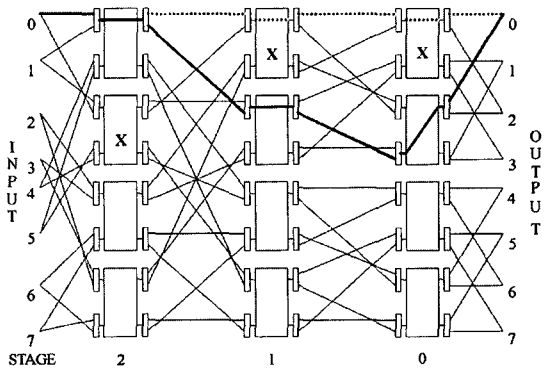


Figure 1. Structure of ECN

Also, because each input and output port of a switch is connected to two links now, a multiplexer and demultiplexer is required to put in each place. Notice that the original links are connected to the upper input position of each mux/demux, while the conjugate links are connected to the lower input position except the links to output terminals. This arrangement of links is for correctly routing messages using a bit of the destination address, as explained in the section for routing.

As shown in Table I of structure comparisons in Section 4.1, the hardware requirement of ECN is much smaller than DR and Gamma network. It is also said to be at least comparable with ESC, which has one extra column of switches. Note that mux/demux in our ECN operate in a different fashion from ESC such that it is not used to bypass switches but simply select one link out of two. Thus each network cycle is uniform, and no extra delay occurs because of extra stage between communicating pairs like ESC.

Figure 1 shows an example how the connection between input port 0 and output port 0 is realized in ECN with some faults. Here *X* denotes the faulty switch nodes. Note that, with this fault distribution, ESC can not allow the connection for this pair. As shown in this example, ECN will survive from much more significant number of faults than ESC.

3.3 Performance Evaluation of ECN

• Terminal Reliability

Redundancy graph of ECN is shown in Figure 2. There exist four links between adjacent stages and eight links between stage 1 to 0 for any input and output pair in ECN. However, to employ a practical simple digit controlled routing scheme, the number of available links between any stage is fixed to four.

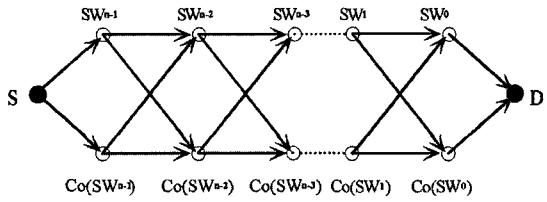


Figure 2. Redundancy graph of ECN

For the explanation of redundancy graph of Figure 2, let's denote SW_j to any switch box in *Stage j*. I also assume R_j to be the probability that at least one fault-free path is available from SW_{j+1} to the destination D, given that SW_{j+1} is fault-free. R^c_j is assumed in the same way using $Co(SW_{j+1})$. Here, j is from n-2 to 0. Because destinations are nonfaulty, $R_{-1} = R^c_{-1} = 1$ Also, by the symmetry of the redundancy graph, $R_j = R^c_j$ is obtained. The following recurrence relationship can then be obtained from Figure 2.

R_j = Prob{SW_j is nonfaulty} R_{j-1} + Prob{SW_j is faulty} Prob{Co(SW_j)is nonfaulty}R^c_{j-1}

$$R_j = pR_{j-1} + p(1-p)R^c_{j-1}$$
$$R^c_j = pR^c_{j-1} + p(1-p)R_{j-1}$$

TR_{ECN} can be obtained as follows.

$$\begin{aligned} TR_{ECN} &= \Pr ob\{SW_{n-1} is\ nonfaulty\}R_{n-2} \\ &\quad + \Pr ob\{SW_{n-1} is\ faulty\}\Pr ob \\ &\quad \{Co(SW_{n-1}) is\ faulty\}R^c_{n-2} \\ &= pR_{n-2} + (1-p)pR^c_{n-2} \\ &= \{p + (1-p)p\}^n \end{aligned}$$

• Network Reliability and MTTF

ECN is alive as long as no conjugate pair of switch boxes are faulty while there exist $nN/4$ conjugate pairs. Thus network reliability is simply obtained as follows.

$$NR_{ECN} = \left\{1 - (1-p)^2\right\}^{\frac{nN}{4}}$$

Tabel I. Comparison of structures of the fault tolerant MINs.

Features	G. Cube	SEN+	ESC	DR	Gamma	ABN-1	ECN
Number of switch boxes	nN/2	(n+1)N/2	(n+1)N/2	(n+1)(N+S)	(n+1)N	(n-1)N/2	nN/2
Size of switch	2x2	2x2	2x2	3x3	3x3	5x3,3x3,3x5	2x2
Number of links	(n-1)N	nN	nN	3n(N+S)	3nN	(3n/2-1)N	2(n+1)N
Extra hardware	None	None	Mux, Demux	None	None	Aux. link	Mux, Demux

MTTF of ECN is given by

$$MTTF_{ECN} = \frac{1}{\lambda} 2^k \left[\sum_{i=0}^{k} \frac{1}{k+i} \left\{ {k \atop i} \right\} (-2)^i \right]$$

Here, $k = \frac{nN}{4}$.

• Network Survivability

In an $N \times N$ ECN, there exist $nN/2$ number of switch boxes and $nN/4$ number of conjugate pairs. If at least one switch box is good in any conjugate pair, the network can survive. Let's assume that there exist f number of faulty switch boxes. The total number of combinations of selecting f switch boxes is $\left\{ {nN/2 \atop f} \right\}$, and $\left\{ {nN/2 \atop f} \right\}$ number of ways of selecting conjugate pairs exists. In each pair selected, there exist two choices of switch boxes. Therefore NS of ECN can be obtained as follows.

$$NS_{ECN} = \frac{\left\{ {nN/4 \atop f} \right\} 2^f}{\left\{ {nN/2 \atop f} \right\}}$$

4. Comparison of ECN with Other Designs

4.1 Comaprison of Structures

Table I summarizes and compares the structures of several fault-tolerant MIN including ECN. Observe from the table that ECN requires smaller number of switch nodes than ESC, while links are about twice. Also, it needs more multiplexers and demultiplexers. However, links and mux/demuxs (pure combinational circuit) are much simpler and reliable than switching elements. Therefore, ECN can be said to require at least a comparable structural overhead as ESC. ABN-1 requires slightly fewer switch nodes and links than ECN, while much more complex and nonhomogeneous switch nodes are used. Again, the overheads in terms of hardware complexity are about the same. However, more importantly, structure of ECN is completely regular and simple, while the other two are not. The property of regularity is very important for MINs because it directly determines the complexity of message routing control, and eventually the performance of the system. Gamma and DR networks require significantly higher hardware overhead.

4.2 Comparison of Terminal Reliabilities

Figure 3 compares the terminal reliability of various networks when p=0.9. The TR of ESC is consistently higher than SEN+ as expected because ESC network can still be alive even when switch boxes in either the first or last stage fail, by utilizing multiplexers and demultiplexers.

As expected, ECN displays the highest terminal reliability among the designs studied. When p=0.9, the terminal reliability changes from 0.970 to 0.886 if the network size grows from 8x8 to 40964096. Observe that the TR of ECN is consistently higher than ABN-1 as well as ESC, and the difference is

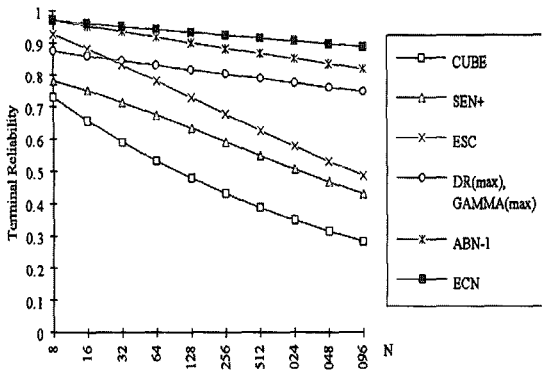


Figure 3. Comparisons of Terminal Reliability.

more significant for relatively large networks. With switch boxes of higher reliability(p=0.99) ESC performs better. However, ECN still outperforms ESC even in this case. Note that the effectiveness of TR as a mea sure for network reliability and performance varies depending on the network topology (path sharability among pairs).

4.3 Comparison of Network Reliabilities

Table II compares the network reliabilities of the networks when p=0.9. Improvement of network reliability of ECN is even more significant than for terminal reliability. Network reliability of ESCub and ECN are 0.203 and 0.851, respectively when N is 16 and p=0.9. The improvement increases geometrically as the network size grows.

ABN-1 requires much more complex switches than ours while slightly fewer links are used. Because of its operational characteristic of looping between switches, switches and the overall network structure are not uniform. As mentioned this is expected to cause routing and switching control very difficult. Thus, ECN seems to be more practical.

ECN is expected to even improve the network performance because messages can be rerouted using

Tabel II. Network reliabilities of fault tolerant MINs (p=0.9)

Network Size(N)	Generalized Cube	SEN+$_{lb}$	SEN+$_{ub}$	ESC$_{lb}$	ESC$_{ub}$	Gamma	DR(S=3)
4	6.56×10^{-1}	6.50×10^{-1}	6.50×10^{-1}	8.99×10^{-1}	8.99×10^{-1}	2.82×10^{-1}	6.45×10^{-1}
8	2.82×10^{-1}	3.80×10^{-1}	4.14×10^{-1}	5.74×10^{-1}	6.08×10^{-1}	3.43×10^{-2}	1.40×10^{-1}
16	3.43×10^{-2}	8.98×10^{-2}	1.64×10^{-1}	1.29×10^{-1}	2.03×10^{-1}	2.18×10^{-4}	1.75×10^{-3}
32	2.18×10^{-4}	2.32×10^{-3}	2.49×10^{-2}	2.67×10^{-3}	2.53×10^{-2}	0.0	0.0
64	0.0	0.0	5.28×10^{-4}	0.0	5.28×10^{-4}	0.0	0.0
128	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Network Size(N)	ABN-1$_{ub}$	ABN-1$_{lb}$	ECN
4	9.90×10^{-1}	9.90×10^{-1}	9.80×10^{-1}
8	9.61×10^{-1}	9.44×10^{-1}	9.41×10^{-1}
16	8.86×10^{-1}	8.29×10^{-1}	8.51×10^{-1}
32	7.25×10^{-1}	5.94×10^{-1}	6.69×10^{-1}
64	4.48×10^{-1}	2.63×10^{-1}	3.81×10^{-1}
128	1.45×10^{-1}	3.83×10^{-2}	1.05×10^{-1}
256	1.11×10^{-2}	4.52×10^{-4}	5.82×10^{-3}
512	3.39×10^{-5}	1.90×10^{-8}	9.37×10^{-6}
1024	8.78×10^{-11}	3.40×10^{-18}	6.70×10^{-14}
2048	4.49×10^{-23}	9.44×10^{-40}	2.62×10^{-29}
4096	6.84×10^{-50}	5.95×10^{-87}	2.32×10^{-64}

Tabel III. Comparison of network survivabilities of ECN and ESC

	ESC				ECN		
Network Size(N)	f=2	f=3	f=4 (Lower Bound)	f=4 (Upper Bound)	f=2	f=3	f=4
8	0.2333	0.0286	0.0022	0.0220	0.9091	0.7273	0.4848
16	0.3282	0.0890	0.0172	0.0648	0.9677	0.9032	0.8098
32	0.4105	0.1644	0.0464	0.1204	0.9873	0.9620	0.9245
64	0.4792	0.2409	0.0831	0.1809	0.9948	0.9843	0.9687

conjugate paths in the case of congestion. A number of schemes such as packet combining and buffer bypassing have been proposed for the performance enhancement of MIN with nonuniform traffic. One shortcoming of these schemes is the nontrivial hardware and software implementation overhead. Our ECN design will be a good candidate for solving this problem.

4.4 Comparison of Network Survivabilities

As shown in [19], network survivability(NS) of ESC is much better than that of DR. Therefore, here, comparisons between ESC and ECN are presented in Table III. From Table III, it can be seen that ECN is incomparably better than ESC in tolerating faulty components. ECN, thus, can be said to be much more preferable than ESC for an application where the system is required to fully operate with some faulty components.

5. Conclusions

To develop an efficient design for fault tolerant MINs, it is important to study the effect of implemented redundancy. Based on conjugate pair concept an efficient fault tolerant network called ECN is developed. It displays a very high terminal and network reliability. Structure of ECN is also very simple, and thus routing is straightforward. Because of this property of high reliability and structural simplicity, ECN is expected to be very effective for large scale parallel system especially in VLSI environment. One important aspect relevant with fault tolerant MINs is that the extra hardware implemented for fault tolerance can positively or negatively influence the performance of the original network.

References

[1] C.L. Wu and T.Y. Feng, "On a class of multistage interconnection networks," IEEE Trans. Computers, vol. C-29, pp. 694~702, Aug. 1980.

[2] H.J. Siegel, R.J. McMillen, "The multistage cube: A versatile interconnection network," Computer, vol. 14, pp. 65~76, Dec. 1981.

[3] J.H. Patel, "Performance of processor-memory interconnections for multiprocessors," IEEE Trans. Computers, vol. C-30, pp. 771~780, Oct. 1981.

[4] D.H. Lawrie, "Access and alignment of data in an array processor," IEEE Trans. Computers, vol. C-24, pp. 1145-1155, Dec. 1975.

[5] L.R. Goke and G.J. Lipovski, "Banyan networks for partitioning multimicroprocessor systems," 1st Symp. Computer Architecture, pp. 21~28, Dec. 1973.

[6] G.B. Adams III and H.J. Siegel, "The Extra Stage Cube: A fault-tolerant interconnection network for supersystems," IEEE Trans. Computers, vol. C-31, pp. 443~454, May 1982.

[7] J.T. Blake and K.S. Trivedi, "Multistage inter-

connection network reliability," IEEE Trans. Computers, vol. 38, pp. 1600~1604, Nov. 1989.

[8] M. Jeng and H.J. Siegel, "A fault-tolerant multistage interconnection network for multiprocessor systems using dynamic redundancy," 6th Int'l Conf. Distributed Computing Systems, Computer Society Press, Silver Spring, Md., pp. 70~77, 1986.

[9] D.S. Parker and C.S. Raghavendra, "The Gamma network: A multiprocessor interconnection network with redundant paths," Proceedings of the 9th Annual Symp. on Computer Architecture, pp. 73~80, April 1982.

[10] L.Ciminiera and A. Serra, "A connecting network with fault tolerance capabilities," IEEE Tans. Computers, pp. 578~580, June 1986.

[11] K.Padmanabhan and D.H. Lawrie, "A class of redundant path multistage interconnection networks," IEEE Trans. Computers, pp. 1099~1108, Dec. 1983.

[12] D.M. Dias and J.R. Jump, "Augmented and pruned NlogN multistage networks: Topology and performance," 1982 Int'l Conf. Parallel Processing, Computer Society Press, Silver Spring, Md., pp. 10~11, 1982.

[13] S.M. Reddy and V.P. Kumar, "On fault-tolerant multistage interconnection networks," 1984 Int'l Conf. Parallel Processing, Computer Society Press, Silver Spring, Md., pp. 155~164, 1984.

[14] C.S. Raghavendra and A. Varma, "INDRA: A class of interconnection networks with redundant paths," 1984 Realtime Systems Symp., Computer Society Press, Silver Spring, Md., pp. 1153~164, 1984.

[15] J.P. Shen and J.P. Hayes, "fault-tolerance of Dynamic-Full-Access Interconnection Networks," IEEE Trans. Computers. pp. 241~248, Mar. 1985.

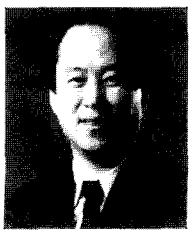
[16] G.B. Adams and H.J. Siegel, "Modifications to improve the fault tolerance of the extra stage cube interconnection network," 1984 Int'l Conf. Parallel Processing, pp. 169~173, Aug. 1984.

[17] K. Padmanabhan and D.H. Lawrie, "Fault tolerance schemes in shuffle-exchange type interconnection networks," Proceedings of the 1983 Int'l Conf. on Parallel Processing, pp. 71~74, Aug. 1983.

[18] V. P. Kumar and S. M. Reddy, "Design and analysis of fault-tolerant multistage interconnection networks with low link complexity," Proceedings of the 12 Annual Symp. on Computer Architecture, pp. 376~386, 1985.

[19] Y. Mun and H.Y. Youn, "On performance evaluation of fault tolerant multistage interconnection networks," in Proc. 1992 ACM Symp. on Applied Computing, pp. 1~8, March, 1992.

◐ 저 자 소 개 ◑



문 영 성

1983년 연세대학교 전자공학과 졸업(학사)
1986년 알버타대학교 대학원 전자공학과 졸업(석사)
1999년 텍사스대학교 대학원 컴퓨터공학과 졸업(박사)
1994년~현재 : 숭실대학교 컴퓨터학부 부교수
관심분야 : Mobile IP, IPv6, GRID, QoS, 성능분석, 이동단말 보안, Honeypot
E-mail : mun@computing.ssu.ac.kr