Support for Multiple Classes of Traffic in Multicomputer Routers *

Jennifer Rexford and Kang G. Shin
Real-Time Computing Laboratory
Department of Electrical Engineering and Computer Science
The University of Michigan
Ann Arbor, MI 48109-2122

Abstract. Emerging parallel real-time and multimedia applications broaden the range of performance requirements imposed on the interconnection network. This communication typically consists of a mixture of different traffic classes, where guaranteed packets require bounds on latency or throughput while good average performance suffices for the best-effort traffic. This paper investigates how multicomputer routers can capitalize on low-latency routing and switching techniques for best-effort traffic while still supporting guaranteed communication. Through simulation experiments, we show that certain architectural features are best-suited to particular performance requirements. Based on these results, the paper proposes and evaluates a router architecture that tailors low-level routing, switching, and flow-control policies to the unique needs of best-effort and guaranteed traffic. Careful selection of these policies, coupled with fine-grain arbitration between the classes, allows the guaranteed and best-effort packets to share network bandwidth without sacrificing the performance of either class.

1 Introduction

Although multicomputer router design has traditionally emphasized providing low-latency communication, modern parallel applications require additional services from the interconnection network [1, 2]. Multimedia and real-time applications, such as scientific visualization and process control, necessitate control over delay variance and throughput, in addition to low average latency [3]. While guaranteed traffic necessitates deterministic or probabilistic bounds on throughput or end-to-end delay, best-effort service often suffices for the remaining traffic. For example, control or audio/video messages may mandate explicit performance guarantees, while data transfer may tolerate delay variability in exchange for improved average latency.

* The work reported in this paper was supported in part by the National Science Foundation under Grant MIP-9203895 and an Office of Naval Research graduate fellowship. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.
Handling this mixture of disparate traffic classes affects the suitability of architectural features in multicomputer routers. While the router alone cannot satisfy application performance requirements, design decisions should not preclude the system from providing necessary guarantees. Servicing guaranteed traffic requires control over network access time and bandwidth allocation, so the router should bound the influence best-effort packets have on these parameters. The software, or even the hardware, can then utilize these bounds to satisfy quality-of-service requirements through packet scheduling and resource allocation for communicating tasks. Additionally, the design should not unduly penalize the performance of best-effort packets.

Modern parallel routers significantly reduce average latency by avoiding unnecessary packet delay at intermediate nodes; however, these low-latency techniques often impinge on control over packet scheduling. In particular, cut-through switching [4, 5] decentralizes bandwidth allocation and packet scheduling by allowing an incoming packet to proceed directly to the next node in its route if a suitable outgoing link is available. Other multicomputer router features, such as FIFO queueing and adaptive routing, further complicate the effort to provide predictable or guaranteed service.

Research in networking considers techniques for the effective mixing of multiple traffic classes in a communication fabric [6, 7]. However, the design trade-offs for parallel machines differ significantly from those in a heterogeneous, distributed environment. In parallel machines, router design trade-offs reflect the large network size and the tight coupling between nodes. Speed and area constraints motivate single-chip solutions, including designs that integrate the processing core and the communication subsystem [8–11]. Regular topologies facilitate efficient offset-based routing, avoiding the costs of implementing and maintaining a table-driven scheme at each node.

While these implementation constraints restrict some router design options, the tighter coupling between nodes enables multicomputer routers to consider more diverse routing and switching techniques for handling different traffic classes. The shorter, wider communication links in most parallel machines result in much lower packet transmission delays, compared to distributed systems. These low-latency channels broaden the spectrum of flow-control schemes that can be implemented efficiently. The fine-grain interaction between and within the nodes necessitates effective mapping of application tasks onto the interconnection network [12]. Effective router techniques for handling multiple traffic classes should provide useful software abstractions to parallel applications.

The remainder of this paper is organized as follows. Section 2 presents a simulation model for studying the impact of routing and switching on interconnection network performance. Using this model, Section 3 investigates how switching schemes affect the network's ability to service multiple traffic classes. Based on these results, Section 4 proposes and evaluates a router architecture that allows best-effort packets to capitalize on low-latency routing and switching techniques without compromising the performance of guaranteed traffic.

This architecture uses virtual channels [13] to logically partition the inter-