Advances in Engineering Research, volume 149

2nd International Conference on Mechanical, Electronic, Control and Automation Engineering (MECAE 2018)

Research into the Impact of HQoS and GTS Rate-Limiting Strategies on Data Transmission

Hongyuan Liu [a], Qiang Han [b]

China Satellite Maritime Tracking and Control Department, Jiangyin 214431, China;

[a]SDLHY123@163.com, [b]1510603630@qq.com

Keywords: Speed limit strategy, HQoS, GTS, Multicast data.

Abstract. This paper introduces the principles of two rate-limiting strategies--HQoS and GTS, and researches the impact of the two strategies on the data transmission of different types of services. By analyzing data transmission mechanism and comparing experimental data, this paper lays emphasis on the analysis of the causes why the speech distortion occurs under the GTS strategy during multicast data transmission while it does not happen under HQoS, and concludes that the EF queue multicast dataunder GTS strategy shows a larger buffer which could effectively smooth multicast data and prevents packet loss, but would influence the transmission of EF-priority unicast messages. The conclusion of this research provides a technical reference in regard to the choosing of router-based rate-limiting strategies for networks of different service transmission needs.

1. Introduction

The WAN circuit of IP network adopts the satellite communication link. Compared with the internal gigabit network architecture, the output bandwidth of the WAN is smaller, which is the bottleneck of network transmission. Therefore, a rate-limiting strategy needs to be configured at the egress of router to reduce risk of packets loss of high-priority services. This paper takes multicast data and voice data as examples, researching into and analyzing the IP network router speed strategy.

2. Introduction to the Rate-Limiting Strategies

2.1 Deploy to the Rate-Limiting Strategies.

IP network rate-limiting strategies involve CAR at the access layer (Committed Access Rate), and HQoS(Hierarchical QoS)or GTS(Generic Traffic Shaping) at the WAN egress of router[1,2].Speed limit policy deployment is shown in Figure 1.

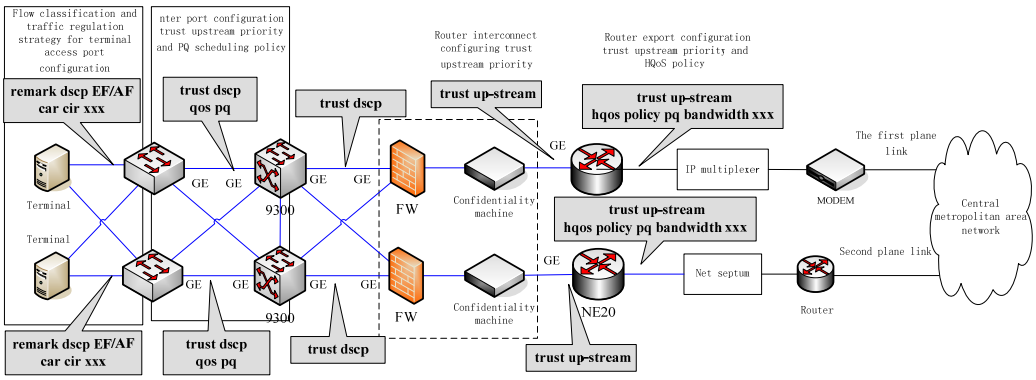

Figure 1 speed limit strategy map of IP network

The IP network rate-limiting strategiesare based on the principle of service flow classification--the configuration priority. The priority strategies mainly include the following aspects:

(1) For WAN information flows in real-time transmission, quintuple is adopted to identify and prioritize service flows.

(2) For information flow and other information in non-real-time transmission, because of their unfixed destination address, triple is adopted at the service access port (source IP, transport protocol, destination port number) to identify and prioritize service flows.

(3) Service flows are classified into five priority categories, respectively, code-named as 5,4,3,2 and 0 of which 5 is of the highest priority.

(4) All the unacknowledged flow prioritiesare marked as 0.

Priority partition rules are shown in Table 1

Table 1 priority partition rules

Serial number	Business priority level	Type and description of task information	Priority code
1	Special priority	Network platform self protocol data information	6
2	Highest priority	Real-time measurement control information based on application layer protocol exchange Voice communication data information	5
3	High priority	The operation of management information, including real-time control, fault alarm, equipment etc.	4
4	Secondary high priority	Non real-time measurement and control data using FEP protocol	3
5	Middle priority	Meteorological information and service service	2
6	Sub low priority	Retain	1
7	Lowest	All unidentified traffic flows	0

To analyze the rate-limiting strategies of IP network router, it is necessary to analyze the data policing, priority trust and PQ queue at the access layer in accordance to the service flow directions.

2.2 Traffic Regulation Strategy.

Because of the asymmetry between LAN and WAN flow rate, if there is virus in the intranet and continuous malicious data is incurred, it will cause the congestion in satellite communication link. To avoid such problems, the data policing strategy is adopted at the access port to monitor the data rate entering the network and to "process" the excess data so as to limit the data to a reasonable range [1, 3]. Data policing uses token bucket to evaluate data, depending on whether the number of tokens in the token bucket is sufficient for message forwarding. If there are enough tokens in the bucket to forward messages, the datashall be named as complying with or conforming to the agreed value, otherwise known as non-compliance or out-of-limits.The token bucket evaluation flow is shown in Figure 2.

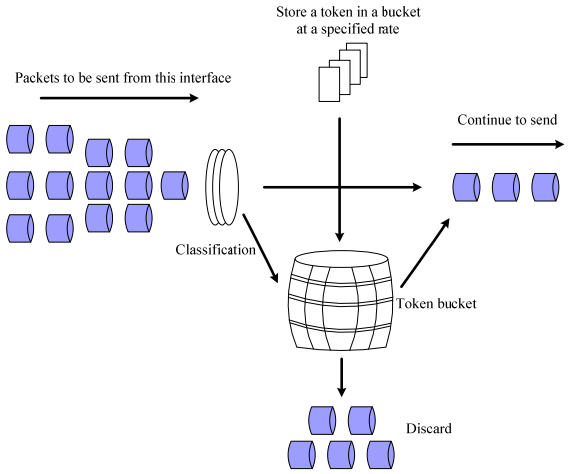

Figure 2 token bucket assessment flow diagram

IP network HUAWEI switch uses double token bucket mechanism to realize traffic regulation, one C barrel and one P barrel, which mainly includes four parameters:

(1) cir(committed information rate) : Represents the rate of placing the token in the C barrel, the average speed of the data flow allowed by the C barrel.

(2) cbs(committed burst size) : It represents the capacity of the C barrel, that is, the maximum flow size allowed by a burst of C barrels per time.

(3) pir(peak information rate) : Represents the rate of placing the token in the P barrel, the average speed of the data flow allowed by the P barrel.

(4) pbs(peak burst size) : The capacity of the P barrel, that is, the maximum flow size allowed by a burst of P barrels per time

Every time the incoming message is measured, if the C bucket has enough token, the message is labeled green. If the C bucket token is insufficient, but the P bucket has enough token, the message is marked as yellow; if the C bucket and the P bucket do not have enough tokens, the message is labeled red. Traffic supervision is based on different evaluation results. By default, green and yellow are forwarded and red messages are discarded, as shown in Figure 3.

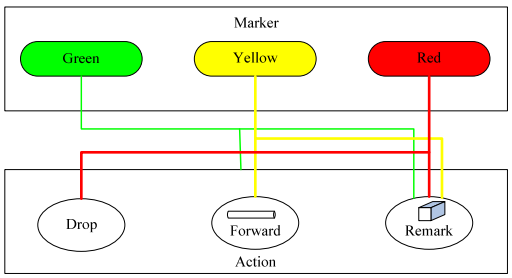

Figure 3 Schematic diagram of flow regulation

2.3 Priority Trust and Queue Scheduling.

To ensure the consistency of message priority in the network, downstream devices trust upstream traffic priorities along the traffic flow direction.

When congestion management, an absolute priority queue (PQ queue) is used to schedule [4] for the message, the process is as follows:

According to the DSCP value of the IP message, all the messages are divided into 8 classes, which correspond to the 8 priority queues of the PQ.

Send a message to the corresponding queue according to the priority category of the message.

The message is queued in the following order: PQ first sends the message in the high priority queue until the message in the high priority queue is sent, and then sends the message in the rest of the low priority queue. If the message in the high priority queue is not sent out, the message in the low priority queue will never be sent.

3. NE20E-8 Rate-Limiting Strategy

The electrical interface board of IP Network NE20E-8 router WAN 100 Mbps is a high-rate board. Depending on the features of in-network data service, HQoS and GTS can be used to limit the port rate. However, HQoS and GTS could not be used instantaneously.

3.1 HqoS Strategy.

HQoS strategy is implemented by hierarchical scheduling. Level 1 is scheduled to ports to allocate fixed bandwidth so as to control the total bandwidth of the ports. When the data exceeds the fixed bandwidth, the network is congested. And, then, level 2 scheduling is enabled, and the PQ queue is used. The messages of high-priority in the queue are sent firstly, and messages in the low-priority queue could not be sent until the messages in the high-priority queue are all sent[5,11].

HQoS strategy processes messages through the token bucket mechanism.as shown in Figure 2. Therefore, HQoS allows a certain degree of burst, and the maximum burst data volume is proportional to the size of the token bucket.

3.2 GTS Strategy.

GTS strategy is a data management technique that limits data of a certain type sent by ports, and that is to shape data on every queue so as to realize rate-limiting at ports[6]. GTS uses the leaky bucket algorithm, as shown in Figure 4. the purpose of which is to control the rate at which data is

injected into the network, smooth the burst of data on the network in order to provide a steady stream of data to the network.

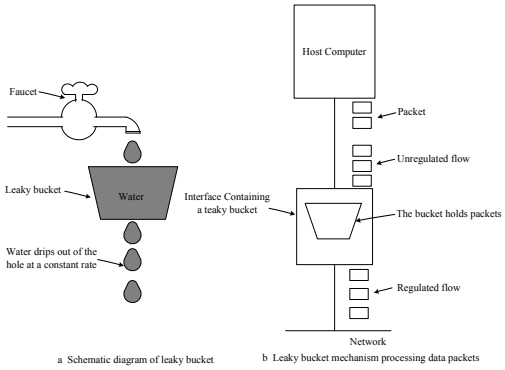

Figure 4 Schematic diagram of leaky bucket algorithm

The GTS process is shown in Figure 5, in which the GTS queue is used for caching the message, the process of processing:

(1) when the message arrives, first the message is classified. If the message does not need to be processed by GTS, it will continue to be sent without token bucket processing. If the message needs to be processed by GTS, it will be compared with the token in the token bucket.

(2) to compare the number of message and token token bucket in GTS treatment, if there is enough message sends the token token bucket, message is sent directly, at the same time the token number according to the packet length accordingly reduced; if the token is insufficient, the message will be cached in the GTS queue.

(3) the token in the token bucket is used, and after the number is reduced, the system will place the token in the token bucket at the rate set by the user (CIR).

(4) when there are messages in the GTS queue, GTS fetches the message from the queue in a certain period, and sends it to the token number every time. Until the token number is reduced to the queue, the message can no longer be sent or the message is sent to the end.

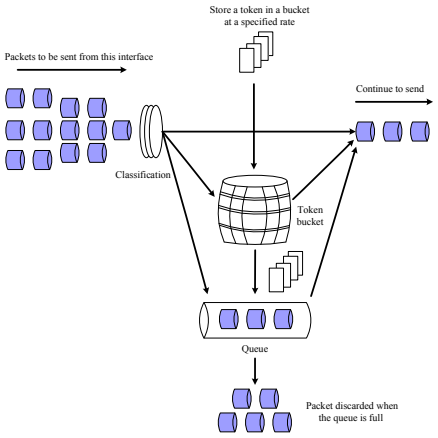

Figure 5 GTS processing flow

4. Rate-Limiting Case Analysis

4.1 Example Analysis of GTS Strategy.

The first plane router (hereinafter referred to as router 1)WAN of IP network uses GTS strategy to perform rate-limiting in the data flow of EF-priority and AF-priority. The experiment testing shows that the multicast data transmission is normal when the multicast data is sent outward, but the outward voice data is distorted. Take one segment of data from both aggregation switch and router, (sending data in a period of time before and after sending multicast data). and the start sequence and end sequence of TCP and UDP messages of the multicast data are the same with those of the voice data in the two segments of data are consistent. The number of data statistics is compared as shown in Table

2. It is shown from the table that 5203 packet is lost from voice UDP data, which is in line with the voice distortion phenomenon.

Table 2 packet loss statistics

data type	Number of packets for converging switches	number of routers	packet loss values	Remarks
Multicast data	152770	152769	1	≤0.1% Packet loss
Voice data TCP	1394	753	641	
Voice data UDP	16638	11435	5203	

Analyze the message of the converging switch 1, the multicast data and the voice communication data are EF priority. And the multicast data that is sent outward is not sent strictly according to the packet frequency of the 20ms/ packet, and will burst into a continuous message at a certain time point. Through the analysis of the 1 router packet capture data, can be seen in the multicast data transmission period, TCP voice data and UDP packet flow decreased, as Figure 6 shows, 7.

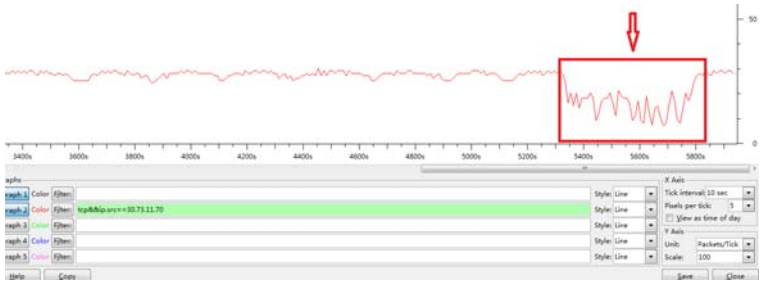

Figure 6 TCP packet frequency graph of speech data

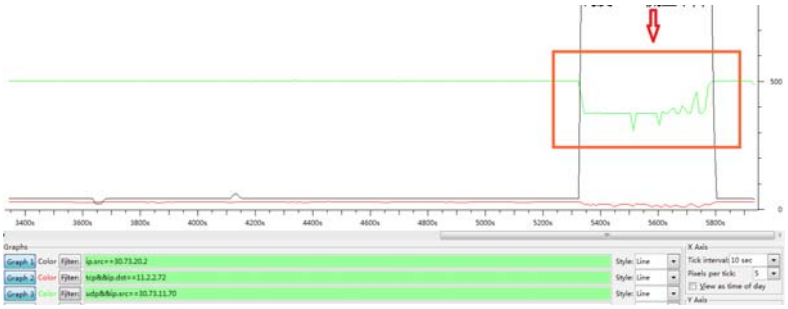

Figure 7 UDP packet frequency graph of speech data

After TCP is lost, the voice data send TCP retransmission packets repeatedly. In the specified RTO time, the voice data do not receive TCP ACK response packet from the opposite end, and the voice data will retransmit TCP message [7]. The new TCP messages will keep entering into cache, which will further increase RTT, or continue being discarded by the router, which will lead to a gradual increase of RTO [8]. After analysis of the experimental data, it is found that there are 253 retransmitted TCP packets in a section of data with 2626 TCP packets. By analyzing the voice data TCP packets both during normal data and data burst, it is discovered that RTO increases from 0.6 seconds at normal datato 234.6 seconds at data burst. When voice data is retransmitted repeatedly to TCP packets, the link will be judged to be in exception, and a part of voice UDP packet will not be sent, which will further lead to speech distortion.

The GTS strategy uses PQ queue scheduling for different caching queues of BE, AF, and EF priority, giving priority to high priority caching queues. At the same priority of business processing, the unicast and multicast services for EF priority are processed with different caching queues. The size of the EF queue cache is not related to the speed limit value. The unicast queue cache is about 9KByte, and the multicast queue cache is about 380KByte. Obviously, the multicast business cache is much larger than the unicast service. When the multicast and unicast traffic of this priority is transmitted simultaneously, if the network is congested, the system will give priority to the multicast

service [9]. If multicast business bursts continuously, it may bring packet loss of unicast service. The longer the burst multicast message is, the more the token is needed, the greater the impact on the unicast business. This is why multicast data in the experiment is almost unlost, while the same priority of unicast data (voice TCP and UDP) is lost.

4.2 Example Analysis of HQoS Strategy.

The router's wide area network port is changed to the speed limit of the HQoS strategy. Through experimental test, it is normal to send the multicast data to the outside, and the external voice communication is normal.

Under HQoS strategy, the rate limit is 9792kbps. When multicast data burst, CBS initial value of EF queue has a capacity of about 98KByte. TCP messages of the burst multicast data and the voice data will be forwarded because there is token bucket in the token bucket, which does not cause the RTT of TCP to increase [10]. In a short time, when data burst leading to network congestion, PQ queue will send messages in high-priority EF queue, and messages in the low-priority queue could not be sent until the messages in the high-priority queue are all sent, which means equivalently that 9792kbps bandwidth can be used to send EF priority messages, and it is greater than 8576kbps bandwidth under EF queue rate-limiting in GTS strategy, so the message token of multicast data and voice data can be forwarded firstly [11]. When multicast data is sent under HQoS strategy, data flow curve is examined. BE data drops, which indicates that the HQoS strategy leaves bandwidth as much as possible for the use of high priority messages.

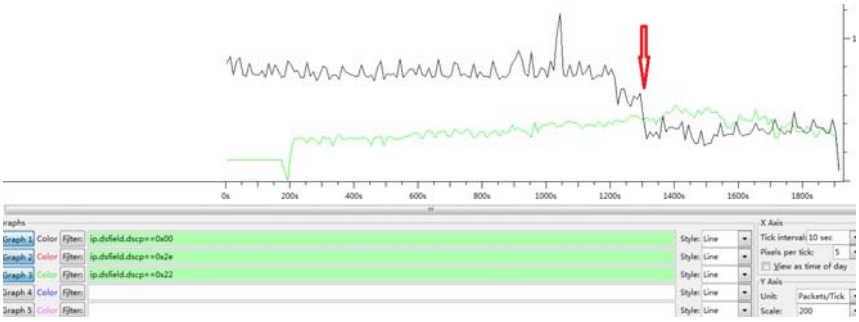

Figure 8 BE data flow graph under HQOS policy

However, the use of GTS strategy, and send multicast data, check data flow curve through the capture software, as shown in figure 9. BE traffic is almost constant, indicating that the GTS policy is set against the priority queue, and the EF queue does not use the bandwidth of the AF and BE queues.

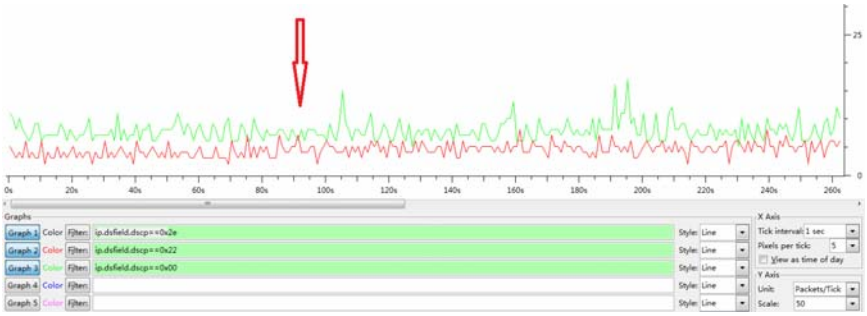

Figure 9 BE data flow graph under GTS policy

5. Summary

The GTS strategy of NE20E-8 router is of leaky bucket mechanism. It buffers burst data firstly and then send them out evenly, which reduces packet loss rate, but increases message transmission delay. And, it may cause TCP messages to be timeout and retransmitted. In addition, under GTS strategy, cache of EF queue multicast service is of large size, which can effectively smooth the burst services such as multicast data category, and it is not easy to cause packet loss, but may affect unicast messages of EF priority. Therefore, it is necessary to avoid dealing with EF priority multicast and unicast services in the meantime under GTS strategy.

The initial token bucket capacity of the HQoS strategy can satisfy a certain degree of burst services, and EF priority messages will have priority in the using of bandwidth when the network is in congestion. Therefore, when selecting the rate-limiting strategies of router, we should choose appropriate ones according to the features of unit network services so as to ensure the stability and reliability of the network data transmission.

References

[1]. Nam D S,Youn C H,Lee B H,et al. Qo S-constrained resource al-location for a grid -based multiple source electrocardiogram appli-cation[C]//LNCS 3043: ICCSA 2004, 2004: 352-359.

[2]. Lv Yan, Wang Dandan. QoS test method [J]. Silicon Valley, 2010(24):172 - 176.

[3]. Lin Tao. Characteristics of QoS network equipment research D test method. Beijing: Beijing University of Posts and Telecommunications, 2013.

[4]. Guo Tianchen. Using low HQoS performance access card to realize user differential protection scheme [J]. Telecommunication technology, 2016,11(13) : 53-55

[5]. Yu Hongzeng, Chen Haixiao. Test Method and Application of Line Rate [J]. radio engineering, 2014,44(10) : 1-4

[6]. Cai Huijuan, Jiang Wenxian. Performance analysis and configuration optimization of GTS in IEEE802.15.4 multi time gap [J]. Computer application, 2012,32(12) : 3499-3504

[7]. Sun Haofeng. Network operation and maintenance and management [M]. Electronic Industry Press, 2014.06 P.38-39

[8]. Wu Gongyi. Computer network [M]. Tsinghua University press, 2011.06 P.292-293

[9]. Zhuge Jianwei, Chen Lin, Xu Weilin translation. Wireshark data packet analysis real battle [M]. people post and Telecommunications Press, 2013.03 P.113-116

[10]. Ardagna D,Giunta G,Ingraffia N,et al.Qo S-driven Web services se-lection in autonomic grid environments[C]//LNCS 4276: OTM 2006, 2006: 1273-1289.