Fair Queueing Architectures for High-Speed Networks

Albert G. Greenberg
Network Services Research Laboratory
AT&T Bell Laboratories

Abstract
High-speed links in emerging communication networks require simple and efficient algorithms for apportioning bandwidth to competing connections. A single link may carry traffic for thousands of connections with different quality-of-service parameters, requiring scalable switch architectures that avoid performing operations for every connection or cell during each transmission slot. Still, effective cell scheduling algorithms should provide delay and bandwidth guarantees for a wide variety of applications, even in the presence of bursty traffic. With Jennifer Rexford and Flavio Bonomi, I have investigated a collection of self-clocked fair queueing (SCFQ) architectures amenable to efficient hardware implementation in network switches. These event-driven architectures perform a small, bounded amount of work in response to each cell arrival and departure, independent of the number of connections. Exact and approximate implementations of SCFQ efficiently handle a moderate range of connection bandwidth parameters, while hierarchical arbitration schemes scale to a large range of throughput requirements. Simulation experiments demonstrate that these architectures divide link bandwidth fairly on a small time scale, preserving connection bandwidth and burstiness properties.