ELSEVIER

Computer Networks 37 (2001) 391–406

COMPUTER NETWORKS

www.elsevier.com/locate/comnet

Comparative performance analysis of directed flow control for real-time SCI

Robert W. Todd, Matthew C. Chidester, Alan D. George *

High-performance Computing and Simulation (HCS) Research Laboratory, Department of Electrical and Computer Engineering, University of Florida, 327 Larsen Hall, P.O. Box 116200, Gainesville, FL 32611-6200, USA

Received 3 March 2000; received in revised form 22 February 2001; accepted 19 April 2001

Responsible Editor: I. Nikolaidis

Abstract

The distributed nature of routing and flow control in a register-insertion ring topology complicates priority enforcement for real-time systems. Two divergent approaches for priority enforcement for ring-based networks are reviewed: a node-oriented scheme called preemptive priority queue and a ring-wide arbitration approach dubbed TRAIN. This paper introduces a hybrid protocol named directed flow control that combines node- and ring-oriented flow control to yield greater performance. A functional comparison of the three protocols as implemented on the scalable coherent interface is presented, followed by performance results obtained through high-fidelity modeling and simulation.

Keywords: Discrete-event simulation; High-performance networks; Real-time protocols; Scalable coherent interface

1. Introduction

The information revolution has created a world in which data is a valuable commodity. In the past, the worth of a network was judged by its ability to deliver data reliably and rapidly. With the emergence of a commercial market for video- and audio-on-demand and an increasing reliance on computer control of mission-critical applications in the military and industrial sectors, networks must also deliver data within a guaranteed maximum response time. As the number of users in the global network grows, the capacity and performance requirements for such real-time networks are rapidly increasing. The combination of a high-bandwidth, low-latency network protocol with real-time capabilities is necessary to meet these needs.

The protocols in this study implement real-time capabilities on a ring-based network. In addition to the cost-effectiveness of ring-based topologies for large-scale networks, chip-multiprocessor architectures frequently include rings as the on-chip interconnect [5]. Though the protocols in this paper are generically applicable to any ring-based network, the scalable coherent interface (SCI) is used as the underlying transport [8]. As IEEE Standard 1596-1992, SCI provides a bus-like interface over unidirectional, point-to-

* Corresponding author. Tel.: +1-352-392-9034; fax: +1-352-392-8671.

E-mail address: george@hcs.ufl.edu (A.D. George).

1389-1286/01/$ - see front matter
PII: S 1 3 8 9 - 1 2 8 6 (0 1) 0 0 2 2 1 - 3

point links to implement distributed shared memory. The protocol calls for link speeds of 1 GB/s and sub-microsecond latencies over distances of tens of meters. However, the base SCI protocol does not attempt to address real-time issues.

Providing real-time capability in a network requires enforcement of packet priorities. Such a priority scheme is required for implementation of algorithms such as rate-monotonic scheduling (RMS) [10]. To support RMS, all packets are assigned a priority and the network must enforce transmission of higher-priority packets ahead of those with lower priorities.

Two protocols were originally proposed to provide real-time capabilities for SCI: preemptive priority queue (PPQ) and TRAIN. Simulation shows neither solution to be optimal in sufficient numbers of test cases to be acceptable for a generalized standard. This paper presents a third solution to the SCI/RT problem called directed flow control (DFC). DFC is a hybrid that combines selected aspects of the PPQ and TRAIN protocols to provide performance greater than either protocol alone.

Traditionally, real-time protocols for ring-based networks have taken a token-passing approach [2,16]. However, such a scheme is unable to capitalize on the concurrent bandwidth that inherently exists in a register-insertion ring network. Other flow-control schemes allow the concurrent bandwidth to be used efficiently but do not adequately address real-time performance needs [11,12]. The real-time protocols presented in this paper provide for both concurrent communication to make effective use of available ring bandwidth and priority-level enforcement for real-time communications.

This paper begins with an overview of the requirements for a real-time extension to SCI. Next, the three protocols are presented from a functional standpoint, illustrating the novel ways in which each protocol addresses priority inversions in a ring-oriented network. The performance of the protocols is then compared using simulation results from several common test cases.

2. Real-time protocol extensions to SCI

While the three protocols presented in this paper were targeted specifically for SCI, each could be easily adapted to provide real-time capabilities to any ring-oriented network. For the purposes of direct comparison, all three protocols were simulated with SCI as the underlying network medium. Therefore, it is important to understand the requirements of a real-time protocol extension to SCI.

2.1. Real-time issues in SCI

In any physical network, it is often impractical to guarantee that high-priority packets will never be preempted by low-priority packets. When a high-priority packet is delayed while a low-priority packet is freely transmitted, a priority inversion is said to occur. Real-time functionality is achieved by avoiding priority inversions wherever contention for a shared resource may occur.

Fig. 1 shows the basic structure of an SCI network interface. Like any register-insertion ring protocol, SCI requires input, bypass, and output queues. The address decode logic determines whether an incoming packet is destined for the local node or for some other node. The packet is then routed to an input queue or the bypass queue, respectively. In order to output a packet, there must be sufficient free space in the bypass queue to

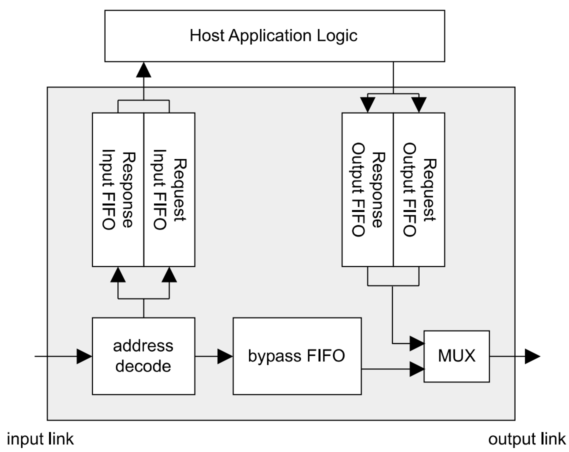

Fig. 1. High-level structure of an SCI node.