Guest Editors' Introduction: Challenges in Designing Fault-Tolerant Routing in Networks

Jie Wu, *Senior Member, IEEE*, and Dharma P. Agrawal, *Fellow, IEEE*

THERE are several means of interconnecting the components of a computer system. Components can be connected through *direct* or *indirect* connections. In a direct network, channels directly connect each processor. Neighboring nodes communicate with each other by sending messages across the channels. In an indirect network, processors are connected through switches and/or buses. Depending on how components are connected, we have bus-based networks, router-based networks, switch-based networks, and cluster-based (hybrid of the other three networks) networks. Networks can also be classified based on *topologies,* which can be either *regular* or *irregular.* Another classification is based on *applications*—there are local area networks (LANs), networks of workstations (NOWs), multiprocessors, and mobile networks.

In a computer network, the efficiency of a routing process is critical to the system performance. A routing process deals with moving data between processors in a given network. When a routing process involves one source and one destination, it is called *unicast* or *point-to-point* communication. When such a process involves more than one source and/or one destination, it is called a *collective communication process.* Examples of collective communication include *multicast* (one-to-many, one source and many destinations), *broadcast* (one-to-all), and *gossip* (all-to-all).

More recently, the lower cost NOWs have become accepted as a replacement for expensive supercomputers. Routing in NOWs is done primarily using a large LAN. Thus, routing in such networks has become increasingly critical for high performance computing applications. The router design and routing techniques are altogether different for NOWs and LANs, as compared to conventional multiprocessors. Support for PVM-type message communication mechanisms and their effective use has become increasingly important. Such schemes differ greatly in mobile networks, because the signals have to pass through the noisy atmosphere. Also, the conversion from medium access control in the air to ATM-type protocol for the backbone network has become a necessity in the communication world.

As the number of processors in a network increases, the probability of processor failure also increases. Therefore, it is important to design a routing process with fault tolerance capability to guarantee a successful routing in the presence of faults. However, techniques used to achieve fault tolerance are often at the expense of considerable performance degradation and reduction in adaptivity. In some other cases, a fault-tolerant routing is subject to deadlock and livelock if it is not carefully designed. Virtual channels and virtual networks can be used to maintain a certain degree of routing adaptivity and eliminate the deadlock and livelock situation. However, these mechanisms complicate the routing process and increase the circuit complexity of the router.

Fault-tolerant routing techniques can also be classified as *hardware-based* and *software-based.* In a hardware-based fault tolerant routing, fault-tolerant routers are customized to support a selected approach, such as the addition of virtual channels and enforcement of routing restrictions. However, when the fault rates are relatively low, a software-based approach is more justifiable. In addition, the information on fault distribution also plays an important role in fault-tolerant routing. Most of these models assume that each node either knows only the neighbors' status or the status of all the nodes. A model that uses the former assumption is called *local-information-based,* while a model that uses the latter assumption is called *global-information-based.* Normally, a global-information-based model can obtain an optimal or suboptimal result; however, it requires a complex process that collects global information. A *coded-information-based* model has been used, which is a compromise between local-information and global-information-based approaches.

Fault-tolerant routing also depends on many other factors, such as switching techniques, the type and nature of faults, network topology, system port models, and support for deadlock-free routing. It also involves a set of design choices: special vs. general purpose, minimal vs. nonminimal, deterministic vs. adaptive, and redundant vs. nonredundant. The goal of the special section of this issue is to put together some of recent results on fault-tolerant routing in networks. Eight papers are included which cover various aspects of fault-tolerant routing.

The first two papers deal with fault-tolerant routing in traditional networks with regular topology: hypercubes, 2D meshes, and high-dimensional hypercubes.

In Q.-P. Gu and S.-T. Peng's article, a routing algorithm is proposed for unicast in k-safe hypercubes. A hypercube is k-safe if each node in the cube has at least k nonfaulty neighbors. This fault model is a

For information on obtaining reprints of this article, please send e-mail to: tpds@computer.org, and reference IEEECS Log Number 110012.

1045-9219/99/$10.00 © 1999 IEEE

special case of forbidden faulty sets, where components in the system cannot be faulty at the same time. Three algorithms are proposed with optimal time complexities for different given numbers of faults in the system. It is also shown that these algorithms for unicast in the 1-safe and 2-safe hypercubes are optimal. An open problem is also given for finding an optimal path in k-safe hypercubes, with $k > 2$.

B.A. AlMohammad and B. Bose's paper studies fault-tolerant routing in k-ary n-dimensional hypercubes. The types of collective communication include one-to-all broadcasting, all-to-all broadcasting, one-to-all personalized communication, and all-to-all personalized broadcasting. Algorithms for these types of communication can tolerate $2n - 2$ node faults and they can be extended to cover $2n - 1$ node faults with an increase in path length. The communication complexities are also analyzed under different switching techniques.

Recently, the interest in research on routing networks with irregular topology has increased greatly because of the greater flexibility and scalability of irregular networks and the widespread use of Internet applications. Irregular networks can be implemented by switched networks, such as wormhole-switched networks, or optimal networks, such as optimal WDM (Wavelength Division Multiplexing) networks. The next three papers of this special section deal with fault-tolerant routing in networks with irregular topology.

W. Jia, W. Zhao, D. Xuan, and G.-C. Xu study fault-tolerant multicast routing protocol for internet applications. The widely used Core-Based Tree techniques (CBT) is extended with fault-tolerance capability and with better efficiency and effectiveness. The proposed scheme predefines backup paths to bypass faulty components. When a fault is detected, only the routers in the neighborhood need to be reconfigured to ensure the scalability of the approach. The performance evaluation shows that the proposed approach almost matches the best possible method that requires global tree reformation, while utilizing much less runtime overhead and implementation cost.

V. Halwan, F. Ozguner, and A. Dogan describe a deadlock-free routing in a special NOWs called "wormhole-switched clustered network," with several clusters connected to a central network. Global routing based on local algorithms (within each cluster) is proposed and is deadlock-free with two virtual channels. The use of the proposed approach in achieving fault-tolerant routing in 2D meshes is also shown.

In H. Shen, F. Chin, and Y. Pan's paper, the problem of efficient communication in unreliable optical networks is addressed. The optical networks under consideration are supported by WDM routing in which a communication path is established by chaining together several optical channels of possible different wavelengths. A set of algorithms in an unreliable WDM network on a proposed cost model is presented. These algorithms support multiple unicast and several collective communications, such as multicast and multiple multicast.

The last three papers in this special issue deal with hardware-based routing through special routers, performance analysis, and testbed of fault-tolerant routing.

Boppana and Chalasani propose a technique to incorporate fault tolerance into networks with partitioned dimension-order routers using a multichip. This technique works with local knowledge of faults, handles multiple faults, and guarantees livelock- and deadlock-free routing using four virtual channels. The simulation results show that the proposed technique exhibits similar graceful degradation of performance as mesh networks with crossbar-based dimension-order routers.

In M.P. Haynos and Y. Yang's paper, there is an analytical model on the blocking probability of the Clos network with link failures. Routing in a Clos network (a special type of multistage networks) is done through a connection request. The blocking probability is the probability that a connection request fails. Therefore, a low blocking probability corresponds to a high probability of a successful routing. The analytical and simulation results show a relatively small interstage link failure probability, and indicate that the Clos network has a good fault-tolerant capability.

The effectiveness of fault-tolerant routing techniques are mostly evaluated by either analytical models or simulation with synthetic workloads. However, it is more important to have a detailed analysis of a design with realistic workloads. In the last paper, A.S. Vaidya, C.R. Das, and A. Sivasubramaniam present an evaluation testbed for interconnection networks and routing algorithms using real applications. This testbed is illustrated with one fault-tolerant algorithm and four shared memory applications with different fault conditions.

List of Referees:

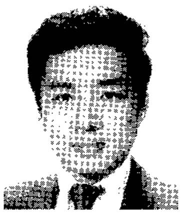
B.A. AlMohammad,
A. Arora,
D. Avresky,
R. Boppana,
B. Bose,
G.-H. Chen,
D.-L. Dai,
C. Das,
J. Duato,
E.B. Fernandez,
P. Fraigniaud,
R.I. Greenberg,
Q.-P. Gu,
S. Gupta,
C. Hamacher,
S. Han,
D.F. Hsu,
W.-J. Jia,
J. Kim,
Y. Lan,
S. Latifi,
R. Liberskind-Hadas,
X. Lin,
Y. Liu,

D.H. Linder,
D. Medhi,
R. Melhem,
S. Morinaga,
S. Murthy,
J. Ng,
S. Olariu,
Y. Oruc,
F. Ozguner,
T. Pinkston,
C. Qiao,
D.S. Reeves,
G. Rouskas,
D.K. Panda,
S.-T. Peng,
H. Shen,
T.K. Shih,
A. Sivasubramaniam,
N. Speirs,
P.K. Srimani,
D. Taylor,
K.G. Shin,
C. Stunkel,
Y. Tseng,
S. Varadarajan,
N.H. Vaidya,
S.-D. Wang,
L. Zakrevski,
W. Zhao,
A. Zomaya,
S. Yalamanchili,
Y. Yang,
H. Yoon,
B. Yurcik,
S.Q. Zheng

Acknowledgments

We would like to express our sincere thanks to all the authors of submitted papers and reviewers of this special issues. Special thanks are given to Professor John A. Stankovic, Editor-in-Chief, for his encouragement and guidance in editing this special issue.

The work of Jie Wu was supported in part by U.S. National Science Foundation grant CCR 9900646. The work of Dharma P. Agrawal was supported in part by the Ohio Board of Regents funds.



Jie Wu received the BS degree in computer engineering in 1982 and the MS degree in computer science in 1985 from the Shanghai University of Science and Technology. He received his PhD in computer engineering in 1989 from Florida Atlantic University, Boca Raton, and currently works there as a professor and director of the graduate programs for the Department of Computer Science and Engineering. Dr. Wu has published more than 100 papers in various journals and conference proceedings. His research interests are in the area of parallel and distributed systems, fault-tolerant computing, interconnection networks, Petri net applications, and software engineering. He serves on the editorial board of the *International Journal on Computers and Applications* and is the author of the text *Distributed System Design*. Dr. Wu is a recipient of the 1996-1997 Researcher of the Year Award at Florida Atlantic University, and of the 1998 Outstanding Achievements Award from IASTED. He is an IEEE Computer Society Distinguished Visitor, a member of ACM and ISCA, and a senior member of the IEEE.

Dharma P. Agrawal has pulished a number of papers in the areas of mobile networks, parallel system architecture, multicomputer networks, routing techniques, parallelism detection and scheduling techniques, reliability of real-time distributed systems, modeling of C-MOS circuits, and computer arithmetic. He is currently the Ohio Board of Regents' Distinguished Professor, and director of the distributed and mobile computing laboratory at the University of Cincinnati. He has served as an editor for *Computer, IEEE Transactions on Computers*, and the IEEE Computer Society Press. At present, he is an editor of the *Journal of Parallel and Distributed Computing* and the *International Journal of High-Speed Computing*. Dr. Agrawal is a fellow of the ACM and the IEEE.