

# S Series Switch QoS Technology White Paper

Issue 01  
Date 2013-05-25

**Copyright © Huawei Technologies Co., Ltd. 2013. All rights reserved.**

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

## **Trademarks and Permissions**



and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.

All other trademarks and trade names mentioned in this document are the property of their respective holders.

## **Notice**

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

## **Huawei Technologies Co., Ltd.**

Address: Huawei Industrial Base  
Bantian, Longgang  
Shenzhen 518129  
People's Republic of China

Website: <http://enterprise.huawei.com>

---

# Contents

---

<b>1 Introduction to QoS .....</b>	<b>1</b>
1.1 What Is QoS? .....	1
1.2 QoS Specifications .....	2
1.2.1 Bandwidth/Throughput .....	2
1.2.2 Delay .....	3
1.2.3 Delay Variation (Jitter) .....	3
1.2.4 Packet Loss Rate .....	4
1.3 Common QoS Specifications .....	5
<b>2 Technology Description .....</b>	<b>7</b>
2.1 QoS Service Models .....	7
2.1.1 Best-Effort .....	7
2.1.2 IntServ .....	7
2.1.3 DiffServ .....	8
2.1.4 Comparison Between DiffServ and IntServ Models .....	12
2.1.5 Components in the DiffServ Model .....	13
2.2 Traffic Classification and Marking .....	14
2.2.1 Simple Traffic Classification .....	14
2.2.2 Complex Traffic Classification .....	21
2.2.3 Traffic Marking .....	28
2.2.4 Application of Traffic Classification and Marking .....	31
2.3 Traffic Policing and Traffic Shaping .....	32
2.3.1 Traffic Policing .....	32
2.3.2 What Is a Token Bucket .....	32
2.3.3 CAR .....	37
2.3.4 Traffic Shaping .....	45
2.3.5 Comparison Between Traffic Policing and Traffic Shaping .....	50
2.4 Congestion Management and Congestion Avoidance .....	50
2.4.1 Background .....	50
2.4.2 Congestion Management .....	52
2.4.3 Congestion Avoidance .....	59
<b>3 Application Scenarios .....</b>	<b>66</b>
3.1 User-based Differentiated Services .....	66

3.1.1 Networking Requirements .....	66
3.1.2 Configuration Roadmap.....	67
3.1.3 Procedure .....	67
3.2 Service-based Differentiated Services .....	70
3.2.1 Networking Requirements .....	70
3.2.2 Configuration Roadmap.....	71
3.2.3 Procedure .....	71
<b>4 Troubleshooting Cases.....</b>	<b>73</b>
4.1 Packets Enter Incorrect Queues .....	73
4.2 Priority Mapping Results Are Incorrect.....	75
4.3 Traffic Policy Does Not Take Effect.....	76
<b>5 FAQ.....</b>	<b>79</b>
5.1 Does the S9700 Collect Traffic Statistics Based on Packets or Bytes?.....	79
5.2 What Are the Differences Between Interface-based CAR and Global CAR?.....	79
5.3 How Does Level-2 CAR Take Effect?.....	79
5.4 A Traffic Policy Contains an ACL Rule Defining TCP or UDP Port Number Range. When the Traffic Policy Is Delivered, the System Displays the Message "Add rule to chip failed." Why?.....	80
5.5 CAR Is Incorrect. Why? .....	80
5.6 An ACL Applied to the Outbound Direction Cannot Define the Port Number Range. Why? .....	81
5.7 Can 802.1p Re-marking and Traffic Statistics Be Configured in a Traffic Policy Simultaneously on the S9700?.....	81
5.8 When Both QinQ and Traffic Policy-based VLAN Stacking Are Configured on an Interface, Which Configuration Takes Effect? .....	81
5.9 Why ACL Rule Update May Cause Instant Traffic Interruption?.....	81
5.10 After an ACL or QoS Is Configured, the Configuration Is Invalid for Mirroring Packets. Why? .....	81
5.11 Why a Traffic Policy Containing Traffic Filtering or CAR Is Invalid for Incoming Packets on an S9700?.....	82
5.12 Why PQ+DRR Configured on an S9700 Interface Does Not Take Effect? .....	82
5.13 Why Priorities in Outgoing Mirroring Packets Are Not Changed After Priority Mapping Is Configured? .....	82
5.14 When You Configure a Deny Rule in a Traffic Policy Containing Flow Mirroring, Normal Service Traffic Is Affected. Why?.....	82
5.15 When a Traffic Policy Containing Flow Mirroring Is Applied to an Interface, the Global Traffic Policy Becomes Invalid. Why? .....	83
5.16 What Is the Relationship Between an ACL and a Traffic Policy?.....	83
5.17 How Are Packets Forwarded Using PBR on S Series Switches? .....	84
<b>6 Appendix .....</b>	<b>85</b>
6.1 Common Service Priorities.....	85
6.2 Port Numbers of Common Application Services .....	85
6.3 Common Queue Scheduling Solution.....	86
6.4 Recommended WRED Parameter Setting .....	86
6.4.1 Color-based WRED Parameter Setting.....	86
6.4.2 Queue-based WRED Parameter Setting .....	87
6.5 Video Service Bandwidth Usage.....	87
6.5.1 Coding-based Video Bandwidth .....	87

---

6.5.2 HD-based Video Bandwidth .....	88
6.5.3 Video Conference Bandwidth .....	88
6.6 Audio Bandwidth Usage .....	89
6.6.1 Audio Bandwidth Based on Codec Technologies .....	89

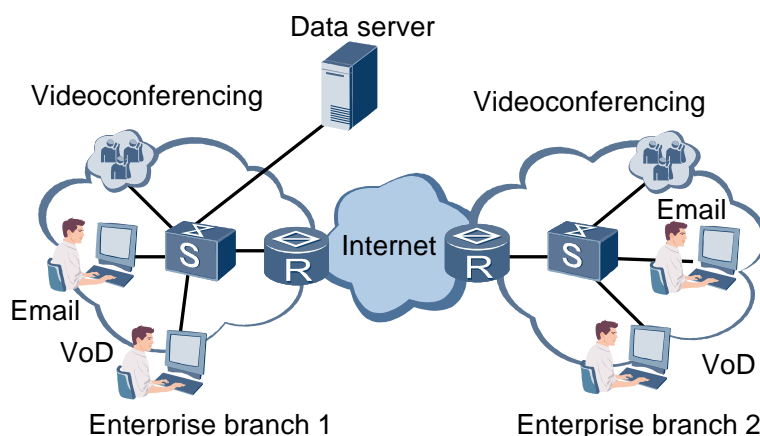
# 1 Introduction to QoS

## 1.1 What Is QoS?

As network technologies rapidly develop, services on the Internet become increasingly diversified. Apart from traditional applications such as WWW, email, and File Transfer Protocol (FTP), the Internet has expanded to encompass other services such as IP phones, e-commerce, multimedia games, e-learning, telemedicine, videophones, videoconferencing, video on demand (VoD), and online movies. In addition to web page browsing, new enterprises require services including identity authentication of employees and visitors, remote video conferencing, emails, video, FTP file upload and download, and Telnet services on special devices in working hours.

These new services have special requirements on the bandwidth, delay, and delay variation. For example, videoconferencing and VoD services demand high bandwidth, short delay, and low delay variation. Key tasks such as transaction processing and Telnet require short delay and preferential handling when congestion occurs, although such tasks do not necessarily demand high bandwidth.

**Figure 1-1** Enterprise new services



Diversified services enrich people's lives but also increase the risk of traffic congestion on the Internet. When traffic congestion occurs, services encounter long delays or even packet loss. As a result, services deteriorate or even become unavailable. Therefore, a solution to resolve traffic congestion on the IP network is urgently needed.

The best way to limit traffic congestion is to increase network bandwidths. However, increasing network bandwidths is not feasible due to the high operation and maintenance costs.

The most cost-effective way is to use a "guarantee" policy to management traffic congestion. This method is quality of service (QoS). QoS provides end-to-end service guarantee for differentiated services and has played an overwhelmingly important role on the Internet. Without QoS, service quality cannot be guaranteed.

## 1.2 QoS Specifications

QoS provides customized service guarantee for key services based on the following specifications:

- Bandwidth/Throughput
- Delay
- Delay variation (jitter)
- Packet loss rate

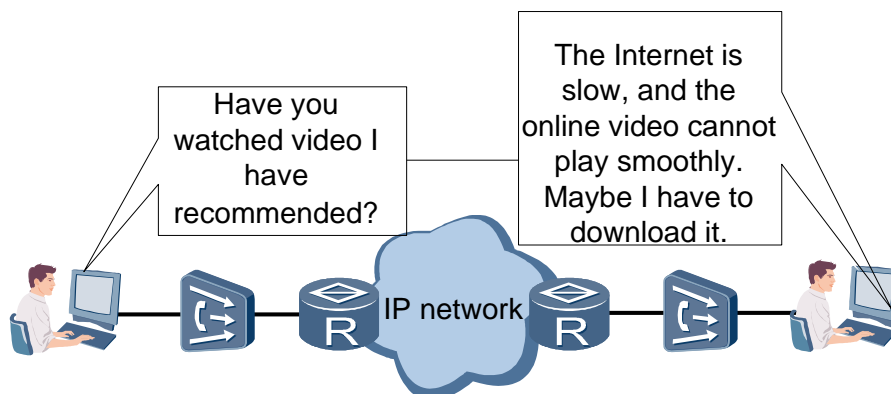
### 1.2.1 Bandwidth/Throughput

Bandwidth, also called throughput, refers to the maximum number of bits allowed to transmit between two ends within 1 second or the average rate at which specific data flows are transmitted between two network nodes. Bandwidth is expressed in bit/s.

The water supply network is used to help you understand bandwidth. The diameter of a water supply pipe measures the capability to carry water. The diameter of the water supply pipe is similar to bandwidth and water is similar to data. A thick pipe indicates higher bandwidth and greater capability to transmit data.

As services become increasingly diversified, Internet citizens expect higher bandwidths so they can not only browse the Internet for news but also experience any number of popular applications. The epoch-making information evolution continually delivers new and attractive applications, such as new-generation multimedia, video transmission, database, and IPTV, all of which demand extremely high bandwidths. Therefore, bandwidth is always the major focus of network planning and provides an important basis for network analysis.

**Figure 1-2** Insufficient bandwidth



 **NOTE**

Two concepts, upstream rate and downstream rate, are relevant to bandwidth. The upstream rate refers to the rate at which users send information to the network, and the downstream rate refers to the rate at which the network sends data to users. For example, the rate at which users upload files to the network through FTP is determined by the upstream rate, and the rate at which users download files is determined by the downstream rate.

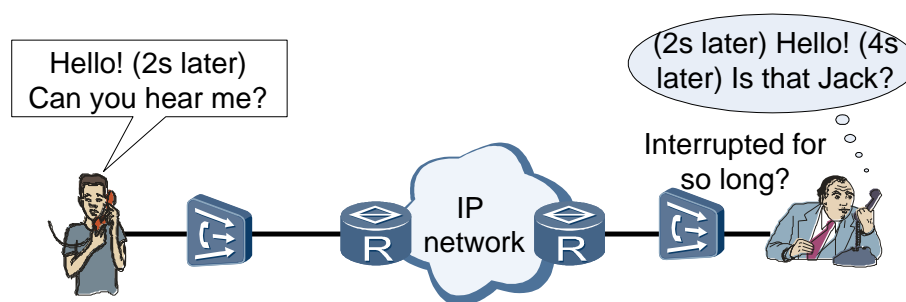
## 1.2.2 Delay

A delay refers to the period of time during which a packet is transmitted from a source to its destination.

Use voice transmission as an example. A delay refers to the period during which words are spoken and then heard. If the delay is too long, voices become unclear or interrupted.

Most users are insensitive to a delay of less than 100 ms. If a delay ranging from 100 ms to 300 ms occurs, the speaker can sense slight pauses in the responder's reply, which can seem annoying to both. If a delay greater than 300 ms occurs, both the speaker and responder sense an obvious 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.

**Figure 1-3** Long delay



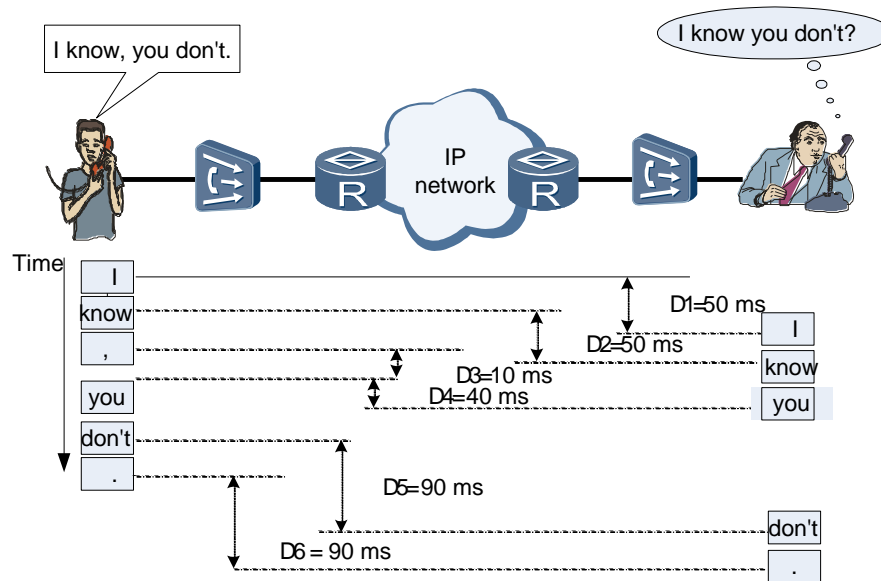
## 1.2.3 Delay Variation (Jitter)

Jitter refers to the difference in delays of packets in the same flow. If the period before a packet that has reached a device is sent by the device differs from one packet to another in a flow, jitters occur, and service quality is affected.

Specific services, especially voice and video services, are zero-tolerant of jitters, because jitter will interrupt voice or video services.



**Figure 1-4** High jitter



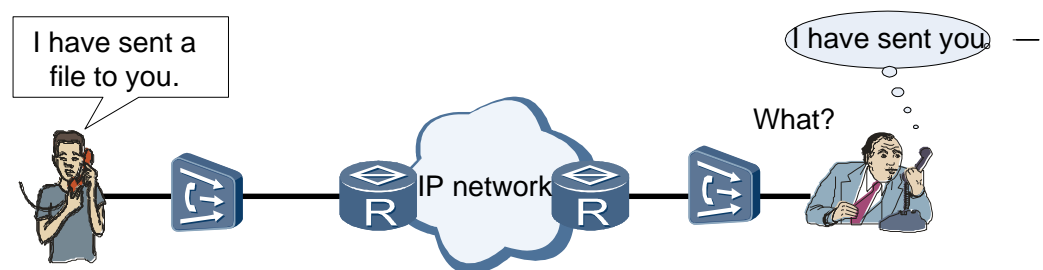
Jitters also affect protocol packet transmissions. Specific protocol packets are transmitted at a fixed interval. If high jitters occur, such protocols flap, adversely affecting quality.

Jitter thrives on networks but service quality will not be affected if jitters do not exceed a specific tolerance. Buffers can alleviate excess jitters but prolong delays.

## 1.2.4 Packet Loss Rate

Packet loss occurs when one or more packets traveling across a network fail to reach their destination. Slight packet loss does not affect services. For example, users are unaware of the loss of a bit or a packet in voice transmission. If a bit or a packet is lost in video transmission, the image on the screen becomes momentarily garbled but the image recovers very quickly. Even if TCP is used to transmit data, slight packet loss is not a problem because TCP instantly retransmits the packets that have been lost. If severe packet loss does occur, packet transmission efficiency is affected. The packet loss rate indicates the severity of service interruptions on networks and concerns users.

**Figure 1-5** High packet loss rate



## 1.3 Common QoS Specifications

Internet users have different requirements for the bandwidth, delay, jitter, and packet loss rate for different services on the IP network. Table 1-1 and Table 1-2 list QoS specifications for different services. Table 1-3 list QoS specifications defined by the Metro Ethernet Forum (MEF), including availability, delay, jitter, loss, and restoration time.

**Table 1-1** QoS specifications for common services

Enterprise Service Type	Bandwidth/Throughput	Delay	Jitter	Packet Loss Rate
Videoconferencing	High	Very low	Very low	Low and predictable
E-commerce	Medium	Low	Low	Low
Streaming media	High	Low	Low	Low and predictable
Emails and file transfer	Low	Not important	Not important	Not important
HTML web page browsing	Not specific	Medium	Medium	Not important
FTP client/server	Medium	Low	Low	Low

**Table 1-2** Reference values of QoS specifications for common services

Enterprise Service Type	Delay	Jitter	Packet Loss Rate
Videoconferencing	≤50 ms	≤10 ms	≤0.1%
E-commerce	≤200 ms	≤100 ms	TCP guarantee
Streaming media	≤1s	≤200 ms	≤0.1%
Emails and file transfer	N/A	N/A	TCP guarantee
HTML web page browsing	N/A	N/A	NA
FTP client/server	N/A	N/A	TCP guarantee

**Table 1-3** QoS specifications defined by the MEF

Service Class	Service Characteristics	Service Performance
Premium	Real-time IP telephony or IP video applications	Availability > 99.99% Delay < 40 ms Jitter < 1 ms Loss < 0.1% Restoration time: 50 ms
Silver	Burst mission-critical data applications requiring low loss and delay such as storage	Availability > 99.99% Delay < 50 ms Jitter: N/A Loss < 0.1% Restoration time: 200 ms
Bronze	Burst data applications requiring bandwidth assurances	Availability > 99.90% Delay < 500 ms Jitter: N/A Loss: N/A Restoration time: 2s
Standard	Best effort service	Availability > 97.00% Delay: N/A Jitter: N/A Loss: N/A Restoration time: 5s

# 2 Technology Description

---

## 2.1 QoS Service Models

Network applications require end-to-end communication. Traffic may traverse multiple switches on one network or even multiple networks before reaching the destination host. Therefore, to provide end-to-end QoS guarantee, an overall network deployment is required. Service models are used to provide an end-to-end QoS guarantee based on specific requirements.

QoS provides the following types of service models:

- Best-Effort
- Integrated service (IntServ)
- Differentiated service (DiffServ)

### 2.1.1 Best-Effort

Best-Effort is the default service model on the Internet and applies to various network applications, such as FTP and email. It is the simplest service model. Without network notification, an application can send any number of packets at any time. The network then makes its best attempt to send the packets but does not provide any guarantee for performance such as delay and reliability.

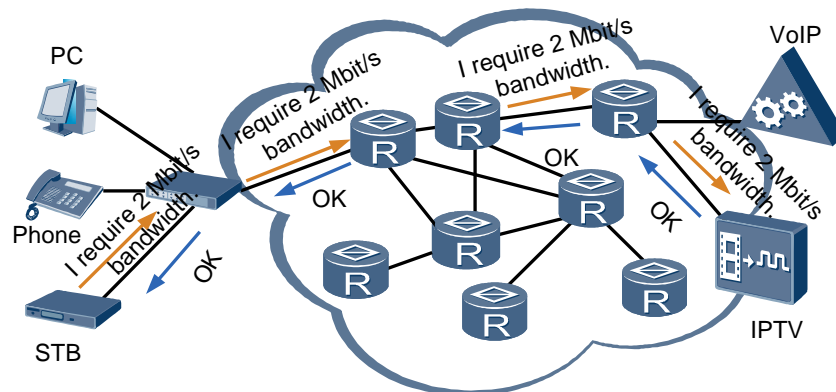
The Best-Effort model applies to services that do not require low delay and high reliability.

### 2.1.2 IntServ

Before sending a packet, IntServ uses signaling to apply for a specific level of service from the network. The application first notifies the network of its traffic parameters and specific service qualities, such as bandwidth and delay. After receiving a confirmation that sufficient resources have been reserved, the application sends the packets. The network maintains a state for each packet flow and executes QoS behaviors based on this state to fulfill the promise made to the application. The packets must be controlled within the range described by the traffic parameters.

IntServ uses the Resource Reservation Protocol (RSVP) as signaling, which is similar to Multiprotocol Label Switching Traffic Engineering (MPLS TE). RSVP 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 and is periodically updated using RSVP messages. Each network element checks whether sufficient resources can be reserved based on these RSVP messages. The path is available only when all involved network elements can provide sufficient resources.

**Figure 2-1** IntServ model



The IntServ model provides end-to-end guarantee, but has the following limitations:

- MPLS TE is feasible because it is deployed on the core network and the network scale is controllable. The IntServ model involves end-to-end services at the core, aggregation, and access layers, and more network elements. The complex network limits its development.
- IntServ must be supported by all network nodes. Core, aggregation, and access devices have different performances, and some of them may not support the IntServ model.

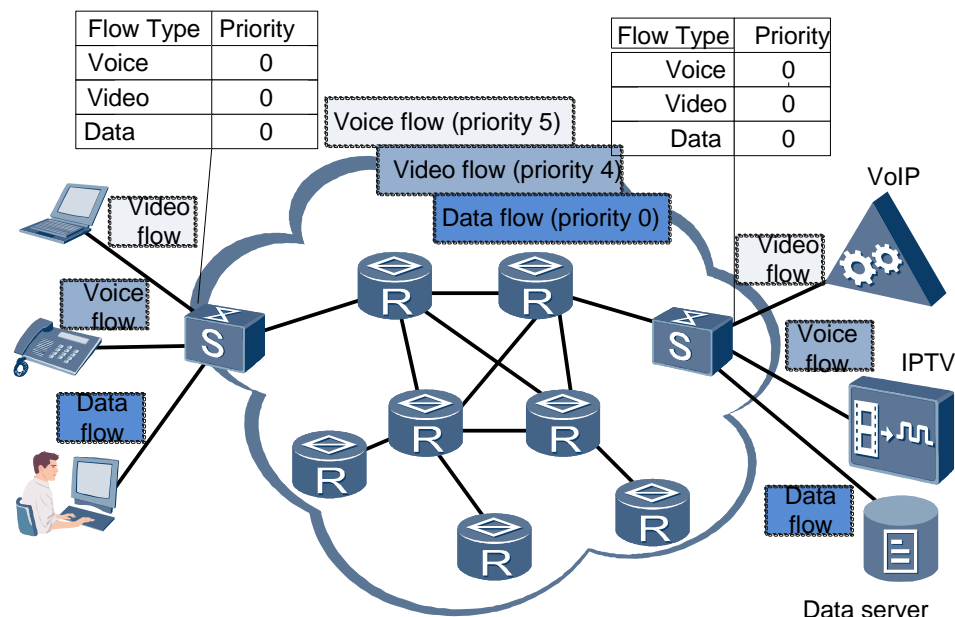
The IntServ model cannot be widely applied to the Internet backbone network.

### 2.1.3 DiffServ

DiffServ classifies packets on the network into multiple classes for differentiated processing. When traffic congestion occurs, classes with a higher priority are given preference. This function allows packets to be differentiated and to have different packet loss rates, delays, and jitters. Packets of the same class are aggregated and sent as a whole to ensure the same delay, jitter, and packet loss rate.

In the DiffServ model, edge nodes classify and aggregate traffic. Edge nodes classify packets based on a combination of fields, such as the source and destination addresses of packets, precedence in the ToS field, and protocol type. Edge nodes also re-mark packets with different priorities, which can be identified by other nodes for resource allocation and traffic control. Therefore, DiffServ is a flow-based QoS model.

Figure 2-2 DiffServ model



Different from IntServ, DiffServ requires no signaling. In the DiffServ model, an application does not need to apply for network resources before transmitting packets. Instead, the application notifies the network nodes of its QoS requirements by setting QoS parameters in packets. The network does not need to maintain a state for each data flow but provides differentiated services based on the QoS parameters of each data flow.

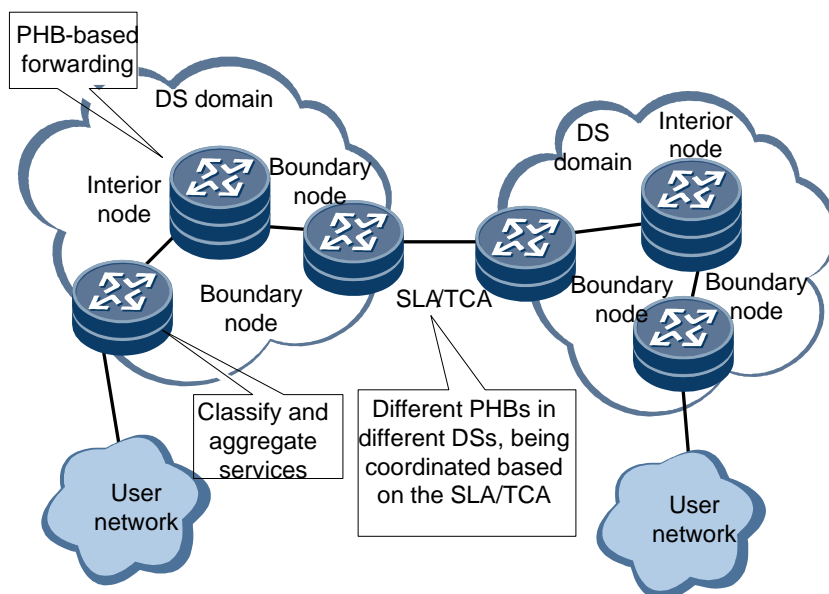
DiffServ classifies incoming packets on the network edge and manages packets of the same class as a whole to ensure the same transmission rate, delay, and jitter. DiffServ processes flows of each type separately.

Network edge nodes mark packets with a specific service class in packet headers, and then apply traffic management policies to the packets based on the service class. Interior nodes perform specific behaviors for packets based on packet information.

DiffServ takes full advantage of network flexibility and extensibility of the IP network and transforms information in packets into per-hop behaviors, greatly reducing signaling operations. Therefore, DiffServ not only adapts to Internet service provider (ISP) networks but also accelerates IP QoS applications on live networks. It is the mainstream model on networks.

## Entities in the DiffServ Model

Figure 2-3 Entities in the DiffServ model



- DS node: a network node that implements the DiffServ function. All network elements in Figure 2-3 are DS nodes.
- DS domain: a set of contiguous DS nodes that adopt the same service policy and per-hop behavior (PHB). One DS domain covers one or more networks under the same administration. For example, a DS domain can be an ISP's network or an organization's intranet. For an introduction to PHB, see the next section.



### NOTE

A PHB describes the externally observable forwarding treatment applied to a DS node.

- DS boundary node: connects to another DS domain or a non-DS-aware domain. The DS boundary node classifies and manages incoming traffic.
- DS interior node: connects to DS boundary nodes and other interior nodes in one DS domain. DS interior nodes implement simple traffic classification based on DSCP values, and manage traffic.
- SLA/TCA: The SLA refers to the services that the ISP promises to provide for individual users, enterprise users, or adjacent ISPs that need intercommunication. The SLA covers multiple dimensions, including the accounting protocol. The service level specification (SLS) provides technique description for the SLA. The SLS focuses on the traffic control specification (TCS) and provides detailed performance parameters, such as the committed information rate (CIR), peak information rate (PIR), committed burst size (CBS), and peak burst size (PBS).
- DS region: consists of one or more adjacent DS domains. Different DS domains in one DS region may use different PHBs to provide differentiated services. The SLA and traffic conditioning agreement (TCA) are used to allow for differences between PHBs in different DS domains. The SLA or TCA specifies how to maintain consistent processing of the data flow from one DS domain to another.

## PHB

An action taken for packets on each DS node is called PHB. PHB is a description of the externally observable forwarding treatment applied to a DS node. You can define PHB based on priorities or QoS specifications such as the delay, jitter, and packet loss ratio. The PHB defines some forwarding behaviors but does not specify the implementation mode.


Currently, the IETF defines four types of PHBs: Class Selector (CS), Expedited Forwarding (EF), Assured Forwarding (AF), and best-effort (BE). BE is the default PHB.

RFC 2597 classifies AF into four classes: AF1 to AF4. RFC 2474 classifies CS into CS6 and CS7. There are eight types of PHBs. Each PHB corresponds to a Class of Service (CoS) values. Different CoS values determine different congestion management policies. In addition, each PHB is assigned three drop priorities, also called colors (green, yellow, and red). Different drop priorities determine congestion avoidance policies of different flows.


For details about CoS values and colors, see Priority Mapping. For details about congestion management and congestion avoidance, see section 2.4 "Congestion Management and Congestion Avoidance."

Table 2-1 describes standard PHBs and their usage.

**Table 2-1** Standard PHBs and usage

PHB	Description	Sub-PHB	Usage
CS (RFC 2474)	The CS PHB indicates the same service class as the IP precedence value. The CS PHB is of the highest priority among standard PHBs.	CS7	CS6 and CS7 PHBs are used for protocol packets by default, such as STP, LLDP, and LACP packets. If these packets are not forwarded, protocol services are interrupted.
		CS6	
EF (RFC 2598)	The EF PHB defines that the rate at which packets are sent from any DS node must be higher than or equal to the specified rate. The EF PHB cannot be re-marked in the DS domain but can be re-marked on the edge nodes.  The EF PHB applies to real-time services that require a short delay, low jitter, and low packet loss rate, such as video, voice, and video conferencing.	-	EF PHB is used for voice services. Voice services require a short delay, low jitter, and low packet loss rate, and are second only to protocol packets in terms of importance.   <b>NOTE</b> The bandwidth dedicated to EF PHB must be restricted so that other services can use the bandwidth.
AF (RFC 2597)	The AF PHB defines that traffic exceeding	AF4	AF4 PHB is used for signaling of voice services.



PHB	Description	Sub-PHB	Usage
	<p>the specified bandwidth (as agreed to by users and an ISP) can be forwarded. The traffic that does not exceed the bandwidth specification is forwarded as required, and the traffic that exceeds the bandwidth specification is forwarded at a lower priority.</p> <p>The AF PHB applies to services that require a short delay, low packet loss rate, and high reliability, such as e-commerce and VPN services.</p>		<p> <b>NOTE</b></p> <p>Signaling is used for call control, during which a seconds-long delay is tolerable, but no delay is allowed during a conversation. Therefore, the processing priority of voice services is higher than that of signaling.</p>
		AF3	AF3 PHB is used for Telnet and FTP services. The services require medium bandwidth and reliable transmission, but are sensitive to the delay and jitter.
		AF2	AF2 PHB is used for live programs of IPTV and ensures smooth transmission of online video services. Live programs are real-time services, requiring continuous bandwidth and a large throughput guarantee. They allow less packet loss.
		AF1	AF1 PHB is used for common data services such as emails. Common data services require only zero packet loss, and do not require high real-time performance and jitter.
BE (RFC 2474)	The BE PHB focuses only on whether packets can reach the destination, regardless of the transmission performance. Any switch must support BE PHB.	-	BE PHB applies to best-effort services on the Internet, such as HTTP web page browsing services.

## 2.1.4 Comparison Between DiffServ and IntServ Models

**Table 2-2** Comparison between DiffServ and IntServ models

Item	DiffServ	IntServ
End-to-end QoS guarantee	Implements end-to-end QoS guarantee by connecting multiple DS domains.	Directly implements end-to-end QoS guarantee.
Network scale	Applies to various networks, and applies to large-scale networks using multiple DS domains.	Is inapplicable to large-scale networks.

Item	DiffServ	IntServ
Network cost	Has no extra cost because DiffServ model notifies other devices of packet priorities using packet precedence fields.	Has extra cost because the IntServ model uses RSVP to notify other devices and periodically update network resources.
Network element cost	Has low cost because resources do not need to be reserved for network elements.	Has high cost because resources need to be reserved for network elements.

## 2.1.5 Components in the DiffServ Model

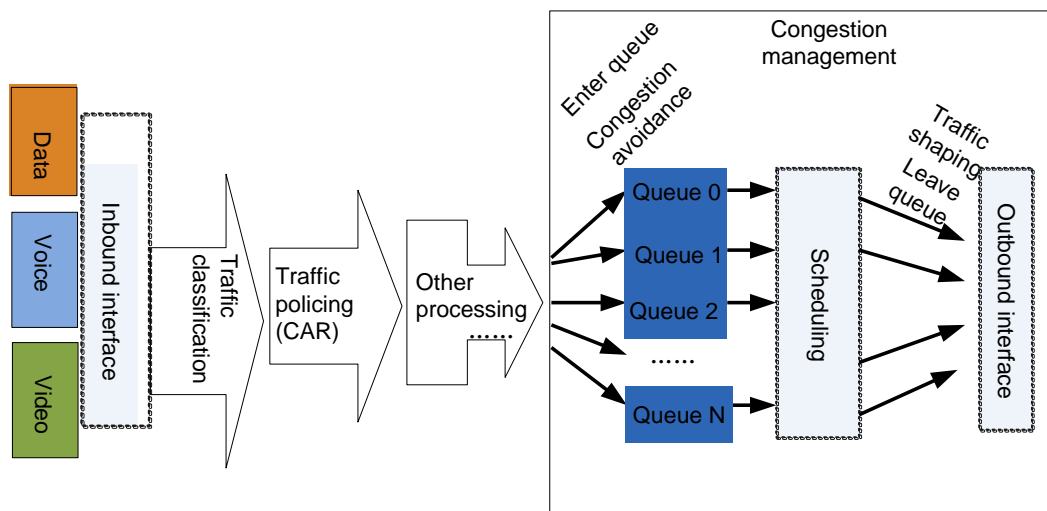
The DiffServ model consists of four QoS components:

- **Traffic classification and marking:** Traffic classification classifies packets while keeping the packets unchanged. Traffic marking sets different priorities for packets and therefore changes the packets.
- **Traffic policing and shaping:** Limit the traffic rate. When traffic exceeds the specified rate, traffic policing drops excess traffic, and traffic shaping buffers excess traffic.
- **Congestion management and avoidance:** Congestion management buffers packets in queues when traffic congestion occurs and determines the forwarding order based on a specific scheduling algorithm. Congestion avoidance monitors network resources. When network congestion aggravates, the device drops packets to regulate traffic so that the network is not overloaded.
- **Port mirroring and traffic mirroring:** Mirroring copies packets on a specified interface to the mirroring destination interface that is connected to a data monitoring device. Then you can use the data monitoring device to analyze the packets copied to the destination interface, and monitor the network and troubleshoot faults.

Traffic classification and marking are the basis for implementing differentiated services. Traffic policing, traffic shaping, congestion management, and congestion avoidance control network traffic and allocated resources.

Packets are processed by the components in sequence, as shown in Figure 2-4.

**Figure 2-4** Processing of QoS components



The four QoS components are implemented at different locations on a network according to the DiffServ model and service development. Traffic classification, traffic marking, and traffic policing are performed in the inbound direction on an access interface, traffic shaping is performed in the outbound direction on an access interface, and congestion management and congestion avoidance are performed in the outbound direction on a network-side interface. If services with different CoS values are transmitted on an access interface, queue scheduling and a packet drop policy must be configured in the outbound direction on the access interface.

## 2.2 Traffic Classification and Marking

Traffic classification technology allows a device to classify packets that enter a DiffServ domain so that other applications or devices learn about the packet service type and apply any appropriate action upon the packets.

Packets can be classified based on QoS priorities, or packet information such as the source IP address, destination IP address, MAC address, IP protocol, and port number, or specifications in an SLA.

After packets are classified on the DiffServ domain edge, internal nodes provide differentiated services for the packets that are classified. A downstream node can resume the classification result calculated on an upstream node or perform another traffic classification based on its own criteria.

Traffic classification is classified into simple traffic classification and complex traffic classification. For details, see section 2.2.1 "Simple Traffic Classification" and section 2.2.2 "Complex Traffic Classification."

### 2.2.1 Simple Traffic Classification

Simple traffic classification classifies packets based on simple rules, for example, 802.1p priorities in VLAN packets, ToS values in IP packets, TC values in IPv6 packets, EXP values in MPLS packets, to identify traffic with different priorities or CoS values and implement mapping between external and internal priorities.

Simple traffic classification trusts priorities in upstream packets on an interface and performs priority mapping. That is, simple traffic classification maps QoS priorities in upstream packets to CoS values and colors, and maps CoS values and colors in downstream packets to QoS priorities.

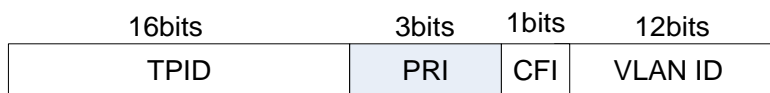
Simple traffic classification is deployed on DS interior nodes.

## QoS Priority Fields

DiffServ provides differentiated services for packets that carry different QoS information in specific fields. The fields are described as follows:

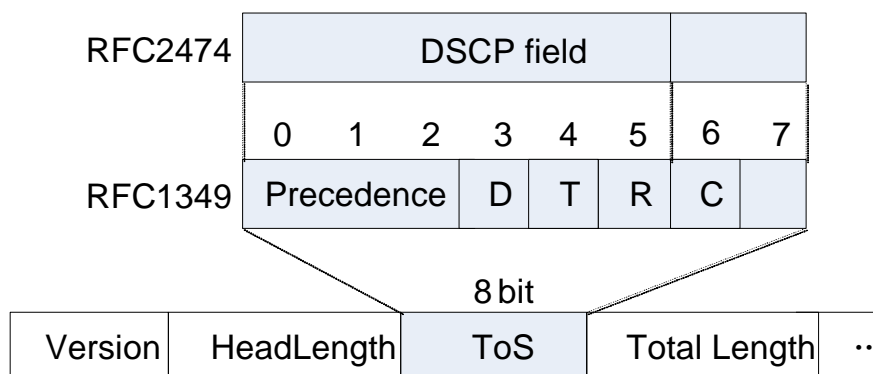
- 802.1p priority  
 VLAN packets are classified based on the 802.1p priority (PRI) in the packets. The PRI field in a VLAN packet header identifies the QoS requirement. The PRI field is 3 bits long and indicates precedence. The value ranges from 0 to 7 with a larger value reflecting a higher precedence.

**Figure 2-5** 802.1p priority in a VLAN packet



- ToS field in an IP packet  
 In an IPv4 packet header, the three leftmost bits (IP precedence) in the ToS field or the six leftmost bits (DSCP field) in the ToS field are used to identify a QoS priority. The IP precedence classifies packets into a maximum of eight classes, and the DSCP field classifies packets into a maximum of 64 classes.

**Figure 2-6** ToS field in an IPv4 packet header



RFC 1349 defines bits in the ToS field as follows:

- Bits 0 to 2 refer to the precedence. The value ranges from 0 to 7 with a larger value reflecting a higher precedence. The ToS field in IP packets is similar in function to the 802.1p priority in VLAN packets.
- The D bit refers to the delay. The value 0 indicates no specific requirement for the delay and the value 1 indicates that the network is required to minimize the delay.

- The T bit refers to the throughput. The value 0 indicates no specific requirement for the throughput and the value 1 indicates that the network is required to maximize the throughput.
- The R bit refers to reliability. The value 0 indicates no specific requirement for reliability and the value 1 indicates that the network demands high reliability.
- The C bit refers to the monetary cost. The value 0 indicates no specific requirement for the monetary cost and the value 1 indicates that the network is required to minimize the monetary cost.
- Bits 6 and 7 are reserved.

RFC 2474 defines bits 0 to 6 as the DSCP field, and the three leftmost bits indicate the class selector code point (CSCP) value, which identifies a class of DSCP. The DSCP value ranges from 0 to 7 with a larger value reflecting a higher precedence. The DSCP value in IP packets is similar in function to the 802.1p priority in VLAN packets. The three rightmost bits are seldom used and are not mentioned here.

- EXP field in an MPLS packet header

Multiprotocol Label Switching (MPLS) packets are classified based on the EXP field value. The EXP field is 3 bits long and indicates precedence. The value ranges from 0 to 7 with a larger value reflecting a higher precedence. The EXP field in MPLS packets is similar in function to the ToS field or DSCP field in IP packets.

**Figure 2-7** EXP field in an MPLS packet header

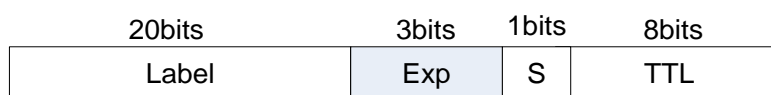


Table 2-3 describes the mapping from the IP precedence, EXP, and 802.1p values to the DSCP value.

**Table 2-3** Mapping from the IP precedence, EXP, and 802.1p values to the DSCP value

IP Precedence	MPLS EXP Value	802.1p Priority	DSCP Value
0	0	0	0
1	1	1	8
2	2	2	16
3	3	3	24
4	4	4	32
5	5	5	40
6	6	6	48
7	7	7	56

Table 2-4 describes the mapping from the DSCP value to 802.1p, EXP, and IP precedence values.

**Table 2-4** Mapping from the DSCP value to 802.1p, EXP, and IP precedence values

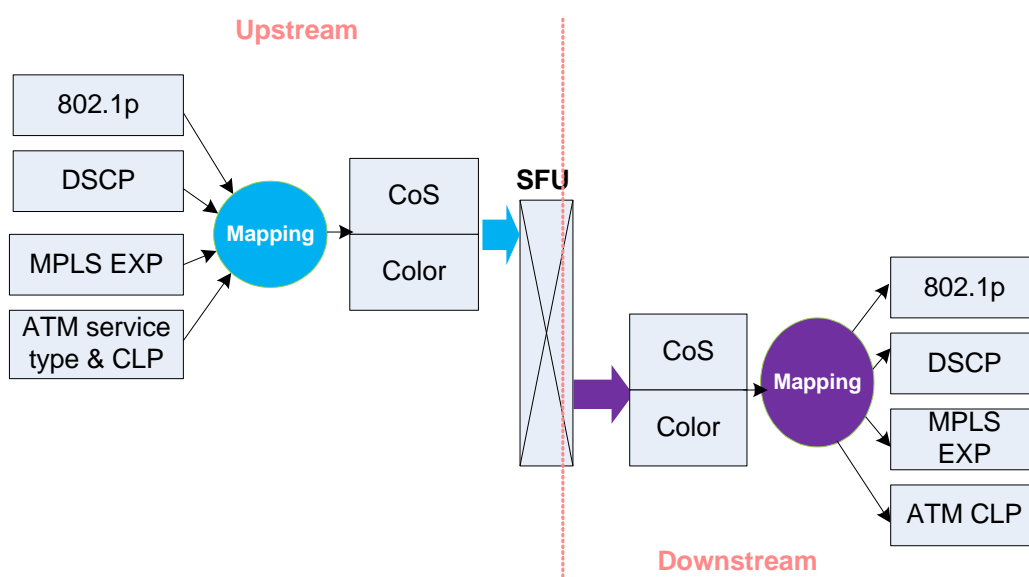
DSCP Value	IP Precedence	MPLS EXP Value	802.1p Priority
0-7	0	0	0
8-15	1	1	1
16-23	2	2	2
24-31	3	3	3
32-39	4	4	4
40-47	5	5	5
48-55	6	6	6
56-63	7	7	7

## Priority Mapping

The priority field in a packet varies with network type. For example, a packet carries the 802.1p field on a VLAN, the DSCP field on an IP network, and the EXP field on an MPLS network. To provide differentiated services for different packets, the switch maps the QoS priority of incoming packets to the CoS value (also called scheduling precedence) and drop precedence (also called color), and then performs congestion management based on the CoS value and congestion avoidance based on the color. Before forwarding packets out, the device maps the CoS value and color back to the QoS priority so that other devices can process the packets based on the QoS priority.

A device maps the QoS priority to the CoS value and color for incoming packets and maps the CoS value and color back to the QoS priority for outgoing packets, as shown in Figure 2-8.

**Figure 2-8** Priority mapping



- CoS

The CoS refers to the internal service class of packets. Eight CoS values are available: Class Selector 7 (CS7), CS6, Expedited Forwarding (EF), Assured Forwarding 4 (AF4), AF3, AF2, AF1, and Best Effort (BE). CoS determines the type of queues to which packets belong.

The priority of queues with different CoS values depends on the scheduling algorithms used:

- If queues with eight CoS values all use priority queuing (PQ), the priority of queues is: CS7 > CS6 > EF > AF4 > AF3 > AF2 > AF1 > BE.
- If the BE queue uses PQ scheduling (rarely on live networks) and all the other seven queues use weighted fair queuing (WFQ), the BE queue has the highest priority.
- If all the eight queues use WFQ scheduling, the priority is irrelevant to WFQ scheduling.



**NOTE**

For details about queue scheduling, see Queue Scheduling.

- Color

Color, also referred to as drop precedence of packets on a device, determines the order in which packets in a queue are dropped when traffic congestion occurs. As defined by the Institute of Electrical and Electronics Engineers (IEEE), the color of a packet can be green, yellow, or red.

Drop precedence is determined by the configured parameters. For example, if a maximum of 50% of the buffer size is configured for packets colored Green, whereas a maximum of 100% of the buffer size is configured for packets colored Red, packets colored Green have a higher drop precedence than packets colored Red. Packet priorities depend on the QoS configuration.

- Trusting the priority of received packets

As described in section 2.2 Traffic Classification and Marking, after packets are classified on the DiffServ domain edge, internal nodes provide differentiated services for the packets that are classified. A downstream node can resume the classification result calculated on an upstream node or perform another traffic classification based on its own criteria. If the downstream node resumes the classification result calculated on an upstream node, the downstream node trusts the QoS priority (DSCP, IP precedence, 802.1p, or EXP) of packets received on the interface connecting to the upstream node. This is called the mode of trusting the interface.

The switch trusts the following priorities:

- 802.1p priority

The switch classifies packets based on 802.1p priorities and searches the mapping table of 802.1p priorities and CoS values. The switch classifies untagged packets based on the default 802.1p priority of an interface. Then the switch maps CoS values to 802.1p priorities and provides differentiated services.

- DSCP priority

The switch classifies packets based on DSCP priorities and searches the mapping table of DSCP priorities and CoS values. Then the switch maps CoS values to DSCP priorities and provides differentiated services.

The switch implements priority mapping according to the priority mapping table. In the DiffServ model, different DS domains allow different PHB mappings, so the device needs to allow an administrator to define DS domains and set different mappings in different DS domains.

 **NOTE**

Huawei switch allows an administrator to define DS domains. In addition, the system defines the default DS domain **default**. You can modify mappings in the DS domain **default**, but cannot delete it.

Huawei switch provides the following priority mapping modes:

- DiffServ domain: S9700, S7700, S5700HI, S5710EI, S5710HI, and S6700
  - Priority mapping table: S5700SI, S5700EI, S5700LI, S5700S-LI, and S2750
- If the mapping table is used, priority mapping implements mapping between packet priorities and PHBs, but cannot implement mapping between packet priorities and colors. All packets are green by default.

 **NOTE**

When the DiffServ domain is used, run the **display diffserv domain name default** command to view the default mappings. When the mapping table is used, run the **display qos map-table** command to view the default mappings.

The following tables show the mappings in the DiffServ domain:

- Table 2-5 describes the mappings from 802.1p priorities to PHBs and colors.
- Table 2-6 describes the mappings from DSCP priorities to PHBs and colors.
- Table 2-7 describes the mappings from precedences to PHBs and colors.
- Table 2-8 describes the mappings from EXP priorities in MPLS packets to PHBs and colors.

**Table 2-5** Mappings from 802.1p priorities to PHBs and colors

802.1p Priority	PHB	Color
0	BE	Green
1	AF1	Green
2	AF2	Green
3	AF3	Green
4	AF4	Green
5	EF	Green
6	CS6	Green
7	CS7	Green

**Table 2-6** Mappings from DSCP priorities to PHBs and colors

DSCP	PHB	Color	DSCP	PHB	Color
0-7	BE	Green	28	AF3	Yellow
8	AF1		29	BE	Green
9	BE		30	AF3	Red
10	AF1		31	BE	Green



DSCP	PHB	Color	DSCP	PHB	Color
11	BE		32	AF4	
12	AF1	Yellow	33	BE	
13	BE	Green	34	AF4	
14	AF1	Red	35	BE	
15	BE	Green	36	AF4	Yellow
16	AF2		37	BE	Green
17	BE		38	AF4	Red
18	AF2		39	BE	Green
19	BE		40	EF	
20	AF2	Yellow	41-45	BE	
21	BE	Green	46	EF	
22	AF2	Red	47	BE	
23	BE	Green	48	CS6	
24	AF3		49-55	BE	
25	BE		56	CS7	
26	AF3		57-63	BE	
27	BE				

**Table 2-7** Mappings from precedences to PHBs and colors

IP Precedence	PHB	Color
0	BE	Green
1	AF1	Green
2	AF2	Green
3	AF3	Green
4	AF4	Green
5	EF	Green
6	CS6	Green
7	CS7	Green

**Table 2-8** Mappings from EXP priorities in MPLS packets to PHBs and colors

Exp	PHB	Color
0	BE	Green
1	AF1	Green
2	AF2	Green
3	AF3	Green
4	AF4	Green
5	EF	Green
6	CS6	Green
7	CS7	Green

## 2.2.2 Complex Traffic Classification

As networks rapidly develop, services on the Internet become increasingly diversified. Various services share limited network resources, so simple traffic classification can hardly meet requirements. Network devices must possess a high degree of awareness for services and support in-depth packet analysis to parse any packet field at any layer. Complex traffic classification meets the requirement to a certain degree.

Complex traffic classification classifies packets in fine-grained manner based on rules such as the source MAC address, destination MAC address, inner and outer tags, source IP address, source port number, destination IP address, and destination port number. Complex traffic classification is deployed on edge nodes.

Huawei switches provide various traffic classifiers and traffic behaviors. Traffic classifiers can be associated with traffic behaviors to form a traffic policy. To implement complex traffic classification, apply the traffic policy to an interface, a VLAN, or the system. The traffic policy based on complex traffic classification is also called class-based QoS.

A traffic policy based on complex traffic classification is configured using a profile, which allows batch configuration or modification.

A QoS profile defines the following items:

- Traffic classifier: defines a service type. The if-match clauses are used to set traffic classification rules.
- Traffic behavior: defines actions for classified traffic.
- Traffic policy: associates traffic classifiers with traffic behaviors. After a traffic policy is configured, apply it to an interface, a VLAN, or the system.

### Traffic Classifier

A traffic classifier identifies packets of a certain type by using matching rules so that differentiated services can be provided for these packets. A traffic classifier can contain matching rules that do not conflict.

If a traffic classifier has multiple matching rules, the AND/OR logic relationships between rules are described as follows:

- OR: Packets that match any of the if-match clauses configured in a traffic classifier match this traffic classifier.
- AND: If a traffic classifier contains ACL rules, packets match the traffic classifier only when the packets match one ACL rule and all the non-ACL rules. If a traffic classifier does not contain ACL rules, packets match the traffic classifier only when the packets match all the non-ACL rules.

On the Huawei switch, the default logic is OR and a traffic classifier can define matching rules based on the following items:

- Outer VLAN ID
- Inner and outer VLAN IDs in QinQ packets
- 802.1p priority in VLAN packets
- Inner 8021p priority of QinQ packets
- Outer VLAN ID or inner and outer VLAN IDs of QinQ packets
- Double tags of QinQ packets
- Destination MAC address
- Source MAC address
- Protocol type field encapsulated in the Ethernet frame header
- All packets
- DSCP priority in IP packets
- IP precedence in IP packets
- Layer 3 protocol type
- Inbound interface
- Outbound interface
- ACL rule
- Matching order of ACL rules

After a traffic classifier is configured, the system matches packets against an ACL as follows:

- Checks whether the ACL exists (traffic classifiers can reference non-existent ACLs).
- Matches packets against rules in the order in which the rules are displayed. When packets match one rule, the system stops the match operation.

An ACL can contain multiple rules and each rule specifies different packet ranges. ACL rules are matched according to the following matching modes:

- Config: ACL rules are matched according to the sequence in which they were configured.
- Auto: ACL rules are matched based on the depth-first principle.

## Traffic Behavior

A traffic behavior is the action to be taken for packets matching a traffic classifier and is the prerequisite to configuring a traffic policy. Table 2-9 describes traffic behaviors that can be implemented individually or jointly for classified packets on a Huawei switch.

**Table 2-9** Traffic behaviors

Traffic Behavior	Description	Usage
Marking	Sets or modifies the packet priority, such as 802.1p priority in VLAN packets and DSCP/internal priority in IP packets, to relay QoS information to the next device. Modifying packet priorities is also called re-marking.	Voice services, video services, and data services have QoS requirements in descending order of priority.
Traffic policing	Limits network traffic and controls the usage of network resources by monitoring the traffic rate on a network. According to the configured traffic policing action, the device performs traffic policing for packets matching traffic classification rules, and discards excess packets or re-marks colors or CoS values of the excess packets.	On an enterprise network, an aggregation switch often connects to multiple access switches. You can configure traffic policing on the inbound interfaces of the aggregation switch to limit traffic.
Traffic statistics	According to the configured traffic statistics action, the device collects statistics on packets matching traffic classification rules. The statistics on forwarded and discarded packets after a traffic policy is applied help you check whether the traffic policy is correctly applied and locate faults.	The enterprise NMS often provides this function, and monitors traffic based on services or users.
Packet filtering	Is the basic traffic control method. The device determines whether to drop or forward packets based on traffic classification results.	It has the following functions: Limits resources accessed by some users. Filters out packets matching blacklist entries to protect the enterprise network.
Redirection	Determines the packet forwarding path based on traffic classification results. According to the configured redirection action, the device redirects the packets matching traffic classification rules to the CPU, specified next hop address, or specified interface.  The traffic policy that contains the redirection action can only be applied to the inbound direction of the system, an interface, or a VLAN.	If there is a backup link in the outbound direction, configure redirection to the next hop address so that the device redirects high-priority services such as voice and video services to a higher-bandwidth or more stable link.
Flow mirroring	Copies the packets of an observed flow and then sends the copy to a specified observing interface.	Using this action, you can collect incoming and outgoing packets on an interface for fault analysis.

## Traffic Policy

You can apply a traffic policy bound to traffic behaviors and traffic classifiers to the system, an interface, or a VLAN so that the device can provide differentiated services.

When creating a traffic policy on a Huawei switch, you can specify the matching order of traffic classifiers in the traffic policy. The matching order includes the auto order and config order.

**Auto order:** The matching order depends on priorities of traffic classifiers. The traffic classifiers based on the following information are in descending order of priority: Layer 2 and Layer 3 information, Layer 2 information, and Layer 3 information. A traffic classifier with the highest priority is matched first.

If the config order is used, traffic classifiers are matched in the sequence in which traffic classifiers were bound to the traffic policy. A traffic classifier that was bound to the traffic policy first is matched first.

- Applying a traffic policy globally

Only one traffic policy can be applied to the system or slot (stack on chassis or box switches) in one direction. A traffic policy cannot be applied to the same direction in the system and slot simultaneously.

In a stack composed of box switches, a traffic policy that is applied to the system takes effect on all the interfaces and VLANs of all the member switches in the stack. The system then performs traffic policing for all the incoming or outgoing packets that match traffic classification rules on all the member switches. A traffic policy that is applied to a specified LPU takes effect on all the interfaces and VLANs of the member switch with the specified stack ID. The system then performs traffic policing for all the incoming or outgoing packets that match traffic classification rules on this member switch.

On a box switch in a non-stack scenario, a traffic policy that is applied to the system takes effect on all the interfaces and VLANs of the box switch. The system then performs traffic policing for all the incoming or outgoing packets that match traffic classification rules on the box switch. Traffic policies applied to the LPU and system have the same functions.

When a traffic policy is applied globally on a chassis switch, the chassis switch performs traffic policing for all the incoming or outgoing packets that match traffic classification rules.

When a traffic policy is applied to an LPU on a chassis switch, the chassis switch performs traffic policing for all the incoming or outgoing packets that match traffic classification rules.

- Applying a traffic policy to an interface

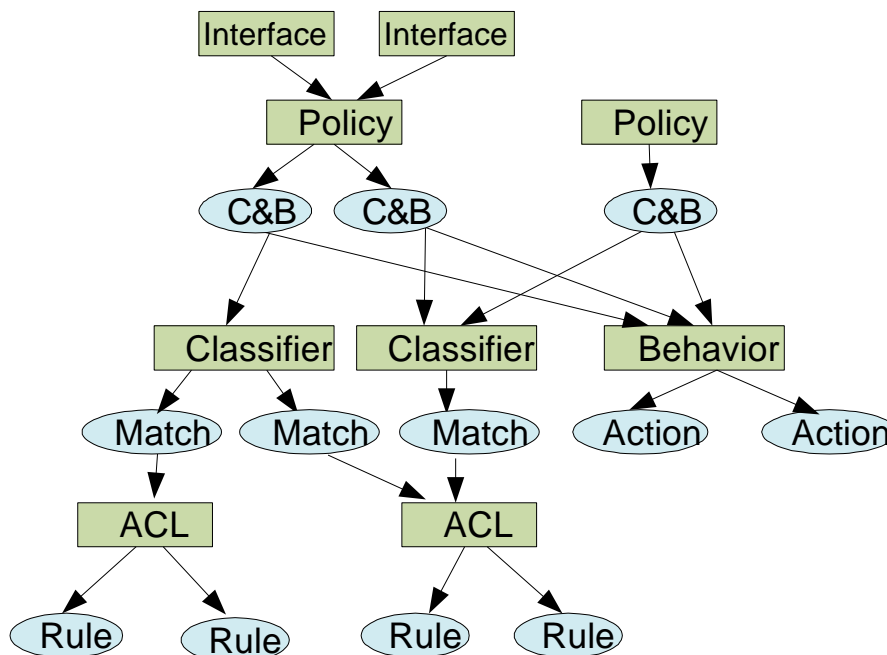
On a Huawei switch, a traffic policy can be applied to only one direction on an interface, but a traffic policy can be applied to different directions on different interfaces. After a traffic policy is applied to an interface, the system performs traffic policing for all the incoming or outgoing packets that match traffic classification rules on the interface.

- Applying a traffic policy to a VLAN

On a Huawei switch, a traffic policy can be applied to only one direction in a VLAN.

Figure 2-9 shows relationships between an interface, traffic policy, traffic behavior, traffic classifier, and ACL.

**Figure 2-9** Relationships between an interface, traffic policy, traffic behavior, traffic classifier, and ACL



The relationships are as follows:

- A traffic policy can be applied to different interfaces.
- One or more pairs of traffic classifiers and traffic behaviors can be configured in a traffic policy. A pair of a traffic classifier and a traffic behavior can be configured in different traffic policies.
- One or more if-match clauses can be configured in a traffic classifier, and each if-match clause can specify an ACL. An ACL can be defined in different traffic classifiers and contains one or more rules.
- One or more actions can be configured in a traffic behavior.

Example: Configure two pairs of traffic classifiers and traffic behaviors in a traffic policy.

```

acl 3001
rule permit ip source 1.1.1.1 0
rule permit ip source 1.1.10.1 0
acl 4001
rule permit vlan-id 10
rule permit source-mac 1111-1111-1111
traffic classifier 11
if-match acl 3001
traffic classifier 12
if-match acl 4001
    
```

Create a traffic policy and apply the traffic policy to an interface.

```

traffic policy 1
classifier 11 behavior 11
classifier 12 behavior 12 (The traffic behavior configuration is not mentioned here.)
    
```

Examples:

- When receiving packets with source IP address 1.1.1.1, source MAC address 2222-2222-2222, and VLAN 100, the system matches the packets with classifier 11 and applies behavior11 to the packets.
- When receiving packets with source IP address 1.1.10.1, source MAC address 1111-1111-1111, and VLAN 10, the system first matches the packets with classifier 11 and applies behavior11 to the packets, and then matches the packets with classifier 12 and applies behavior12 to the packets.
- When receiving packets with source IP address 1.1.11.1, source MAC address 1111-1111-1111, and VLAN 10, the system first matches the packets with classifier 12 and applies behavior12 to the packets.



**NOTE**

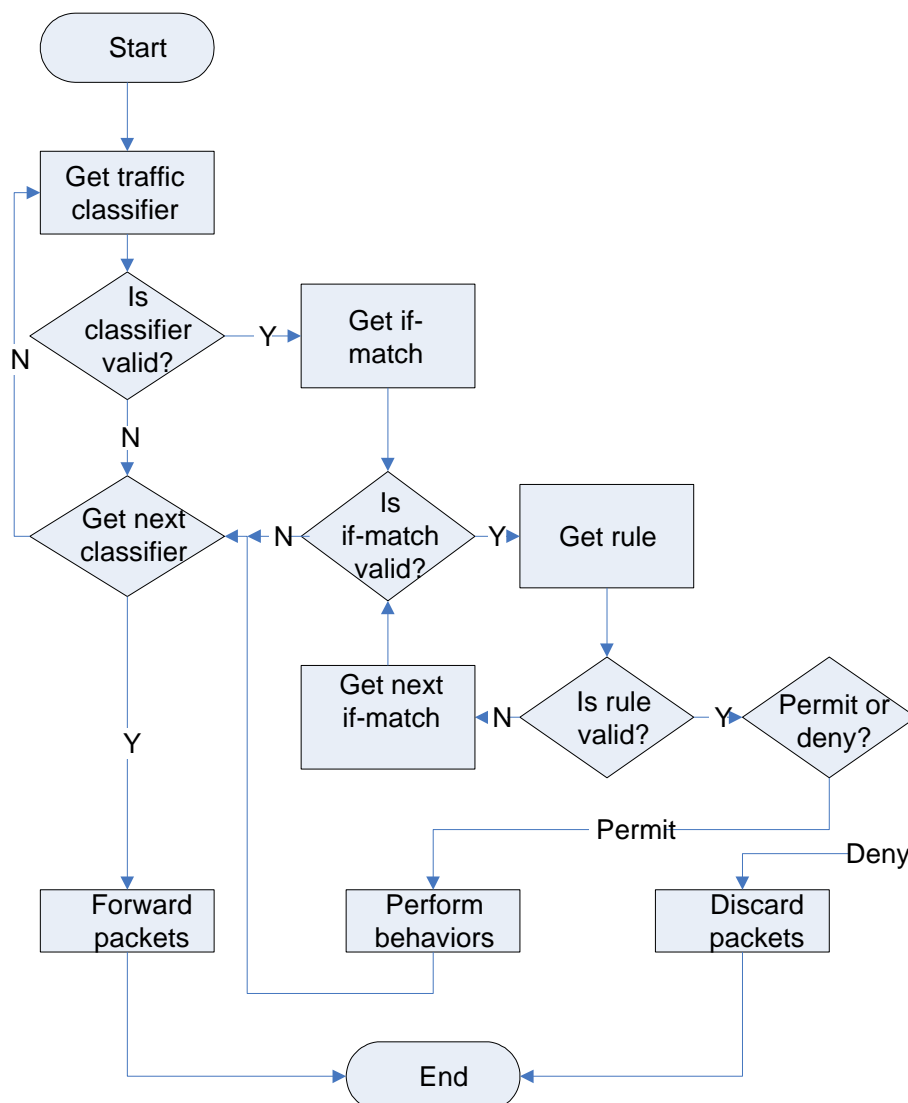
On different models of Huawei switches, if traffic classifiers based on different information are used, traffic policies are executed in a different manner.

On other switches except S2300, if two traffic classifiers are defined based on information at the same layer (for example, Layer 2 or Layer 3, classifier 11 is based on Layer 3 information and classifier 12 is based on Layer 2 information), the system applies only the first matched traffic behavior to packets. If two traffic classifiers are defined based on information at different layers, the system applies the two traffic behaviors to packets.

The S2300 applies only the first matched traffic behavior to packets.

As shown in Figure 2-10, when complex traffic classification needs to be performed for a received packet, the system matches the packet against traffic classifiers of a traffic policy in the sequence in which the traffic classifiers were configured. If the packet matches a traffic classifier, no further match operation is performed. If not, the packet is matched against the following traffic classifiers one by one. If the packet matches no traffic classifier at all, the packet is forwarded with no traffic policy executed.

**Figure 2-10** Traffic policy execution



If multiple if-match clauses are configured for a traffic classifier, the packet is matched against them in the sequence in which the clauses were configured. If an ACL is specified in an if-match clause, the packet is matched against multiple rules in the ACL. The system first checks whether the ACL exists. (A traffic classifier can reference a non-existent ACL.) If the packet matches a rule in the ACL, no further match operation is performed.

When packets match traffic classifiers of different traffic policies that are applied to different objects, either of the following situations occurs:

- When actions in traffic policies do not conflict, all actions are taken.
- If actions in traffic policies conflict, traffic policies take effect as follows:
  - If traffic classification rules in the traffic policies are of the same type, that is, the rules are all user-defined ACL rules, Layer 2 rules, or Layer 3 rules, only one traffic policy takes effect. The traffic policy that takes effect depends on the object that the traffic policy has been applied. The traffic policies applied to the interface, VLAN, and system take effect in descending order of priority.



- If traffic classification rules in the traffic policies are of different types and actions do not conflict, all the traffic policies take effect. If actions conflict, the traffic policy that takes effect is relevant to rules. The rule priority is as follows: user-defined ACL rule > Layer 2 rule + Layer 3 rule > Layer 2 rule > Layer 3 rule.

**NOTE**

A conflicting traffic policy is used only when conflicts occur. Do not use the conflicting traffic policy during network deployment and configuration.

## 2.2.3 Traffic Marking

Traffic marking, also called re-marking, sets or modifies the packet priority to relay QoS information to the connected device.

Priority Mapping maps original priorities of packets to internal priorities and is implemented in the inbound direction. Traffic marking maps internal priorities to packet priorities and is implemented in the outbound direction.

The following tables show the mappings in the DiffServ domain:

- Table 2-10 describes the mappings from PHBs and colors to 802.1p priorities.
- Table 2-11 describes the mappings from PHBs and colors to DSCP priorities.
- Table 2-12 describes the mappings from PHBs and colors to precedences.
- Table 2-13 describes the mappings from PHBs and colors to EXP priorities in MPLS packets.

**Table 2-10** Mappings from PHBs and colors to 802.1p priorities

PHB	Color	802.1p
BE	Green, yellow, and red	0
AF1	Green, yellow, and red	1
AF2	Green, yellow, and red	2
AF3	Green, yellow, and red	3
AF4	Green, yellow, and red	4
EF	Green, yellow, and red	5
CS6	Green, yellow, and red	6
CS7	Green, yellow, and red	7

**Table 2-11** Mappings from PHBs and colors to DSCP priorities

PHB	Color	DSCP
BE	Green, yellow, and red	0
AF1	Green	10
AF1	Yellow	12
AF1	Red	14

PHB	Color	DSCP
AF2	Green	18
AF2	Yellow	20
AF2	Red	22
AF3	Green	26
AF3	Yellow	28
AF3	Red	30
AF4	Green	34
AF4	Yellow	36
AF4	Red	38
EF	Green, yellow, and red	46
CS6	Green, yellow, and red	48
CS7	Green, yellow, and red	56

**Table 2-12** Mappings from PHBs and colors to precedences

PHB	Color	IP Precedence
BE	Green, yellow, and red	0
AF1	Green, yellow, and red	1
AF2	Green, yellow, and red	2
AF3	Green, yellow, and red	3
AF4	Green, yellow, and red	4
EF	Green, yellow, and red	5
CS6	Green, yellow, and red	6
CS7	Green, yellow, and red	7

**Table 2-13** Mappings from PHBs and colors to EXP priorities in MPLS packets

PHB	Color	MPLS EXP
BE	Green, yellow, and red	0
AF1	Green, yellow, and red	1
AF2	Green, yellow, and red	2
AF3	Green, yellow, and red	3

PHB	Color	MPLS EXP
AF4	Green, yellow, and red	4
EF	Green, yellow, and red	5
CS6	Green, yellow, and red	6
CS7	Green, yellow, and red	7

Table 2-14 describes the mappings from DSCP priorities to 802.1p priorities and drop priorities, and Table 2-15 describes the mappings from IP precedences to 802.1p priorities, when the mapping table is used.

**Table 2-14** Mappings from DSCP priorities to 802.1p priorities and drop priorities

Input DSCP	Output 802.1p	Output DP
0-7	0	0
8-15	1	0
16-23	2	0
24-31	3	0
32-39	4	0
40-47	5	0
48-55	6	0
56-63	7	0

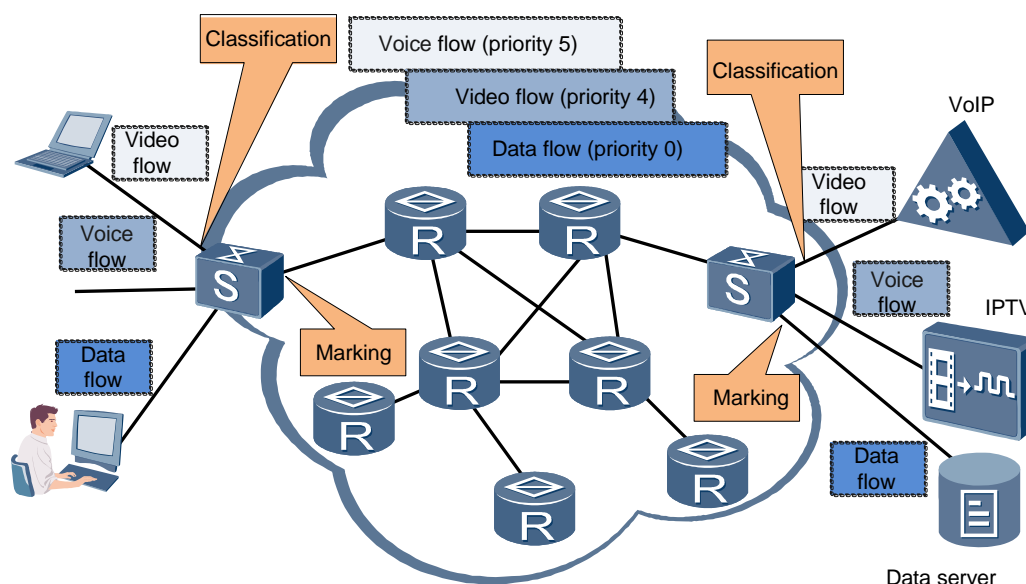
**Table 2-15** Mappings from IP precedences to 802.1p priorities

Input IP Precedence	Output 802.1p	Output Precedence
0	0	0
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7

## 2.2.4 Application of Traffic Classification and Marking

Traffic classification and marking are the basis for implementing differentiated services, and are deployed at the DiffServ domain edge.

**Figure 2-11** Application of traffic classification and marking



The following traffic classification configuration is recommended:

- If packets are classified by service, Layer 2 services are classified by VLAN ID and Layer 3 services are classified by DSCP priority. This is because the services have different requirements of the bandwidth and delay.

Data services are classified by application such as email and BT download or port number. For details about port numbers for common applications, see section 6.2 "Port Numbers of Common Application Services."

- If packets are classified by region (for example, enterprise headquarters and branches have different rights), packets are classified by IP address. This is because different regions use different network segments.

The following traffic priority configuration is recommended:

- Different services have different requirements of the bandwidth and delay, so different priorities are recommended. For details about recommended priorities, see chapter 6 "Appendix."

## 2.3 Traffic Policing and Traffic Shaping

Traffic policing and traffic shaping limit the traffic and resource usage by monitoring the rate limit.

If the transmit rate is larger than the receive rate or the rate of an interface on a downstream device is lower than the rate of an interface on an upstream device, traffic congestion may occur. A network will be congested if traffic sent by users is not limited. To make use of limited network resources and provide better user services, limit the user traffic.

Traffic policing and traffic shaping limit traffic and resources used by the traffic by monitoring the traffic rate.

### 2.3.1 Traffic Policing

Traffic policing controls the rate of incoming packets to ensure that network resources are properly allocated. If the traffic rate of a connection exceeds the specifications on an interface, traffic policing allows the interface to drop excess packets or re-mark the packet priority to maximize network resource usage and protect operators' profits. An example of this process is restricting the rate of HTTP packets to 50% of the network bandwidth.

Traffic policing implements QoS requirements defined in the service level agreement (SLA). The SLA contains parameters, such as the Committed Information Rate (CIR), Peak Information Rate (PIR), Committed Burst Size (CBS), and Peak Burst Size (PBS) to monitor and control incoming traffic. The device performs Pass, Drop, or Markdown action for the traffic exceeding the rate limit. Markdown means that packets are marked with a lower CoS value or a higher drop precedence so that these packets are preferentially dropped when traffic congestion occurs. This measure ensures that the packets conforming to the SLA can have the services specified in the SLA.

Traffic policing uses committed access rate (CAR) to control traffic. CAR uses token buckets to meter the traffic rate. Then preset actions are implemented based on the metering result. These actions include:

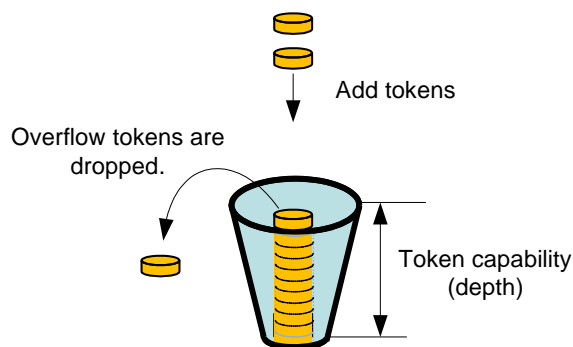
- Pass: forwards the packets whose traffic rate does not exceed the CIR.
- Discard: drops the packets whose traffic rate exceeds the PIR.
- Re-mark: re-marks the packets whose traffic rate is between the CIR and PIR with a lower priority and allows these packets to be forwarded.

### 2.3.2 What Is a Token Bucket

A token bucket is a commonly used mechanism that measures traffic passing through a device. A token bucket is considered as an interior storage pool, and tokens are considered as virtual packets put into a token bucket at a given rate.

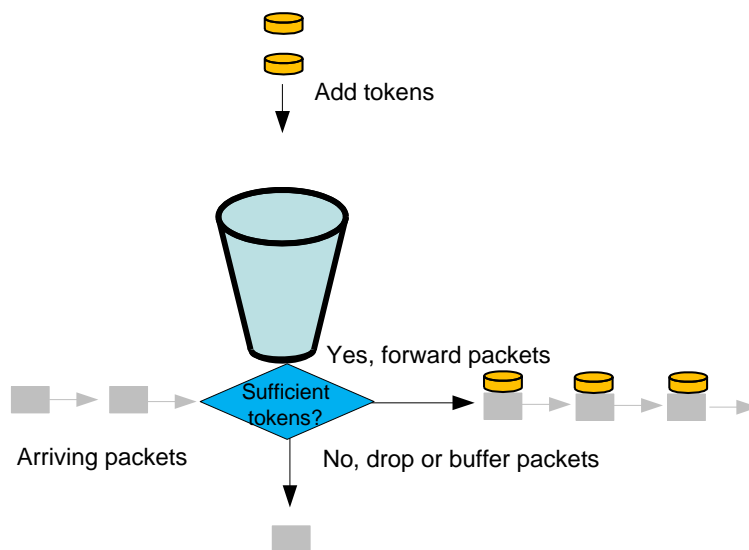
As shown in Figure 2-12, a token bucket can be considered a container of tokens, which has a pre-defined capacity. The system places tokens into a token bucket at the configured rate. If the token bucket is full, excess tokens overflow. A token bucket measures traffic but does not filter packets or perform any action, such as dropping packets.

**Figure 2-12** Token bucket



As shown in Figure 2-13, when data packets reach a device, the device fetches tokens from the TB for transmitting data packets. One token is required for one data packet. If the token bucket does not have enough tokens to send the packet, the packet is discarded or buffered. This feature limits packets to be sent at a rate less than or equal to the rate at which tokens are generated.

**Figure 2-13** Processing packets using token buckets



RFC standards define two token bucket algorithms:

- Single rate three color marker (srTCM), defined by RFC 2697, focuses on the burst packet size.
- Two rate three color marker (trTCM), defined by RFC 2698, focuses on the burst traffic rate.

srTCM and trTCM mark packets in red, yellow, and green based on the assessment result. QoS sets drop priorities based on packet colors. The two algorithms can work in color-aware and color-blind modes.

## srTCM

- srTCM parameters

Committed Information Rate (CIR): rate at which tokens are put into a token bucket. The CIR is expressed in bit/s.

Committed Burst Size (CBS): committed volume of traffic that an interface allows to pass through, also the depth of a token bucket. The CBS is expressed in bytes. The CBS must be greater than or equal to the size of the largest possible packet in the stream. Note that sometimes a single packet can consume all the tokens in the token bucket. The larger the CBS is, the greater the traffic burst can be.

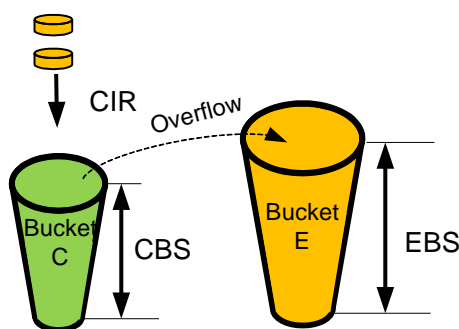
Extended burst size (EBS): maximum size of burst traffic before all traffic exceeds the CIR. The EBS is expressed in bytes.

- srTCM mechanism

srTCM uses two token buckets: bucket C and bucket E. The maximum size of bucket C is the CBS, and the maximum size of bucket E is the EBS. If burst traffic is prevented, the EBS is 0.

When the EBS is not 0, two token buckets are used and packets are marked either green, yellow, or red. When the EBS is 0, no token is added in bucket E. Therefore, only bucket C is used for srTCM. When only bucket C is used, packets are marked either green or red.

Figure 2-14 srTCM mechanism



- Method of adding tokens for srTCM

In srTCM, tokens are put into bucket C at the CIR. When the capacity of bucket C reaches the CBS, tokens are put into bucket E at the CIR (tokens in bucket E are used to transmit excess burst traffic). When the capacity of bucket E reaches the EBS, new tokens are discarded.

Both buckets C and E are initially full.

- srTCM rules

When receiving a packet, the system compares the packet with the number of tokens in the token bucket. If there are sufficient tokens, the system forwards the packet (a token indicates 1-bit forwarding capability). If there are no sufficient tokens, the packet is discarded or buffered.

T<sub>c</sub> and T<sub>e</sub> refer to the number of tokens in buckets C and E. The initial values of T<sub>c</sub> and T<sub>e</sub> are the CBS and EBS.

In color-blind mode, the following rules apply when a packet of size B arrives:

- When one token bucket is used:
  - If  $T_c - B \geq 0$ , the packet is marked green, and T<sub>c</sub> is decremented by B.
  - If  $T_c - B < 0$ , the packet is marked red, and T<sub>c</sub> remains unchanged.

- When two token buckets are used:
  - If  $Tc - B \geq 0$ , the packet is marked green, and  $Tc$  is decremented by  $B$ .
  - If  $Tc - B < 0$  but  $Te - B \geq 0$ , the packet is marked yellow, and  $Te$  is decremented by  $B$ .
  - If  $Te - B < 0$ , the packet is marked red, and neither  $Tc$  nor  $Te$  is decremented.

In color-aware mode, the following rules apply when a packet of size  $B$  arrives:

- When one token bucket is used:
  - If the packet has been pre-colored as green and  $Tc - B \geq 0$ , the packet is re-marked green, and  $Tc$  is decremented by  $B$ .
  - If the packet has been pre-colored as green and  $Tc - B < 0$ , the packet is re-marked red, and  $Tc$  remains unchanged.
  - If the packet has been pre-colored as yellow or red, the packet is re-marked red regardless of the packet length and  $Tc$  remains unchanged.
- When two token buckets are used:
  - If the packet has been pre-colored as green and  $Tc - B \geq 0$ , the packet is re-marked green, and  $Tc$  is decremented by  $B$ .
  - If the packet has been pre-colored as green and  $Tc - B < 0$  but  $Te - B \geq 0$ , the packet is marked yellow, and  $Te$  is decremented by  $B$ .
  - If the packet has been pre-colored as yellow and  $Te - B \geq 0$ , the packet is re-marked yellow, and  $Te$  is decremented by  $B$ .
  - If the packet has been pre-colored as yellow and  $Te - B < 0$ , the packet is re-marked red, and  $Te$  remains unchanged.
  - If the packet has been pre-colored as red, the packet is re-marked red regardless of the packet length. The  $Tc$  and  $Te$  values remain unchanged.

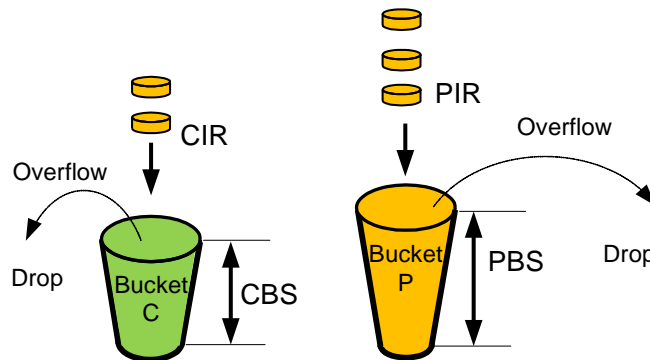
## trTCM

- trTCM parameters
  - CIR: rate at which tokens are put into a token bucket. The CIR is expressed in bit/s.
  - CBS: committed volume of traffic that an interface allows to pass through, also the depth of a token bucket. The CBS is expressed in bytes. The CBS must be greater than or equal to the size of the largest possible packet entering a device.
  - PIR: maximum rate at which an interface allows packets to pass and is expressed in bit/s. The PIR must be greater than or equal to the CIR.
  - PBS: maximum volume of traffic that an interface allows to pass through in a traffic burst.
- trTCM mechanism

The trTCM uses two token buckets and focuses on the burst traffic rate. The trTCM uses two token buckets, C and P, with rates CIR and PIR, respectively. The maximum size of bucket C is the CBS, and the maximum size of bucket P is the PBS.



**Figure 2-15** trTCM mechanism



- Method of adding tokens for trTCM  
Both buckets C and P are initially full. Tokens are put into buckets C and P at the rate of CIR and PIR, respectively. When one bucket is full of tokens, any subsequent tokens for the bucket are dropped, but tokens continue being put into the other bucket if it is not full.
- trTCM rules  
trTCM focuses on the traffic burst rate and checks whether the traffic rate is conforming to the specifications. Therefore, traffic is measured based on bucket P and then bucket C. trTCM can work in color-aware and color-blind modes.  $T_c$  and  $T_p$  refer to the numbers of tokens in buckets C and P, respectively. The initial values of  $T_c$  and  $T_p$  are respectively the CBS and PBS.

In color-blind mode, the following rules apply when a packet of size B arrives:

- If  $T_p - B < 0$ , the packet is marked red, and  $T_c$  and  $T_p$  values remain unchanged.
- If  $T_p - B \geq 0$  but  $T_c - B < 0$ , the packet is marked yellow, and  $T_p$  is decremented by B.
- If  $T_c - B \geq 0$ , the packet is marked green and both  $T_p$  and  $T_c$  are decremented by B.

In color-aware mode, the following rules apply when a packet of size B arrives:

- If the packet has been pre-colored as green, and  $T_p - B < 0$ , the packet is re-marked red, and neither  $T_p$  nor  $T_c$  is decremented.
- If the packet has been pre-colored as green and  $T_p - B \geq 0$  but  $T_c - B < 0$ , the packet is re-marked yellow,  $T_p$  is decremented by B, and  $T_c$  remains unchanged.
- If the packet has been pre-colored as green and  $T_c - B \geq 0$ , the packet is re-marked green, and both  $T_p$  and  $T_c$  are decremented by B.
- If the packet has been pre-colored as yellow and  $T_p - B < 0$ , the packet is re-marked red, and neither  $T_p$  nor  $T_c$  is decremented.
- If the packet has been pre-colored as yellow and  $T_p - B \geq 0$ , the packet is re-marked yellow, and  $T_p$  is decremented by B and  $T_c$  remains unchanged.
- If the packet has been pre-colored as red, the packet is re-marked red regardless of what the packet length is and  $T_p$  and  $T_c$  values remain unchanged.

### 2.3.3 CAR

Traffic policing uses committed access rate (CAR) to control traffic. CAR uses token buckets to measure traffic and determines whether a packet is conforming to the specification.

CAR has the following functions:

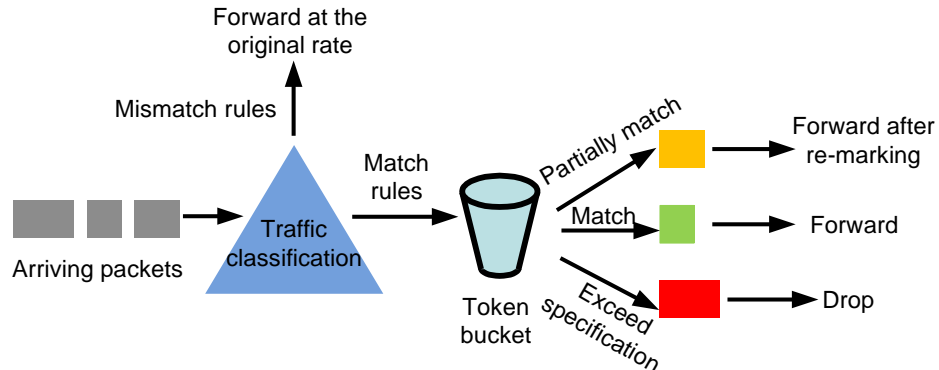
- Rate limit: Only packets allocated enough tokens are allowed to pass in a period of time so that the traffic rate is restricted.
- Traffic classification: Packets are marked internal priorities, such as the CoS value and drop precedence, based on the measurement performed by token buckets.

Huawei switches provide the following two implementation modes:

- Interface-based CAR: If users' traffic is not limited, continuous burst data from many users makes the network congested. You can configure traffic policing to limit the traffic within a specified range and to protect network resources as well as the enterprise users' interests.
- Flow-based CAR: To limit traffic of a specified type in the inbound or outbound direction on an interface, configure flow-based traffic policing. A traffic policy can be applied to different interfaces. When the receive or transmit rate of packets matching traffic classification rules exceeds the rate limit, the packets are discarded. Flow-based traffic policing can implement differentiate services using complex traffic classification.

Figure 2-16 shows the CAR process.

Figure 2-16 CAR process



Huawei switches conform to RFC 2697 and RFC 2698 to implement CAR. In CAR, buckets are added when packets are received. The number of added tokens is the CIR multiplied by the difference between the current time and last time tokens were added. After tokens are put into a bucket, the system determines whether there are sufficient tokens to transmit a packet.

CAR supports srTCM with single bucket, srTCM with two buckets, and trTCM. This section provides examples of the three marking methods in color-blind mode. The implementation in color-aware mode is similar to that in color-bind mode.

#### srTCM with Single Bucket

This example uses the CIR of 1 Mbit/s, the CBS of 2000 bytes, and the excess burst size (EBS) of 0. EBS 0 indicates that only bucket C is used. Bucket C is initially full of tokens.

- If the first packet arriving at an interface is 1500 bytes long, the packet is marked green because the number of tokens in bucket C is greater than the packet length. The number of tokens in bucket C then decreases by 1500 bytes, with 500 bytes remaining.
- Assume that the second packet arriving at the interface after a delay of 1 ms is 1500 bytes long. Additional 125-byte tokens are put into bucket C ( $CIR \times \text{time period} = 1 \text{ Mbit/s} \times 1 \text{ ms} = 1000 \text{ bits} = 125 \text{ bytes}$ ). Bucket C now has 625-byte tokens, which are not enough for the 1500-byte second packet. Therefore, the second packet is marked red.
- Assume that the third packet arriving at the interface after a delay of 1 ms is 1000 bytes long. Additional 125-byte tokens are put into bucket C ( $CIR \times \text{time period} = 1 \text{ Mbit/s} \times 1 \text{ ms} = 1000 \text{ bits} = 125 \text{ bytes}$ ). Bucket C now has 750-byte tokens, which are not enough for the 1000-byte third packet. Therefore, the third packet is marked red.
- Assume that the fourth packet arriving at the interface after a delay of 20 ms is 1500 bytes long. Additional 2500-byte tokens are put into bucket C ( $CIR \times \text{time period} = 1 \text{ Mbit/s} \times 20 \text{ ms} = 20000 \text{ bits} = 2500 \text{ bytes}$ ). This time 3250-byte tokens are destined for bucket C, but the excess 1250-byte tokens over the CBS (2000 bytes) are dropped. Therefore, bucket C has 2000-byte tokens, which are enough for the 1500-byte fourth packet. The fourth packet is marked green, and the number of tokens in bucket C decreases by 1500 bytes to 500 bytes.

Table 2-16 illustrates the preceding process.

**Table 2-16** srTCM with single bucket

No.	Time (ms)	Packet Length (Bytes)	Delay	Token Addition	Tokens in Bucket C Before Packet Processing	Tokens in Bucket C After Packet Processing	Marking
-	-	-	-	-	2000	2000	-
1	0	1500	0	0	2000	500	Green
2	1	1500	1	125	625	625	Red
3	2	1000	1	125	750	750	Red
4	22	1500	20	2500	2000	500	Green

## SrTCM with Two Buckets

This example uses the CIR of 1 Mbit/s and the CBS and EBS both of 2000 bytes. Buckets C and E are initially full of tokens.

- If the first packet arriving at an interface is 1500 bytes long, the packet is marked green because the number of tokens in bucket C is greater than the packet length. The number of tokens in bucket C then decreases by 1500 bytes, with 500 bytes remaining. The number of tokens in bucket E remains unchanged.
- Assume that the second packet arriving at the interface after a delay of 1 ms is 1500 bytes long. Additional 125-byte tokens are put into bucket C ( $CIR \times \text{time period} = 1 \text{ Mbit/s} \times 1 \text{ ms} = 1000 \text{ bits} = 125 \text{ bytes}$ ). Bucket C now has 625-byte tokens, which are not enough for the 1500-byte second packet. Bucket E has 2000-byte tokens, which are enough for the second packet. Therefore, the second packet is marked yellow, and the

number of tokens in bucket E decreases by 1500 bytes, with 500 bytes remaining. The number of tokens in bucket C remains unchanged.

- Assume that the third packet arriving at the interface after a delay of 1 ms is 1000 bytes long. Additional 125-byte tokens are put into bucket C ( $CIR \times \text{time period} = 1 \text{ Mbit/s} \times 1 \text{ ms} = 1000 \text{ bits} = 125 \text{ bytes}$ ). Bucket C now has 750-byte tokens and bucket E has 500-byte tokens, neither of which is enough for the 1000-byte third packet. Therefore, the third packet is marked red. The number of tokens in buckets C and E remain unchanged.
- Assume that the fourth packet arriving at the interface after a delay of 20 ms is 1500 bytes long. Additional 2500-byte tokens are put into bucket C ( $CIR \times \text{time period} = 1 \text{ Mbit/s} \times 20 \text{ ms} = 20000 \text{ bits} = 2500 \text{ bytes}$ ). This time 3250-byte tokens are destined for bucket C, but the excess 1250-byte tokens over the CBS (2000 bytes) are put into bucket E instead. Therefore, bucket C has 2000-byte tokens, and bucket E has 1750-byte tokens. Tokens in bucket C are enough for the 1500-byte fourth packet. Therefore, the fourth packet is marked green, and the number of tokens in bucket C decreases by 1500 bytes, with 500 bytes remaining. The number of tokens in bucket E remains unchanged.

Table 2-17 illustrates the preceding process.

**Table 2-17** srTCM with two buckets

No.	Time (ms)	Packet Length (Bytes)	Delay	Token Addition	Tokens in Bucket C Before Packet Processing	Tokens in Bucket E Before Packet Processing	Tokens in Bucket C After Packet Processing	Tokens in Bucket E After Packet Processing	Marking
-	-	-	-	-	2000	2000	2000	2000	-
1	0	1500	0	0	2000	2000	500	2000	Green
2	1	1500	1	125	625	2000	625	500	Yellow
3	2	1000	1	125	750	500	750	500	Red
4	22	1500	20	2500	2000	1750	500	1750	Green

## trTCM

This example uses the CIR of 1 Mbit/s, the PIR of 2 Mbit/s, and the CBS and EBS both of 2000 bytes. Buckets C and P are initially full of tokens.

- If the first packet arriving at the interface is 1500 bytes long, the packet is marked green because the number of tokens in both buckets P and C is greater than the packet length. Then the number of tokens in both buckets P and C decreases by 1500 bytes, with 500 bytes remaining.
- Assume that the second packet arriving at the interface after a delay of 1 ms is 1500 bytes long. Additional 250-byte tokens are put into bucket P ( $PIR \times \text{time period} = 2 \text{ Mbit/s} \times 1 \text{ ms} = 2000 \text{ bits} = 250 \text{ bytes}$ ) and 125-byte tokens are put into bucket C ( $CIR \times \text{time period} = 1 \text{ Mbit/s} \times 1 \text{ ms} = 1000 \text{ bits} = 125 \text{ bytes}$ ). Bucket P now has 750-byte tokens, which are not enough for the 1500-byte second packet. Therefore, the second packet is marked red, and the number of tokens in buckets P and C remain unchanged.

- Assume that the third packet arriving at the interface after a delay of 1 ms is 1000 bytes long. Additional 250-byte tokens are put into bucket P ( $\text{PIR} \times \text{time period} = 2 \text{ Mbit/s} \times 1 \text{ ms} = 2000 \text{ bits} = 250 \text{ bytes}$ ) and 125-byte tokens are put into bucket C ( $\text{CIR} \times \text{time period} = 1 \text{ Mbit/s} \times 1 \text{ ms} = 1000 \text{ bits} = 125 \text{ bytes}$ ). Bucket P now has 1000-byte tokens, which equals the third packet length. Bucket C has only 625-byte tokens, which are not enough for the 1000-byte third packet. Therefore, the third packet is marked yellow. The number of tokens in bucket P decreases by 1000 bytes, with 0 bytes remaining. The number of tokens in bucket C remains unchanged.
- Assume that the fourth packet arriving at the interface after a delay of 20 ms is 1500 bytes long. Additional 5000-byte tokens are put into bucket P ( $\text{PIR} \times \text{time period} = 2 \text{ Mbit/s} \times 20 \text{ ms} = 40000 \text{ bits} = 5000 \text{ bytes}$ ), but excess tokens over the PBS (2000 bytes) are dropped. Bucket P has 2000-byte tokens, which are enough for the 1500-byte fourth packet. Bucket C has 625-byte tokens left, and additional 2500-byte tokens are put into bucket C ( $\text{CIR} \times \text{time period} = 1 \text{ Mbit/s} \times 20 \text{ ms} = 2000 \text{ bits} = 250 \text{ bytes}$ ). This time 3250-byte tokens are destined for bucket C, but excess tokens over the CBS (2000 bytes) are dropped. Bucket C then has 2000-byte tokens, which are enough for the 1500-byte fourth packet. Therefore, the fourth packet is marked green. The number of tokens in both buckets P and C decreases by 1500 bytes, with 500 bytes remaining.

Table 2-18 illustrates the preceding process

**Table 2-18** trTCM

No.	Time (ms)	Packet Length (Bytes)	Delay	Token Addition	Tokens in Bucket C Before Packet Processing	Tokens in Bucket P Before Packet Processing	Tokens in Bucket C After Packet Processing	Tokens in Bucket P After Packet Processing	Marking
-	-	-	-	-	2000	2000	2000	2000	-
1	0	1500	0	0	2000	2000	500	500	Green
2	1	1500	1	125	500	750	500	750	Red
3	2	1000	1	125	625	1000	625	0	Yellow
4	22	1500	20	2500	2000	2000	500	500	Green

## Usage Scenarios for the Three Marking Methods

srTCM focuses on the traffic burst size and has a simple token-adding method and packet processing mechanism. trTCM focuses on the traffic burst rate and has a complex token-adding method and packet processing mechanism.

srTCM and trTCM have their own advantages and disadvantages. They vary from each other in performance, such as the packet loss rate, burst traffic processing capability, hybrid packet forwarding capability, and data forwarding smoothing capability. The three markers fit for traffic with different features as follows:

- To control the traffic rate, use srTCM with single bucket.
- To control the traffic rate and distinguish traffic marked differently and process them differently, use srTCM with two buckets. Note that traffic marked yellow must be

processed differently from traffic marked green. Otherwise, the implementation result of srTCM with two buckets is the same as that of the srTCM with single bucket.

- To control the traffic rate and check whether the traffic rate is less than the CIR or is greater than CIR but less than the PIR, use trTCM. Note that traffic marked yellow must be processed differently from traffic marked green. Otherwise, the implementation result of trTCM is the same as that of srTCM with single bucket.

**Table 2-19** Comparison between three marking methods

Marking Method	Advantage	Disadvantage	Usage Scenario
srTCM with single bucket	Limits bandwidth with simple configuration.	Does not reserve any bandwidth for burst traffic exceeding the single bucket capacity.	Discards low-priority services such as HTTP traffic, and excess traffic.
srTCM with two buckets	Limits bandwidth, allows some burst traffic, and distinguishes burst and normal services.	Is much complex compared with srTCM with single bucket because the capacity of bucket E needs to be considered.	Reserves bandwidth for burst traffic or important services (for example, email data is one of important services).
trTCM	Allocates bandwidth in a fine-grained manner, and determines whether the bandwidth is less than the CIR or is greater than CIR but less than the PIR.	Considers the CIR, CBS, PIR, and PBS before deployment and distinguishes these parameters for different services.	Is recommended for important services because it better monitors burst traffic and guides traffic analysis.

## CAR Parameter Setting

The CIR is the key to determine the volume of traffic allowed to pass through a network. The larger the CIR is, the higher the rate at which tokens are generated. The more the tokens allocated to packets, the greater the volume of traffic allowed to pass through. The CBS is also an important parameter. A larger CBS results in more accumulated tokens in bucket C and a greater volume of traffic allowed to pass through.

The CBS must be greater than or equal to the maximum packet length. For example, the CIR is 100 Mbit/s, and the CBS is 200 bytes. If a device receives 1500-byte packets, the packet length always exceeds the CBS, causing the packets to be marked red or yellow even if the traffic rate is lower than 100 Mbit/s. This leads to an inaccurate CAR implementation.

The EBS is expressed in bytes. On Huawei switches, the CBS and EBS are values of buckets C and E respectively, that is, the CBS is irrelevant to the EBS. If burst traffic is not allowed, set the EBS to 0. If the token bucket is required to transmit burst traffic, set the EBS larger than 0.

The bucket depth (CBS, EBS, or PBS) is set based on actual rate limit requirements. The bucket depth is calculated based on the following conditions:

1. The bucket depth must be greater than or equal to the MTU.

2. The bucket depth must be greater than or equal to the allowed burst traffic volume.

Condition 1 is easy to meet. Condition 2 is difficult to operate, and the following formula is used on a Huawei switch:

- When the bandwidth is lower than or equal to 100 Mbit/s: Bucket depth (bytes) = Bandwidth (kbit/s) x 1000 (ms)/8.
- When the bandwidth is larger than 100 Mbit/s: Bucket depth (bytes) = 100,000 (kbit/s) x 1000 (ms)/8.

Assume that an interface connected to 100-channel VoIP phones uses 10 Mbit/s bandwidth, and voice traffic occupies less bandwidth (100 kbit/s in G.711) on the interface. According to the preceding formula, the reserved bucket depth is 1250 Mbytes. If the interface connected to 10-channel video services uses 20 Mbit/s bandwidth and video services use 2 Mbit/s bandwidth, the reserved bucket depth is 2500 Mbytes.

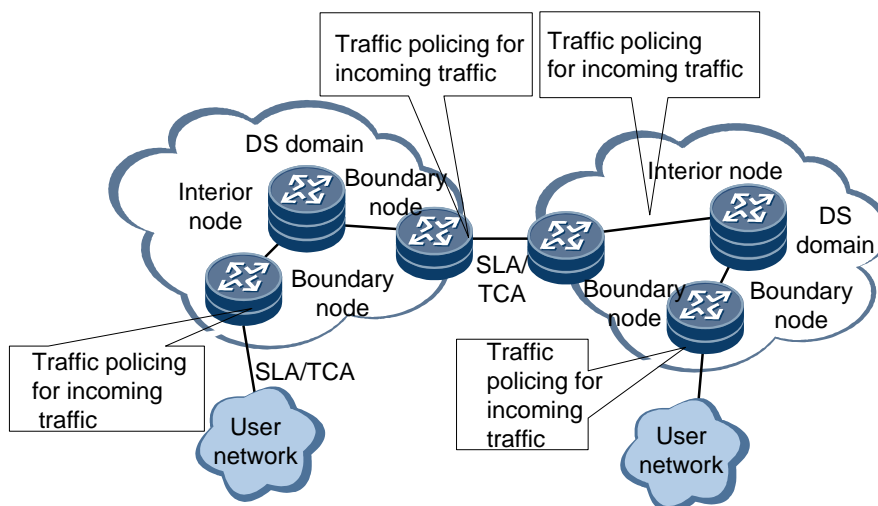
The PIR is often 1.5 times the CIR. Large PIR causes high load on a device.

For voice services, the CIR and PIR are 100 kbit/s and 150 kbit/s. For video services, the CIR is 2 Mbit/s and the PIR is 3 Mbit/s.

## Traffic Policing Applications

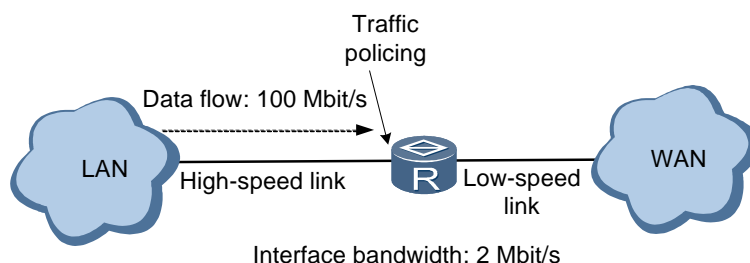
Traffic policing mainly applies to the network edge. The switch performs Pass, Drop, or Markdown action for the traffic exceeding the SLA. This measure ensures that the packets conforming to the SLA can have the services specified in the SLA and core devices can normally process data. Figure 2-17 shows typical networking.

Figure 2-17 Application 1



Enterprise users connect the WAN and LAN through access switches. LAN bandwidth (100 Mbit/s) is often higher than LAN bandwidth (2 Mbit/s or less). When LAN users send a large amount of data through the WAN, congestion may occur at the network edge. You can perform traffic policing at the edge of the network edge switch to limit the rate of data, as shown in Figure 2-18.

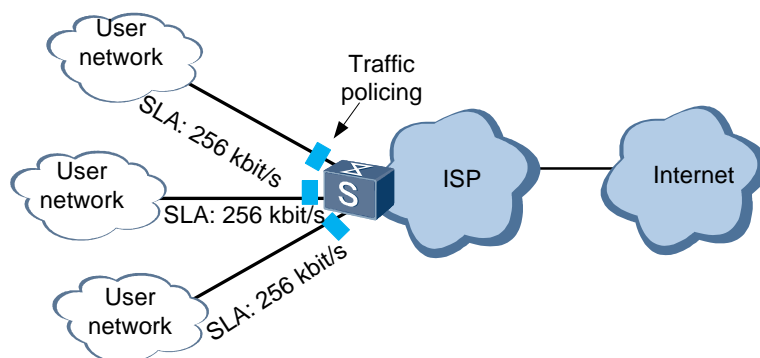
**Figure 2-18** Application 2



- Interface-based traffic policing

Interface-based traffic policing controls all traffic that enters an interface and does not identify the packet types. As shown in Figure 2-19, an edge switch connects to networks of three departments. The SLA defines that each user can send traffic at a maximum rate of 256 kbit/s. However, burst traffic is sometimes transmitted. Traffic policing can be configured on the inbound interface of the edge switch to limit the traffic rate to a maximum of 256 kbit/s. All excess traffic over 256 kbit/s will be dropped.

**Figure 2-19** Interface-based traffic policing



- Class-based traffic policing

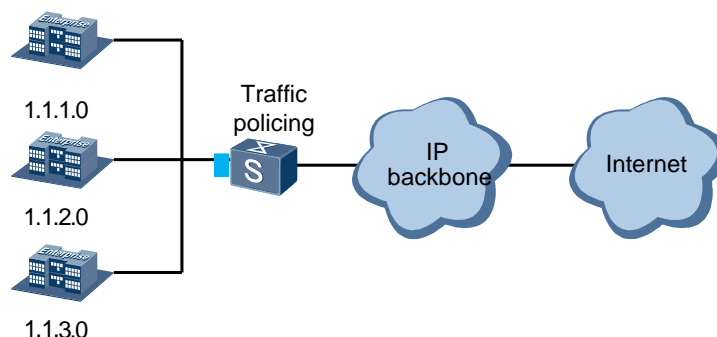
The class-based traffic policy controls the rate of one or more types of packets that enter an interface but not all types of packets.

As shown in Figure 2-20, traffic from the three users at 1.1.1.1, 1.1.1.2, and 1.1.1.3 is aggregated on one switch. The SLA defines that each user can send traffic at a maximum rate of 256 kbit/s. However, burst traffic is sometimes transmitted. When a user sends a large amount of data, services of other users may be affected even if they send traffic at a rate within 256 kbit/s. To resolve this problem, configure traffic classification and traffic policing based on source IP addresses on the inbound interface of the switch to control



the rate of traffic sent from different users. The switch drops excess traffic when the traffic rate of a certain user exceeds 256 kbit/s.

**Figure 2-20** Class-based traffic policing

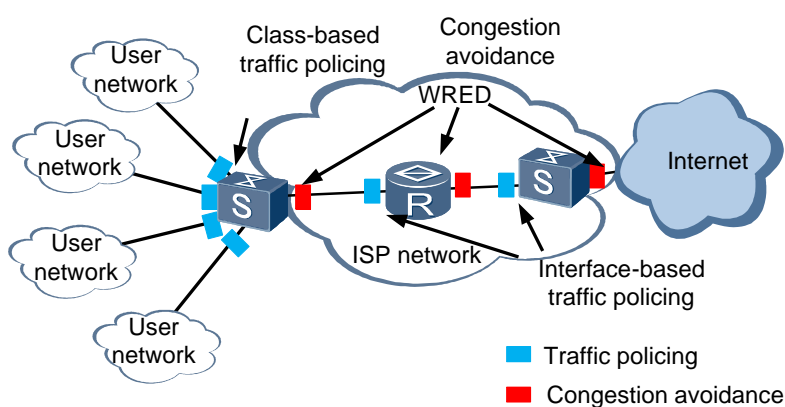


## Combination of Traffic Policing and Other QoS Policies

Traffic policing is often used together with QoS policies such as congestion avoidance and re-marking to guarantee QoS network-wide.

Figure 2-21 shows how traffic policing works with congestion avoidance to control traffic. In this networking, four user networks connect to a switch at the ISP network edge. The SLA defines that each user can send FTP traffic at a maximum rate of 256 kbit/s. However, burst traffic is sometimes transmitted at a rate even higher than 1 Mbit/s. When a user sends a large amount of FTP data, FTP services of other users may be affected even if they send traffic at a rate within 256 kbit/s. To resolve this problem, configure class-based traffic policing on each inbound interface of the switch to monitor the FTP traffic and re-mark the DSCP values of packets. The traffic at a rate lower than or equal to 256 kbit/s is re-marked AF11. The traffic at a rate ranging from 256 kbit/s to 1 Mbit/s is re-marked AF12. The traffic at a rate higher than 1 Mbit/s is re-marked AF13. Weighted Random Early Detection (WRED) is configured as a drop policy for these types of traffic on outbound interfaces to prevent traffic congestion. WRED drops packets based on the DSCP values. Packets in AF13 are first dropped, and then AF12 and AF11 in sequence.

**Figure 2-21** Combination of traffic policing and congestion avoidance

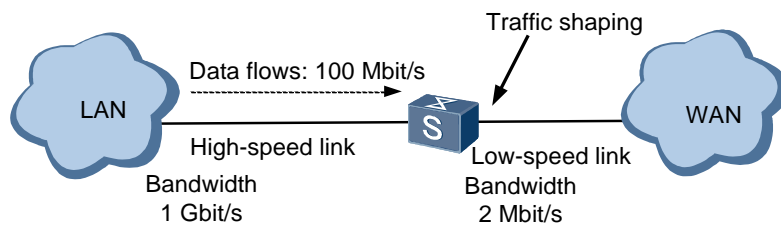


## 2.3.4 Traffic Shaping

Traffic shaping controls the rate of outgoing packets so that packets are sent at an even rate.

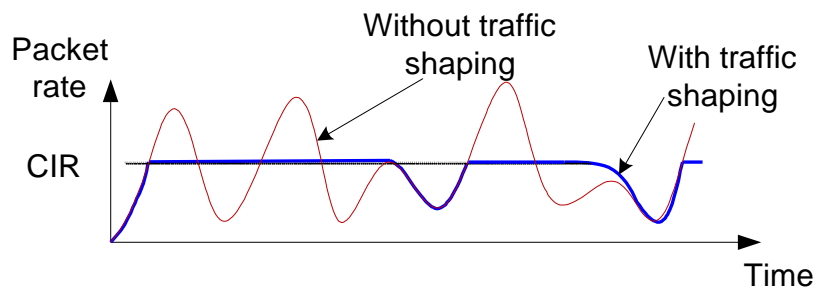
Traffic shaping controls the rate of outgoing packets to allow the traffic rate to match that on the downstream device. When traffic is transmitted from a high-speed link to a low-speed link or a traffic burst occurs, the inbound interface of the low-speed link is prone to severe data loss. To prevent this problem, traffic shaping must be configured on the outbound interface of the device connected to the low-speed link, as shown in Figure 2-22.

**Figure 2-22** Data transmission from the high-speed link to the low-speed link



As shown in Figure 2-23, traffic shaping can be configured on the outbound interface of an upstream device to make irregular traffic transmitted at an even rate, preventing traffic congestion on the downstream device.

**Figure 2-23** Effect of traffic shaping



Traffic shaping uses the buffer and token bucket to control traffic. Traffic shaping buffers overspeed packets and uses token buckets to transmit these packets afterward at an even rate.

Traffic shaping is implemented for packets that have been implemented with queue scheduling and are leaving the queues. For details about queues and queue scheduling, see section 2.4 "Congestion Management and Congestion Avoidance."

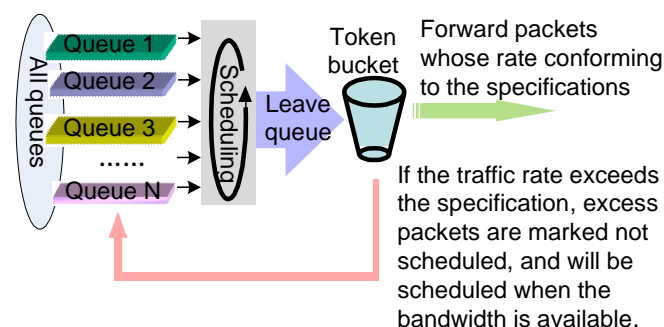
## Classification and Comparison

There are two traffic shaping modes: interface-based traffic shaping and queue-based traffic shaping.

- Interface-based traffic shaping, also called line rate (LR), is used to limit the rate at which all packets (including burst packets) are transmitted. Interface-based traffic shaping takes effect on the entire outbound interface, regardless of packet priorities. Figure 2-24 shows how interface-based traffic shaping is implemented:

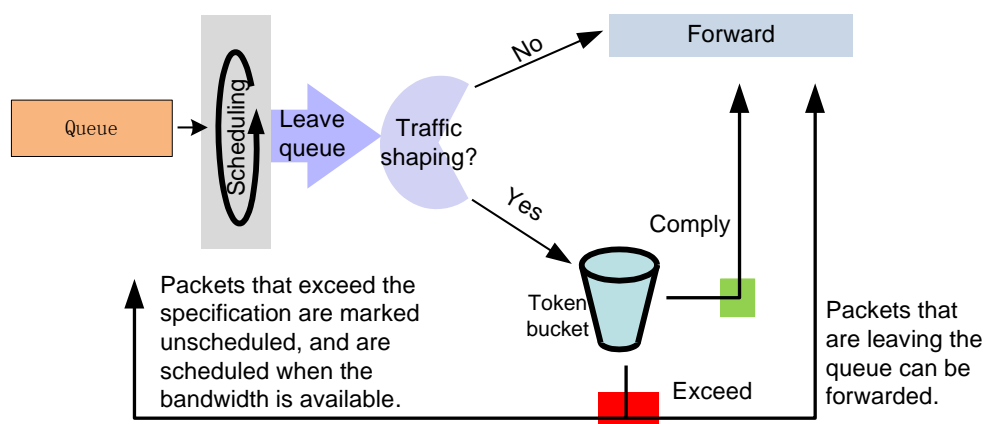
- When packets have been implemented with queue scheduling and are leaving queues, all queues are measured against token buckets.
- After queues are measured against token buckets, if packets in a queue are transmitted at a rate conforming to the specifications, the queue is forwarded. If packets in a queue are transmitted at a rate exceeding the specifications, the queue is marked unscheduled and will be scheduled when the bandwidth is available.

Figure 2-24 Interface-based traffic shaping



- Queue-based traffic shaping applies to each queue on an outbound interface. Figure 2-25 shows how queue-based traffic shaping is implemented:
  - When packets have been implemented with queue scheduling and are leaving queues, the packets that do not need traffic shaping are forwarded; the packets that need traffic shaping are measured against token buckets.
  - After queues are measured against token buckets, if packets in a queue are transmitted at a rate conforming to the specifications, the packets in the queue are marked green and forwarded. If packets in a queue are transmitted at a rate exceeding the specifications, the packet that is leaving the queue is forwarded, but the queue is marked unscheduled and can be scheduled after new tokens are added to the token bucket. After the queue is marked unscheduled, more packets can be put into the queue, but excess packets over the queue capacity are dropped. Therefore, traffic shaping allows traffic to be sent at an even rate but does not provide a zero-packet-loss guarantee.

Figure 2-25 Queue-based traffic shaping



Interface-based traffic shaping and queue-based traffic shaping have advantages and disadvantages and are used according to actual networking.

**Table 2-20** Comparison between traffic shaping modes

Traffic Shaping Mode	Advantage	Disadvantage	Usage Scenario
Interface-based traffic shaping	Has simple configuration.	Cannot differentiate services.	An interface transmits single services such as financial data on bank transaction networks and voice services on the 110 police platform.
Queue-based traffic shaping	Differentiates services.	Has complex configuration.	An interface transmits hybrid services and needs to shape traffic of different services.
Interface- and queue-based traffic shaping (hierarchical traffic shaping)	Shapes traffic based on services and considers the entire bandwidth to implement hierarchical management.	If both queue-based traffic shaping and interface-based traffic shaping are configured on an interface, the CIR of interface-based traffic shaping cannot be smaller than the sum of CIR values of all the queues on the interface; otherwise, traffic shaping may be incorrect.	If the sum of PIR values exceeds the maximum bandwidth of an interface or allowed bandwidth, configure interface-based traffic shaping to ensure that the total traffic volume is within the specified rate limit. This combination can also be used in common scenarios. (Recommended)

## Traffic Shaping Parameter Setting

Interface-based traffic shaping supports only bucket C, and queue-based traffic shaping supports both buckets C and P.

Traffic shaping parameters need to be set considering buckets C and P on the inbound interface, interface buffer capability, and SLA specifications.

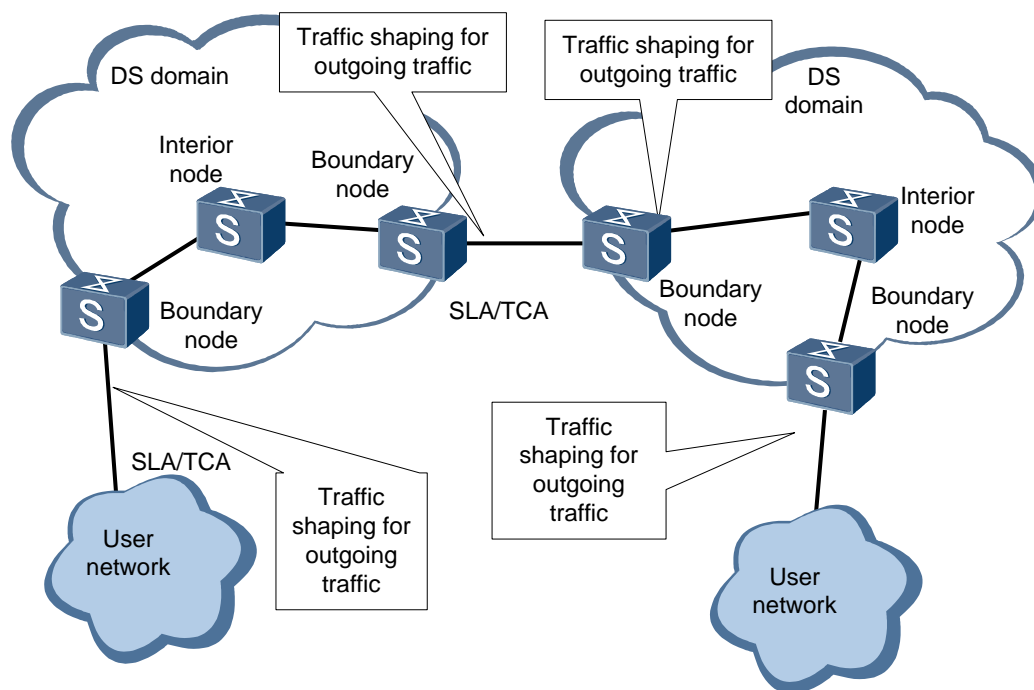
Enterprise LANs have sufficient bandwidth, so traffic shaping is not configured on enterprise LANs. Traffic shaping is often deployed on the LAN outbound interface connected to the Internet and bandwidth is allocated to each service according to the SLA.

For example, an enterprise leases 5 Mbit/s bandwidth, and allocates bandwidths of 2.5 Mbit/s, 1 Mbit/s, and 1.5 Mbit/s to video, voice, and data services respectively. The CIR, CBS, PIR, and PBS are calculated according to CAR Parameter Setting.

## Traffic Shaping Applications

Traffic shaping controls traffic output based on SLA specifications of traffic policing on the downstream node, to minimize packet loss.

**Figure 2-26** Traffic shaping application

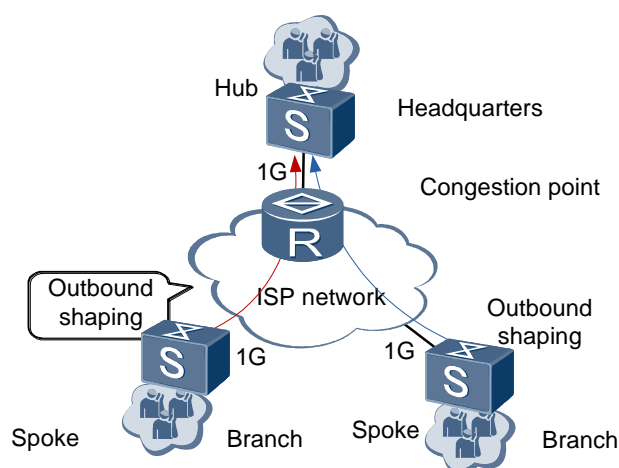


- Interface-based traffic shaping

Interface-based traffic shaping shapes traffic of all packets passing the interface.

As shown in Figure 2-27, the enterprise headquarters is connected to branches through leased lines on an ISP network in Hub-Spoke mode. The bandwidth of each leased line is 1 Gbit/s. If all branches send data to headquarters, traffic congestion occurs on the nodes connected to headquarters at the ISP network edge. To prevent packet loss, configure traffic shaping on outbound interfaces of the nodes at the branch network edge.

**Figure 2-27** Interface-based traffic shaping



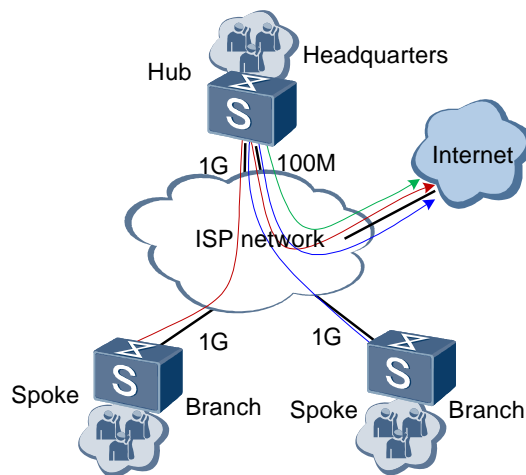
- Queue-based traffic shaping

Queue-based traffic shaping shapes traffic of the packets of a certain type passing the interface. The packets are classified based on simple traffic classification. Queue-based traffic shaping shapes traffic based on service types, such as audio, data, and video services.

As shown in Figure 2-28, the enterprise headquarters is connected to branches through leased lines on an ISP network in Hub-Spoke mode. The bandwidth of each leased line is 1 Gbit/s. Branches access the Internet through headquarters, but the link bandwidth between headquarters and the Internet is only 100 Mbit/s. If all branches access the Internet at a high rate, the rate of web traffic sent from headquarters to the Internet may exceed 100 Mbit/s, causing web packet loss on the ISP network.

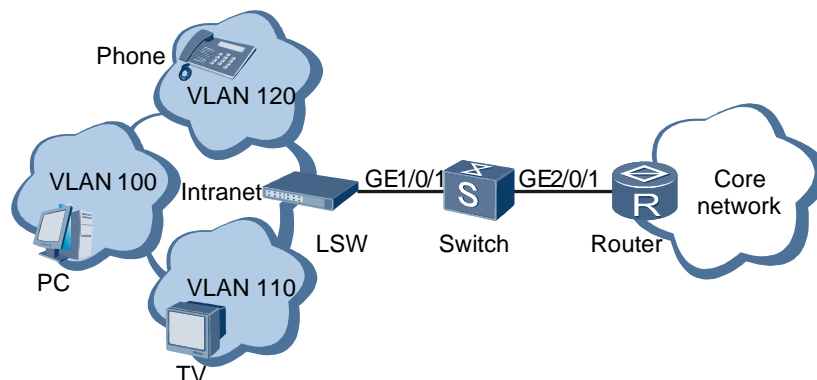
To prevent web packet loss, configure queue-based traffic shaping for web traffic on outbound interfaces of branches and outbound interfaces connecting to the Internet on headquarters.

Figure 2-28 Queue-based traffic shaping



- Interface- and queue-based traffic shaping (hierarchical traffic shaping)

Figure 2-29 Hierarchical traffic shaping



The switch connects to the router through GE0/0/2, and voice, video, and data services from the enterprise LAN reach the Internet through the switch and router, as shown in Figure 2-29. Because the traffic rate from the enterprise LAN is larger than the rate of GE0/0/2, jitter may

occur on GE0/0/2. To reduce jitter and meet service requirements, configure hierarchical traffic shaping on GE0/0/2.

Interface-based traffic shaping is first configured to ensure that traffic on the outbound interface is within the allowed bandwidth, and then queue-based traffic shaping is performed.

## 2.3.5 Comparison Between Traffic Policing and Traffic Shaping

- Similarities
  - Traffic policing and traffic shaping share the following features:
    - Limit the network traffic rate.
    - Use token buckets to measure the traffic rate.
    - Apply to the network edge.
- Differences

**Table 2-21** Differences between traffic policing and traffic shaping

Traffic Policing	Traffic Shaping
Applies to the inbound direction.	Applies to the outbound direction.
Drops excess traffic over SLA specifications or re-marks such traffic with a lower priority.	Buffers excess traffic over a policy or SLA specifications.
Consumes no additional memory resources and brings no delay or jitter.	Consumes memory resources for excess traffic buffer and brings delay and jitter.
Packet loss may result in packet retransmission.	Packet loss seldom occurs, so packets are seldom retransmitted.
Supports traffic re-marking.	Does not support traffic re-marking.

## 2.4 Congestion Management and Congestion Avoidance

### 2.4.1 Background

Traffic congestion occurs when multiple users compete for the same resources (such as the bandwidth and buffer) on the shared network. For example, a user on a LAN sends data to a user on another LAN through a WAN. The WAN bandwidth is generally higher than the LAN bandwidth. Therefore, data cannot be transmitted at the same rate on the WAN as that on the LAN. Traffic congestion occurs on the edge switch connecting the LAN and WAN, as shown in Figure 2-30.

**Figure 2-30** Traffic congestion

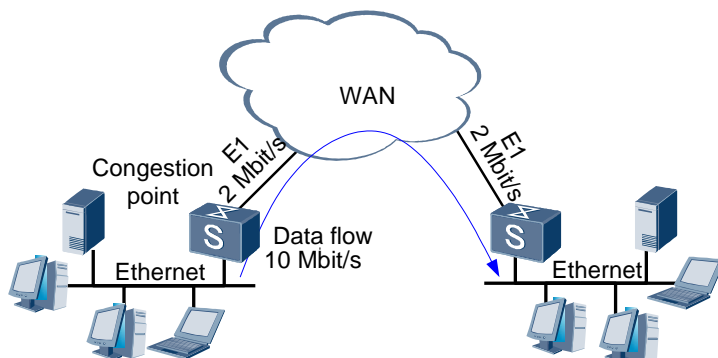
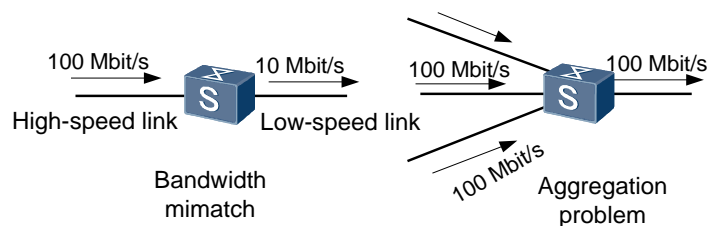


Figure 2-31 shows common traffic congestion causes:

- Traffic rate mismatch: Packets are transmitted to a device through a high-speed link and are forwarded out through a low-speed link.
- Traffic aggregation: Packets are transmitted from multiple interfaces to a device and are forwarded out through a single interface without enough bandwidth.

**Figure 2-31** Link bandwidth bottleneck





Traffic congestion is derived not only from link bandwidth bottleneck but also from any resource shortage, such as available processing time, buffer, and memory resource shortage. In addition, traffic is not satisfactorily controlled and exceeds the capacity of available network resources, also leading to traffic congestion.

Traffic congestion has the following adverse impacts on network traffic:

- Increases the delay and delay jitter of packet transmission.
- Causes packet retransmission due to overlong delays.
- Lowers the network throughput.
- Occupies a large number of network resources, especially the storage resource. Improper resource allocation may cause resources to be locked and the system to go Down.

Therefore, traffic congestion is the main cause of service deterioration. Since traffic congestion prevails on the PSN network, traffic congestion must be prevented or effectively controlled.

A solution to traffic congestion is a must on every network. A balance between limited network resources and user requirements is required so that user requirements are satisfied and network resources are fully used.

Congestion management and congestion avoidance are commonly used to relieve traffic congestion:

- Congestion management: provides means to manage and control traffic when traffic congestion occurs. Packets sent from one interface are placed into multiple queues that are marked with different priorities. The packets are sent based on the priorities. Different queue scheduling mechanisms are designed for different situations and lead to different results.
- Congestion avoidance: is a flow control technique used to relieve network overload. A system configured with congestion avoidance monitors network resources such as queues and memory buffers. When congestion occurs or aggravates, the system discards packets. Congestion avoidance prevents queue overflow due to line congestion.

## 2.4.2 Congestion Management

### Queue Overview

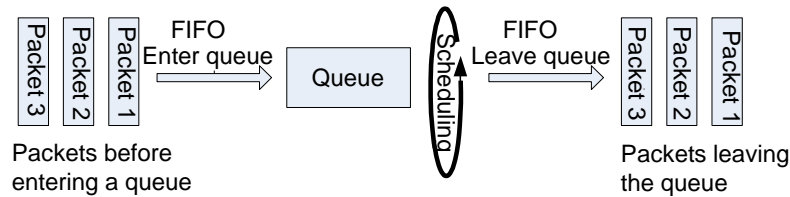
Congestion management defines a policy that determines the order in which packets are forwarded and specifies drop principles for packets. Queuing technology is generally used.

Queuing technology orders packets in the buffer. When the packet rate exceeds the interface bandwidth or the bandwidth allocated to the queue that buffers packets, the packets are buffered in queues and wait to be forwarded. The queue scheduling algorithm determines the time and order in which packets are leaving a queue and the relationships between queues.

Each interface on a Huawei switch stores eight downstream queues, which are called class queues (CQs) or port queues. The eight queues are BE, AF1, AF2, AF3, AF4, EF, CS6, and CS7.

The first in first out (FIFO) mechanism is used to transfer packets in a queue.

Figure 2-32 FIFO process



## Queue Scheduling

Huawei switches support various queue scheduling modes.

- Priority queuing (PQ)

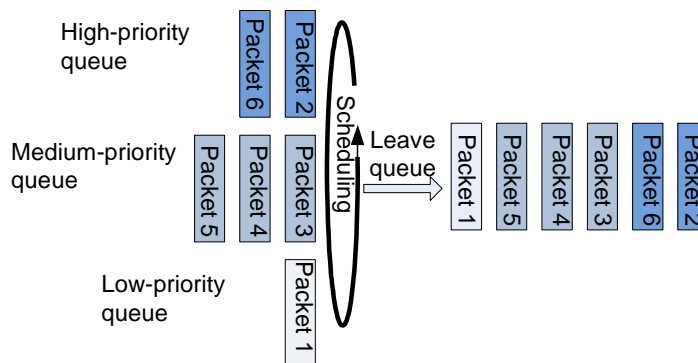
PQ schedules packets in descending order of priority. Packets in queues with a low priority can be scheduled only after all packets in queues with a high priority have been scheduled.

By using PQ scheduling, the device puts packets of delay-sensitive key services into queues with higher priorities and packets of other services into queues with lower priorities so that packets of key services can be transmitted first.

The disadvantage of PQ is that the packets in lower-priority queues are not processed until all the higher-priority queues are empty. As a result, a congested higher-priority queue causes all lower-priority queues to starve out.

As shown in Figure 2-33, three queues with a high, medium, and low priority respectively are configured with PQ scheduling. The number indicates the order in which packets arrive.

Figure 2-33 PQ scheduling

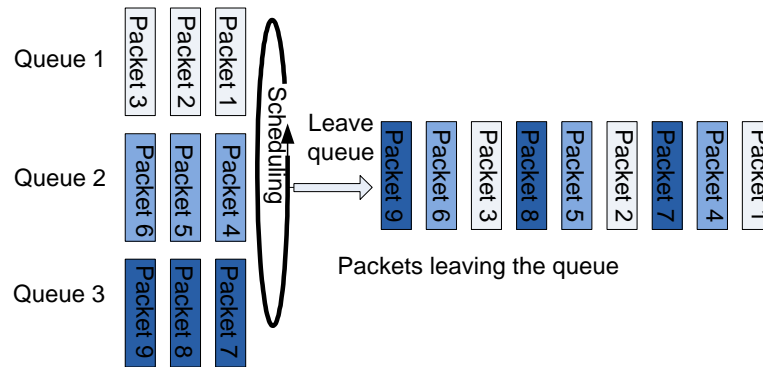


When packets leave queues, the device forwards the packets in descending order of priority. Packets in the higher-priority queue are forwarded preferentially. If packets in the higher-priority queue come in between packets in the lower-priority queue that is being scheduled, the packets in the high-priority queue are still scheduled preferentially. This implementation ensures that packets in the higher-priority queue are always forwarded preferentially. As long as there are packets in the high queue, no other queue will be served.

- Round robin (RR)

RR schedules multiple queues in ring mode. If the queue on which RR is performed is not empty, the scheduler takes one packet away from the queue. If the queue is empty, the queue is skipped, and the scheduler does not wait.

**Figure 2-34** RR scheduling



In RR scheduling, each queue has the same scheduling chance.

RR cannot differentiate queue priorities, and processes key services and non-key services in the same manner. As a result, key services cannot be processed in a timely manner.

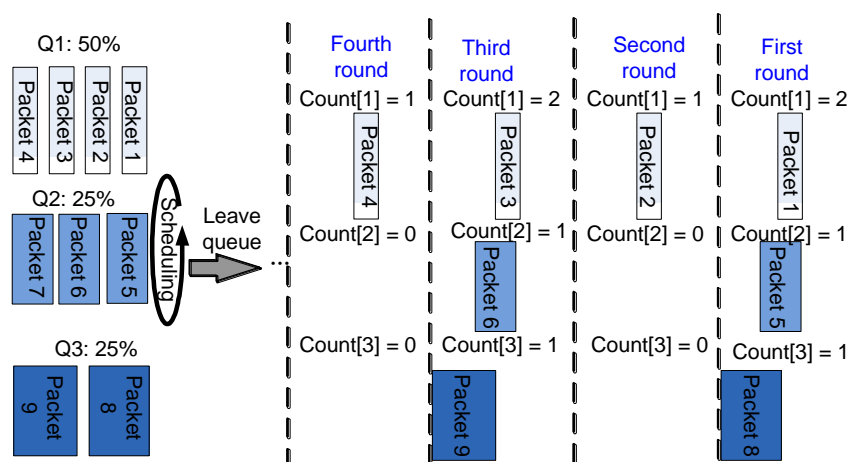
- Weighted round robin (WRR)

Compared with RR, WRR can set the weights of queues. During the WRR scheduling, the scheduling chance obtained by a queue is in direct proportion to the weight of the queue. RR scheduling functions the same as WRR scheduling in which each queue has a weight of 1.

WRR configures a counter for each queue and initializes the counter based on weights. Each time a queue is scheduled, a packet is taken away from the queue and being transmitted, and the counter decreases by 1. When the counter becomes 0, the device stops scheduling the queue and starts to schedule other queues with a non-0 counter.

When the counters of all queues become 0, all these counters are initialized again based on the weight, and a new round of WRR scheduling starts. In a round of WRR scheduling, the queues with the larger weights are scheduled more times.

**Figure 2-35** WRR scheduling



In an example, three queues with the weight 50%, 25%, and 25% respectively are configured with WRR scheduling.

The counters are initialized first: Count[1] = 2, Count[2] = 1, and Count[3] = 1.

– First round of WRR scheduling

Packet 1 is taken from queue 1, with Count[1] = 1. Packet 5 is taken from queue 2, with Count[2] = 0. Packet 8 is taken from queue 3, with Count[3] = 0.

– Second round of WRR scheduling

Packet 2 is taken from queue 1, with Count[1] = 0. Queues 2 and 3 do not participate in this round of WRR scheduling because Count [2] = 0 and Count[3] = 0.

Then, Count[1] = 0; Count[2] = 0; Count[3] = 0. The counters are initialized again: Count[1] = 2; Count[2] = 1; Count[3] = 1.

– Third round of WRR scheduling

Packet 1 is taken from queue 3, with Count[1] = 1. Packet 6 is taken from queue 2, with Count[2] = 0. Packet 9 is taken from queue 3, with Count[3] = 0.

– Fourth round of WRR scheduling

Packet 4 is taken from queue 1, with Count[1] = 0. Queues 2 and 3 do not participate in this round of WRR scheduling because Count [2] = 0 and Count[3] = 0.

Then, Count[1] = 0; Count[2] = 0; Count[3] = 0. The counters are initialized again: Count[1] = 2; Count[2] = 1; Count[3] = 1.

The statistics show that the number of times packets are scheduled in each queue is in direct ratio to the weight of this queue. A higher weight indicates more times packets are scheduled. If the interface bandwidth is 100 Mbit/s, the queue with the lowest weight can obtain a minimum bandwidth of 25 Mbit/s, preventing packets in the lower-priority queue from being starved out when SP scheduling is implemented.

During WRR scheduling, the empty queue is directly skipped. Therefore, when the rate at which packets arrive at a queue is low, the remaining bandwidth of the queue is used by other queues based on a certain proportion.

WRR scheduling has two disadvantages:

- WRR schedules packets based on the number of packets. Therefore, each queue has no fixed bandwidth. With the same scheduling chance, a long packet obtains higher bandwidth than a short packet. Users are generally sensitive to the bandwidth. When the average lengths of the packets in the queues are the same or known, users can obtain expected bandwidth by configuring WRR weights of the queues; however, when the average packet length of the queues changes, users cannot obtain expected bandwidth by configuring WRR weights of the queues.
- Delay-sensitive services, such as voice services, cannot be scheduled in a timely manner.

- Deficit round robin (DRR)

DRR implementation is similar to RR implementation.

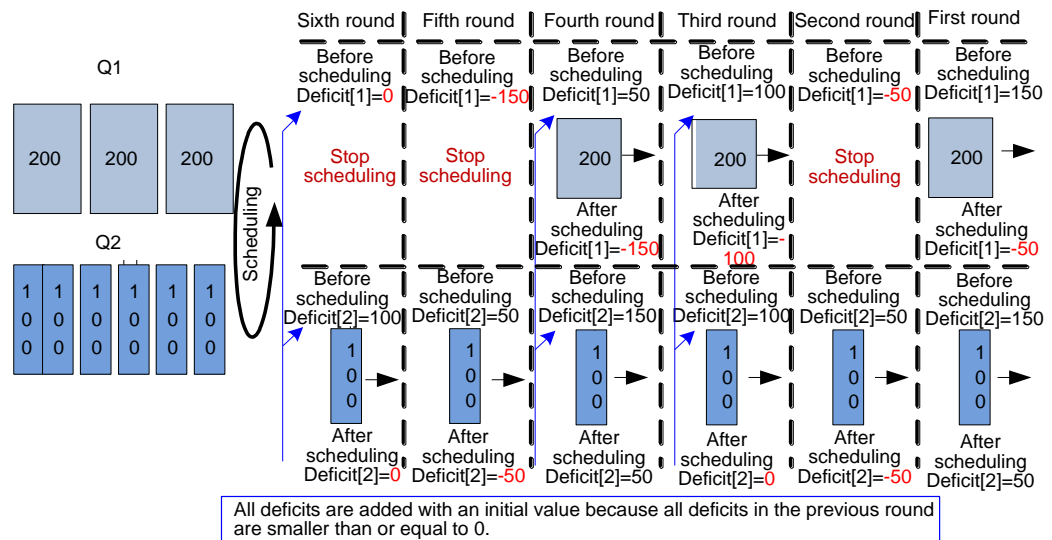
RR schedules packets based on the number of packets, whereas DRR schedules packets based on the packet length.

DRR configures a counter, which implies the number of excess bytes over the threshold (deficit) in the previous round for each queue. The counter is initialized as the maximum number of bytes (generally the interface MTU) allowed in a round of DRR scheduling. Each time a queue is scheduled, a packet is taken away from the queue, and the counter decreases by packet length. If a packet is too long for the queue scheduling capacity, DRR allows the counter to be a negative. This ensures that long packets can be

scheduled. In the next round of scheduling, however, this queue will not be scheduled. When the counter becomes 0 or a negative, the device stops scheduling the queue and starts to schedule other queues with a positive counter. When the counters of all queues become 0 or negatives, all these counters are initialized, and a new round of DRR scheduling starts.

In an example, the MTU of an interface is 150 bytes. Two queues Q1 and Q2 use DRR scheduling. Multiple 200-byte packets are buffered in Q1, and multiple 100-byte packets are buffered in Q2. Figure 2-36 shows how DRR schedules packets in these two queues.

Figure 2-36 DRR scheduling



As shown in Figure 2-37, after six rounds of DRR scheduling, three 200-byte packets in Q1 and six 100-byte packets in Q2 are scheduled. The output bandwidth ratio of Q1 to Q2 is actually 1:1.

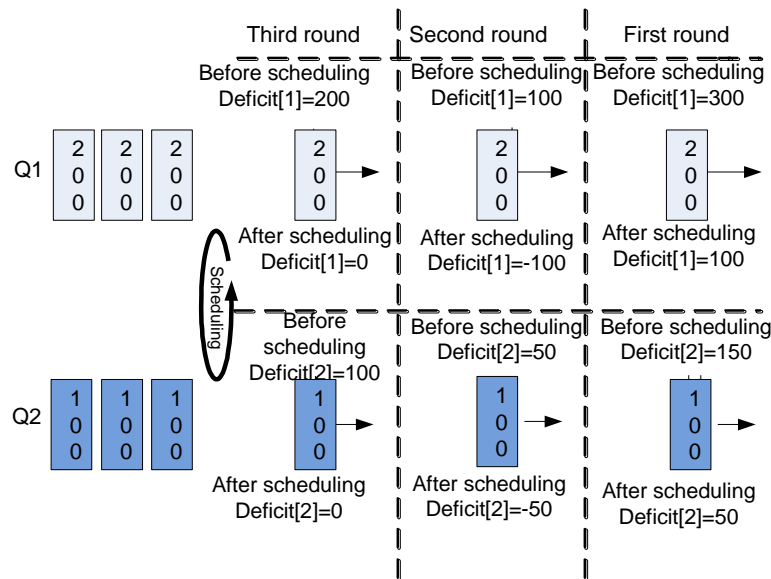
Unlike PQ scheduling, DRR scheduling prevents packets in low-priority queues from being starved out. However, DRR scheduling cannot set weights of queues and cannot schedule delay-sensitive services such as voice services in a timely manner.

- Deficit weighted round robin (DWRR)

Compared with DRR, DWRR can set the weights of queues. DRR scheduling functions the same as DWRR scheduling in which each queue has a weight of 1.

DWRR configures a counter, which implies the number of excess bytes over the threshold (deficit) in the previous round for each queue. The counter is initialized as the weight multiplied by the MTU. Each time a queue is scheduled, a packet is taken away from the queue, and the counter decreases by packet length. When the counter becomes 0, the device stops scheduling the queue and starts to schedule other queues with a non-0 counter. When the counters of all queues become 0, all these counters are initialized as the weight multiplied by the MTU, and a new round of DWRR scheduling starts.

Figure 2-37 DWRR scheduling



In an example, the MTU of an interface is 150 bytes. Two queues Q1 and Q2 use DRR scheduling. Multiple 200-byte packets are buffered in Q1, and multiple 100-byte packets are buffered in Q2. The weight ratio of Q1 to Q2 is 2:1. Figure 2-37 shows how DWRR schedules packets.

– First round of DWRR scheduling

The counters are initialized as follows: Deficit[1] = weight1 x MTU = 300 and Deficit[2] = weight2 x MTU=150. A 200-byte packet is taken from Q1, and a 100-byte packet is taken from Q2. Then, Deficit[1] = 100 and Deficit[2] = 50.

– Second round of DWRR scheduling

A 200-byte packet is taken from Q1, and a 100-byte packet is taken from Q2. Then, Deficit[1] = -100 and Deficit[2] = -50.

– Third round of DWRR scheduling

The counters of both queues are negatives. Therefore, Deficit[1] = Deficit[1] + weight1 x MTU = -100 + 2 x 150 = 200 and Deficit[2] = Deficit[2] + weight2 x MTU = -50 + 1 x 150 = 100.

A 200-byte packet is taken from Q1, and a 100-byte packet is taken from Q2. Then, Deficit[1] = 0 and Deficit[2] = 0.

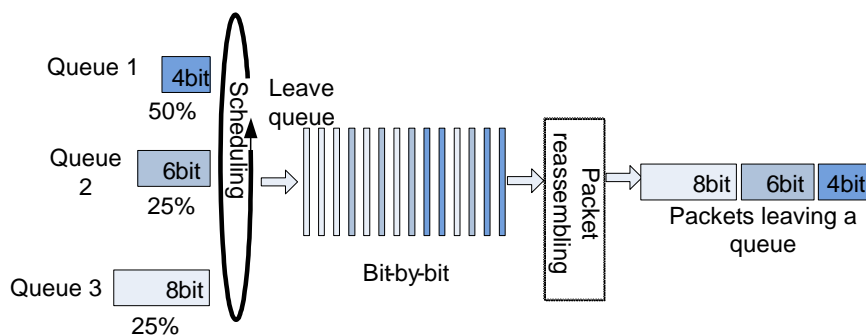
As shown in Figure 2-37, after three rounds of DWRR scheduling, three 200-byte packets in Q1 and three 100-byte packets in Q2 are scheduled. The output bandwidth ratio of Q1 to Q2 is actually 2:1, which conforms to the weight ratio.

DWRR scheduling prevents packets in low-priority queues from being starved out and allows bandwidths to be allocated to packets based on the weight ratio when the lengths of packets in different queues vary or change greatly. However, DWRR scheduling does not schedule delay-sensitive services such as voice services in a timely manner.

- Weighted fair queuing (WFQ)

WFQ allocates bandwidths to flows based on the weight. In addition, to allocate bandwidths fairly to flows, WFQ schedules packets in bits. Figure 2-38 shows how bit-by-bit scheduling works.

**Figure 2-38** WFQ scheduling



The bit-by-bit scheduling mode shown in Figure 2-38 allows the device to allocate bandwidths to flows based on the weight. This prevents long packets from preempting bandwidths of short packets and reduces the delay and jitter when both short and long packets wait to be forwarded.

The bit-by-bit scheduling mode, however, is an ideal one. A Huawei switch performs WFQ scheduling based on a certain granularity, such as 256 B and 1 KB. Different boards support different granularities.

WFQ has the following advantages:

- Different queues obtain the scheduling chances fairly, balancing delays of flows.
- Short and long packets obtain the scheduling chances fairly. If both short and long packets wait in queues to be forwarded, short packets are scheduled preferentially, reducing jitters of flows.
- The lower the weight of a flow is, the lower the bandwidth the flow obtains.



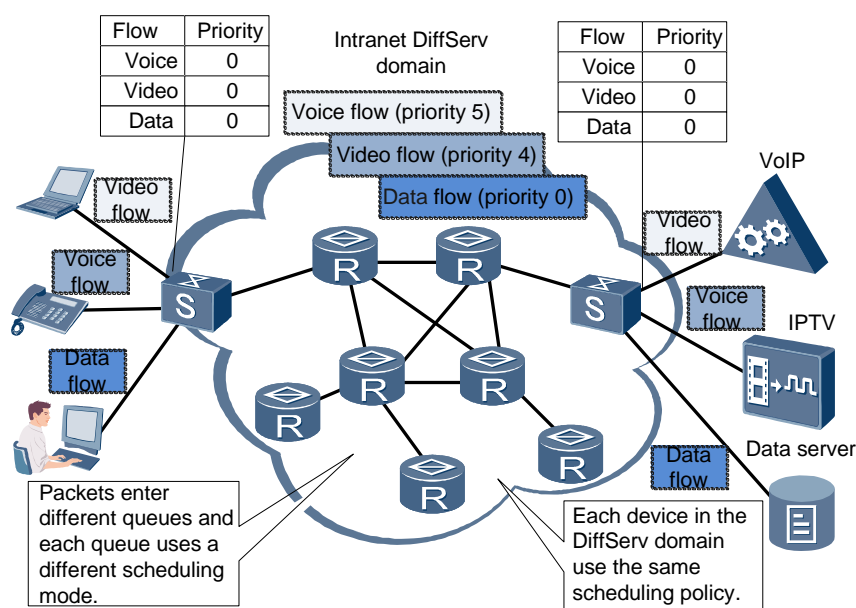
**NOTE**

Only WAN boards on the S9700 and S7700 support WFQ scheduling.

## Congestion Management Application

Congestion management is also called queue scheduling and is commonly used in QoS solutions.

**Figure 2-39** Queue scheduling



In a DS domain, different services are marked with different internal and external priorities and put into different queues. Different queues use different scheduling policies. The device provides differentiated services for packets based on queue scheduling policies.

Generally, high-priority queues use PQ so that the queues can be scheduled preferentially and are not affected by low-priority queues. Low-priority queues use round scheduling, which prevents low-priority queues from starving out. When WRR and DWRR are not configured, WRR is used on a Huawei switch by default and the queue weight is 1.

## 2.4.3 Congestion Avoidance

Congestion avoidance is a flow control technique used to relieve network overload. A system configured with congestion avoidance monitors network resources such as queues and memory buffers. When congestion occurs or aggravates, the system discards packets.

Huawei switches support two drop policies:

- Tail drop
- Weighted Random Early Detection (WRED)

### Tail Drop

Tail drop is the traditional congestion avoidance mechanism that processes all packets equally without classifying the packets into different types. When congestion occurs, packets at the end of a queue are discarded until the congestion problem is solved.

Tail drop causes global TCP synchronization. In tail drop mechanisms, all newly arrived packets are dropped when congestion occurs, causing all TCP sessions to simultaneously enter the slow start state and the packet transmission to slow down. Then all TCP sessions restart their transmission at roughly the same time and then congestion occurs again, causing another burst of packet drops, and all TCP sessions enters the slow start state again. The behavior cycles constantly, severely reducing the network resource usage.



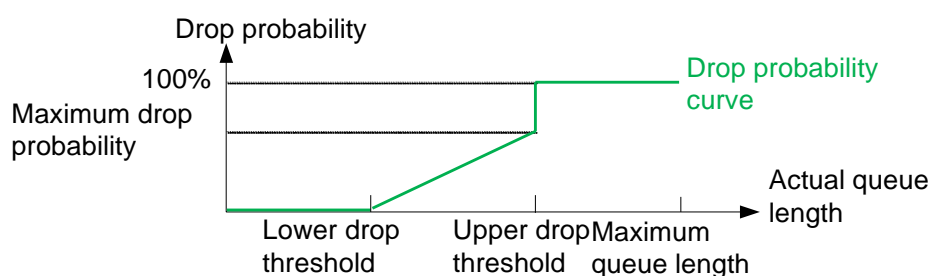
## WRED

WRED is a congestion avoidance mechanism used to drop packets before the queue overflows. WRED resolves global TCP synchronization by randomly dropping packets to prevent a burst of TCP retransmission. If a TCP connection reduces the transmission rate when packet loss occurs, other TCP connections still keep a high rate for sending packets. The WRED mechanism improves the bandwidth use efficiency.

WRED sets lower and upper drop thresholds for each queue and defines the following rules:

- When the length of a queue is lower than the lower drop threshold, no packet is dropped.
- When the length of a queue exceeds the upper drop threshold, all newly arrived packets are tail dropped.
- When the length of a queue ranges from the lower drop threshold to the upper drop threshold, newly arrived packets are randomly dropped, but a maximum drop probability is set. The maximum drop probability refers to the drop probability when the queue length reaches the upper drop threshold. Figure 2-40 is a drop probability graph. The longer the queue, the larger the drop probability.

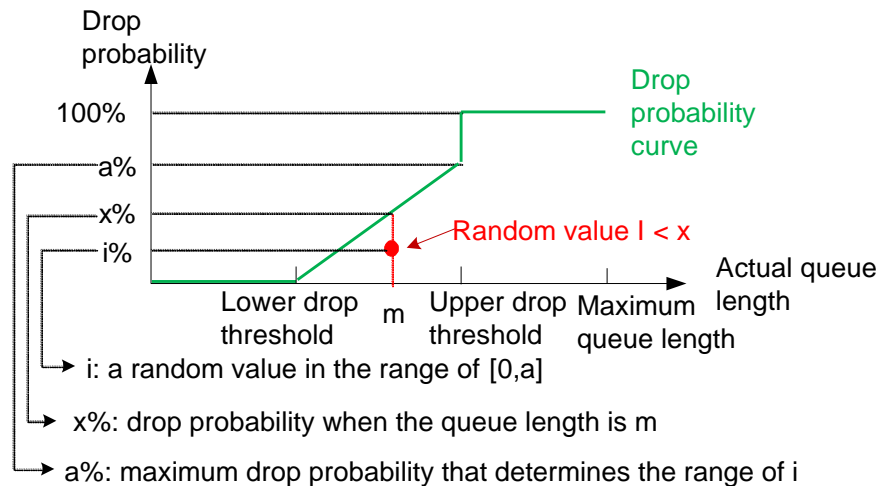
**Figure 2-40** WRED drop probability



As shown in Figure 2-41, the maximum drop probability is  $a\%$ , the length of the current queue is  $m$ , and the drop probability of the current queue is  $x\%$ . WRED delivers a random value  $i$  to each arrived packet, ( $0 < i\% < \text{maximum drop probability}$ ), and compares the random value with the drop probability of the current queue. If the random value  $i$  ranges from 0 to  $x$ , the newly arrived packet is dropped; if the random value ranges from  $x$  to  $a$ , the newly arrived packet is not dropped.

An example is that the lower drop threshold is 40%, upper drop threshold is 80%, current queue length is 50% of the total queue length, and maximum drop probability is 5%. If a packet enters the queue, a random value in the range of 0 to 20 is assigned to the packet. If the random value ranges from 0 to 5, the packet is discarded. If the random value ranges from 5 to 20, the packet enters the queue. Because the current queue length is much less than the upper drop threshold, the drop probability of the packet is low. If the current queue length is 75% of the total queue length, the drop probability of the packet is 18%. When the random value ranges from 0 to 18, the drop probability of the packet becomes high.

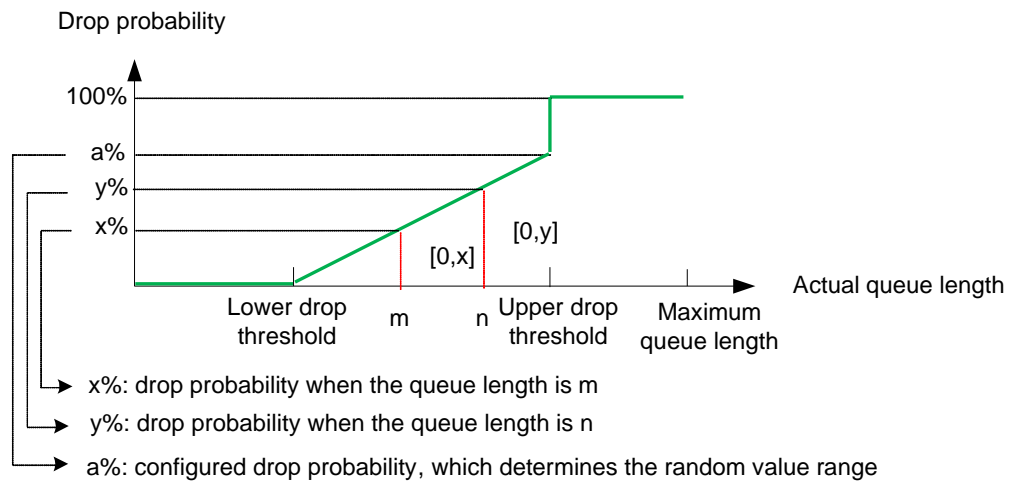
**Figure 2-41** WRED implementation



As shown in Figure 2-42, the drop probability of the queue with the length  $m$  (lower drop threshold  $< m <$  upper drop threshold) is  $x\%$ . If the random value ranges from 0 to  $x$ , the newly arrived packet is dropped. The drop probability of the queue with the length  $n$  ( $m < n <$  upper drop threshold) is  $y\%$ . If the random value ranges from 0 to  $y$ , the newly arrived packet is dropped. The range of 0 to  $y$  is wider than the range of 0 to  $x$ . There is a higher probability that the random value falls into the range of 0 to  $y$ . Therefore, the longer the queue, the higher the drop probability.

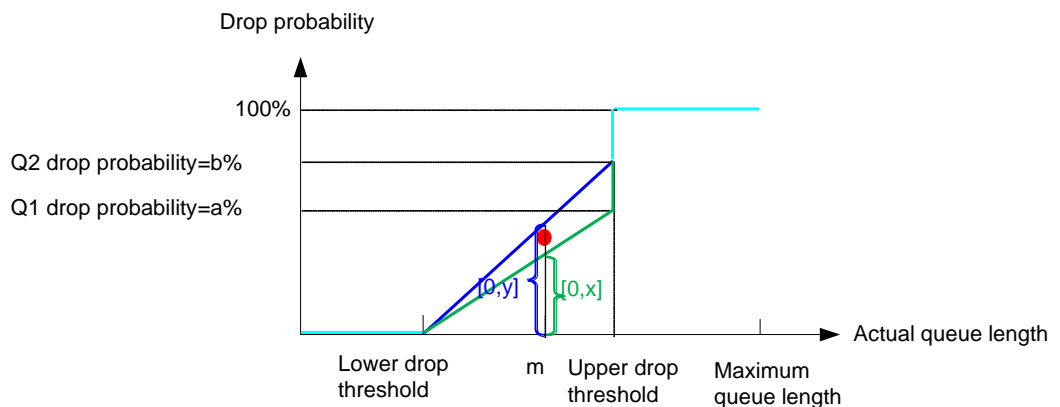
An example is that the drop probability is 12% when the queue length is 50% of the total queue length. When the random value ranges from 0 to 12, the newly arrived packet is discarded. When the queue length is 60%, the drop probability is 15%. When the random value ranges from 0 to 15, the newly arrived packet is discarded. The range of 0 to 15 is wider than the range of 0 to 12. There is a higher probability that the random value falls into the range of 0 to 15. Therefore, the longer the queue, the higher the drop probability.

**Figure 2-42** Drop probability change with the queue length



As shown in Figure 2-43, the maximum drop probabilities of two queues Q1 and Q2 are a% and b%, respectively. When the length of Q1 and Q2 is m, the drop probabilities of Q1 and Q2 are respectively x% and y%. If the random value ranges from 0 to x, the newly arrived packet in Q1 is dropped. If the random value ranges from 0 to y, the newly arrived packet in Q2 is dropped. The range of 0 to y is wider than the range of 0 to x. There is a higher probability that the random value falls into the range of 0 to y. Therefore, when the queue lengths are the same, the higher the maximum drop probability, the higher the drop probability.

**Figure 2-43** Drop probability change with the maximum drop probability



## WRED Parameter Setting

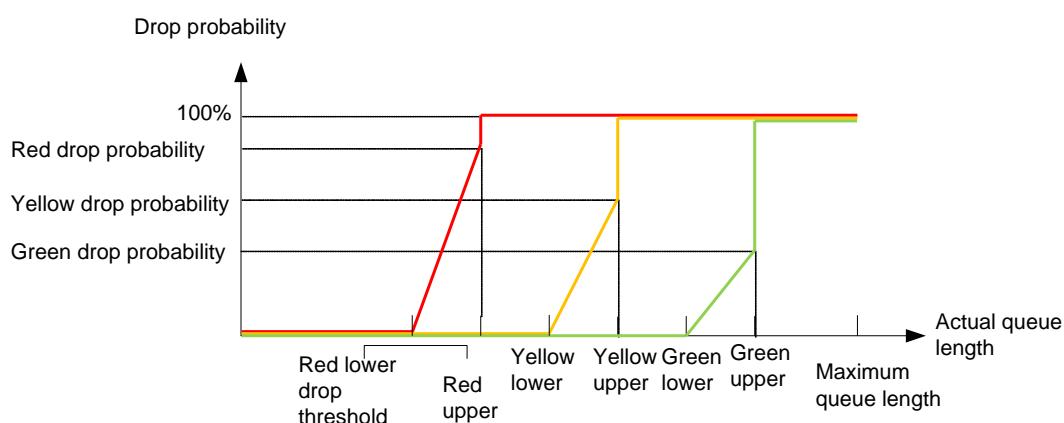
Tail drop applies to services that have high requirements for real-time performance. This is because packets of such services require bandwidth guarantee. Tail drop drops packets only when the queue overflows. In addition, PQ queues preempt bandwidths of other queues. Therefore, when traffic congestion occurs, highest bandwidths can be provided for real-time services.

WRED applies to WFQ queues. WFQ queues share bandwidth based on the weight and are prone to traffic congestion. Using WRED for WFQ queues effectively resolves global TCP synchronization when traffic congestion occurs.

- **WRED lower and upper drop thresholds and drop probability**

In real-world situations, it is recommended that the WRED lower drop threshold starts from 50% and changes with the drop precedence. As shown in Figure 2-44, a lowest drop probability and highest lower and upper drop thresholds are recommended for green packets; a medium drop probability and medium lower and upper drop thresholds are recommended for yellow packets; a highest drop probability and lowest lower and upper drop thresholds are recommended for red packets. When traffic congestion aggravates, red packets are first dropped due to lower drop threshold and high drop probability. As the queue length increases, the device drops green packets at last. If the queue length reaches the upper drop threshold for red/yellow/green packets, red/yellow/green packets respectively start to be tail dropped.

**Figure 2-44** WRED drop probability for three drop precedences



- **Maximum queue length setting**

The maximum queue length can be set using the **qos queue queue-index length length-value** command on Huawei switches. When traffic congestion occurs, packets accumulate in the buffer and are delayed. The delay is determined by the queue buffer size and the output bandwidth allocated to a queue. When the output bandwidths are the same, the shorter the queue, the lower the delay.

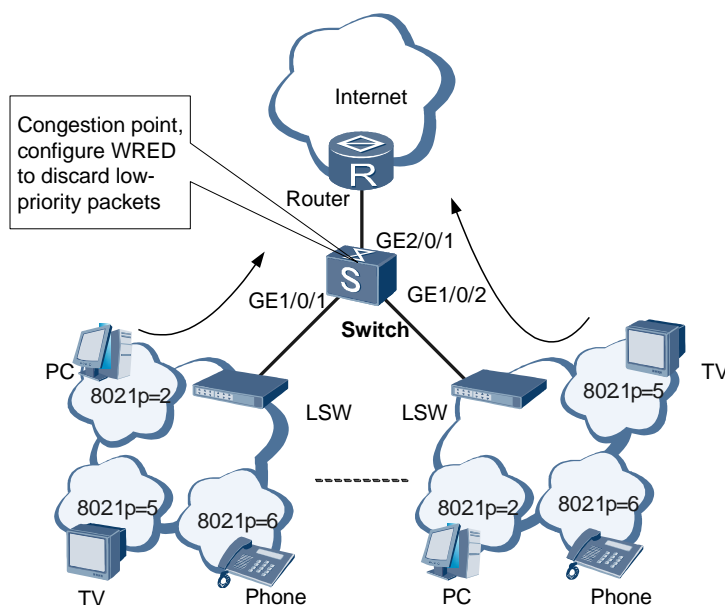
The queue length cannot be set too small. If the length of a queue is too small, the buffer is not enough even if the traffic rate is low. As a result, packet loss occurs. The shorter the queue, the less the tolerance of burst traffic.

The queue length cannot be set too large. If the length of a queue is too large, the delay increases along with it. In particular, when a TCP connection is set up, one end sends a packet to the peer end and waits for a response. If no response is received within the timer timeout period, the TCP sender retransmits the packet. If a packet is buffered for a long time, the packet has no difference with the dropped ones.

## Congestion Avoidance Application

2.4.1 Background describes the scenario where congestion occurs. Generally, WRED is configured on an access switch interface connected to an aggregation device, aggregation switch, or router interface connected to an upstream device. As shown in Figure 2-45, WRED applies to GE2/0/1 on the switch.

**Figure 2-45** Congestion avoidance



WRED is configured in the outbound direction on an interface and applies to queues. Different drop parameters are set for packets of different colors. The upper and lower drop thresholds of important packets are higher than those non-important packets, whereas the maximum drop probability of important packets is lower.

The following WRED parameter settings are recommended for packets of different colors.

**Table 2-22** Recommended WRED parameter settings

Queue (PHB)	Lower Drop Threshold (%)	Upper Drop Threshold (%)	Maximum Drop Probability
Green	80	100	10
Yellow	60	80	20
Red	40	60	30

WRED takes effect after queues are scheduled, so the preceding settings are recommended for each queue.

**Table 2-23** Recommended WRED parameter settings based on queues

<b>Queue (PHB)</b>	<b>Lower Drop Threshold (%)</b>	<b>Upper Drop Threshold (%)</b>	<b>Maximum Drop Probability</b>
High priority (CS7, CS6)	80	100	10
Medium priority (EF, AF1-AF4)	60	80	20
Low priority (BE)	40	60	30

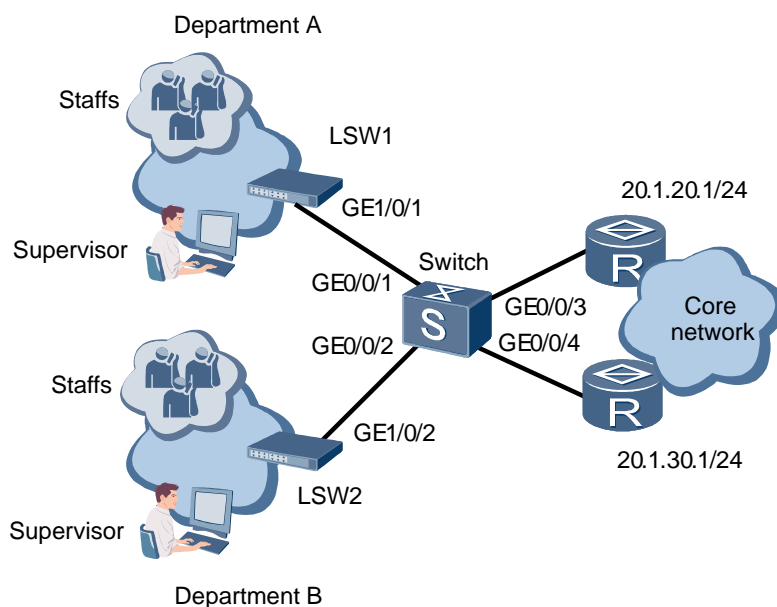
# 3 Application Scenarios

## 3.1 User-based Differentiated Services

### 3.1.1 Networking Requirements

As shown in Figure 3-1, an enterprise has two departments: A and B. Each department has its supervisor and staffs. Supervisors and staffs work together. Enterprise users are dual-homed to external network devices through the switch. Among the two links, one link is the low-speed link and the gateway address is 20.1.20.1/24, and the other link is the high-speed link and the gateway address is 20.1.30.1/24.

**Figure 3-1** User-based differentiated services



The requirements are as follows:

- Supervisors use the high-speed link to access the network, and staffs use the low-speed link to access the network.

- The staffs in department A have higher priority than those in department B, and the supervisor in department A has higher priority than that in department B. The priority of supervisors is higher than that of staffs.

## 3.1.2 Configuration Roadmap

The two departments belong to a large-scale LAN but on different network segments. 802.1p priority re-marking and redirection can be configured to implement PBR so that differentiated services can be provided. The configuration roadmap is as follows:

- Make an overall plan.
  - Assign a VLAN to each department: VLAN 10 to department A and VLAN 20 to department B.
  - The MAC addresses of supervisors' PCs in departments A and B are 0001-0001-0001 and 0002-0002-0002 respectively.
  - Create VLANIF interfaces on the switch and assign IP addresses to the VLANIF interfaces to implement interworking.
- Make Layer 2 plan.
  - Create VLANs on LSW1 and LSW2, and configured interfaces to implement Layer 2 interworking of departments.
  - Configure traffic classifiers on LSW1 and LSW2 to classify packets based on MAC addresses.
  - Configure traffic behaviors on LSW1 and LSW2 to re-mark packets from supervisors' packets with higher 802.1p priorities.
  - Configure traffic policies on LSW1 and LSW2, bind configured traffic classifiers and traffic behaviors to the traffic policies, apply the traffic policies to GE1/0/1 and GE1/0/2 respectively.
- Make Layer 3 plan.
  - Create VLANs and configure interfaces on the switch so that enterprise branches can communicate and access the network through the switch.
  - Configure traffic classifiers on the switch to classify packets based on 802.1p priorities.
  - Configure traffic behaviors on the switch to redirect packets from supervisors' PCs to 20.1.30.1/24 and packets from staffs' PCs to 20.1.20.1/24 and to mark different packets with different IP precedences.
  - Configure traffic policies on the switch, bind configured traffic classifiers and traffic behaviors to the traffic policies, apply the traffic policies to the inbound direction on GigabitEthernet 0/0/1 and GigabitEthernet 0/0/2 respectively to implement PBR.

## 3.1.3 Procedure

**Step 1** Create VLANs and configure interfaces.

# Create VLAN 10 on LSW1 and add GE1/0/1 to VLAN 10.

```
<Quidway> system-view
[Quidway] sysname LSW1
[LSW1] vlan 10
[LSW1-vlan10] quit
[LSW1] interface GigabitEthernet 1/0/1
[LSW1-GigabitEthernet1/0/1] port link-type trunk
[LSW1-GigabitEthernet1/0/1] port trunk allow-pass vlan 10
```



```
[LSW1-GigabitEthernet1/0/1] quit

# Create VLAN 20 on LSW2 and add GE1/0/2 to VLAN 20. The configuration of LSW2 is
similar to the configuration of LSW1, and is not mentioned here.

# Create VLAN 10, VLAN 20, VLAN 100, and VLAN 200 on the switch.

<Quidway> system-view
[Quidway] sysname Switch
[Switch] vlan batch 10 20 100 200

# Configure GE0/0/1, GE0/0/2, GE0/0/3, and GE0/0/4 on the switch as trunk interfaces and
add them to VLANs.

[Switch] interface GigabitEthernet 0/0/1
[Switch-GigabitEthernet0/0/1] port link-type trunk
[Switch-GigabitEthernet0/0/1] port trunk allow-pass vlan 10
[Switch-GigabitEthernet0/0/1] quit
[Switch] interface GigabitEthernet 0/0/2
[Switch-GigabitEthernet0/0/2] port link-type trunk
[Switch-GigabitEthernet0/0/2] port trunk allow-pass vlan 20
[Switch-GigabitEthernet0/0/2] quit
[Switch] interface GigabitEthernet 0/0/3
[Switch-GigabitEthernet0/0/3] port link-type trunk
[Switch-GigabitEthernet0/0/3] port trunk allow-pass vlan 100
[Switch-GigabitEthernet0/0/3] quit
[Switch] interface GigabitEthernet 0/0/4
[Switch-GigabitEthernet0/0/4] port link-type trunk
[Switch-GigabitEthernet0/0/4] port trunk allow-pass vlan 200
[Switch-GigabitEthernet0/0/4] quit

# Create VLANIF 10, VLANIF 20, VLANIF 100, and VLANIF 200 and configure IP
addresses for them.

[Switch] interface vlanif 10
[Switch-Vlanif10] ip address 192.168.10.1 24
[Switch-Vlanif10] quit
[Switch] interface vlanif 20
[Switch-Vlanif20] ip address 192.168.20.1 24
[Switch-Vlanif20] quit
[Switch] interface vlanif 100
[Switch-Vlanif100] ip address 20.1.20.2 24
[Switch-Vlanif100] quit
[Switch] interface vlanif 200
[Switch-Vlanif200] ip address 20.1.30.2 24
[Switch-Vlanif200] quit
```

## Step 2 Configure traffic policies on LSW1 and LSW2.

# Configure a traffic classifier on LSW1.

```
[LSW1] traffic classifier lsw1
[LSW1-classifier-lsw1] if-match source-mac 0001-0001-0001
[LSW1-classifier-lsw1] quit
```

# Configure a traffic behavior on LSW1.

```
[LSW1] traffic behavior lsw1
[LSW1-behavior-lsw1] remark 8021p 5
```

```
[LSW1-behavior-lsw1] quit

# Configure a traffic policy on LSW1 and apply it in the inbound direction of GigabitEthernet
1/0/1.

[LSW1] traffic policy lsw1
[LSW1-trafficpolicy-lsw1] classifier lsw1 behavior lsw1
[LSW1-classifier-lsw1] quit
[LSW1] interface GigabitEthernet 1/0/1
[LSW1-GigabitEthernet1/0/1] traffic-policy lsw1 inbound
[LSW1-GigabitEthernet1/0/1] quit

# The configuration of LSW2 is similar to the configuration of LSW1, and is not mentioned
here.
```

### Step 3 Configure traffic policies on the switch.

```
# Configure traffic classifiers on the switch.

[Switch] traffic classifier switch1
[Switch-classifier-switch1] if-match 8021p 5
[Switch-classifier-switch1] quit
[Switch] traffic classifier switch2
[Switch-classifier-switch2] if-match any
[Switch-classifier-switch2] quit

# Configure traffic behaviors on the switch.

[Switch] traffic behavior switch1
[Switch-behavior-switch1] remark local-precedence af1
[Switch-behavior-switch1] redirect ip-nexthop 20.1.20.1
[Switch-behavior-switch1] quit
[Switch] traffic behavior switch2
[Switch-behavior-switch2] remark local-precedence af2
[Switch-behavior-switch2] redirect ip-nexthop 20.1.20.1
[Switch-behavior-switch2] quit
[Switch] traffic behavior switch3
[Switch-behavior-switch3] remark local-precedence af3
[Switch-behavior-switch3] redirect ip-nexthop 20.1.30.1
[Switch-behavior-switch3] quit
[Switch] traffic behavior switch4
[Switch-behavior-switch4] remark local-precedence af4
[Switch-behavior-switch4] redirect ip-nexthop 20.1.30.1
[Switch-behavior-switch4] quit

# Configure traffic policies on the switch and apply them in the inbound direction of
GigabitEthernet 0/0/1 and GigabitEthernet 0/0/2.

[Switch] traffic policy switch1 match-order config
[Switch-trafficpolicy-switch1] classifier switch1 behavior switch1
[Switch-trafficpolicy-switch1] classifier switch2 behavior switch3
[Switch-trafficpolicy-switch1] quit
[Switch] traffic policy switch2 match-order config
[Switch-trafficpolicy-switch2] classifier switch1 behavior switch2
[Switch-trafficpolicy-switch2] classifier switch2 behavior switch4
[Switch-trafficpolicy-switch2] quit
[Switch] interface GigabitEthernet 0/0/1
[Switch-GigabitEthernet0/0/1] traffic-policy switch1 inbound
[Switch-GigabitEthernet0/0/1] quit
```

```
[Switch] interface GigabitEthernet 0/0/2  
[Switch-GigabitEthernet0/0/2] traffic-policy switch2 inbound  
[Switch-GigabitEthernet0/0/2] quit
```

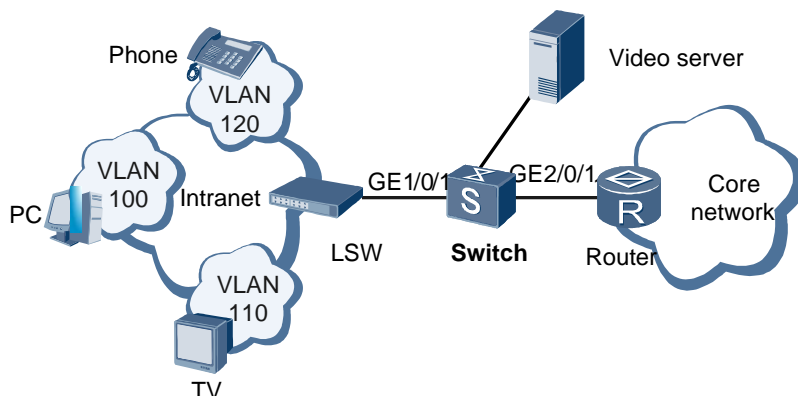
----End

## 3.2 Service-based Differentiated Services

### 3.2.1 Networking Requirements

Figure 3-2 shows the network of a small-scale enterprise. In working hours (8:00 to 18:00 from Monday to Friday), email and voice services are mainly transmitted, and employees are not allowed to access external networks and browse entertainment videos. In off-work hours, employees are allowed to access external networks and browse entertainment videos with limited bandwidth. In addition, email and voice services cannot be affected.

**Figure 3-2** Service-based differentiated services



The total leased bandwidth of each service is within 10000 kbit/s.

Voice, video, and data service are transmitted in VLAN 120, VLAN 110, and VLAN 100 respectively.

Traffic policing needs to be configured on the switch to police packets of different services so that traffic is limited in a range and bandwidth of each service is guaranteed.

Voice services, data services, and video and HTTP services have QoS requirements in descending order of priority. The switch needs to re-mark DSCP priorities in different service packets so that the downstream router processes them based on priorities, ensuring QoS of different services.

## 3.2.2 Configuration Roadmap

QoS parameters are set for different services.

**Table 3-1** QoS parameters for different services

Traffic Type	CIR (kbit/s)	PIR (kbit/s)	DSCP Priority	Time Range
Voice	2000	10000	46	All
Email	2000	10000	26	All
Internal video	3000	10000	18	Off-work hours
External data	3000	10000	12	Off-work hours

The configuration roadmap is as follows:

1. Create VLANs and configure interfaces so that the enterprise can access the network through the switch.
2. Configure a CAR profile to limit traffic within 10000 kbit/s.
3. Configure traffic classification rules based on VLAN IDs on the switch. Voice, video, and data service are transmitted in VLAN 120, VLAN 110, and VLAN 100 respectively, and email services are differentiated based on SMTP port numbers.
4. Configure time ranges for video and non-email services on the switch.
5. Configure traffic behaviors on the switch to limit the rate of packets and re-mark DSCP priorities of packets.
6. Configure a traffic policy on the switch, bind traffic behaviors and traffic classifiers, and apply the traffic policy to the inbound direction of the interface on the LSW connected to the switch.

## 3.2.3 Procedure

**Step 1** Create VLANs and configure interfaces.

```
# Create VLAN 100, VLAN 110, and VLAN 120 on the switch and add GE1/0/1 and GE2/0/1 to VLANs. The configuration details are not mentioned here.
```

**Step 2** Configure a CAR profile and define a time range.

```
[Switch] qos car car1 cir 10000
[Switch] time-range work 8:00 to 18:00 working-day
```

**Step 3** Configure a traffic policy on the switch.

```
# Configure traffic classifiers on the switch.
```

```
[Switch] traffic classifier Voice
[Switch-classifier-voice] if-match vlan-id 120
[Switch-classifier-voice] quit
[Switch] acl 4001
[Switch-acl-L2-4001] rule deny vlan-id 100 time-range work
[Switch-acl-L2-4001] rule permit vlan-id 100
[Switch-acl-L2-4001] quit
[Switch] acl 4002
```

```
[Switch-acl-L2-4002] rule deny vlan-id 110 time-range work
[Switch-acl-L2-4002] rule permit vlan-id 110
[Switch-acl-L2-4002] quit
[Switch] acl 3001
[Switch-acl-adv-3001] rule permit tcp destination-port eq smtp
[Switch-acl-adv-3001] quit
[Switch] traffic classifier Mail operator and
[Switch-classifier-mail] if-match vlan-id 100
[Switch-classifier-mail] if-match acl 3001
[Switch-classifier-mail] quit
[Switch] traffic classifier Data
[Switch-classifier-data] if-match acl 4001
[Switch-classifier-data] quit
[Switch] traffic classifier Video
[Switch-classifier-video] if-match acl 4002
[Switch-classifier-video] quit
```

# Configure traffic behaviors on the switch.

```
[Switch] traffic behavior Voice
[Switch-behavior-voice] car cir 2000 pir 10000 green pass
[Switch-behavior-voice] car car1 share
[Switch-behavior-voice] remark dscp 46
[Switch-behavior-voice] quit
[Switch] traffic behavior Mail
[Switch-behavior-mail] car cir 2000 pir 10000 green pass
[Switch-behavior-mail] car car1 share
[Switch-behavior-mail] remark dscp 26
[Switch-behavior-mail] quit
[Switch] traffic behavior Data
[Switch-behavior-data] car cir 3000 pir 10000 green pass
[Switch-behavior-data] car car1 share
[Switch-behavior-data] remark dscp 12
[Switch-behavior-data] quit
[Switch] traffic behavior Video
[Switch-behavior-video] car cir 3000 pir 10000 green pass
[Switch-behavior-video] car car1 share
[Switch-behavior-video] remark dscp 18
[Switch-behavior-video] quit
```

# Configure a traffic policy and apply it to an interface.

```
[Switch] traffic policy Switch match-order config
[Switch-trafficpolicy-switch] classifier Voice behavior Voice
[Switch-trafficpolicy-switch] classifier Mail behavior Mail
[Switch-trafficpolicy-switch] classifier Data behavior Data
[Switch-trafficpolicy-switch] classifier Video behavior Video
[Switch-trafficpolicy-switch] quit
[Switch] interface gigabitethernet 1/0/1
[Switch-GigabitEthernet1/0/1] traffic-policy Switch inbound
[Switch-GigabitEthernet1/0/1] quit
```

----End

# 4 Troubleshooting Cases

## 4.1 Packets Enter Incorrect Queues

### Common Causes

This fault is commonly caused by one of the following:

- Priority mapping configured in the DiffServ domain trusted by the inbound interface is incorrect.
- There are configurations affecting the queues that packets enter on the inbound interface.
- There are configurations affecting the queues that packets enter in the VLAN that packets belong to.
- There are configurations affecting the queues that packets enter in the system.



#### NOTE

On the ES1D2X40SFC0 and ES1D2L02QFC0 of the S7700 and EH1D2X40SFC0 and EH1D2L02QFC0 of the S9700, only queues 0, 1, 2, and 6 are available for unknown unicast packets, and queues 0 to 7 are available for known unicast packets.

### Procedure

**Step 1** Check whether the priority mapping configuration is correct.

Run the **display this** command in the inbound interface view and check the configuration of the **trust upstream** command. Then run the **display diffserv domainname domain-name** command to check whether the priority mapping configured in the DiffServ domain trusted by the inbound interface is correct.

- If not, run the **ip-dscp-inbound**, **mpls-exp-inbound**, or **8021p-inbound** command to correctly configure priority mapping.
- If so, go to step 2.

**Step 2** Check whether there are configurations affecting the queues that packets enter on the inbound interface.

The following configurations affect the queues that packets enter on the inbound interface:

- If the **port vlan-stacking**, **port vlan-stacking 8021p**, or **port vlan-stacking vlan 8021p** command is used with **remark-8021p** specified, packet priorities are re-marked. The mapping between 802.1p priorities and local priorities may be incorrect and queues may enter incorrect queues.

- If the **port vlan-mapping 8021p**, **port vlan-mapping vlan 8021p**, or **port vlan-mapping vlan map-vlan** command is used with **remark-8021p** specified, packet priorities are re-marked. The mapping between 802.1p priorities and local priorities may be incorrect and queues may enter incorrect queues.
- If the **traffic-policy** command that defines **remark local-precedence** is used, the system sends packets to queues based on the re-marked priority.
- If the **traffic-policy** command that defines **remark 8021p** or **remark dscp** is used, the system maps the re-marked priorities of packets to the local priorities and sends the packets to queues based on the mapped priorities.
- If the **traffic-policy** command that defines **nest top-most** is used, the system adds an outer VLAN tag to received tagged packets and maps priorities in the original VLAN tag of packets. The system adds an outer VLAN tag to received untagged packets. Then the system maps packets based on the default priority of the interface and sends the packets to queues based on the mapped priority.
- If the **trust upstream none** command is used, priorities of all the incoming packets are not mapped and the packets enter queues based on the default priority of the interface.
- If the **port link-type dot1q-tunnel** command is used but the **trust 8021p inner** command is not used on the interface, all the incoming packets enter queues based on the default priority of the interface.

Run the **display this** command in the inbound interface view to check whether there are configurations affecting the queues that packets enter on the inbound interface.

- If so, delete or modify the configuration.
- If not, go to step 3.

**Step 3** Check whether there are configurations affecting the queues that packets enter in the VLAN that packets belong to.

The following configurations affect the queues that packets enter:

- If the **traffic-policy** command that defines **remark local-precedence** is used, the system sends packets to queues based on the re-marked local priorities.
- If the **traffic-policy** command that defines **remark 8021p** or **remark dscp** is used, the system maps the re-marked priorities of packets to the local priorities and sends the packets to queues based on the mapped priorities.
- If the **traffic-policy** command that defines **nest top-most** is used, the system adds an outer VLAN tag to received tagged packets and maps priorities in the original VLAN tag of packets. The system adds an outer VLAN tag to received untagged packets. Then the system maps packets based on the default priority of the interface and sends the packets to queues based on the mapped priority.

Run the **display this** command in the view of the VLAN that packets belong to and check whether the configurations affecting the queues that packets enter are performed in the VLAN.

- If so, delete or modify the configuration.
- If not, go to step 4.

**Step 4** Check whether there are configurations affecting the queues that packets enter in the system.

The following configurations affect the queues that packets enter:

- If the **qos local-precedence-queue-map** command is used, the system sends packets to queues based on the mapping between local priorities and queues specified by this command.



**NOTE**

The ES1D2X40SFC0 and ES1D2L02QFC0 of the S7700, and EH1D2X40SFC0 and EH1D2L02QFC0 of the S9700 do not support the **qos local-precedence-queue-map** command.

- If the **traffic-policy global** command that defines **remark local-precedence** is used, the system sends packets to queues based on the re-marked priority.
- If the **traffic-policy global** command that defines **remark 8021p** or **remark dscp** is used, the system maps the re-marked priorities of packets to the local priorities and sends the packets to queues based on the mapped priorities.
- If the **traffic-policy global** command that defines **nest top-most** is used, the system adds an outer VLAN tag to received tagged packets and maps priorities in the original VLAN tag of packets. The system adds an outer VLAN tag to received untagged packets. Then the system maps packets based on the default priority of the interface and sends the packets to queues based on the mapped priority.

Run the **display current-configuration** command to check whether there are configurations affecting the queues that packets enter in the system. If so, delete or modify the configuration.



**NOTE**

A traffic policy is applied to an interface, a VLAN, and the system in descending order of priority.

----End

## 4.2 Priority Mapping Results Are Incorrect

### Common Causes

This fault is commonly caused by one of the following:

- On the inbound interface, packets do not enter queues corresponding to the priority of packets.
- The priority type trusted by the inbound or outbound interface is incorrect.
- The priority mapping configured in the DiffServ domain trusted by the inbound or outbound interface is incorrect.
- There are configurations affecting priority mapping on the inbound or outbound interface.

### Procedure

**Step 1** Check whether packets enter correct queues on the outbound interface.

Run the **display qos queue statistics interface** *interface-type interface-number* command to check whether packets enter correct queues on the outbound interface.

- If not, locate the fault according to section 4.1 Packets Enter Incorrect Queues.
- If so, go to step 2.

**Step 2** Check whether the priority type trusted by the inbound or outbound interface is correct.

Run the **display this** command in the view of the inbound or outbound interface to check whether the trusted priority type set by using the **trust** command on the inbound or outbound interface is correct. (If the **trust** command is not used, the system trusts the 802.1p priority in the outer VLAN tag by default.)



- If not, run the **trust** command to correctly configure the priority type trusted by the inbound/outbound interface.
- If so, go to step 3.

**Step 3** Check that priority mapping configured in the DiffServ domain trusted by the inbound or outbound interface is correct.

Run the **display this** command in the view of the inbound or outbound interface to check whether the **trust upstream** command is used. If the **trust upstream** command is not used, the system trusts the default DiffServ domain by default.

Then run the **display diffserv domainname** *domain-name* command to check whether the mapping between local priorities and packet priorities is correct.



**NOTE**

The local priority refers to the mapped priority of the inbound interface.

- If not, run the **ip-dscp-outbound**, **mpls-exp-outbound**, or **8021p-outbound** command to correctly configure the mapping between local priorities and packet priorities.
- If so, go to step 4.

**Step 4** Check whether there are configurations affecting priority mapping on the inbound or outbound interface.

The following configurations affect the queues that packets enter:

- If the **undo qos phb marking enable** command is used, the system does not map outgoing packets to PHBs on an interface.
- If the **trust upstream none** command is used, the system does not map outgoing packets to PHBs on an interface.
- If the **traffic-policy** command that defines **remark 8021p** or **remark dscp** is used in the inbound or outbound direction, the re-marked priority is the packet priority.

Run the **display this** command in the view of the inbound or outbound interface to check whether there are configurations affecting priority mapping. If so, delete or modify the configuration.

----End

## 4.3 Traffic Policy Does Not Take Effect

### Fault Symptom

After a traffic policy is applied, the device cannot implement pre-defined QoS action for classified traffic.

### Procedure

**Step 1** Check that the traffic policy is applied correctly.

Run the **display traffic-policy applied-record** *policy-name* command to check the traffic policy record.

- If the value of the Policy total applied times field is 0, the traffic policy is not applied. Run the **traffic-policy** *policy-name* **global** { **inbound** | **outbound** } [ **slot** *slot-id* ] or

**traffic-policy** *policy-name* { **inbound** | **outbound** } command to apply the traffic policy to the system, an LPU, an interface, or a VLAN.

- If the value of the slot field is **success**, check whether the traffic policy is applied to a correct direction. The traffic policy must be applied to the inbound direction if the traffic policy matches packets received by the device, and must be applied to the inbound direction if the traffic policy matches packets sent from the device.
  - If the traffic policy is applied to an incorrect direction, run the **undo traffic-policy** [ *policy-name* ] **global** { **inbound** | **outbound** } [ **slot** *slot-id* ] or **undo traffic-policy** [ *policy-name* ] { **inbound** | **outbound** } command to unbind the traffic policy from the system, LPU, interface, or VLAN. Then run the **traffic-policy** *policy-name* **global** { **inbound** | **outbound** } [ **slot** *slot-id* ] or **traffic-policy** *policy-name* { **inbound** | **outbound** } command to re-apply the traffic policy to the system, LPU, interface, or VLAN.
  - If the traffic policy is applied to a correct direction, go to step 2.
- If the value of the state field is fail, the traffic policy fails to be applied. If the traffic policy fails to be applied, the system displays an error message. Run the **undo traffic-policy** [ *policy-name* ] **global** { **inbound** | **outbound** } [ **slot** *slot-id* ] or **undo traffic-policy** [ *policy-name* ] { **inbound** | **outbound** } command to unbind the traffic policy from the system, LPU, interface, or VLAN. Then run the **traffic-policy** *policy-name* **global** { **inbound** | **outbound** } [ **slot** *slot-id* ] or **traffic-policy** *policy-name* { **inbound** | **outbound** } command to re-apply the traffic policy to the system, LPU, interface, or VLAN. Rectify the fault identified in the error message.

#### Step 2 Check whether packets match rules in the traffic classifier.

Run the **display traffic policy statistics** { **global** [ **slot** *slot-id* ] | **interface** *interface-type* *interface-number* | **vlan** *vlan-id* } { **inbound** | **outbound** } [ **verbose** { **classifier-base** | **rule-base** } [ **class** *classifier-name* ] ] command to check traffic statistics. If the value of each field is 0, packets do not match rules in the traffic classifier. If the value of each field is not 0, packets match rules in the traffic classifier.



#### NOTE

Before viewing traffic statistics, ensure that the **statistic enable** command has been used in the traffic behavior view to enable the traffic statistics function.

- If packets do not match the rules in the traffic classifier, go to step 3.
- If packets match the rules in the traffic classifier, go to step 4.

#### Step 3 Check whether packet characteristics match rules in the traffic classifier.

View information (such as the IP address, MAC address, DSCP priority, VLAN ID, and 802.1p priority) in packets, run the **display traffic policy user-defined** [ *policy-name* [ **classifier** *classifier-name* ] ] command to view the traffic classifier in the traffic policy, and run the **display traffic classifier user-defined** [ *classifier-name* ] command to view rules in the traffic classifier. Check whether packet characteristics match rules in the traffic classifier.

- If not, modify the rules to match packet characteristics.
- If so, go to step 4.

#### Step 4 Check that the traffic behavior associated with the traffic classifier is configured correctly.

Run the **display traffic behavior user-defined** [ *behavior-name* ] command to check whether the traffic behavior associated with the traffic classifier is configured correctly.

- If not, run the **traffic behavior** *behavior-name* command to enter the traffic behavior view and correctly configure a traffic behavior.

----End

# 5 FAQ

## 5.1 Does the S9700 Collect Traffic Statistics Based on Packets or Bytes?

The S9700s equipped with E-series and F-series boards can collect traffic statistics based on packets and bytes, but the S9300s equipped with S-series boards collect traffic statistics only based on packets.

## 5.2 What Are the Differences Between Interface-based CAR and Global CAR?

Interface-based CAR limits the rate of traffic on an interface, whereas global CAR limits the total rate of all interfaces on a device. For example, the CAR is 5000 kbit/s:

- If CAR is applied to an interface, the interface can send or receive packets at a maximum rate of 5000 kbit/s.
- If CAR is applied globally, the total rate of all the interfaces cannot exceed 5000 kbit/s.

## 5.3 How Does Level-2 CAR Take Effect?

Level-2 CAR can be configured on an S9700. The following example describes how to configure level-2 CAR:

```
[Quidway] qos car car1 cir 16000
[Quidway] traffic classifier 1
[Quidway-classifier-1] if-match vlan-id 100
[Quidway-classifier-1] quit
[Quidway] traffic classifier 2
[Quidway-classifier-2] if-match vlan-id 101
[Quidway-classifier-2] quit
[Quidway] traffic behavior 1
[Quidway-behavior-1] car cir 6000 pir 8000
[Quidway-behavior-1] car car1 share
[Quidway-behavior-1] quit
[Quidway] traffic behavior 2
```

```
[Quidway-behavior-2] car cir 8000 pir 10000
[Quidway-behavior-2] car car1 share
[Quidway-behavior-2] quit
[Quidway] traffic policy 1
[Quidway-trafficpolicy-1] classifier 1 behavior 1
[Quidway-trafficpolicy-1] classifier 2 behavior 2
[Quidway-trafficpolicy-1] quit
```

Traffic policy 1 is applied to an interface, and flows in VLAN 100 and VLAN 101 are tested. When the rate limits for flows in VLAN 100 and VLAN 101 are smaller than 6 Mbit/s and 8 Mbit/s respectively, the bandwidth of the two flows is ensured and no packet is lost. When the rate limits for flows in VLAN 100 and VLAN 101 are greater than 6 Mbit/s and 8 Mbit/s respectively, packets are discarded. The total bandwidth of the two flows is 16 Mbit/s. That is, the two flows share CAR. According to the preceding information, level-2 CAR does not limit flows whose level-1 CAR is smaller than the CIR. Level-2 CAR limits only the flows whose rate is greater than the CIR but smaller than the PIR. When you configure level-2 CAR, the CAR value must be in the range of the CIR and PIR. Otherwise, level-2 CAR does not take effect.

## 5.4 A Traffic Policy Contains an ACL Rule Defining TCP or UDP Port Number Range. When the Traffic Policy Is Delivered, the System Displays the Message "Add rule to chip failed." Why?

The causes are as follows:

- The traffic policy is applied to the outbound direction.  
When a traffic policy is applied to the outbound direction, its ACL rule cannot define the port number range.
- The number of ACL rules defining the port number ranges has reached or exceeded the maximum.

## 5.5 CAR Is Incorrect. Why?

The switch counts the inter-frame gap and VLAN tag when limiting the packet rate. During the test, you are advised to use packets with more than 1000 bytes.

If an untagged packet enters the chip, for example, 64-byte packet, the actual packet length is 88 bytes (20 inter-frame gap + 4-byte VLAN tag + 64-byte packet length). As a result, the rate limit is inaccurate. If long packets are used, the inter-frame gap and VLAN tag occupy less percentage of the total packet length. The impact on the rate limit is small and the rate limit is accurate.

## 5.6 An ACL Applied to the Outbound Direction Cannot Define the Port Number Range. Why?

When you define the port number range in an ACL and apply the ACL to the outbound direction, the system displays a failure message.

You can define the port number range in an ACL and apply the ACL to the inbound direction. A maximum of 16 port number ranges can be defined on a common LPU. A maximum of 32 port number ranges can be defined in the inbound direction.

## 5.7 Can 802.1p Re-marking and Traffic Statistics Be Configured in a Traffic Policy Simultaneously on the S9700?

If you configure **remark 8021p** and traffic statistics in a traffic policy simultaneously, the system displays a message indicating that the configuration fails. To only schedule packets on the switch, use **remark local-precedence**.

## 5.8 When Both QinQ and Traffic Policy-based VLAN Stacking Are Configured on an Interface, Which Configuration Takes Effect?

When both QinQ and traffic policy-based VLAN stacking are configured on an interface, traffic policy-based VLAN stacking takes effect.

## 5.9 Why ACL Rule Update May Cause Instant Traffic Interruption?

When a traffic policy contains more routing policies, the switch deletes the ACL rule that has become valid if you edit a referenced ACL rule. In this case, the traffic cannot match the ACL and the switch fails to redirect traffic to the next hop. Some packets are lost. When you reconfigure an ACL rule, traffic can match the ACL rule. Therefore, traffic can be restored.

## 5.10 After an ACL or QoS Is Configured, the Configuration Is Invalid for Mirroring Packets. Why?

The switch processes outgoing packets as follows:

- Buffers outgoing packets.
- Performs Layer 2 and Layer 3 processing.
- Mirrors packets.

- Performs ACL/QoS processing.
- Because ACL/QoS processing is performed after mirroring, the ACL/QoS configuration is invalid for mirroring packets.

## 5.11 Why a Traffic Policy Containing Traffic Filtering or CAR Is Invalid for Incoming Packets on an S9700?

Check whether a static or dynamic binding table exists. The traffic policy containing traffic filtering or CAR is invalid for packets matching binding entries.

Run the **display dhcp { snooping | static } user-bind { interface *interface-type* *interface-number* | ip-address *ip-address* | mac-address *mac-address* | vlan *vlan-id* } \*** [ **verbose** ] command to view the static or dynamic DHCP snooping binding table.

Run the **display dhcp { snooping | static } user-bind all [ verbose ]** command to view the static or dynamic DHCP snooping binding table.

## 5.12 Why PQ+DRR Configured on an S9700 Interface Does Not Take Effect?

Queues are scheduled only after traffic of different services must enter different queues.

To send packets to different queues, modify 802.1p priorities in packets on the upstream device or configure a traffic policy on the inbound interface of the switch and run the **remark local-precedence** command.

## 5.13 Why Priorities in Outgoing Mirroring Packets Are Not Changed After Priority Mapping Is Configured?

The system first mirrors packets, and then maps priorities in packets, so priorities in outgoing mirroring packets are not changed.

## 5.14 When You Configure a Deny Rule in a Traffic Policy Containing Flow Mirroring, Normal Service Traffic Is Affected. Why?

When you configure a deny rule in a traffic policy containing flow mirroring, the system applies the deny rule to matching packets and mirrors them. It is recommended that the permit rule be used in this situation.

## 5.15 When a Traffic Policy Containing Flow Mirroring Is Applied to an Interface, the Global Traffic Policy Becomes Invalid. Why?

The traffic policies applied to the interface, VLAN, and system take effect in descending order of priority. If a traffic policy has been applied to the system or VLAN, when you apply a traffic policy containing flow mirroring to an interface, mirroring packets that match the traffic policy on the interface cannot match the traffic policy in the system or VLAN. As a result, services are abnormal.

## 5.16 What Is the Relationship Between an ACL and a Traffic Policy?

An ACL is often used with a traffic policy. A traffic policy defines the traffic classifier matching an ACL and a traffic behavior such as permit/deny associated with the traffic classifier.

The permit/deny actions in an ACL and a traffic behavior in the traffic policy are used as follows.

ACL	Traffic Behavior in the Traffic Policy	Final Action Taken for Matching Packets
permit	permit	permit
permit	deny	deny
deny	permit	deny
deny	deny	deny



### NOTE

A switch permits packets by default. To reject packets between subnets, define the packets to be rejected in the ACL. If the **rule permit all** command is used, all packets match the rule. If the traffic behavior defines the deny action, all packets are filtered, causing service interruption.



## 5.17 How Are Packets Forwarded Using PBR on S Series Switches?

The switch forwards packets according to the destination IP address if the next hop address is unavailable.

Starting from V1R6, the switch supports multiple next-hop IP addresses. When there are multiple next-hop IP addresses, the switch redirects packets in active/standby mode. A maximum of four next-hop IP addresses can be configured in a traffic behavior. The device determines the primary path and backup paths according to the sequence in which next-hop IP addresses were configured. The next-hop IP address that was configured first has the highest priority and this next hop is used as the primary path. Other next hops are used as backup paths. When the primary link becomes Down, a next hop with higher priority is used as the primary link.

# 6 Appendix

## 6.1 Common Service Priorities

**Table 6-1** Common service priorities

Service	802.1p	DSCP	PHB
Voice	5	46	EF
Video	4	34	AF4
Enterprise important data	3	26	AF3
Common data	0	0	BE

## 6.2 Port Numbers of Common Application Services

**Table 6-2** Port numbers of common application services

Application Service	Port Number	Application Service	Port Number
FTP	21 (T)	Telnet	23 (T)
SMTP	25 (T)	DNS	53 (T)
DHCP	67 & 68 (U)	TFTP	69 (U)
HTTP	80 (T)	HTTPS	443 (U)
SNMP	161 (U)	WWW	8080 (T)
QQ chat client	4000 (U)	QQ chat server	8000 (U)
MSN	1863 (T)	BitSpirit	16881 (U)
Thunder software	3076/3077/ 5200/6200 (U)	eMule	4662 (U)



**NOTE**

(T) indicates TCP transmission and (U) indicates UDP transmission.

## 6.3 Common Queue Scheduling Solution

**Table 6-3** Common queue scheduling solution

PHB	Queue	Weight	Service
CS6 & CS7	PQ	-	Protocol packets (PQ is used by default for protocol packets)
EF	PQ/WRR	-/35%	Enterprise voice services and services that are sensitive to the delay and jitter
AF4	WRR	25%	Enterprise voice services, services that are sensitive to the delay and jitter, and enterprise key information services
AF3	WRR	15%	Telnet and FTP services and services that are sensitive to the delay and packet loss
AF2	WRR	12%	Enterprise IPTV services and services that are insensitive to the delay and packet loss
AF1	WRR	8%	Enterprise email services and services demanding low bandwidth, delay, and jitter
BE	WRR	5%	Enterprise web page browsing services

## 6.4 Recommended WRED Parameter Setting

### 6.4.1 Color-based WRED Parameter Setting

**Table 6-4** Color-based WRED parameter setting

Queue (PHB)	Lower Drop Threshold (%)	Upper Drop Threshold (%)	Maximum Drop Probability
Green	80	100	10
Yellow	60	80	20
Red	40	60	30

## 6.4.2 Queue-based WRED Parameter Setting

**Table 6-5** Queue-based WRED parameter setting

Queue (PHB)	Lower Drop Threshold (%)	Upper Drop Threshold (%)	Maximum Drop Probability
High priority (CS7, CS6)	80	100	10
Medium priority (EF, AF1-AF4)	60	80	20
Low priority (BE)	40	60	30

## 6.5 Video Service Bandwidth Usage

Video resolution can be simply understood as HD and SD, 480P, 720P, and 1080P, or 352\*288 and 1024\*720. Video encoding is a compression of the original video, and MPEG and H.26X are often used.

Before estimating the bandwidth of a video service, understand the service resolution. This is because different resolutions determine required bandwidth of video services. A high resolution indicates higher bandwidth required. The resolution 1280\*720, also called 720P HD video, is used as an example. Video pixels occupy 900 kbit/s (1280\*720). Each pixel is differentiated by 8-bit color (256 colors) and occupies 7200 kbit/s (900 kbit/s x 8). A device must transmit at least 24 frames per second to ensure smoothness of video image. Generally, 30 frames are transmitted per second (60 frames per second for super HD video). The length of the packet protocol header is 0.3 times the packet content, so the bandwidth per second is 274 Mbit/s (7200 kbit/s\*30\*1.3).

Video bandwidth exceeds the allowed range, so video coding technologies are important. Original video processed by different video coding technologies (some low-end coding technologies cannot process HD video) and compression ratio (ratio of original video to compressed video) are different, so the compressed bandwidth is also different. Reserved bandwidth for different video services needs to be calculated separately.

Video services are not enterprise key services, video conference requires only the resolution of 480P and 2 Mbit/s bandwidth. For HD services, enterprise can configure corresponding bandwidth.

### 6.5.1 Coding-based Video Bandwidth

**Table 6-6** Coding-based video bandwidth

Coding Technology	Common Compression Ratio	Video Resolution	Common Video	Recommended CIR (bit/s)
MPEG-1	20-30	352x240x30 352x288x25	352x288x25	1.5 M

Coding Technology	Common Compression Ratio	Video Resolution	Common Video	Recommended CIR (bit/s)
MPEG-2	30-40	352x288– 1920x1152	1280x720x30 (720P)	8 M
MPEG-4	60	>=176x144	1280x720x30 (720P)	5 M
H.261	20	352x288	352x288x25	2 M
H.264	80-100	Various resolutions	1920x1080x30 (1080P)	10 M

## 6.5.2 HD-based Video Bandwidth

0 shows the video classification and guaranteed bandwidth in an enterprise network deployment solution. SD low-bit-rate video refers to video of 360P or lower, SD high-bit-rate video refers to 720P video, and super HD video refers to video of 1080P or higher.

**Table 6-7** HD-based video bandwidth

HD	Guaranteed Bandwidth (kbit/s)
Super HD video (1080P and higher)	120M
SD high-bit-rate video (720P)	5M
SD low-bit-rate video (360P and lower)	1.5M

## 6.5.3 Video Conference Bandwidth

Video conference often uses 480P resolution, and HD video conference uses 720P resolution. Video conferences are often applied to fixed scenarios such as meeting rooms, so incremental data (B and P frames) is mainly transmitted. Incremental data is 10% to 40% of the original bandwidth. In variable conference scenarios, 40% bandwidth is used. Bandwidth of video conferences is 10% to 40% of bandwidth of video code streams.

Assume that 720P is used and the original bandwidth is 8 Mbit/s. The reserved bandwidth of 1 to 3.5 Mbit/s is recommended.

## 6.6 Audio Bandwidth Usage

### 6.6.1 Audio Bandwidth Based on Codec Technologies

Codec Technology	Compression Rate (kbit/s)	Ethernet Bandwidth Variable	Actual Bandwidth (kbit/s)	Recommended CIR (kbit/s)
G.711	64	1.41	90.4	100
G.729	8	0.54	34.6	40
G.723.1	5.3	0.32	20.8	25
G.726	24	0.73	47.2	50
G.728	16	0.49	31.5	35

Note: The codec technology used depends on the IP phone.