Network Congestion Avoidance Through Speculative Reservation

Nan Jiang, Daniel U. Becker, George Michelogiannakis, William J. Dally

Concurrent VLSI Architecture Group, Stanford University

Background

As a shared communication medium, the presence of congestion in a network has a global impact on system performance.

With sufficient bisection bandwidth and adaptive routing, network congestion occurs almost entirely at the edge of the network due to hot-spot destinations.

In lossless networks, the congested traffic remains in the network until it is delivered. As a result, congested traffic backs up into the rest of the network in a condition called tree saturation.

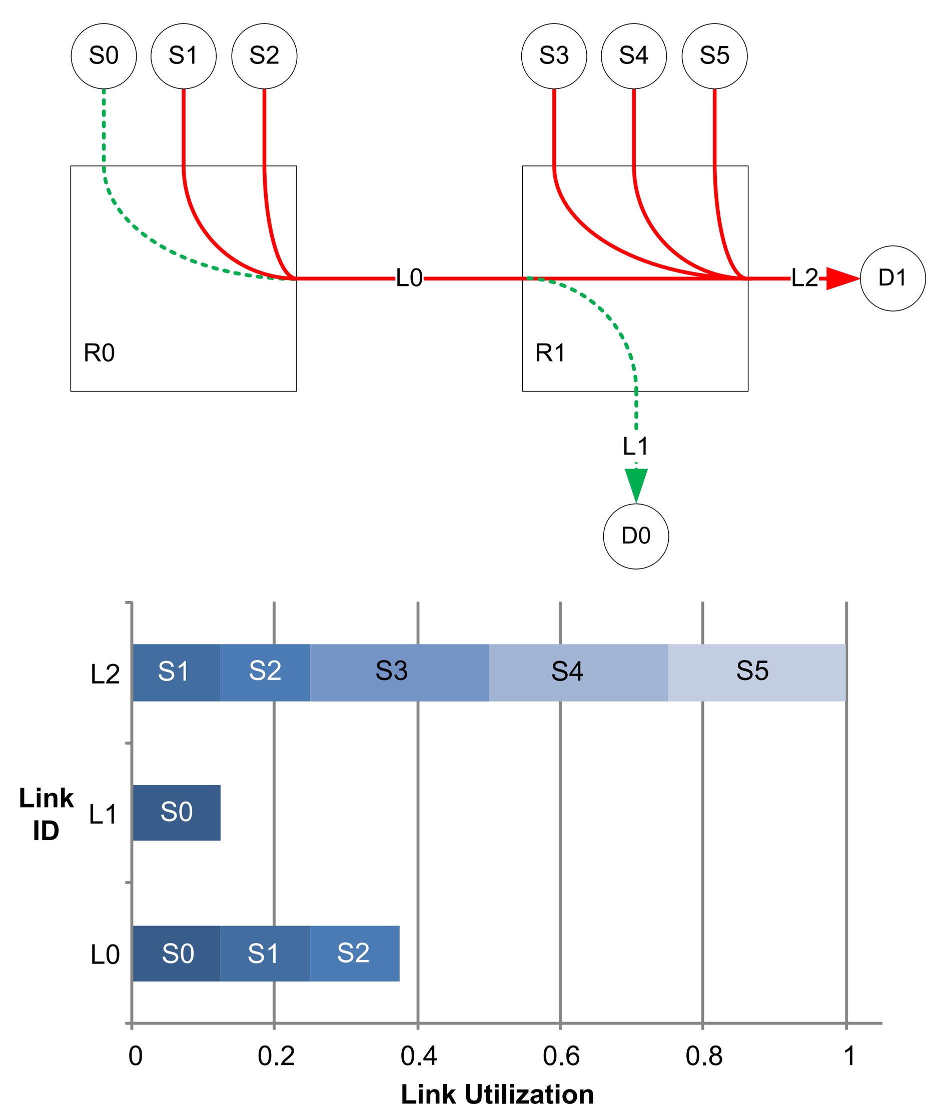

Figure 1 – Tree saturation caused by the hot-spot traffic to D1 prevents S0 from utilizing available bandwidth on the L0 and L1 channels.

Existing Congestion Solution

Explicit congestion notification (ECN) is used in many networking systems. The method is reactive, triggering only after congestion has already formed.

ECN detects congestion through buffer occupancy levels and marks packets traveling through the bottleneck. Marked packets then generate notifications to signal traffic senders to reduce their injection rate.

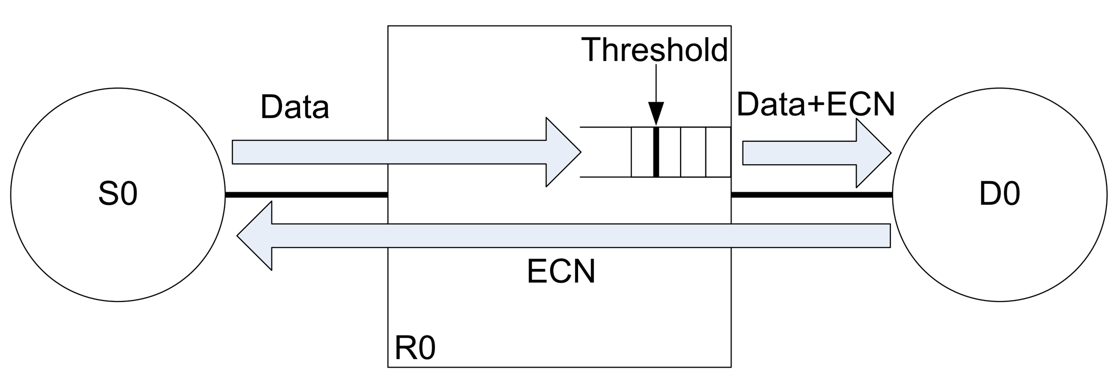

Figure 2 – Example of ECN in operation..

Speculative Reservation

The Speculative Reservation Protocol (SRP) operates on the principle of congestion avoidance. SRP requires a reservation-grant handshake between communicating nodes to avoid over-loading any network destination.

To reduce the latency overhead associated with the reservation handshake, the sender can begin transmitting packets speculatively before the reservation grant returns. These speculative packets have a limited time-to-live inside the network and can be dropped—and later retransmitted—when congestion begins to form.

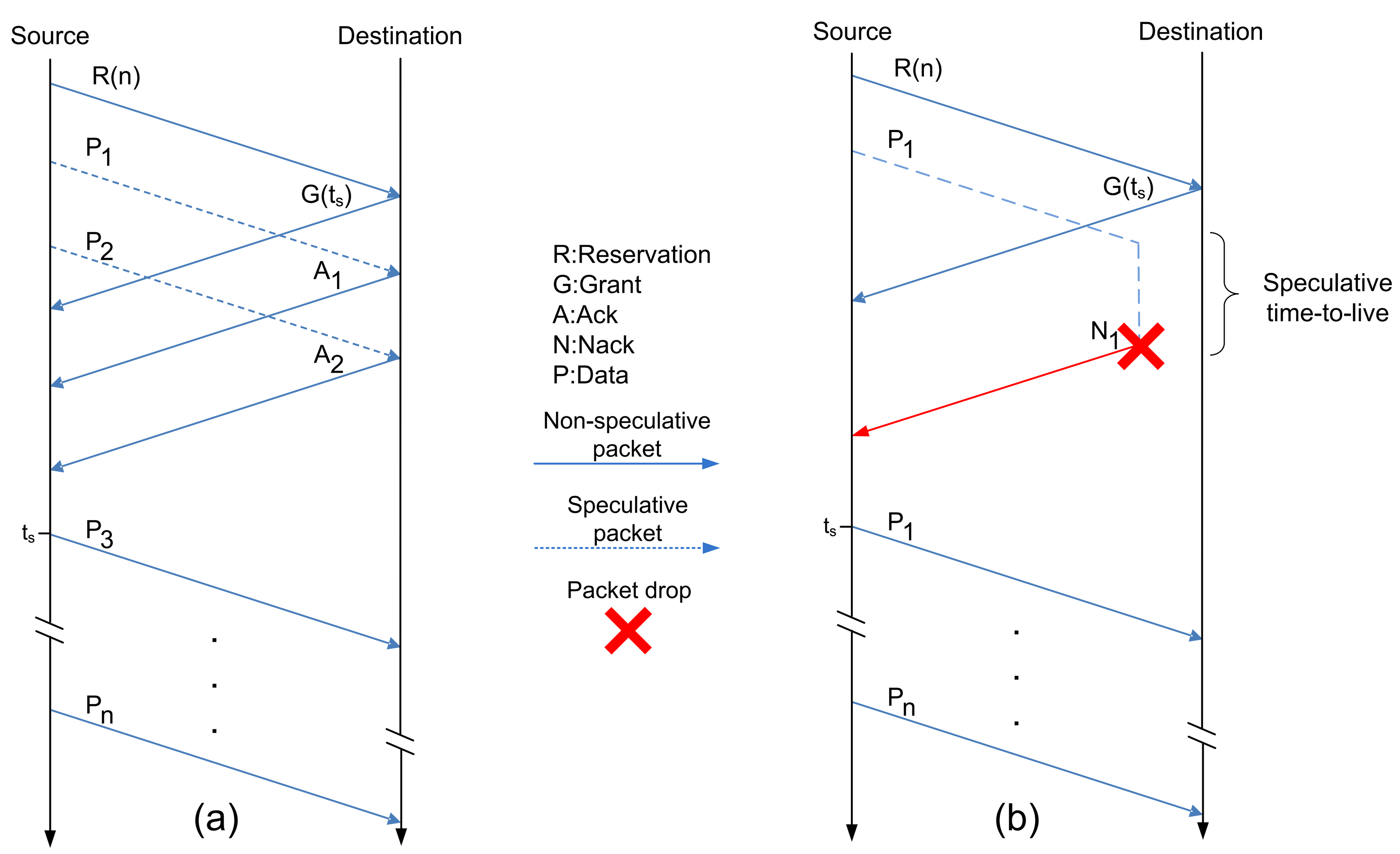

Figure 3 – Time diagrams of SRP scenarios: (a) hot-spot free network (b) network hot-spot at node D.

Improved Hot-spot Transient Response

A key advantage of SRP is its fast response to the onset of congestion. As soon as the hot-spot traffic is initiated, it becomes regulated by the reservation protocol, and any congestion in the network is completely avoided.

In contrast, packets sent during the ECN activation period are destined to create congestion in the network, causing large latency/throughput spikes.

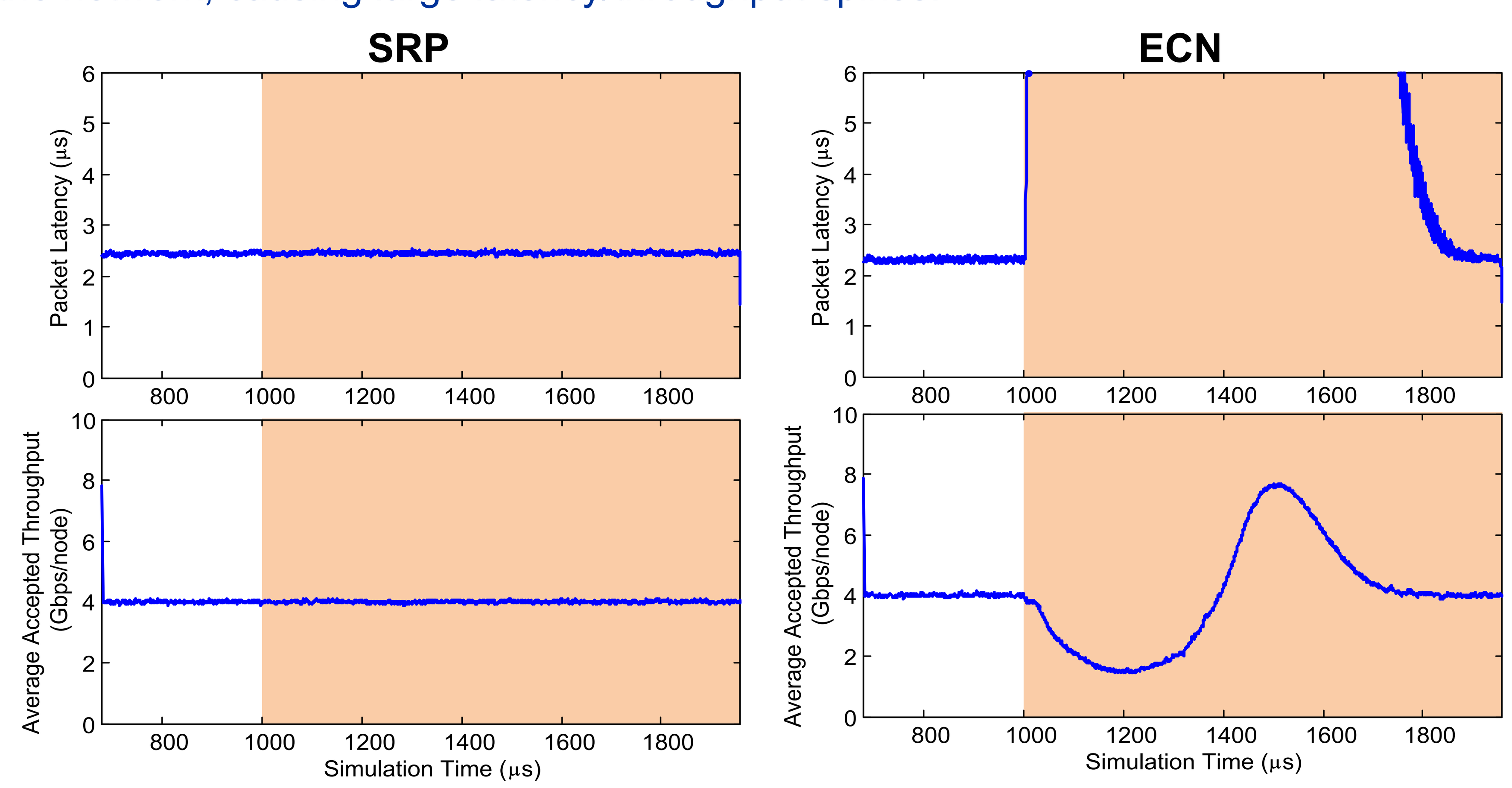

Figure 4 – Transient hot-spot response of 256-node networks. Both are warmed up with uniform random traffic for 1ms before the addition of a 40:1 hot-spot traffic pattern. 2KB message size.

Improved Bursty Traffic Throughput

Large message size increases traffic burstiness in a network. When multiple messages converge on a destination, a temporary hot-spot is formed. By segmenting traffic bursts into multiple reservations, SRP improves network throughput. SRP also outperforms ECN due to its rapid transient response.

SRP's reservation schedule is well-behaved when interacting with multiple message sizes.

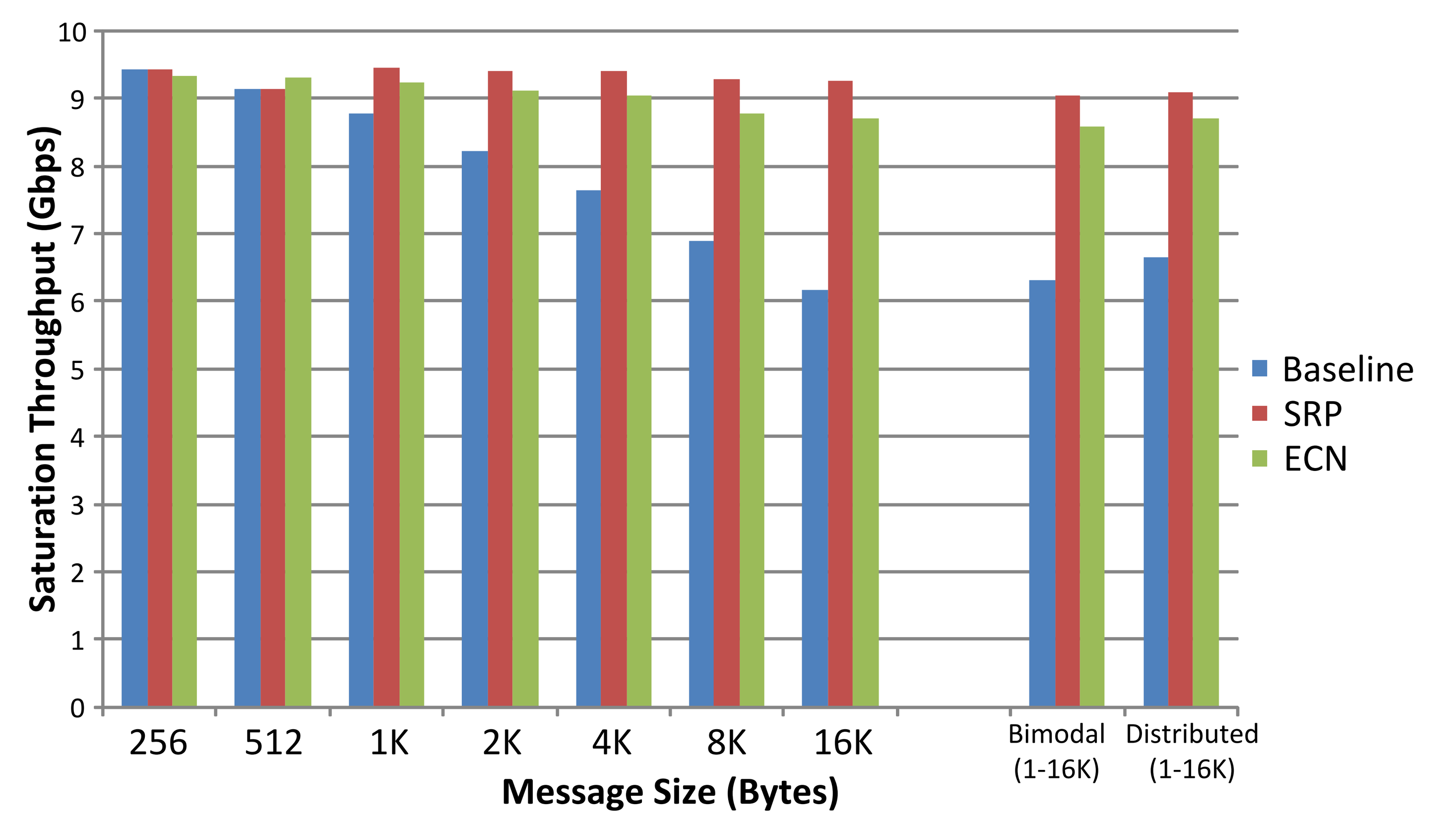

Figure 5 – Saturation throughput of 256-node networks running uniform random traffic with different message sizes. SRP maintains high network throughput for both benign and bursty traffic.

Low Protocol Overhead

Latency overhead created by the reservation is masked by speculative packet transmissions. Observable latency overhead only occurs at very high load.

Dropped speculative packets account for the bulk of bandwidth overhead. The drop rate is a function of network latency, speculative time-to-live, and reservation size.

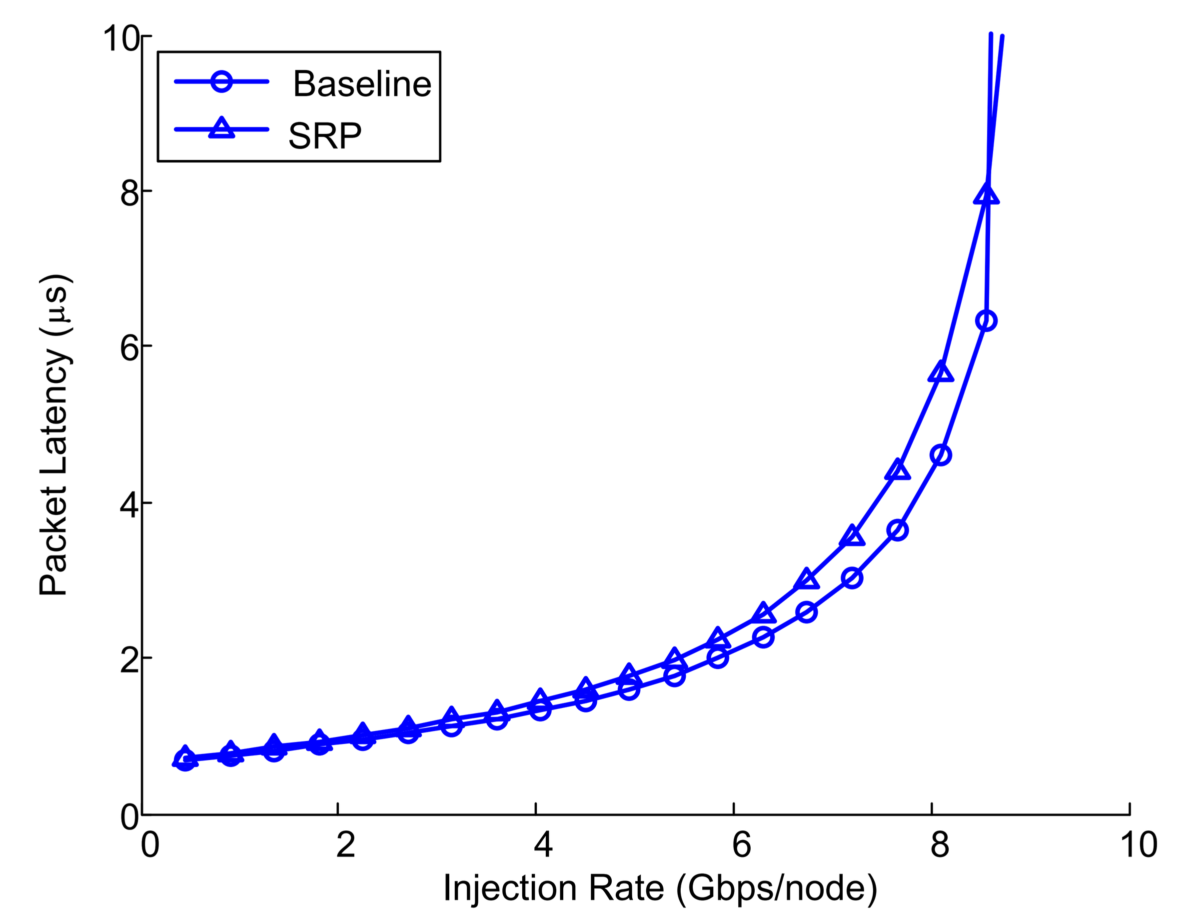

Figure 6 – 256-node SRP network running uniform random traffic with 1KB messages.