# Quality of Service in Multiservice Networks for Digital Economy

**R. Jayanthan** (*jayan@iee.org*), **Kithsiri Gunasakera** (*kithsiri@bccomputersltd.com*)
*BC Software Systems (Pvt) Ltd.*

***Abstract***

The role of a network have shifted from being a facilitator in business to a media over which business is carried out. The digital communication revolution has enabled the networks; be it the public Internet or private corporate network to carry a growing mix of traffic. Assurance of Quality of Service (QoS) in multiservice network is the key element in the success of Digital Economy. This paper discuss the current developments and standards established to provide QoS in local area networks and wide area networks giving special focus for telephony applications.

## 1. Introduction

Quality of Service (QoS) is an ambiguous term with multiple interpretations, many of which depend on the uniqueness of the problems facing the networking community. It may mean different thing to different people and at different scenarios. To examine the concept of QoS, first examine the two operative words: quality and service. Both words can be equally ambiguous.

Quality can encompass many properties in networking, but people generally use quality to describe the process of delivering data in a reliable manner or even somehow in a manner better than normal. This method include the aspect of data loss, minimal (or no) induced delay or latency, consistence delay characteristics (also know as *jitter*), and the capability to determine the most efficient use of network resources (such as the shortest distance between two end points or maximum efficiency of circuit bandwidth). Quality also can mean a distinctive trait or distinguishing property, so people also use quality to define particular characteristics of specific networking application or protocols.

The term service also introduces ambiguity; depending on how an organization or business is structured, service may have several meaning. People generally use service to describe something offered to the end-users of any network, such as end-to-end communications or client-server applications. Service can cover a broad range of offerings, from electronic mail to desktop video, from web browsing to chat rooms. In multiprotocol network, service also can have several other definitions. In a Novell Netware network, each SAP (Service Advertising Protocol) advertisement is considered an individual service. In other cases, service may be categorized according to the various protocol suits, such as SNA, DECnet, Apple Talk, and so forth. In this fashion, you can bring a finer level of granularity to the process of classifying service.

Another side of QoS is the service guarantees to their subscribers, most of which are contractual. Network availability, for example, is one of the more prevalent and traditional measurements for a  Service Level Agreement (SLA) between a provider and a subscriber.   Here access to the network is the basic service. If the network is inaccessible, the quality aspect of the service is clearly questionable.

## 2. Elements of QoS

QoS refers to the performance of data packet, frame or cell flow through one or more networks. The aim is to deliver end-to-end QoS to user traffic. As mentioned above QoS is characterized by a small set of metrics, including service availability, delay, delay variation (*jitter*) throughput and packet loss rate. A brief description of each of these measurable parameters follows.

***Availability:*** **-** the reliability of the user's connection to the network and services.

***Delay:*** - also known as latency, refers to the interval between transmitting and receiving data (packets/frames/cells) between two reference points. Four types network delays can contribute to the overall delay.

- ***Propagation Delay:*** An inherent delay associated with signals travelling on any physical medium. In the case of fibre optics, propagation delay is some what more than the speed of light delay (the theoretical minimum).
- ***Link speed delay:*** Data transfer rate is determined by the bit rate of the link. A fast link will obviously transfer data much faster than a slower link, so the slower link introduces a relative delay. Link speed delay is independent of propagation delay and is by far the grater of the two components.
- ***Queuing delay:*** Every switch and router employs queues, where packets can be stored until capacity is available to transfer them out to the link. Time spent in queues constitutes queuing delay, which accumulates with each device traversed.
- ***Hop Count:*** Each switch or router traversed by a packet is considered a hop. Queuing delay grows as hop count increases, so hop count is an important metric to control.

***Delay variation:*** - also called jitter, refers to the variation in time duration between all packets in a stream taking the same route.

***Throughput:*** - the rate at which the packets are transmitted in a network; can be expressed as an average or peak rate.

***Packet loss rate:*** - the maximum rate at which packets can be discarded during transfer through a network; packet loss typically results from congestion.

## 3. Quality of Service and Class of Service

Rather than just sending all traffic in the order in which it arrived at a switch or router, QoS requires that we send some of the traffic immediately, and cause other traffic to wait (or in severe congestion, drop some packets). This function is called scheduling.

There are many different scheduling techniques used in network devices to implement QoS. Well-known mechanisms include prioritization, weighted fair queuing (WFQ), weighted random early detection (WRED), bandwidth reservation, traffic shaping, flow control etc. The major challenge for QoS is not how to implement scheduling techniques, but determining what traffic to apply. The capability to differentiate between traffic or service type, so that the devices treat one or more class of traffic differently than other types requires defining of Class of Service.

Many class definitions are possible. The traffic can be classified by the type application or device. For example, different classes could be defined for voice traffic, video traffic, mission-critical transaction processing, large file transfers, web traffic \, email, etc.

Alternatively, traffic class could be defined by type of user. Its not hard to envision that in some enterprises senior executives will expect special treatment of their traffic. Or network administrators themselves might expect to belong to a supper class that has the highest priority; network managers may have a critical need to access network devices in the event of severe congestion or network failure. Some enterprises may decide to use a combination of application based and user-based class definitions.

Once the classes and their membership have been defined, network switches and routers must have a way to recognize the class associated with each received packet, (again, so they know how to schedule the packet transmission). Application/server, device or user class could be identified by examining source/destination Ethernet (MAC) address, IP address, TCP/UDP port number, or other information included in the data packet (eg. Web URL).

Today it is possible to define classes (and associated QoS policies) separately in individual routers or switches, but this is often impractical in large networks. Increasingly enterprises are expected to define class using a centralized policy-management system, with class definitions and memberships stored in enterprise directory system.

## 4. QoS Techniques

The following discussion details some of the currently available quality of service techniques.

***Prioritization:*** Traffic can be prioritized in each router/switch giving selected applications priority access to the outbound links when contention for band width occurs.

However, with a simple prioritization scheme, some low-priority traffic can be "starved" for bandwidth, causing session time outs and retransmissions.  To avoid this a "fair queuing" feature can be added which gives at least  some service to all applications or, getting more sophisticated, we can assign bandwidth in proportion to the traffic's priority. The latter technique is called "weighted fair queuing."

For LANs (local area networks) standard like IEEE 802.1p and 802.1q describes the implementation of prioritization schemes.  For WANs (wide area networks) technologies like Frame Relay and ATM has their own QoS features built in the specifications.  This is implemented by the DE (discard eligible)  marking on low priority data in Frame Relay and by assigning CLP (cell loss priority) for data cells in the ATM transmissions.

***Bandwidth Reservation:*** While prioritization can provide a mechanism for determining the *relative* service provided to different types of network traffic, it doesn't always guarantees absolute performance, especially when bandwidth is limited.  There are certain applications, such as interactive voice or video, that require bounded network delays (e.g., latency < 100 milliseconds) or limited delay variation.  For this type of traffic we may need to reserve network link bandwidth.

Unlike time-division multiplexers, routers/switches that support bandwidth reservation don't use static bandwidth allocation.  Instead, they set aside bandwidth only when certain type of traffic (e.g., voice, video) are being transmitted; otherwise the "reserved" bandwidth is available to other applications.

Resource Reservation Protocol (RSVP) is a signaling protocol that can be used by applications to request special treatment by the network, or routers to signal each other. Today, most routers that support RSVP requests do so by reserving bandwidth.

In Frame Relay and ATM the concept of Switched Virtual Circuits (SVC) does similar function.

***Congestion Control:***  Network nodes can become congested if too much traffic converge on the same device and must traverse the same outbound link.  When this happens and the buffers of the node device is exhausted, data packets re dropped.  This causes significant number of retransmissions. The results will be poor overall network performance.

A solution this problem is the "random early discard" (RED) algorithm in the router/switch, which randomly drops data packets when the node gets busy, but long before the buffers become exhausted.  This keeps the traffic "wave" behavior from occurring.  The enhanced version of RED is called "weighted RED".

Frame Relay handles congestion control by sending FECN (forward explicit congestion notification) and BECN (backward explicit congestion notification) messages between switches controlthe data flow in such away that buffer overflows are minimized.

***Traffic Shaping:*** The traffic shaping technologies try to decrease the burstiness of the data traffic by identifying the traffic flows and then managing the maximum transmission rates of those flows into the network. This provided a stable network and smooth traffic flow in the network devices.

## 5. Conclusion

In the wide area network, there has always been a recognition that some type of QoS would be useful, due to the need to make the most effective use of relatively expensive and limited bandwidth. But, in the local area, network architects are also becoming interested in QoS, mostly because of an expectation that IP telephony is coming.

Corporate services are the primary focus of QoS, with Service Level Agreements (SLAs) defining the guarantees and responsibilities between subscribers and providers. To forge an agreement that customers can trust, a service provider needs a network with QoS capabilities and a policy management system to configure, control, and maintain performance levels. It important that the today's network designers should pay sufficient attention to the QoS issue make their network future proof, while protecting the investment.

## 6. References


1. Paul Ferguson, Geoff Huston: "**Quality of Service – Delivering QoS on the internet and the corporate networks**", John Wiley & Sons, 1998.
2. David Passmore: "**Quality of Service in the LAN**" Business Communication Review, June 1999.
3. David Passmore: "**Classifying the Traffic**", Business Communication Review, August 1999.
4. David Passmore: "**QoS: What's Good Enough?**", Business Communication Review, August 1998.
5. Uyless Black: "**Frame Relay Networks**", McGraw-Hill, 1998.
6. David E. McDysan, Darren L. Spohn, "**Hands-On ATM**", McGraw-Hill, 1998.