What Is Quality of Service (QoS)?

This chapter defines Quality of Service (QoS) and introduces the QoS background, indicators, service models, and mechanisms in the DiffServ model.

Quality of Service (QoS)

Introduction

This chapter defines Quality of Service (QoS) and introduces the QoS background, indicators, service models, and mechanisms in the DiffServ model.

QoS is a basic feature of Huawei data communications products, including switches, routers, WLAN products, and firewalls.

Understanding QoS

QoS Background

Diverse services result in a sharp increase in network traffic, which may cause network congestion, increase forwarding delay, or even cause packet loss. Any of these situations will cause service quality deterioration or even service interruption. Therefore, real-time services require a solution to prevent network congestion. The best solution is to increase network bandwidth, but increasing network bandwidth is costly. The most cost-effective way is to use a "guarantee" policy to manage traffic congestion.

QoS guarantees end-to-end service quality based on the requirements of different services. It helps improve utilization of network resources and allows different types of traffic to preempt network resources based on their priorities; for example, voice, video, and important data applications can be processed preferentially on network devices.

QoS Indicators

The factors that affect the network service quality need to be learned to improve network quality. Traditionally, factors that affect network quality include link bandwidth, packet transmission delay, jitter, and packet loss rate. To improve the network service quality, ensure the bandwidth of transmission links, and reduce packet transmission delay, jitter, and packet loss rate. These factors that affect the network service quality become QoS indicators.

- **Bandwidth**
  The bandwidth, also called throughput, refers to the maximum number of transmitted data bits between two ends within a specified period (1 second) or the average rate at which specified data flows are transmitted between two network nodes. Bandwidth is expressed in bit/s.

  Generally, data transmission capability and network service quality are accompanied by the bandwidth. In other words, a lane is positive to the traffic flow capacity with low traffic jam in a highway. Network users all expect higher bandwidth; however, the O&M costs are higher. Therefore, bandwidth becomes a serious bottleneck as the Inte… develops rapidly and services become increasingly diversified.
- **Delay**
  The delay refers to the time required to transmit a packet or a group of packets from the transmit end to the rec… end. It consists of the transmission delay and processing delay.

  Voice transmission is used as an example. A delay refers to the period during which words are spoken and ther… heard. Generally, people are insensitive to a delay of less than 100 ms. If a delay ranging from 100 ms to 300 m… occurs, a speaker can sense slight pauses in the responder's reply, which can seem annoying to both. If a dela… longer than 300 ms occurs, both the speaker and responder obviously sense the delay and have to wait for responses. If the speaker cannot wait but repeats what has been said, voices overlap and the quality of the conversation deteriorates severely.
- **Jitter**

Jitter is an important parameter for real-time transmission, especially for real-time services, such as voice and video, which are zero-tolerant of jitters because jitters will cause voice or video interruptions.

Jitters also affect protocol packet transmission. Specific protocol packets are transmitted at a fixed interval. High jitters may cause flapping of the protocols.

Jitters exist on networks but the service quality will not be affected if jitters do not exceed a specific tolerance. The buffer can alleviate excess jitters but prolongs delays.

- **Packet Loss Rate**
  The packet loss rate refers to the ratio of lost packets to total packets. Slight packet loss does not affect services. For example, users are unaware of the loss of a bit or a packet in voice transmission. The loss of a bit or a packet in video transmission may cause the image on the screen to become garbled instantly, but the image can be restored quickly.

  TCP is used to transmit data to handle slight packet loss because TCP instantly retransmits the packets that have been lost. If severe packet loss does occur, the packet transmission efficiency is affected. QoS focuses on the packet loss rate. The network packet loss rate must be controlled within a certain range during transmission.

QoS Service Models

How are QoS indicators defined within proper ranges to improve network service quality? The QoS model is involved. The QoS model is not a specific function, but an E2E QoS scheme. For example, intermediate devices may be deployed between two connected hosts. E2E service quality guarantee can be implemented only when all devices on a network use the same QoS service model. International organizations such as the IETF and ITU-T designed QoS models for their concerned services. The following describes three main QoS service models.

- **Best-Effort**

  Best-Effort is the default service model for the Internet and applies to various network applications, such as the File Transfer Protocol (FTP) and email. It is the simplest service model, in which an application can send any number of packets at any time without notifying the network. The network then tries its best to transmit the packets but provides no guarantee of performance in terms of delay and reliability.

  The Best-Effort model is suitable for services that have low requirements for delay and packet loss rate.

- **Integrated Service (IntServ)**

  In the IntServ model, an application uses a signaling protocol to notify the network of its traffic parameters and apply for a specific level of QoS before sending packets. The network reserves resources for the application based on the traffic parameters. After the application receives an acknowledgement message and confirms that sufficient resources have been reserved, it starts to send packets within the range specified by the traffic parameters. The network maintains a state for each packet flow and performs QoS behaviors based on this state to guarantee application performance.

  The IntServ model uses the Resource Reservation Protocol (RSVP) for signaling. The RSVP protocol reserves resources such as bandwidth and priority on a known path, and each network element along the path must reserve required resources for data flows requiring QoS guarantee. That is, each network element maintains a soft state for each data flow. A soft state is a temporary state that is periodically updated through RSVP messages. Each network element checks whether sufficient resources can be reserved based on these RSVP messages. The path is available only if all involved network elements can provide sufficient resources.

- **Differentiated Service (DiffServ)**

  The DiffServ model classifies packets on a network into multiple classes and takes different actions for each class. When network congestion occurs, packets of different classes are processed based on their priorities, resulting in different packet loss rates, delay, and jitter. Packets of the same class are aggregated and sent as a whole to ensure consistent delay, jitter, and packet loss rate.

  Unlike the IntServ model, the DiffServ model does not require a signaling protocol. In this model, an application does not need to apply for network resources before sending packets. Instead, the application sets QoS parameters in the packets, through which the network can learn the QoS requirements of the application. The network provides differentiated services based on the QoS parameters of each data flow and does not need to maintain a state for each data flow. DiffServ takes full advantage of IP networks' flexibility and extensibility and transforms information in packets into per-hop behaviors (PHBs), greatly reducing signaling operations. DiffServ is the most commonly used QoS model on current networks. QoS implementation described in the subsequent

  sections is based on this model.

Mechanisms in the DiffServ Model

QoS services based on the DiffServ model are supported on Huawei data communications products, including sw routers, WLAN products, and firewalls. The DiffServ model involves the following QoS mechanisms:

- **Traffic classification and marking**

  Traffic classification and marking are prerequisites for differentiated services. Traffic classification divides packe different classes or sets different priorities, and can be implemented using traffic classifiers configured on the M QoS Command-Line Interface (MQC). Traffic marking sets different priorities for packets and can be implement through priority mapping and re-marking.

- **Traffic policing, traffic shaping, and interface-based rate limiting**

  Traffic policing and traffic shaping control the traffic rate within a bandwidth limit. Traffic policing drops excess traffic when the traffic rate exceeds the limit, whereas traffic shaping buffers excess traffic. Traffic policing and traffic shaping

Congestion management buffers packets in queues upon network congestion and uses a scheduling algorithm to determine the forwarding order. Congestion avoidance monitors network resource usage and drops packets to mitigate network overload if congestion worsens.

Traffic classification and marking are the basis of differentiated services. Traffic policing, traffic shaping, interface-based rate limiting, congestion management, and congestion avoidance control network traffic and resource allocation to implement differentiated services.

Related Information

For more information and detailed procedures, refer to the following documents:

S12700 V200R013C00 Configuration Guide - QoS

CloudEngine 12800, 12800E V200R005C10 Configuration Guide - QoS

AR100, AR120, AR160, AR1200, AR2200, AR3200, and AR3600 V300R003 CLI-based Configuration Guide - QoS

Wireless Access Controller (AC and FITAP) V200R010C00 Product Documentation - QoS Configuration Guide

HUAWEI USG6000, USG9500, and NGFW Module V500R005C10 Product Documentation - Configuration - QoS