Congestion Control as a Building Block for QoS

David Harrison, Shivkumar Kalyanaraman and Sthanunathan Ramakrishnan[1]
Dept of ECSE & Dept of CS, Rensselaer Polytechnic Institute, Troy, NY 12180
Email: {*harrisod,shivkuma*}@networks.ecse.rpi.edu
URL: http://www.ecse.rpi.edu/Homepages/shivkuma/

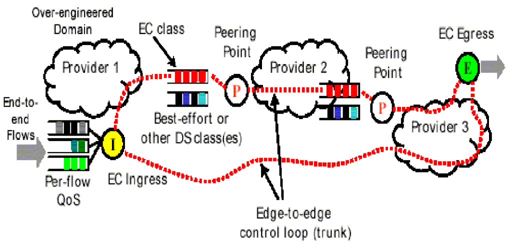


Figure 1: Example Model

In this work, we focus on the question: How to provide QoS for private networks in an inter-domain multi-provider environment. Traditional QoS mechanisms like diff-serv, CSFQ require deployment of the QoS mechanisms at every potential bottleneck. Scalable performance as well as complex coordination and deployment issues have inhibited the large scale deployment of QoS mechanisms. Our solution is to use a new overlay congestion control algorithm (*Riviera*) to push back all congestion to the edge (e.g access router) and hence called an edge-to-edge control algorithm (see Figure 1). Then QoS issues like queueing, bandwidth sharing, packet loss distribution are all pushed to the edge allowing the core to operate using largely stateless mechanisms. Use of congestion control as a QoS building block has two implications:

1. The differentiation between flows occurs in RTT timescales unlike in schedulers where differential service is provided on a packet-by-packet basis.

2. The queues at the interior bottlenecks are all consolidated and distributed at the edges. These smaller queues at the edge are much easier to handle and buffer management schemes may be used to control them more effectively (as opposed to directly controlling the queue at the interior bottleneck).

For such an architecture to be realized we need the algorithm to *transparently* detect congestion, have *near* lossless behavior, not involve upgrades of interior routers and have minimum configuration requirements. Our Riviera algorithm is rate-based since it is easier to map services in a rate-based architecture, and the control traffic is much lesser than window-based schemes.

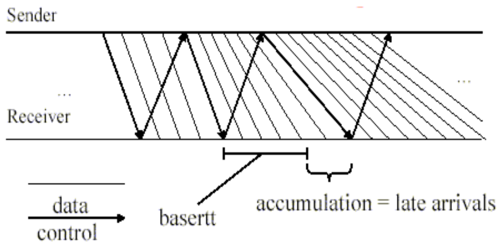
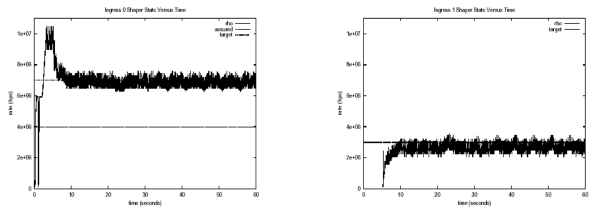


Figure 2: Accumulation Measurement

We realize such an algorithm using the concept of accumulation. Accumulation is the number of packets backlogged in the network and measured as shown in Figure 2. We use a thresholding scheme to detect congestion epochs. When not in congestion we increase additively over the output rate and decrease multiplicatively from the output rate otherwise. This policy causes the queue to increase slowly and to decrease quickly (within an RTT in a single bottleneck case) while maintaining high utilization. We have proved the stability of this scheme in a multi-bottleneck scenario and have also shown that the scheme achieves proportional fairness.



(a) Trunk 1 w/ 4Mbps Assured + 3Mbps Best-Effort (b) Trunk 2 w/ 3Mbps Best-Effort

Figure 3: Assured Service with Closed Loop Blocks

We have implemented basic services like "Better than best effort service", and "Isolating misbehaving flows" and advanced services like assured services (Eg: Figure 3) and quasi-leased line (QLL) by modifying the increase and decrease functions. Advanced services need policing at edges. In the case of a QLL we need admission control to prevent over-subscription. The signaling/configuration complexity of the basic scheme (without the advanced services) is minimal. ISPs just need to provide a single class for all edge-to-edge flows.

Our scheme supports bounded scalability, incremental deployment, simplified interior configuration and a simple set of overlay services. Future work includes understanding the dynamic range of services possible using such building blocks, design of control plane and management plane functions for these services etc.

[1]This work was funded in part by NSF ANIR grants: ANI9806660, ANI9819112 and a grant from Intel Corp.