IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Hershey PA 17033-1117, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITB7354

Chapter VII

Multilayered Quality of Service

Luiz A. DaSilva
Virginia Polytechnic Institute and State University, USA

Today's networks support applications that deliver text, audio, images and video, often in real time and with a high degree of interactivity, using a common infrastructure. More often than not, traffic is carried over packet-switched networks that treat all data the same, under what is known as best-effort service. Packet switching can achieve very high efficiency through statistical multiplexing of data from numerous sources; however, due to the very nature of packet switching, one should expect fluctuations in throughput, delay, reliability, etc., for any given flow. The greater the statistical multiplexing capabilities, the greater the efficiency and also the greater the variability of achieved performance; in this sense, best-effort service provides maximum efficiency with highly unpredictable service quality.

Clearly, not all traffic flows are created equal. Interactive web-based applications tend to be very sensitive to throughput, while real-time voice and video are sensitive to delay and jitter, and traditional data applications such as e-mail and file transfers are fairly insensitive to fluctuations in performance. The concept of quality of service (QoS) has evolved from the realization that in networks that carry heterogeneous traffic it makes sense to treat specific classes of traffic according to their specific needs.

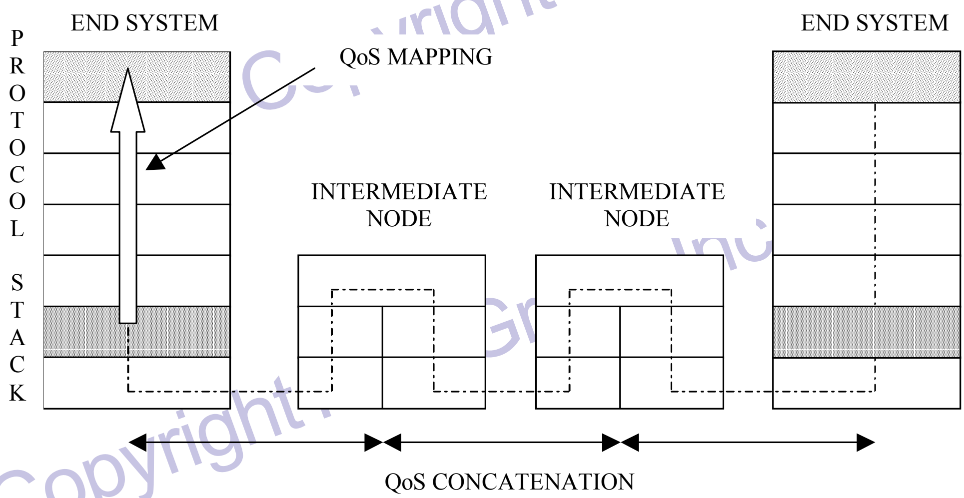
The overwhelming success of the Internet has been instrumental in modifying users' expectations as to what kind of service quality to expect and demand from a packet-switched network. We want our IP telephony applications to sound like the traditional telephone system; we want to play interactive games over the Internet; we want to be able to teleconference with colleagues around the world seamlessly, as if we were in the same room. These and other applications will dictate that tomorrow's networks, from local area networks (LANs) to corporate virtual private networks (VPNs) to the Internet, support differentiated QoS.

Recent advances have given rise to a multilayered structure for QoS architectures, as illustrated in Figure 1. Not only switching can occur in multiple layers, but QoS guarantees may also exist in multiple layers, leading to the need for mapping among these guarantees and, possibly, concatenating them across multiple subnetworks. For instance, it is possible that a single flow may be subject to distinct QoS controls by IP, Ethernet and ATM when traveling from source to destination.

This chapter will introduce the topic and present the main mechanisms that are necessary to support QoS. It will also briefly review QoS guarantees implemented in ATM networks, the standardization efforts currently under way to create a QoS architecture for the Internet Protocol (IP), including Multiprotocol Label Switching, and the issue of QoS in local area networks.

By the end of the chapter, the reader should be able to describe the motivation for QoS architectures and some of the main trade-offs among the various architectures proposed to date. Numerous mechanisms are needed to support QoS, including policies, service class definitions, admission control, traffic policing, service level agreements, appropriate pricing structures, etc.; this chapter provides an appreciation of how these mechanisms come together to form a QoS architecture. We also discuss recent developments in bringing QoS to packet-switched networks, from ATM to IP to LANs, and enumerate the principal challenges facing the widespread deployment of QoS in commercial networks. Finally, we hope to enable the reader to form a vision regarding the role of quality of service in computer networks in the near future, including emerging issues such as the development of 3G wireless networks, coping with ad hoc networks, QoS pricing schemes, etc.

*Figure 1: Layered view of communication between end nodes. QoS guarantees at the lower layer in the protocol stack must be translated into parameters that are meaningful at the higher layers. Note that the mapping may also occur in the reverse direction, i.e., QoS objectives dictated by a higher layer may have to be translated into parameters that a lower layer can understand and control (DaSilva, 2000*a).



Related Content