

Quality of Services for ISP Networks

Qi Guan

SIEMENS AG Austria
Siemensstrasse 88-92
A-1210, Vienna, Austria

Key words: Quality of Service, QoS, ISP Internet, Backbone, VoIP

Abstract: Quality of Service (QoS) is a basic functionality that the Internet needs to be able to cope with when the applications increased in number and varied in use. Internet QoS issues concerning the Internet involves many different areas: end user QoS, ISP network QoS, backbone QoS. Backbone QoS is the problem concerning traffic engineering and bandwidth. ISP network QoS is one of the major problems in the Internet. In this paper, we would like to present an architecture model for QoS enabled ISP networks. This model is based on differentiated services and is connected to QoS or non-QoS enabled backbones for IP upstream and downstream QoS traffic. At the end of paper, an example of VoIP application is introduced.

1. INTRODUCTION

Until now, two separate worlds have existed for telecommunication services and networks – one for voice and one for data.

The voice world is that of switched PSTN / ISDN networks, which generally provide real-time services and features such as call diversions, conference and display services. Switched PSTN / ISDN networks guarantee quality and reliability in the form of constant availability of voice services for their customers. As a result thereof, this quality of service has made voice networks extremely profitable.

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: [10.1007/978-0-387-35522-1_37](https://doi.org/10.1007/978-0-387-35522-1_37)

The other world is that of low-cost and efficient data services such as e-mail, file transfer and access to ‘information’ resources of the World Wide Web (WWW) provided by packet-based networks, mainly the Internet. These data networks are based on commercial servers and widespread programming environments. They offer a degree of flexibility which allows the short term creation of applications.

Using slogans like “The convergence of the network“ and “Everything over IP“, more and more traditional telco services (e.g.: telephony, video conference, radio/TV broadcasting, etc.) are being adapted for Internet usage. This has caused an increase in Internet traffic as the number of users and applications that are constantly being introduced to the Internet.

In addition and due to the multitudes of new applications the character of Internet traffic has also been and still is in the process of changing to meet demands. Some of the new breed of Internet applications, for example IP-telephony, have very strict timing requirements. The end-to-end latency (phone-to-phone) determines whether the conversation sounds interactive or not. A delay of 100 msec is the threshold for phone companies. A delay larger the 200 msec is quite noticeable and will usually cause a switch to half-duplex conversation.

The latency of an Internet transmission depends on:

- network congestion,
- number of routers involved,
- type of prioritisation,
- delay of voice encoding, and
- packetisation.

In order to be able to ensure acceptable IP telephone services the Internet must provide enough bandwidth, low delay. Generally speaking, the Internet must offer the same QoS as conventional phone companies. Here QoS means that a network element (e.g. an application, host or router) to have some levels of assurance that its traffic and service requirements can be satisfied.

Satisfying demands for different types of applications requires that different more flexible methods be used for handling network traffic. Internet applications generate traffic at varying rates and generally require that the network is able to carry traffic at the rate they generated. In addition, applications are more or less tolerant of traffic delays in the network and of variations in traffic delays. Certain applications can tolerate some degree of

traffic loss while others can not. For example, telephone service requires no significant delay and no jitter, data service requires no loss of data.

To be able to guarantee quality in Internet telephony it is absolutely necessary to achieve a so-called end-to-end QoS. This is, however, not always the case. The Internet is composed of many different types of networks (logical or physical) and operated by many independent carriers. For example, telcos provide twisted-pair copper lines to end-users as a means of accessing the network. Here the user can either dial-up a modem for low-band connections or use xDSL for broadband connections to its ISP. The Internet Service Providers (ISP) then provide the Internet services to their users, as well as the connection to the Internet backbones. The Internet backbones are again operated by many different carriers or organisations as wholesalers in different countries.

End-to-end QoS in Internet can be divided into various topics, for example:

- End users QoS
- ISP network QoS
- Internet backbone QoS

QoS problems have their own characteristics depending on the demands they have to meet (e.g.: end user, ISP and backbones). Each of these demands must be solved by special technologies. For example, QoS for Internet backbones is more or less a problem of bandwidth and traffic engineering. If we can provide enough bandwidth in the backbone, QoS is no longer a problem. However, in the ISP network, the connections of end users are concentrated and therefore traffic congestion takes place here. This is the main problems in the ISP network concerning QoS.

In this paper we present an architecture model for QoS for ISP networks. In the following section some QoS technologies for the Internet are described. Section 3 describes QoS applied to ISP networks. An example is shown in section 4.

2. QUALITY OF SERVICE TECHNOLOGIES

This section describes some technologies concerning QoS for IP services. The *Internet Engineering Task Force* (IETF) has proposed many service models and mechanisms designed to meet the demands for Internet QoS.

Among them are the *Integrated Services/RSVP* model, the *Differentiated Services* (DiffServ) model and the *Multi-Protocol Label Switching* (MPLS).

The ATM forum has also proposed ATM QoS services, for example, CBR (Constant Bit Rate), UBR (Unspecified Bit Rate), VBR (Variable Bit Rate) and ABR (Available Bit Rate) etc. ATM can be used in access networks, (e.g. ADSL). However, the applications involved in ATM network are mainly IP rather than ATM. ATM QoS can only be partially used as an Internet QoS. ATM is more suitable for traffic engineering on Internet backbones.

2.1 Integrated Services/RSVP

The integrated services model is characterised by resource reservation. Real-time applications must first set up paths and reserve resources, before data can be transmitted; RSVP (Resource Reservation Setup Protocol) is used for this purpose.

The sender sends a PATH message to the receiver specifying the characteristics of the traffic. Each intermediate router along the path then forwards the PATH message to the next hop determined by the routing algorithm. Upon receiving a PATH message, the receiver responds with a RESV message to request resources for the flow. Each intermediate router along the path can reject or accept the request carried in the RESV message. If a request is rejected, the router will send an error message to the receiver and the signaling process will be terminated. If the request is accepted, resources will be reserved for the flow and the related flow state information will be installed in the router.

RSVP is intended to provide the closest thing to circuit emulation on IP networks. RSVP is the most complex of all the QoS technologies for applications (hosts) and for network elements (routers and switches). As a result, it also represents the biggest departure from standard “best-effort” IP service and provides the highest level of QoS in terms of service guarantees, granularity of resource allocation and detail of feedback to QoS-enabled applications and users.

2.2 Differentiated Services

The differentiated services approach of providing quality of service in networks employs a small, well-defined set of building blocks from which a variety of aggregate behaviors may be built.

In DiffServ, packets are classified and marked in DSCP (DiffServ code point, a field in the header of IP packets) at the network ingress routers to create several packet classes. DSCP is placed in the IPv4 TOS octet or the IPv6 traffic class octet. Packets having different classes receive different services, allowing delay indication and packet drop preferences. Thus, the IP packets can be mapped to several QoS classes which in turn will be handled in different queues within the network elements such as routers and switches.

Using different *classification, policing, shaping* rules and *scheduling* algorithms in the routers, many services can be provided, for example,

- 1) *Premium Service (PS)* with its scheduling priority for applications requiring low delay and low jitter,
- 2) *Assured Service (AS)* with high buffer priority for applications requiring reliable but not fixed delay bounds;
- 3) *Best Effort Service*, which is the same as today's internet service.

Note that the differentiated services model only defines the DS fields. It is the ISPs' responsibility to decide what services must be provided.

The main advantages of differentiated services are the simplicity and flexibility. The simplicity makes it possible for DiffServ networks to interwork easily with other QoS enabled networks. This is very important in an open environment like the Internet.

2.3 MPLS

MPLS (Multi-Protocol Label Switching) is a packet-forwarding scheme. Packets are assigned a label at the ingress of a MPLS-capable domain. Subsequent classification, forwarding, and services for the packets are based on the labels.

MPLS is similar to DiffServ in some respects, as it also marks traffic at ingress boundaries in a network, and unmarks it at the egress points. But unlike DiffServ, which uses the marking to determine priority within a router, MPLS markings (20-bit labels) are primarily designed to determine the next router hop. MPLS is protocol-independent (i.e., "multi-protocol"), so it can be used with network protocols other than IP (like IPX, ATM, PPP or Frame-Relay) or directly via data-link layers as well.

MPLS is more a “traffic engineering” protocol than a QoS protocol. MPLS routing is used to establish “fixed bandwidth pipes” analogous to ATM or frame relay virtual circuits. The difference is arguable since the end-result is service improvement and increased service diversity with more flexible, policy-based network management control, all of which the other QoS protocols also provide.

3. QUALITY OF SERVICE APPLIED FOR ISP NETWORK

As mentioned in section 1, QoS problems may occur at the end user, on ISP networks or on the Internet backbones. In this section, we will take a more detailed look at the QoS in ISP networks.

The ISP network, on the one side, can be seen as being connected to the end users. This interface is called UNI (User Network Interface). On the other side, it is connected to the Internet backbones and is called SNI (Service Node Interface).

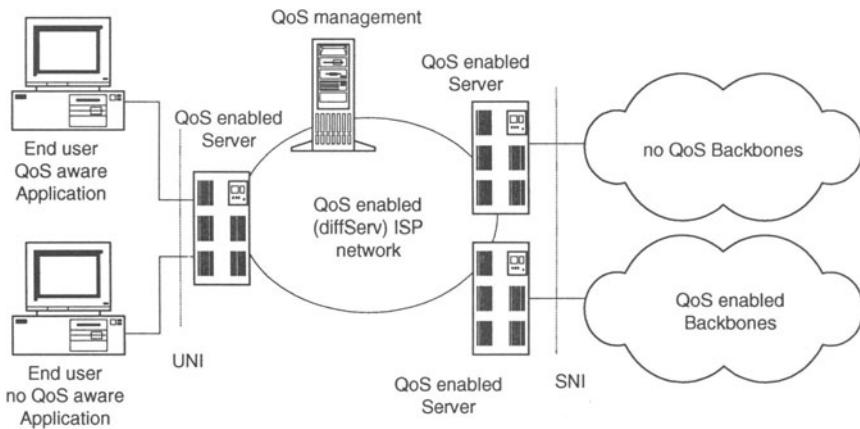


Figure 1. A QoS enabled ISP network

UNI can either be a low-speed access such as a dial-up or ISDN or a high-speed access such as xDSL and coaxial cables. SNI is usually a high-speed connection such as SDH, ATM, etc. Figure 1 shows our configuration for QoS enabled ISP networks.

The QoS enabled ISP is based on DiffServ. The ISP network is defined as a DiffServ domain. It has DiffServ boundary nodes on the UNI and SNI sides. These boundary nodes act as both ingress and egress nodes.

3.1 QoS management

QoS management is used to control traffic within the QoS enabled ISP network. It administers users profile, the load status of each ingress/egress node and applications flow. Based on these information QoS management may cluster (group) users into different service level, assign the application flow different priorities. In order to work with QoS management, the ISP network has some kind of “awareness” functions like

- User awareness,
- Traffic awareness, and
- Application awareness

User awareness means that the ISP network should cluster different user groups. A business user may, for example, be provided with a high level of service, regardless of ISP traffic load.

Traffic awareness is based on the analysis of IP flow. A IP flow can be checked by its IP header – IP address, its TCP/UDP header – port number and SYN/FIN bit. QoS enabled servers on UNI/SNI side may then make intelligent decision which kind service the IP flow can observe: with low delay or no loss of data.

Application awareness is the way to classify applications. A HTTP application may involve real-time or non-real-time applications. The traffic awareness cannot differentiate these applications. A higher layer knowledge is necessary here to optimise the usage of network resource.

3.2 Traffic control on UNI side

For low-speed access via dialup modems or ISDN connections, QoS is necessary for both bandwidth and delay control.

For high-speed access via xDSL or coaxial cables, it seems the QoS would not be necessary if there is enough bandwidth. But as usual there are

congestion points, for example, the connections to backbones, which require QoS to be able to decide what traffic gets first resources at these points.

We use DiffServ servers on the UNI side and provide QoS for the end user.

According to SLA (Service Level Agreement), the end user sends application traffic into the ISP DiffServ network through the ingress node on UNI side. Each transmitted packet will be marked with its appropriate DSCP. Interior nodes within the DiffServ domain use the DSCP to classify packets and apply specific queuing or scheduling behavior (known as a *per-hop behavior* or *PHB*) based on the results of the classification.

A real-time application may, for example, apply the *expedited-forwarding* (or EF) PHB. EF PHB is defined to assure that packets are transmitted from the ingress node on the UNI side to the egress node on SNI side (at some limited rate) with very low latency.

The QoS enabled application must also register itself to the QoS management according to SLA together with other required information.

3.3 Traffic control on SNI side

An ISP network can be connected to QoS or to non-QoS enabled Internet backbones. A QoS enabled backbone with a QoS enabled ISP network provides end-to-end QoS. A non-QoS enabled backbone may provide a certain kind of QoS if there is enough bandwidth and traffic engineering in the backbone.

3.3.1 Connected to QoS enabled backbone

A QoS enabled ISP network connected to a QoS enabled backbone may act as two DiffServ domains. The DiffServ egress node may perform traffic conditioning functions on traffic forwarded to a directly connected peering domain, depending on the details of the TCA (Traffic Conditioning Agreement) between the ISP and backbone DiffServ domains.

The ISP and backbone DiffServ domains may support different PHB groups internally and different code points to the PHB mappings. However, to be able to permit services which span across the domains, the peering DiffServ domains must each establish a peering SLA which defines (either explicitly or implicitly) a TCA which specifies how transit traffic moving

from one DiffServ domain to another is conditioned at the boundary between the two DiffServ domains.

Traffic conditioning performs metering, shaping, policing and/or re-marking to ensure that the traffic entering the DiffServ domain conforms to the rules specified in the TCA, in accordance with the domain's service provisioning policy. The extent of traffic conditioning required is dependent on the specifics of the service, and may range from simple code point re-marking to complex policing and shaping operations.

When packets exit the traffic conditioner of a DiffServ boundary node must set the DS code point of each packet to an appropriate value.

3.3.2 Connected to non QoS Backbones

Packets in a non DiffServ backbone have no settings in their DSCP. When the packets come to the ingress node of an ISP at the SNI, these ingress nodes must perform not only all of the necessary traffic conditioning functions on the incoming traffic but also all of the necessary traffic classifications.

Since there is no default setting in the DSCP of incoming packets, traffic classifications are done by the "awareness" functions in ingress nodes and QoS management. Does the traffic belong to a business user? Is the traffic a http access or large data download? Based on the these analysis and services status on UNI side, traffic conditioning is performed on incoming packets. The DSCP of packets are marked with correspondent service classes.

4. AN EXAMPLE

This section shows an example how to solve the QoS problem related to Voice over IP (VoIP). Our VoIP application provides telephony service between a PC in ISP and a PSTN user. A VoIP gateway is responsible for converting PSTN voice signals into voice IP packets and then sends them to the diverted VoIP terminal (PC) or vice versa. Our VoIP gateway also provides call signalling for PSTN such as SS7 and for the Internet such as H.323 or SIP.

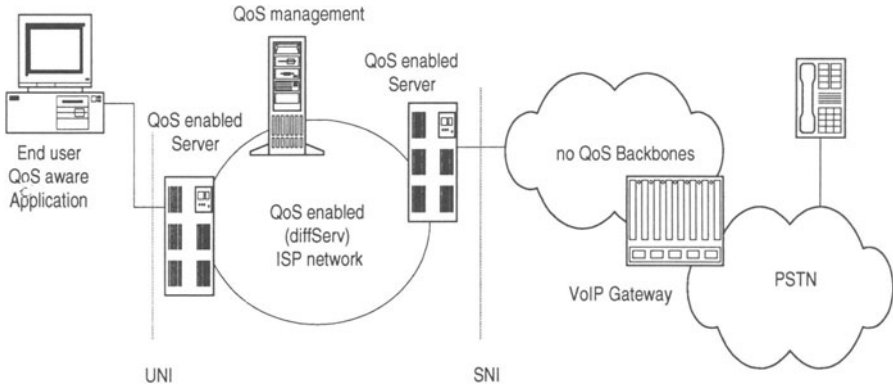


Figure 2. QoS enabled ISP and VoIP Applications

Internet users having a VoIP terminal on their PC can dial an E.164 (a usual telephone number system) code to call a PSTN user by means of Internet telephony signaling H.323 / SIP. The signaling process is like building a VoIP call connection to the nearest VoIP gateway from a called PSTN user. The gateway then sends SS7 signaling to the PSTN user. The PSTN user may pick up the phone and the call connection is established between the caller (the Internet user), and called party (the PSTN user). Up to this time no real-time packet was needed. After the connection is built-up the Internet user may register himself to the QoS management with information such as the IP-address of the VoIP terminal, the port number of the UDP packets (voice packets) and which service class applied to the real-time application.

The VoIP terminal sends the marked voice packet to the ingress node of the ISP. These packets apply the *expedited-forwarding* (or EF) PHB and are transmitted from the ingress node on the UNI side to the egress node on the SNI side with a very low latency. There may be many IP packets at egress node and that is where the congestion usually takes place. Only VoIP packets with a high service class are first transmitted with a high priority through the egress node. When the packets exit the ISP to a non-QoS backbone the DS code point of each packet must be reset.

The PSTN user then sends the voice signal to the VoIP gateway. The gateway converts the voice signal into voice UDP packets. The UDP packets have an IP-address and port number in their header. When the packets arrive at the ingress node of the ISP on SNI side, the node may control the UDP

header and consult the QoS management. If these packets are registered as being part of a real-time application in QoS management, the *expedited-forwarding* (or EF) PHB is applied to them. The packets having a higher service class as compared to other lower service classes get the resource of ingress node first and are transmitted from SNI side to the egress node on the UNI side with a high priority and a very low latency.

5. RELATED WORKS

As mentioned above, IETF has proposed many service models and mechanisms designed to meet the demands for Internet QoS, e.g. RSVP, DiffServ and MPLS etc. Usually RSVP is used in ISP networks, while DiffServ and MPLS are used in Internet backbones. While our QoS enabled ISP network use DiffServ and can be connected to either QoS or non-QoS backbones.

Qbone of Internet2 is an interdomain testbed for DiffServ that seeks to provide the higher-education community with end-to-end services in support of emerging advanced networked applications. In Qbone each network participating is considered a "DS domain". The union of these networks - the QBone itself - is considered a "DS region". QBone participants cooperate to provide one or more interdomain services besides the default, traditional best effort IP service model. Virtual Leased Line Service in Qbone is such kind services as defined in DiffServ: for example "Premium Service". Every QBone DS domain supports the expedited forwarding (EF) per-hop behavior (PHB) [EF] and configure its traffic classifiers and conditioners (meters, markers, shapers, and droppers) to provide a VLL service to EF aggregates. As far as connection to QoS backbone concerned, our QoS enabled ISP network is the same as a DS domain in Qbone. But in Qbone it does not address any problem of connection to non-QoS backbone.

In the 46th IETF Conference an new architecture called DCS (Distributed Call Signalling) was introduced. DCS analyse the use of SIP in a cable environment and is a major push for IP-Telephony. QoS topic in DCS is RSVP and is tight coupling with call signalling. All the problems addressed in DCS are in cable environment. It is than easy to implement End-to-End QoS for VoIP in the domains where RSVP involved. The problems which we discussed are the interdomain VoIP applications. In such situation we must provide QoS for IP-Telephony even when our ISP connected to non-QoS backbones.

6. CONCLUSIONS

The Internet needs QoS for various applications. Internet QoS comprises various topics, i.e. end user QoS, ISP network QoS, backbone QoS. Backbone QoS is more or less a problem of traffic engineering and bandwidth. MLPS and ATM are more suitable for backbone QoS. ISP network QoS is one of major problems concerning to Internet. We presents an architecture model for QoS enabled ISP networks. This model is based on DiffServ and is connected to QoS or non-QoS enabled backbones for upstream and downstream QoS traffic.

7. ACKNOWLEDGMENTS

The author would like to acknowledge Josef Schmitt, Dr.Redha Reda, Wolfgang Aichberger, Albert Nill, Jean Kremlicka, and Josef Hosp for their useful comments, suggestions and improvements to this paper.

8. ABBREVIATIONS

PSTN	Public Switched Telephone Network
SIP	Session Initiation Protocol
VoIP	Voice over IP
QoS	Quality of Service
ISP	Internet Service Provider
CBR	Constant Bit Rate
UBR	Unspecified Bit Rate
VBR	Variable Bit Rate
ABR	Available Bit Rate
ADSL	Digital Subscriber Line
RSVP	Resource Reservation Setup Protocol
DiffServ	Differentiated Services
MPLS	Multi-Protocol Label Switching
SDCP	DiffServ codepoint
PHB	Per-Hop Behaviour
SLA	Service Level Agreement
TCA	Traffic Conditioning Agreement
DCS	Distributed Call Signalling

9. REFERENCES

- [1] Qi Guan, "Voice Services for PSTN and IP Networks", *Intelligence in Networks, SmartNet'99*, November, 1999.
- [2] IETF "Integrated Services" working group.
See <http://www.ietf.org/html.charters/intserv-charter.html>
- [3] IETF "Resource Reservation Setup Protocol " working group.
See [http:// www.ietf.org/ietf/html.charters/rsvp-charter.html](http://www.ietf.org/ietf/html.charters/rsvp-charter.html)
- [4] R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, "Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification", RFC 2205, September 1997.
- [5] IETF "Differentiated Services".
See [http:// www.ietf.org/html.charters/diffserv-charter.html](http://www.ietf.org/html.charters/diffserv-charter.html)
- [6] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998
- [7] IETF "Multiprotocol Label Switching" working group.
See <http://www.ietf.org/html.charters/mppls-charter.html>
- [8] QoS-Forum, "White Paper –QoS protocols & architectures", July 1999.
See <http://www.qosforum.com>
- [9] QoS-Forum, "White Paper – The Need for QoS". July 1999.
See <http://www.qosforum.com>
- [10] Microsoft Corporation. "White Paper – An Overview of QoS", 1999
- [11] Xipeng Xiao and Lionel M. Ni, "Internet QoS: the Big Picture", Department of Computer Science, Michigan State University, 1999.
- [12] Ben Teitelbaum, "Qbone Architecture(v.10)", 1999.
See <http://www.internet2.edu/qos/wg/papers/qbArch/1.0/draft-i2-qbone-arch-1.0.html>
- [13] DCS - Distributed Call Signalling, 1999.
See http://www.packetcable.com/packetcable_specs.html