Quality of Service on the Internet

1. INTRODUCTION

The existing Internet can only provide 'best-effort' service, i.e. it tries its best to forward user traffic. Using existing mechanisms, however, there can be no guarantees regarding loss rate or delays because packets can be dropped indiscrimately during congestion. While this best-effort service works fine for most traditional Interent applications such as email, it is totally unacceptable for emerging real-time multimedia applications such as Internet telephony or video-on-demand (VOD). Random packet-dropping can seriously degrade the audible or visual quality of a streamed movie. In addition, some premium customers are willing to pay higher service charges to obtain better service (e.g. lower delay latency, higher data rate). Thus, there is growing demand to enhance the existing Interent to provide better transmission services – i.e. exceeding 'best-effort'. This leads us to the study of Internet Quality-of-Service (QoS).

This article will describe several models and mechanisms proposed by the Internet Engineering Task Force (IETF) to meet the demand for Internet QoS in the Next Generation Internet. The Resource Reservation Protocol (RSVP), Integrated Service (Intserv) model and Differentiated Service (Diffserv) model will be covered.

These efforts are all emerging technologies to provide Internet QoS at different levels in the OSI model. Intserv and Diffserv represent new service models at the transport layer to extend the current 'best effort' service to accommodate new applications such as real-time video. RSVP is a new signalling protocol for setting up paths and reserving resources to support both Intserv and Diffserv. By applying these technologies together, we can build a framework to address future Internet QoS requirements.

2. WHAT IS QUALITY OF SERVICE

Quality of Service (QoS) is a set of service requirements to be met by the network while transporting a packet stream from a source to a destination (unicast or multicast). The common QoS parameters used in Internet services are:

- **bandwidth**: the average usable and available bandwidth over the link at any time;
- **delay**: the average end-to-end delay caused at network level at any time;
- **delay jitter**: the average difference of the various delay times over the link;

- **packet loss probability**: the average probability of packet loss over the link over a length of time.

Different applications have different requirements for bandwidth, delay, and packet loss probability. Therefore, QoS must also be a measurable level of service delivered to network users. Such QoS can be provided by network service providers in terms of an agreement, known as a **Service Level Agreement (SLA)** made between network users and service providers. QoS features provide the ability to manage traffic intelligently across a network.

Some applications require bandwidth and delay guarantees, referred to as quantitative applications, while others are more qualitative. For example, voice applications have stringent delay requirements and can tolerate minimal packet loss. Conversely, an FTP file transfer may be insentitive to delays but very sensitive to packet drops. To accommodate these differences, QoS generally assigns traffic flows to one of two categories:

- **Guaranteed service** (quantitive) reserves a designated amount of bandwidth from end to end and can guarantee a specified delay tolerance for the exclusive use of an application or even aggregated sessions. This type of service is best for applications that require a specific amount of bandwidth with long duration. E.g. a video conference call may require at least 64 kbps for 30 minutes.

- **Differentitated service** (qualitative) provides simple prioritisation. Applications are detected at the ingress and assigned SLAs. It is the SLA that drive which QoS mechanisms the router will use, for example, which queue will be used to place traffic, and, if congestion requires packets to be dropped, which drop priority will be designated.

After completing our review of these introductory QoS concepts, we will discuss on individual technology, RSVP, to support QoS in next section.

3. RESOURCE RESERVATION PROTOCOL (RSVP)

Although it evolved from a pure data network, the Internet currently carries different type of traffic with diverse characteristics and service requirements. Both real-time and non-real time traffic is sent through the networks using an **Internet Protocol (IP).** The IP bearer service provides no guarantee of service quality in terms of error, delay, or bandwidth. The network will attempt to do the best it can to deliver the message, i.e. best-effort service. If there is sufficient bandwitdh in the network, then all of these service requirement will be met.

Nevertheless, packets will be dropped indiscrimately during congestion. This may not be a problem for traditional data raffic because the **Transport Control Protocol (TCP)**, communication by retransmitting missing packets.

TCP is not suitable, however, for delay-sentistive applications such as audio or video. Alternatively, the **User Data Protocol (UDP)** is used to transport real-time multimedia service. Since UDP does not guarantee delivery of packets, audible or visual quality will be severly affected by dropped packets. To guarantee service quality for a given type of traffic, separate resources can be reserved. For instance, you can reserve 32 kbps of dedicated bandwidth for transmitting voice packets. Using this basic concept, the **Resource Reservation Protocol (RSVP)** is a signalling protocol that can be used by hosts to requrest resource reservation through a network. RSVP can ensure the reserved resources will be available if and when data are actually sent through the network.

How does RSVP work

RSVP is used to set up reservations for network resources. When an application in a host (the data stream receiver) requests a specific quality of service (QoS) for its data stream, it uses RSVP to deliver its request to routers along the data stream paths. RSVP is responsible for the negotiation of connection parameters with these routers. If the reservation is setup, RSVP is also responsible for maintaining router and host states to provide the requested service.

4. INTEGRATED SERVICES (Intserv)

The Integrated Services model is a proposed extension to current Internet architecture and protocols that is intended to provide integrated services that support both real-time and current non-real-time IP service. The basic concept is to foresee a set of service models to be provided on the Internet, moving beyond the best-effort model. Two such service models for real-time applications are defined:

- Guaranteed service provides strict boundaries for the day and loss probabilities for packets from a given flow, provided that the flow complies with a traffic contract.
- Controlled-loaded service provides a QoS comparable to best-effort service under unloaded conditions.

The implementation of the Inteserv model includes mechanisms for resource reservation, admission control and queue management. Applications requiring guaranteed or controlled-load service must set up the path and reserve resources by using RSVP before transmitting their data. The call admission control mechanism determines whether a request for resrouces can then be accepted. A queue management mechanism is used to reschedule packet transmissions according to the QoS requirement. Thus, the Intserv model requires routers along the route to reserve adequate resources and maintain the state of flow.

5. DIFFERENTIATED SERVICES (Diffserv)

Due to the difficulties related to implementing the Intserv model, the IETF recently proposed the Differentiated service (Diffserv) model, which enables either end-to-end or intradomain service discrimination. The Diffserv model can specify and control network traffic by class in wuch a way that certain types of traffic get priority. In constrast to Intserv, user flows in Diffserv are only controlled at the edge of the network and then aggregated into a small set of traffic classes.

In the Diffserv model, traffic is classified and possibly conditioned while entering a network. Based on the result of this classification, classes of traffic are then attributed to different behaviour aggregates. Each of these traffic classes is identified by a field, called a **Diffserv Codepoint (DSCP)**, in the IP header. Within a Diffserv network, packets are forwarded according to the specific queuing behaviour (known as **per-hop behaviour** or **PHB**) associated with their DSCP. Per-hop behaviour defines how an individual router treats an individual packet when sending it over the next hop through the network. In addition, traffic classes from many flows having similar QoS requirements are marked with the same DSCP. These flows can then be aggregated to a common queue. Since the DS field is set at the network boundaries, no per-flow state is necessary at the network core. This eliminates the scalability problem characteristic of the Intserv model.

By establishing a way of delivering differentiated per-hop forwarding behaviour to IP packets, Diffserv allows a shared network to accommodate different QoS levels for traffic streams using the same infrastructure. Diffserv enables ISPs to define *classes of service (CoS)* to support particular traffic requirements and to offer ‘premium’ services for special data types such as voice. In other words, we can define ‘gold’, ‘silver’ or ‘bronze’ services for different users. The standard will also specify components that might be used in usage-based pricing, helping ISPs balance customer demand with available bandwidth.

6. CONCLUSION

In short, providing QoS to the Internet is a complicated issue that requires new protocols to cooperate with each other. The technologies we've discussed are all emerging efforts in an Internet community pushing for better service. The readers can refer to additional literatures in section 7 for the future trend of QoS development on the Internet.

106.

11. XIAO, Xipeng and NI, Lionel M., "Internet QoS: A Big Picture",IEEE Network, March-April 1999, pp.8 – 18.

Related Internet Drafts:

1. "Specification of Committed Rate Quality of Service", Internet Draft, June 1996,

ftp://ds.internic.net/internet-drafts/draft-ietf-intserv-commit-rate-svc-00.txt

2. "Specification of Guaranteed Quality of Service", Internet Draft, Aug.1996,

ftp://ds.internic.net/internet-drafts/draft-ietf-intserv-commit-rate-svc-06.txt

3. "The Use of RSVP with IETF Integrated Services", Internet Draft, June 1996,

ftp://ds.internic.net/internet-drafts/draft-ietf-intserv-rsvp-use-00.txt

4. "A Framework for Differential Services", Internet Draft, May 1998,

ftp://ds.internic.net/internet-drafts/draft-ietf-diffserv-framework-00.txt

5. "Multiprotocol Label Switching Architecture ", Internet Draft, Mar. 1998,

ftp://ds.internic.net/internet-drafts/draft-ietf-mpls-arch-01.txt

Related RFC's

1. RFC 1633, "Integrated Services in the Internet Architecture: An Overview", July 1994, ftp://cls.internic.net/rfc/rfc1633.txt

2. RFC 2205, "Resource ReSerVation Protocol (RSVP) Version 1 Functional Specification", Sept. 1997 ftp://cls.internic.net/rfc/rfc2205.txt

3. RFC 2475, "An Architecture for Differentiated Services", Dec. 1998, ftp://cls.internic.net/rfc/rfc2475.txt

4. RFC 2386, "A Framework for QoS-based Routing in the Internet", Aug. 1998, ftp://cls.internic.net/rfc/rfc2386.txt

5. RFC 2212, "Specification of Guaranteed Quality of Service", Sep. 1997, ftp://cls.internic.net/rfc/rfc2212.txt

6. RFC 2211, "Specification of the Controlled-Load Network Element Service", Sep. 1997, ftp://cls.internic.net/rfc/rfc2211.txt

7. RFC 2474, "Definition of the Differentiated Serivces Field (DS Field) in the Ipv4 and Ipv6 Headers", Dec. 1998, ftp://cls.internic.net/rfc/rfc2474.txt

8. RFC 2309, "Recommendation on Queue Management and Congestion Avoidance in the Internet", Apr. 1998, ftp://cls.internic.net/rfc/rfc2309.txt