A Novel Globally Adaptive Load-Balanced Routing Algorithm for Torus Interconnection Networks

Hong Wang, Du Xu, and Lemin Li

ABSTRACT—A globally adaptive load-balanced routing algorithm for torus interconnection networks is proposed. Unlike previously published algorithms, this algorithm employs a new scheme based on collision detection to handle deadlock, and has higher routing adaptability than previous algorithms. Simulation results show that our algorithm outperforms previous algorithms by 16% on benign traffic patterns, and by 10% to 21% on adversarial traffic patterns.

Keywords—Torus interconnection network, adaptive routing algorithm, load-balance, traffic pattern, deadlock.

I. Introduction

Torus interconnection networks are widely used as processor/memory interconnects in parallel computing systems [1] or packet switching fabrics (PSF) in Internet routers/switches [2]. The performance of a torus network is significantly affected by its traffic patterns, which can be classified into two categories [3], [4]. Some traffic patterns may cause load imbalance in the network if minimal routing is adopted, so they are called *adversarial* patterns; others, called *benign* patterns, do not cause load imbalance.

Torus networks in real systems are often required to perform well on various traffic patterns. For example, PSF in Internet routers should handle various traffic patterns due to the variety of Internet traffic patterns [3]. Previous works [3], [5] have addressed this problem. Recently, globally adaptive load-balanced routing algorithms, such as GAL [4] and CQR [6] have been proposed, which can make routing decisions adaptive to the global state of the network. They are the best solutions by far to achieve high throughput on both benign and adversarial traffic patterns. However, GAL and CQR employ a deadlock-handling scheme that needs three virtual channels (VCs) per physical channel, but only two out of the three can be alternative output VCs simultaneously. In this letter, we propose an algorithm called the globally adaptive load-balanced routing with collision detection (GALR-CD). It employs a new scheme based on collision detection to handle deadlock, and all the VCs on the physical channel can be alternative output VCs. Therefore, it has higher routing adaptability, and outperforms GAL and CQR on both benign and adversarial traffic patterns.

Manuscript received Jan. 08, 2007; revised Mar. 29, 2007.

This work was supported by National Natural Science Foundation of China (60372011), State Key Development Program of Basic Research of China (2007CB307104 of 2007CB307100), and Youth Research Fund of UESTC 2007.

Hong Wang (phone: +86 28 83203008, email: hwang@uestc.edu.cn), Du Xu (email: xudu@uestc.edu.cn), and Lemin Li (phone: +86 28 83202343, email: lml@uestc.edu.cn) are with Key Lab of Broadband & Optical Communications, University of Electronic Science and Technology of China, Chengdu, China.

II. Problem Definition

A *torus* interconnection network is a graph in the n-dimensional space, with totally $k_0 \times k_1 \times \cdots \times k_{n-1}$ nodes, k_i nodes along dimension i, where $k_i \geq 2$ and $0 \leq i \leq n-1$. Each node x in the graph can be identified by a coordinate vector, $(x_0, x_1, \cdots, x_{n-2}, x_{n-1})$, where $0 \leq x_i \leq k_i - 1$ for all i. Each channel traverses a single dimension. Two nodes, x and y, are connected through a channel if and only if $y_i = x_i$ for all i except one, j, where $y_j = (x_j \pm 1) \bmod k_j$. A 4×3 torus is shown in Fig. 1.

In GALR-CD, there are two types of VCs per physical channel: a special VC (s-VC) and at least one normal VC (n-VC). The concatenation of the s-VCs along the same direction forms a ring, which is referred as the *s-VC ring*.

In torus networks, a deadlock will arise if the requests from the packets to their alternative output VCs form a cycle, and each packet cannot advance because all the alternative output VCs are already occupied by other packets [1].

III. Globally Adaptive Load-Balanced Routing with Collision Detection

Collision detection is performed for each s-VC ring, either by a centralized detector or in a distributed manner. The function of collision detection is presented as follows:

- When a packet P requests an s-VC of the s-VC ring R^s, if R^s is not reserved to other packet, the request of P is granted and R^s is labeled as being reserved to P; otherwise, the request is denied.
- The reservation is held until P has left the current dimension, either reaching its destination or turning to another dimension.

An example is shown in Fig. 1. The packet has been granted at node 0 to advance on the s-VC, and then at node 1 it leaves the s-VC, advancing on the n-VC. When the packet requests an s-VC again at node 2, it is granted since the s-VC ring is still reserved to it. During this period, no other packet can advance on this s-VC ring.

In GALR-CD, the packet advances along only one direction of each dimension. The n-VCs can be used without any more constraints. The s-VCs can be used only if the following conditions are satisfied. For a packet whose header flit is at $(x_0, x_1, \cdots, x_{n-1})$, destined to $(y_0, y_1, \cdots, y_{n-1})$:

- The dimension order routing (DOR) scheme is satisfied; that is, dimension j should be corrected if $x_i = y_i$ for any $i < j$, and $x_j \neq y_j$.
- The request for the output s-VC is granted by the collision detection mechanism.

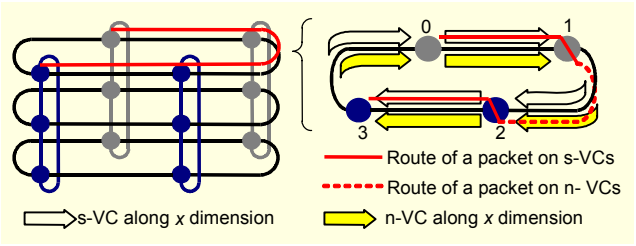

Fig. 1. Collision detection performed on an s-VC ring in 4×3 torus.

Table 1. Number of alternative output VCs supplied by each routing algorithm.

Algorithm	Number of alternative output VCs		
GAL/CQR	If DOR scheme is satisfied		2
	If DOR scheme is not satisfied		1
GALR-CD	If DOR scheme is satisfied	If request for s-VC is granted	3
		If request for s-VC is not granted	2
	If DOR scheme is not satisfied		2

The load-balance scheme introduced in [6] is employed to select one from all the alternative output VCs. The number of alternative output VCs supplied by GALR-CD is listed in Table 1 and compared with that of GAL/CQR, in the case where there is a total of three VCs per physical channel. It is shown that GALR-CD supplies more alternative output VCs.

Many previously proposed collision detection/avoidance schemes, either centralized or distributed, can be used here with very little modification. For example, a centralized scheme could be similar to the token-passing protocol proposed in [7]. But each token should be passed along an s-VC ring, not along a Hamiltonian cycle of the network. For another example, a distributed scheme similar to the CSMA/CD protocol could be implemented with the overhead that all the output ports along an s-VC ring are attached to a shared control channel. Each output port listens to the control channel before transmitting packets, waiting for a random amount of time if collision occurs.

A distributed scheme has better scalability than a centralized scheme, since the later requires a detector for each s-VC ring. Consider the n-D torus with k^n nodes (k nodes along each dimension). The number of s-VC rings (detectors) is nk^{n-1}. In real systems, n is usually much smaller than k, so the cost of detectors is much lower than that of the nodes during scaling.

IV. Deadlock-Freeness

Theorem 1. GALR-CD is deadlock-free.

Proof. In GALR-CD, a packet can always request at least one s-VC as its next hop, so it is not deadlocked if the s-VC does not participate in any deadlock. Therefore it suffices to show that no s-VC can ever participate in any deadlock. We use mathematical induction on the dimension of s-VC.

Basis: We use a proof by contradiction to show that the s-VCs along dimension n-1 cannot participate in deadlock.

Consider an s-VC ring R^s along dimension n-1. Suppose an s-VC of R^s participates in a deadlock, that is, packet P is deadlocked, and a flit of P is on that s-VC. Let $c^s(P)$ denote that s-VC and let $c^s_{next}(P)$ denote the s-VC that is requested by the header flit of P. Then, R^s is reserved to P. Since dimension n-1 is the last dimension, it can be deduced from the DOR scheme that $c^s_{next}(P)$ also belongs to R^s. The request for $c^s_{next}(P)$ will be granted, so P can advance by using $c^s_{next}(P)$ and is not deadlocked. Therefore, the s-VCs along the last dimension cannot participate in any deadlock.

Inductive Step: We prove that if for all $i > t$ the s-VCs along dimension i cannot participate in any deadlock, the s-VCs along dimension t cannot either. Suppose an s-VC along dimension t participates in a deadlock, that is, packet P is deadlocked, and a flit of P is on that s-VC. Suppose $c^s_{next}(P)$ is along dimension i. It can be deduced from the DOR scheme

that $i \geq t$. If $i = t$, the case is the same as *Basis*, so P is not deadlocked. If $i > t$, $c^s_{next}(P)$ cannot participate in any deadlock. Eventually, the header flit of P can advance by using $c^s_{next}(P)$, so P is not deadlocked. Therefore, the s-VCs along dimension t cannot participate in any deadlock.

Consequently, no s-VC can ever participate in any deadlock; therefore, GALR-CD is deadlock-free. □

V. Simulation Results and Analysis

We have made simulations of GAL, CQR, and GALR-CD to compare their throughputs. For each algorithm, five types of traffic patterns are simulated: uniform random (UR), random permutation (RP), transpose (TP), bit-complement (BC), and tornado (TOR). The UR pattern is benign, and the others are adversarial. These patterns have been used in previous works [1], [3]-[6].

For each type of traffic pattern, 40 independent experiments were conducted. The offered/accepted traffic intensity was measured by recording the number of packets injected into/accepted by the network per time unit and then normalized by the network capacity [1]. The results were recorded during steady-state and we averaged the results obtained from 40 experiments. Prior to saturation, the 95% confidence interval width is within 3% of the reported mean values. In cases in which the network is saturated, the 95% confidence interval width occasionally exceeds 6%.

Networks of different scales were simulated, but only the results in 8×8 torus are presented in Fig. 2. The results in other networks follow the same trend. Figure 2(a) presents the accepted traffic of each algorithm on UR and BC patterns. When offered traffic is relatively low, accepted traffic increases with it. After offered traffic reaches the saturation point, the network becomes unstable, and increasing offered traffic reduces accepted traffic. This is typical for fully adaptive routing algorithms [1]. The maximum of accepted traffic represents the throughput. The throughputs of the three algorithms on each traffic pattern are compared in Fig. 2(b). The GAL and CQR algorithms achieve the same throughput on each traffic pattern, as was shown in [6]. GALR-CD outperforms GAL/CQR by 16% on the benign UR pattern, and by 21%, 13%, 10%, and 20% on the four adversarial patterns, respectively.

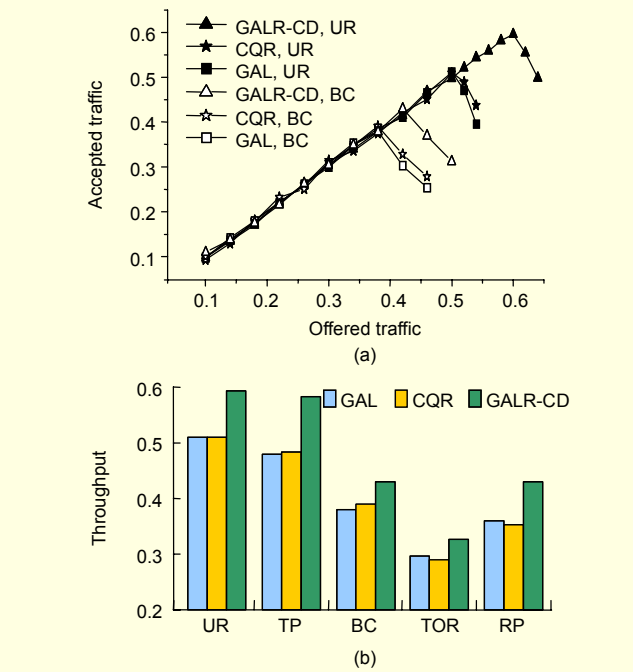

Fig. 2. Throughput of each algorithm on each traffic pattern.

VI. Conclusion

The globally adaptive load-balanced routing algorithm is by far the best solution to achieve high throughput on various traffic patterns in torus networks. However, the performance of previously published algorithms is degraded due to the limited routing adaptability cause by their deadlock-handling scheme. The new GALR-CD algorithm has higher routing adaptability provided by a new scheme based on collision detection to handle deadlock. Simulation results show that GALR-CD outperforms previous algorithms on both benign and adversarial traffic patterns.

References

[1] J. Duato, S. Yalamanchili, and L. Ni, *Interconnection Networks: An Engineering Approach*, revised edition, Morgan Kaufmann, San Francisco, 2002.

[2] W.J. Dally, "Scalable Switching Fabrics for Internet Routers," *Whitepaper*, Avici Systems. http://www.avici.com/technology/whitepapers/

[3] A. Singh, W.J. Dally, and A.K. Gupta, "GOAL: A Load-Balanced Adaptive Routing Algorithm for Torus Networks," *Proc. 30th Annual Int. Symp. Computer Architecture*, 2003, pp. 194-205.

[4] A. Singh, W.J. Dally, and B. Towles, "Globally Adaptive Load-Balanced Routing on Tori," *IEEE Computer Architecture Letters*, vol. 1, 2004, pp. 2-5.

[5] K. Bolding, M.L. Fulgham, and L. Snyder, "The Case for Chaotic Adaptive Routing," *IEEE Trans. Computers*, vol. 12, 1997, pp. 1281-1291.

[6] A. Singh, W.J. Dally, and A.K. Gupta, "Adaptive Channel Queue Routing on *K*-ary *N*-cubes," *Proc. 16th ACM Symp. Parallelism in Algorithms and Architectures*, 2004, pp. 11-19.

[7] K.V. Anjan and T.M. Pinkston, "DISHA: A Deadlock Recovery Scheme for Fully Adaptive Routing," *Proc. 9th Int'l Parallel Processing Symp.*, Apr. 1995, pp. 537-543.