Guest Editorial: Quality of Service in Multimedia Networks

1 Multimedia Applications and QoS

The growth of multimedia applications over wide area networks has increased research interest in quality of service (QoS). Readers are referred to [6] for an overview of providing QoS over the Internet. The communication delay and synchronization needed for voice, data and images are major concerns. Internet telephony (Voice over IP) and other multimedia applications such as video conferencing, video-on-demand, and media streaming require service guarantees and have strict timing requirements. The size and quality of display devices, and resources such as CPU, battery power, and bandwidth are always limited.

Just increasing the amount of resources such as available bandwidth to avoid congestion does not provide proper resource utilization. This over-provisioning solution has been effective in applications such as FTP, HTTP, and email. However, the nature of traffic over the Internet has changed in its characteristics. There are many new types of applications, each with very different operational requirements. The Internet is becoming the backbone of future communications in an entertainment center. Mobile and ad-hoc networks, and hand-held devices such as PDA have different QoS requirements. Multimedia applications require network services beyond what IP delivers.

2 What is QoS?

QoS can be parameterized as throughput, delay, delay variation (jitter), loss and error rates, security guarantees, etc, that are acceptable in an application. As such, QoS is a necessarily application specific. For instance, jitter is an important quality for IP telephony, which can tolerate a certain percentage of packet loss without any degradation of quality. For data transfer, loss is a crucial QoS parameter. Quality of service control requires an understanding of the quantitative parameters at the application, system, and network layers.

2.1 QoS Layering

QoS parameters can be partitioned into two subsets, namely application-dependent parameters and application-independent parameters. They can also be put into three layers: user, application, and system layers as shown in Table 1. The system parameters can be further classified into two categories: network and operating system parameters and device parameters.

For multi-media presentation, the quality of audio and video is important in addition to images, text and numbers. Application parameters describe requirements for application services and are specified in terms of media quality and media relations. Media quality includes source/destination characteristics such as media data unit rate, and transmission characteristics such as response time. Media relations specify relationships among media, such as media conversation, inter-stream synchronization, and intra-stream synchronization. Some of these parameters at a high level can be included in general parameters defined as accuracy, precision, and timeliness. Timeliness, Accuracy, Precision (TAP) can together form a good criterion for QoS. Timeliness is defined as “when an event is to occur”. Maintaining it means meeting a deadline. Accuracy is defined as “the degree to which the output conforms to the semantics and contexts of the applications”. Maintaining it means guaranteeing the correctness of the data. Precision is defined as “the quantity of information provided or processed”. Maintaining it means maintaining the amount of data being processed or transmitted over the network.

System parameters describe communication and operating system requirements that are needed by application QoS. These parameters are specified in quantitative and qualitative terms. Quantitative criteria are those that can be evaluated in terms of concrete measures, such as bits per second, number of errors, task processing time, and data unit size. Qualitative criteria specify expected services, such as inter-stream synchronization, ordered delivery of data, error recovery mechanisms, and scheduling/caching mechanisms.

Network parameters are specified in terms of network load and network performance. Network load refers to ongoing traffic requirements such as packet inter-arrival time. Network performance describes the requirements that must be guaranteed, such as bandwidth, end-to-end delay, and jitter. The network services depend on a traffic model (arrival of connection requests) and perform according to traffic parameters such as peak data rate or burst length. Hence, calculated traffic parameters are dependent on network parameters and are specified in a traffic contract. Device parameters typically specify timing and throughput demands for media data units.

QoS Layer	QoS Parameters
Application	Frame Rate, Frame Size and Resolution, Response Time Throughput, Security, Price and Convenience
System	Buffer Size, Process Priority , Scheduling Policy Caching Policy, Time Quantum
Network	Bandwidth, Throughput, Bit Error Rate, End-to-End Delay Delay Jitter , Peak Duration

Table 1: Quality of Service Layers

2.2 Technical Issues in Multimedia Networks

Network transmission is liable to errors and data loss. Excessive loss in video transmission affect the performance of video quality. Video quality can be enhanced by proper frame synchronization between the video and audio streams. The transmitting a movie over low bandwidth channels can be enhanced by separation of audio and video streams and routed them independently to offload certain channels. Adaptive and dynamic resource management, and reliability issues need to be addressed for continuous media streaming. Emerging continuous media applications have well defined QoS constraints. These applications have stringent resource requirements and can benefit from non-interference to provide forms of progress guarantees. Video stream places high demands for QoS, performance, and reliability on storage servers and communication networks. To support video conferencing, network and application level multicast is a powerful technique. End-to-end congestion control is an important issue for reliable and unreliable multicast transport protocols. Like wired network, resource reservation and rate adaptation can be used to support multimedia services to provide quality of service in wireless networks.

Multimedia networks support real-time data, bulk data, and statistically multiplexed data, which make the traffic management in the network hard. The necessary traffic management components to support QoS are:

- Admission control: The admission control component takes into account resource reservation requests and the available capacity to determine whether to accept a new request with its QoS requirements.
- Scheduling: The scheduling component provides QoS by allocating resources depending on the service requirements. This requires mapping the user-defined QoS requirement to resource allocations for providing the service.
- Resource management: QoS can be provided using over-provisioning of a network, which increases the cost incurred by the provider. Efficient resource management is a cost-effective solution for the provider and it ensures that applications will get the specified QoS during the course of its execution.
- Congestion control: Congestion control is required to avoid anything bad from happening inside a network domain. Some applications may not follow the standard protocol description and try to steal resources, thereby deteriorating the QoS of other applications. Mechanisms are needed to recover from congestion and control flows accordingly.
- Policing/Shaping: Users might send traffic at a rate higher than the agreement. Policing is necessary to monitor these situations, and shaping makes the traffic smooth and reduces its variations over time.

2.3 Network support for QoS

The Internet Engineering Task Force (IETF) has proposed many service models and mechanisms to meet the demand for QoS. Notably among them are the Integrated Services (IntServ) or RSVP models [2], the Differentiated Services (DiffServ) model [1], and MPLS [5]. The IntServ model is characterized by resource reservation. For real-time applications, the applications must first set up paths and reserve resources before data is transmitted. RSVP is a signaling protocol for setting up paths and reserving resources. In DiffServ, packets are marked differently to create several classes. Packets in different classes receive different services. MPLS is a forwarding scheme. Packets are assigned labels at the ingress of an MPLS-capable domain. Subsequent classification, forwarding, and services for the packets are based on those labels. Network and application level multicasting is proposed to support applications that provide services to a large number of receivers such as in video broadcasting, conferencing, and streaming.

2.4 Evaluation

The evaluation of a system that provides QoS is application specific because the quality is application dependent.The monitoring of bandwidth gain, delay, loss or jitter requirements of applications can be studied using simulations or experiments based on implementation in a real deployment environment. QoS parameters should be used to evaluate the quality demanded and achieved by an application.

3 Further Research

Security, survivability, and manageability are important operational issues to deliver multimedia traffic with a wide range of QoS requirements. Existing management schemes suffer from lack of adaptability. Research needs to be done towards designing dynamic and adaptable schemes to account for network dynamics such as change in link status, nodes failure/recovery in addition to the type, duration, timing, extent, and severity of security attacks. QoS-enabled network introduces new threats and vulnerabilities. Attackers can impersonate a legitimate customer by spoofing flow identity (IP addresses, protocol and port numbers). Network filtering at routers in the customer network can detect such spoofing if the attacker and the impersonated customer are on different subnets, but the attacks proceed unnoticed otherwise. Attackers can inject traffic with their own identity and a desired destination. If a customer has a geographically distributed network with multiple entry points into a domain, traffic may be injected into the network such that the extra traffic goes unnoticed at this point, but the aggregate traffic from several different entry points can cause denial of service (DoS) attacks to a downstream domain. This can cause other flows to experience QoS degradation. Readers are referred to [3] for more security issues in QoS networks.

4 This Special Issue

This special issue of Multimedia Tools and Applications brings several important issues to address the QoS problem. The papers cover resource management, admission control, and scheduling. They cover sender-side QoS issues, receiver-side QoS issues, the network behavior, and the transmission of multimedia data. A brief introduction to each paper selected for this issue is given below:

The paper by Aly and Youssef provides mechanism to remedy video loss and to restore synchronization between video and audio streams via quick estimated reconstruction of lost video frames, and their injection at appropriate locations without the need for retransmission or any extra data. Several estimation techniques are considered, namely motion tracking, quadratic interpolation, linear interpolation, two-way duplication, and one-way duplication. The lost frames are estimated using received frames only, and without the existence of any further data. Furthermore, a hybrid estimation system is developed to make best use of the advantages of each technique and to avoid their shortcomings. A decision system is developed to determine the best technique to be used in the estimation of missing frames. The system relies on the amount of motion taking place, instead of the missing frames and the amount of frames lost to come up with the decision.

The paper by Thapliyal, Sidhartha, Li and Kalyanaraman discusses multicast transmission at the network-layer and application-layer, which can be leveraged by future multimedia applications including video-conferencing, multi-player games and distance education. The Internet Engineering Task Force (IETF) has mandated in RFC 2357 [4] that multicast transport protocols in standards track must support viable congestion control algorithms. This is important to ensure the stability of the Internet, and good performance for legacy TCP/IP applications. This paper presents a lightweight, easily deployable single-rate congestion control algorithm. A key feature of this algorithm is that it is purely source-based, and therefore can be quickly implemented in future multimedia applications by a simple upgrade of the multicast source(s). The scheme consists of a simple cascade of filters which operate on loss indications from receivers, and its performance has been shown to be comparable to other heavyweight schemes.

The paper by Chou, Golubchik, Lui, and Chung describes various adaptive and dynamic resource management techniques in conjunction with data placement techniques, for designing a distributed continuous multimedia server. Using these proposed techniques, the authors illustrate how one can improve performance, scalability and reliability of multimedia servers.

The paper by Lu, Lee, and Bhargavan proposes an adaptive quality of service (QoS) design approach via resource reservation and rate adaptation to support multimedia over wireless cellular networks. The paper proposes an adaptive QoS model that seeks to provide QoS assurance within bounds for each packet flow and to make advance resource reservation in order to support seamless mobility. To realize the proposed QoS model, the authors describe a revenue-based resource adaptation design that seeks to maximize network revenue while satisfying the QoS requirements, and a resource reservation protocol that supports both application-initiated resource reservation and network-initiated resource adaptation. The initial implementation and measurements confirm the effectiveness of the proposed design.

The paper by Yau and Bhargava describes an approach to provide differentiated CPU services that are required by diverse application characteristics without necessarily using many different scheduling algorithms customarily assumed in the literature. A fair rate-controlled algorithm uniformly schedules all application classes, and individual classes differ by the amount of over booking of resources that are allowed. This provides a useful spectrum of degrees of resource sharing versus isolation. The experimental results, using real application workloads, demonstrate the versatility of this uniform approach, including highly effective support for soft real-time multimedia applications.

References

[1] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An architecture for Differentiated Services. RFC 2475, December 1998.

[2] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin. Resource ReSerVation Protocol (RSVP). RFC 2205, Sept 1997.

[3] A. Habib and B. Bhargava. On designing scalable edge routers to secure QoS network domains. Technical Report, Department of Computer Sciences, Purdue University, 2002.

[4] A. Mankin and et al. IETF criteria for evaluating reliable multicast transport and application protocols. IETF RFC 2357, June 1998.

[5] E. Rosen, A. Viswanathan, and R. Callon. Multiprotocol label switching architecture. IETF RFC 3031, Jan 2001.

[6] X. Xiao and L. Ni. Internet QoS: A big picture. *IEEE Network*, 13, 2:8–19, March/April 1999.

Acknowledgment

This work was supported by NSF grant CCR-9901712 and CERIAS security center. I thank Ahsan Habib for providing materials for this editorial.

Bharat Bhargava
CERIAS and Department of Computer Sciences
Purdue University, West Lafayette, USA bbcs.purdue.edu