The 7th International Scientific Conference
"DEFENSE RESOURCES MANAGEMENT IN THE 21st CENTURY"
Braşov, November 15th 2012

EVALUATION OF QUEUING DISCIPLINES OVER THE NETWORK PERFORMANCE

Drd. Ioan FIŢIGĂU[1], Prof. Dr. Ing. Gavril TODEREAN[2], Drd. Zoltán NEMES[3]

Department of Communication, Faculty of Electronics, Telecommunications and Information Technology, Technical University of Cluj-Napoca, Str. George Baritiu, nr. 26-28, Cluj-Napoca, Romania

Abstract:
This paper presents a simulation analysis of different queuing disciplines and how they affect the performance and the utilization of the network resources. Some of the common queuing disciplines, first in first out (FIFO), priority queuing (PQ), weighted fair queuing (WFQ) are implemented using OpNet Modeler. Finally, based on the practical results obtained we present the conclusions and the efficiency of the packet delivery for different network services when queuing disciplines are applied.

Key words: QoS, IP, Network, Opnet, FIFO, PQ, WFQ.

1. Introduction

In the Internet and intranets of today, bandwidth is an important subject. The amount of data that is being transmitted through the Internet is increasing exponentially. Multimedia applications, such as IP telephony and videoconferencing systems, need a lot more bandwidth than the applications that were used in the early years of the Internet [1]. While traditional Internet applications, such as www, Ftp, or Telnet, cannot tolerate packet loss but are less sensitive to variable delays, most real-time applications show just the opposite behavior, meaning they can compensate for a reasonable amount of packet loss but are usually very critical toward high variable delays [2]. A QoS can be described as a set of parameters that describe the quality (for example, bandwidth, buffer usage, priority, and CPU usage) of a specific stream of data. Quality of Service (QoS) refers to a set of service requirements to be met by the network while transporting a flow of traffic. The goal of QoS is to provide guarantees on the ability of a network to deliver predictable results of bandwidth, delay, jitter, and loss [3].

2. Queuing disciplines

In a store-and-forward network, routers maintain one or more queues for each output line. These are necessary as a packet may arrive and be destined for a line which is already busy. A queuing policy defines a set of rules for placing packets into the queue and taking them back out.

2.1 First-in, first-out (FIFO)

The traditional policy is First-In, First-Out (FIFO), which is easy to implement and treats all data streams the same. Packets arriving from different flows are treated equally by placing them into the queue, respecting, strictly, their arriving order. Already in the queue they are dispatched in the same order they entered, this means, first packet that comes in is the first packet that goes out. Packets are transmitted from point to point without any guarantee for a special bandwidth or minimum time delay. With the best-effort traffic model, Internet requests are handled with the first come, first serve strategy. This means that all requests have the same priority and are handled one after the other. A packet which has newly arrived is placed at the end of the queue and waits its turn to be sent [1]. FIFO is the queuing discipline most used as default for routers. On Cisco routers, when no other queuing strategies are configured, all interfaces use FIFO queuing discipline by default. The traditional FIFO queuing in common Internet routers provides no service differentiation and can lead to network performance problems. Regular queues invariably employ the FIFO principle: The packet that has been waiting the longest is transmitted first. When the queue is full, and additional packets come in, tail drops happen.

2.2 Priority Queuing (PQ)

A priority queuing system is a combination of a set of queues and a scheduler that empties them in priority sequence [4]. When asked for a packet, the scheduler inspects the highest priority queue and, if there is data present, returns a packet from that queue. Failing that, it inspects the next highest priority queue, and so on. Priority Queuing (PQ) employs four separate queues: high, medium, normal (default) and low. By default each of these queues has 20, 40, 60 and 80 packets capacity. Traffic must be assigned to these queues, usually using access-lists. Packets from the High queue are always processed before packets from the Medium queue. Likewise, packets from the Medium queue are always processed before packets in the Normal queue, etc. The obvious advantage of PQ is that higher-priority traffic is always processed first [5].

2.3 Weighted Fair Queuing (WFQ)

WFQ tries to allocate bandwidth fairly to different conversations (typically TCP sessions) so high-bandwidth sessions don't get to monopolize the connection. WFQ is the default for lower-bandwidth interfaces. In a rate-based queuing system, such as Weighted Fair Queuing (WFQ) [4], the delay that a packet in any given queue will experience depends on the parameters and occupancy of its queue and the parameters and occupancy of the queues it is competing with. A queue whose traffic arrival rate is much less than the rate at which it lets traffic depart will tend to be empty and packets in it will experience nominal delays. A unique feature of this queuing discipline is moving of the real-time interactive traffic to the front of queues and fairly shares the remaining bandwidth among other flows [6]. ToS bits in the IP header is use to identify weight.

3. Simulated network topology

This paper has examined the behavior of different queuing disciplines and how they affect the performance and the utilization of the network resources. The application was realized in OPNET as in Figure 1 and the network consist in computers and servers, connected to router R1 and R2, witch are connected using a link, with a speed rate of 44.736 Mbps. There are 4 workstations and 4 servers connected to routers using Ethernet 100BaseT link.

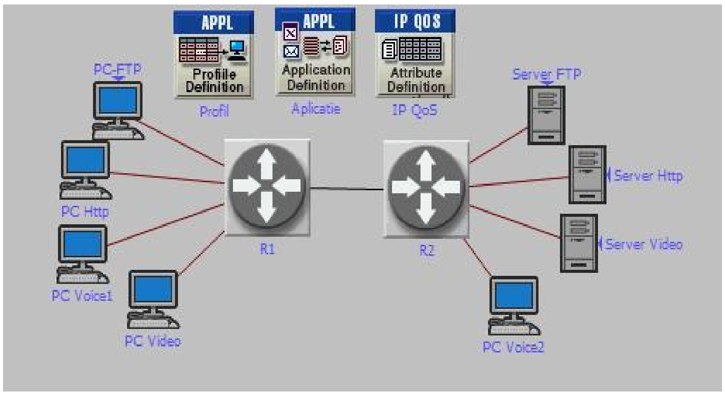

Fig.1 Network components

We set a high load for FTP application with file size (bytes) to a constant value of 1000000. The type of service is set to best effort (0). For Http application the time in seconds between page requests is set to an exponential value of 60 seconds. The start time for a page request is computed by adding the inter-arrival time to the time of the previous page request. We assign PCM quality speech for voice application with type of service set to interactive voice with a value number of 192. It represents a session attribute which allows packets to be processed faster in IP queues. The video application has a low resolution with a frame rate in frames/sec for the incoming and outgoing video streams set to a value of 10 frames/sec. The type of service is interactive multimedia with a value of 160. This is an integer between 0 - 252, 252 being the highest priority.

Profile definition is use to create users profiles to be specified in different nodes on the network. The start time (in seconds) defines when during the simulation the profile session will start, and is set to a constant value of 10 seconds. The operation mode is simultaneous, witch means all the 4 profiles can start all at the same time during the simulation. For the queuing profiles, the default was used. For links between routers across the wan we set the type of service profile to FIFO.

4. Simulation results

The simulation time is set for 3 minute data transfer between computers and servers, applying the three queuing disciplines. In figure 2 is shown the FTP traffic receive by the PC from the FTP server.

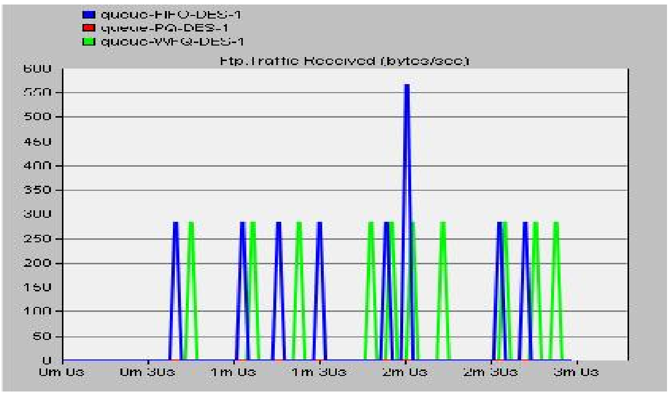

Fig.2 FTP traffic receive

When using priority queuing, the FTP traffic receive is zero, due to heavy congestion in the network or failures leading to exceed maximum TCP retransmission timeouts. Also, in the router R1 and R2, packets have been dropped for congestion reason cause by the congestion attributes such as the maximum number of buffered packets or the queue sizes are too small. The IP traffic dropped is shown in the figure 3:

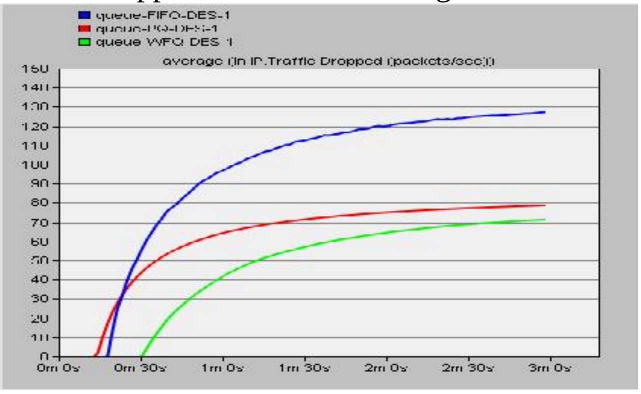

Fig. 3 IP traffic dropped

In figure 3 it can be observed that traffic dropped start around 15 seconds, while for WFQ traffic dropped is zero and by the end of simulation traffic drooped has a lower value.

Voice and video traffic received is shown in the figure 4:

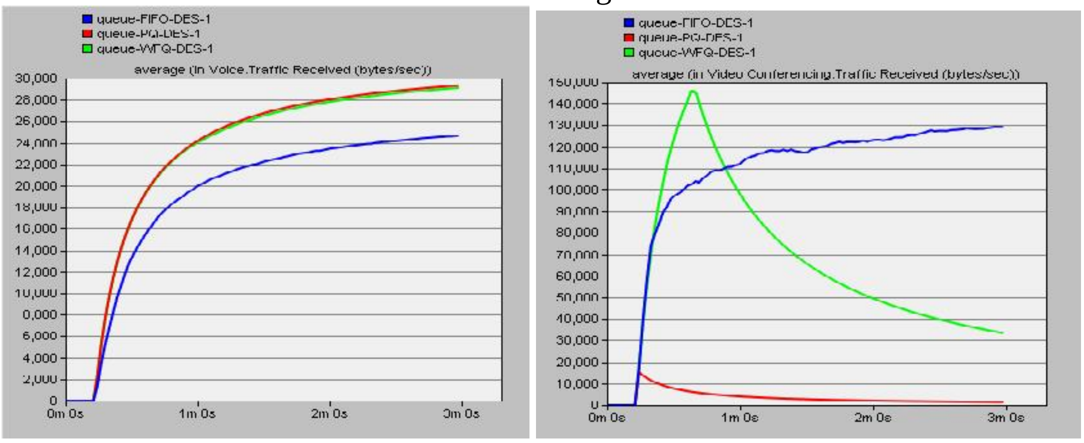

Fig.4 Voice and video traffic receive (bytes/second)

The performance graph line for FIFO is lowered compared to PQ and WFQ, (left side), but in case of video traffic received, the FIFO performance graph has a higher value. The queuing delay and utilization for the link between routers R1 an R2 is shown in figure 5. Queuing delay is a statistic that represents instantaneous measurements of packet waiting times in the transmitter channel's queue. Measurements are taken from the time a packet enters the transmitter channel queue to the time the last bit of the packet is transmitted. Link utilization represents the percentage of the consumption to date of an available channel bandwidth, where a value of 100.0 would indicate full usage.

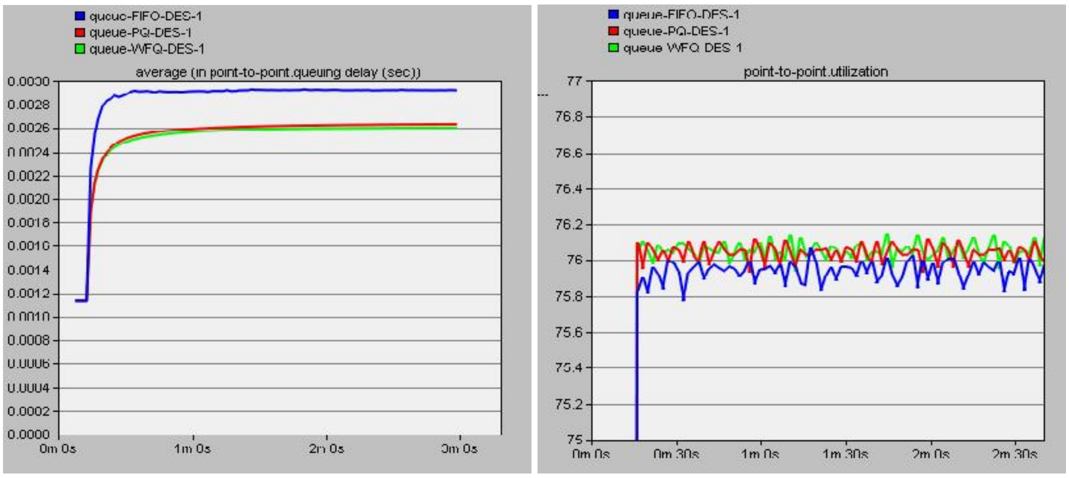

Fig.5 Queuing delay and utilization for R1 R2 link

5. Conclusions

Queues and queue-servicing algorithms are critical elements of traffic handling in providing network QoS. Queuing happens only when the interface is busy. As long as the interface is idle, packets will be transmitted without special treatment. Queuing disciplines roles is to achieve the required quality of service in today's network services. The loss of IP data packets in Priority queue and weighted fair queue discipline is less than FIFO queue. Priority queue has a relatively low overhead for software based routers. User traffic stream like voice, video, data can be easily transferred with its efficient level performance by using Weighted Fair Queue algorithm in routers where the voice, video and data streams are routed to go to their desired destination. The advantage of FIFO is the low computational load on software based routers, but the bursts traffic can consume the buffer and dominate the transmission. For WFO, network bandwidth is allocated to voice and data, network management becomes more complex comparing with the other two. In our days, there is a huge demand for the Internet applications, as the number of services keeps increasing. Each network should maintain a good Quality of Service to provide the best results to the users.

References:

[1] Lydia Parziale, Davis T. Britt, *TCP/Ip tutorial and technical overview*, ibm.com/redbooks, December 2006.
[2] Walran, J., *An introduction to queuing networks*, New Jersey, 1993.
[3] Tanenbaum, Andrew S., *Computer Networks* 4^{Th} *Ed.*, Prentice Hall, 2003.
[4] J.Babiarz, K. Chan, Nortel Networks, *RFC 4594, Configuration guidelines for diffserv service classes*, August 2006.
[5] Aaron Balchunas, *QoS and queuing v1.31*, 2010.
[6] Walran, J., *An introduction to queuing networks*, New Jersey, 1993.
[7] Mitko Gospodinov, *The affects of different queuing disciplinesover ftp, video and voip performance,* CompSysTech' 2004.
[8] J. Kurose and K. Ross, *Computer Networking: A Top-Down Approach Featuring the Internet, 3/e*, Addison Wesley, 2005.
[9] *http://www.opalsoft.net/qos/WhyQos-23.htm.*

[10] *http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/qos.pdf*.
[11] V.Jacobson, G.Minshall, C.Partridge, *RFC 2309, Recommendation on queue management and congestion avoidance in the Internet*, April 1998.

ACKNOWLEDGMENTS

This paper was supported by the project "Improvement of the doctoral studies quality in engineering science for development of the knowledge based society-QDOC" contract no. POSDRU/107/1.5/S/78534, project co-funded by the European Social Fund through the Sectorial Operational Program Human Resources 2007-2013.