A Comparative analysis on traditional Queuing and Hybrid Queuing Mechanism of VoIP's QoS Properties

Md. Zahirul Islam
zahirete@daffodilvarsity.edu.bd
Dept. of ETE
Daffodil International University
102, Sukrabad, Dhanmondi Dhaka-1205, Bangladesh.

Md. Mirza Golam Rashed
mgrashed@daffodilvarsity.edu.bd
Dept. of ETE
Daffodil International University
102, Sukrabad, Dhanmondi Dhaka-1205, Bangladesh.

Abstract

A comparative analysis of basic and hybrid queuing methods and their impact on the VoIP traffic delay within the network have been discussed in this paper. In this experiment the hybrid queuing method is formed by combining three of the most useful basic queuing methods. In basic queuing methods priority queuing (PQ), custom queuing (CQ) and weighted fair queuing (WFQ) are combined with class-based weighted fair queuing (CBWFQ) for hybrid queuing. Hybrid method is compared with the basic method. All the basic and hybrid queuing mechanisms are tested using an OPNET Modeler simulation tool. From this, it is found how queuing combinations affect VoIP traffic quality, especially QoS.

Keywords: PQ, CQ, WFQ, CBWFQ, Hybrid queue, QoS.

1. Introduction

Generally the Internet system is based on Internet protocol (IP) and it supports only best effort services. As the Internet is growing very rapidly day by day and for this IP networks are expected to support not only typical services like ftp and email, but also real-time services and video streaming application. The traffic characteristics of these applications require a certain Quality of Service (QoS) from the network in terms of bandwidth and delay requirements [1].

The internet is intensifying on a daily basis, and the number of network infrastructure components is hastily increasing. Routers are most universally used to interconnect different networks. One of their tasks is to keep the proper quality of service (QoS) level. In case of VoIP the requirement is to deliver packets in less than 150ms. This limit is set to a level where a human ear cannot recognize variations in voice quality. This is one of the main reasons why leading network equipment manufacturers implement the QoS functionality into their solutions. QoS is a very complex and comprehensive system which belongs to the area of priority congestions management. It is implemented by using different queuing mechanisms, which take care of arranging traffic into waiting queues. Time-sensitive traffic should have maximum possible priority provided. However, if a proper queuing mechanism (FIFO, CQ, WFQ, etc.) is not used, the priority loses its initial meaning. It is also a well-known fact that all elements with memory capability involve additional delays during data transfer from one network segment to another, so a proper queuing mechanism and a proper buffer length should be used, or the VoIP quality will deteriorate [2].

All well-known queuing disciplines have both advantages and disadvantages. The main objective in these simulations is oriented towards improving the networks performances in terms of VoIP end-to-end delay and QoS. In typical queuing method different application shows the best performance for different specific techniques but there is no common technique which is best for different application. It is used only typical queuing methods as used in the network equipment (routers), well-known under the CQ, PQ, WFQ, and CBWFQ abbreviations. In order to show how hybrid queuing mechanisms of specific methods affect the VoIP traffic delay within the network, it is created simulations where hybrid queuing disciplines are compared with basic queuing schemes (PQ-CBWFQ with PQ), etc.

The remainder of the paper is organized as follows. Section 2 discusses related work in this field. Section 3 provides overview of QoS. In section 4, typical queuing mechanisms have been discussed. Section 5 contains the hybrid queuing techniques and its salient features. Section 6 and 7 contains the experimental overview, simulation and result. The last section will present conclusions and future research.

2. Related Work

Much similar research using simulations has been done in the area of VoIP's quality improvement; some of it is presented in the following literature: Mansour J. Karam & Fouad A. Tobagi, 2001 [3], Velmurugan T. et al., 2009 [4], and Fischer, M.J. et al., 2007 [5]. VoIP's quality improvement is a very popular research area, mostly focused on queuing aspects, and the problem of decreasing jitter influence as in some case. The hybrid queuing mechanism concept is our original contribution, resulting from the research of the past three years.

Hybrid queuing such as PQ-CBWFQ and WFQ-CBWFQ technique is imposed in VoIP to reduce the end-to-end delay, Ethernet delay, and jitter in [6]. They have compared the Ethernet delay over the typical queuing techniques and as well as various hybrid queuing mechanisms also. They found that for Ethernet delay case WFQ-CBWFQ shows the better result and for the average Voice jitter the typical queuing technique (WFQ) shows the better result rather than hybrid queuing mechanism (WFQ-CBWFQ) [6].

In VoIP, various applications such as FTP, Video and Voice supports different typical queuing techniques [7]. They showed in this paper that different applications give better QoS in different typical queuing techniques. For the case of End to End delay, packet delay and traffic of the video application PQ queuing technique shows a better performance. A. H. Muhamad Amin shows the comparison on different parameters between RTP packets and SIP VoIP communication [8]. It is shown improved QoS parameter on VoIP Communication in this paper.

3. QoS Overview

Quality of Service allows control of data transmission quality in networks, and at the same time improves the organization of data traffic flows, which go through many different network technologies. Such a group of network technologies includes ATM (asynchronous transfer mode), Ethernet and 802.1 technologies, IP based units, etc.; and even several of the abovementioned technologies can be used together. QoS is a network mechanism, which successfully controls traffic flood scenarios, generated by a wide range of advanced network applications. This is possible through the priorities allocation for each type of data stream [2]. QoS mechanism, observed as a whole, roughly represents an intermediate supervising component placed between different networks, or between the network and workstations or servers that may be autonomous or grouped together in local networks. The position of the QoS system in the network is shown in Fig. 1. This mechanism ensures that the applications with the highest priorities (VoIP, Skype, etc.) have priority treatment. QoS architecture consists of the following main elementary parts: QoS identification, QoS classification, QoS congestions management mechanism, and QoS management mechanism, which handle the queue [2].

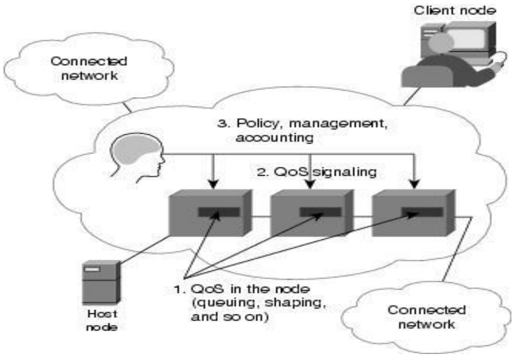
Client node
Connected network
3. Policy, management, accounting
2. QoS signaling
1. QoS in the node (queuing, shaping, and so on)
Host node
Connected network

Figure 1: Quality of Service

4. Typical Queuing Discipline

As part of the resource allocation mechanisms, each router must implement some queuing discipline that governs how packets are buffered while waiting to be transmitted. Various queuing disciplines can be used to control which packets get transmitted (bandwidth allocation) and which packets get dropped (buffer space). The queuing discipline also affects the latency experienced by a packet, by determining how long a packet waits to be transmitted. Examples of the common queuing disciplines are first-in first- out (FIFO) queuing, priority queuing (PQ), and weighted-fair queuing (WFQ) [9].

First in, First out (FIFO): FIFO is an acronym for First In First Out .This expression describes the principle of a queue or first-come first serve behavior: what comes in first is handled first, what comes in next waits until the first is finished etc. Thus it is analogous to the behavior of persons “standing in a line” or “Queue” where the persons leave the queue in the order they arrive. First In First Out (FIFO) is the most basic queuing discipline which is shown in Fig. 2. In FIFO queuing all packets are treated equally by placing them into a single queue, then servicing them in the same order they were placed in the queue. FIFO queuing is also referred to as First Come First Serve (FCFS) queuing [7]. Generally, FIFO queuing is supported on an output port when no other queue scheduling discipline is configured. In some cases, router vendors implement two queues on an output port when no other queue scheduling discipline is configured: a high-priority queue that is dedicated to scheduling network control traffic and a FIFO queue that schedules all other types of traffic.

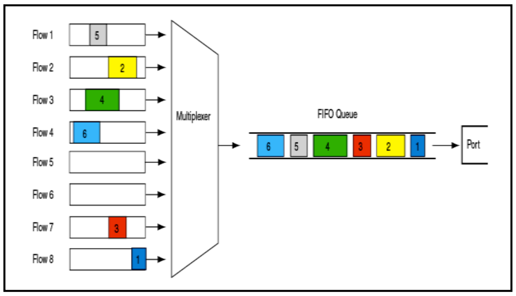

Figure 2: First-in, First-out (FIFO) Queuing

Priority Queuing (PQ): PQ is a simple variation of the basic FIFO queuing. The idea is to mark each packet with a priority; the mark could be carried, for example, in the IP Type of Service (ToS) field. The routers then implement multiple FIFO queues, one for each priority class. Within each priority, packets are still managed in a FIFO manner. This queuing discipline allows high priority packets to cut to the front of the line [9]. Priority queuing (PQ) is the basis for a class of queue scheduling algorithms which is shown in Fig. 3 that are designed to provide a relatively simple method of supporting differentiated service classes. In classic PQ, packets are first classified by the system and then placed into different priority queues. Packets are scheduled from the head of a given queue only if all queues of higher priority are empty. Within each of the priority queues, packets are scheduled in FIFO order [10].

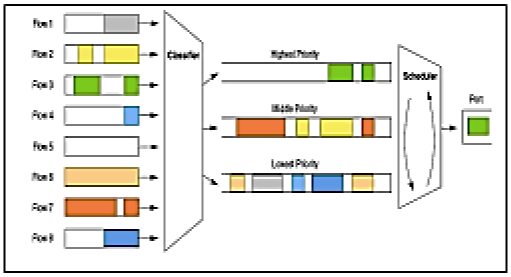

Figure 3.Priority Queuing (PQ)

Fair Queuing (FQ): Fair queuing (FQ) was proposed by John Nagle in 1987. FQ is the foundation for a class of queue scheduling disciplines which is shown in Fig. 4 that are designed to ensure that each flow has fair access to network resources and to prevent a bursty flow from consuming more than its fair share of output port bandwidth. In FQ, packets are first classified into flows by the system and then assigned to a queue that is specifically dedicated to that flow. Queues are then serviced one packet at a time in round-robin order. Empty queues are skipped. FQ is also referred to as per-flow or flow-based queuing [10].

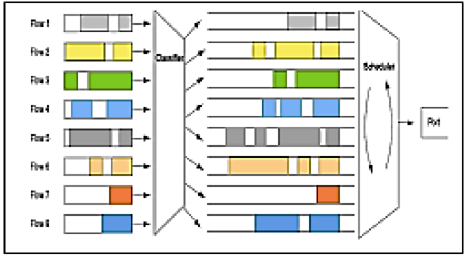

Figure 4: Fair Queuing (FQ)

Weighted Fair Queuing (WFQ): Weighted fair queuing (WFQ) was developed independently in 1989 by Lixia Zhang and by Alan Demers, Srinivasan Keshav, and Scott Shenke. WFQ is the basis for a class of queue scheduling disciplines that are designed to address limitations of the FQ model.

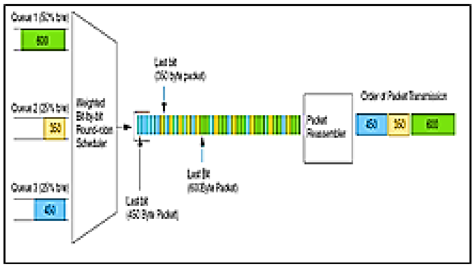
Queue 1 (50% b/w)
600
Queue 2 (25% b/w)
350
Queue 3 (25% b/w)
450
Weighted Bit-by-bit Round-robin Scheduler
Last bit (350 byte packet)
Last bit (600 Byte Packet)
Last bit (450 Byte Packet)
Packet Reassembler
Order of Packet Transmission
450
350
600

Figure 5: A Weighted Bit-by-bit Round-robin Scheduler with a Packet Re-assembler.

The idea of the fair queuing (FQ) discipline is to maintain a separate queue for each flow currently being handled by the router. The router then services these queues in a round-robin manner. WFQ allows a weight to be assigned to each flow (queue). This weight effectively controls the percentage of the link's bandwidth each flow will get. We could use ToS bits in the IP header to identify that weight. [10]

Fig. 5 shows a weighted bit-by-bit round-robin scheduler servicing three queues. Assume that queue 1 is assigned 50 percent of the output port bandwidth and that queue 2 and queue 3 is each assigned 25 percent of the bandwidth. The scheduler transmits two bits from queue 1, one bit from queue 2, one bit from queue 3, and then returns to queue 1. As a result of the weighted scheduling discipline, the last bit of the 600-byte packet is transmitted before the last bit of the 350-byte packet, and the last bit of the 350-byte packet is transmitted before the last bit of the 450-byte packet. This causes the 600-byte packet finishes (complete reassembly) before the 350-byte packet, and the 350-byte packet finishes before the 450-byte packet.

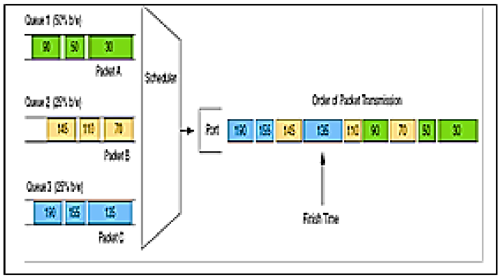

Figure 6: Weighted Fair Queuing (WFQ)—Service According to Packet Finish Time

When each packet is classified and placed into its queue, the scheduler calculates and assigns a finish time for the packet. As the WFQ scheduler services its queues, it selects the packet with the earliest (smallest) finish time as the next packet for transmission on the output port. For example, if WFQ determines that packet A has a finish time of 30, packet B has a finish time of 70, and packet C has a finish time of 135, then packet A is transmitted before packet B or packet C. In Fig 6., observe that the appropriate weighting of queues allows a WFQ scheduler to transmit two or more consecutive packets from the same queue.

Class-based weighted fair queuing (CBWFQ): CBWFQ represents the newest scheduling mechanism intended for handling congestions while providing greater flexibility [11]. It is usable in situations where we want to provide a proper amount of the bandwidth to a specific application (in our case VoIP application). In these cases, the network administrator must provide classes with defined bandwidth amounts, where one of the classes is, for example, intended for a videoconferencing application, another for VoIP application, and so on. Instead of waiting-queue assurance for each individual traffic flow, CBWFQ determines different traffic flows. A minimal bandwidth is assured for each of such classes. One case where the majority of the lower- priority multiple-traffic flow can override the highest- priority traffic flow is video transmission which needs half of the T1-connection bandwidth [12]. A sufficient link bandwidth would be assured using the WFQ mechanism, but only when two traffic data flows are present. In a case where more than two traffic flows appear, the video session suffers the regarding bandwidth, because the WFQ mechanism works on the fairness principle. For example, if nine additional traffic flows make demands of the same bandwidth, the video session will get only 1/10th of the whole bandwidth, and this is insufficient when using a WFQ mechanism. Even if we put an IP priority level of 5 into the ToS field [13] of the IP packet header, the circumstances would not change. In this case, the video conference would only get 6/15 of the bandwidth, and this is not enough because the mechanism must provide half of all the available bandwidth on the T1 link. This can be provided by using the CBWFQ mechanism. The network administrator just defines, for example, the video-conferencing class and installs a video session into that class. The same principle can be used for all other applications which need specific amounts of the bandwidth. Such classes are served by a flow-based WFQ algorithm which allocates the remaining bandwidth to other active applications within the network [11].

5. Hybrid Queuing Discipline

Because different queuing mechanisms have different advantages, combine different queuing mechanisms and join their positive (but also negative) properties into new hybrid queuing methods. The aim of hybrid methods is to concentrate the most possible positive properties of individual methods. Many different hybrid queuing methods are possible. This section provides descriptions of hybrid queuing disciplines for the combinations of CQ-CBWFQ and WFQ-CBWFQ as well as for the known PQ-CBWFQ introduced by Cisco Systems. Each of these methods was evaluated with simulations.

CQ-CBWFQ hybrid waiting queue: This hybrid method combines the properties of the custom queuing (CQ) and the CBWFQ mechanisms (Fig. 7). In the first phase the custom queuing allocates the available bandwidth among all active network applications so that overcrowding cannot appear. In the CQ step traffic is managed by assigning weighted amounts, and is arranged into 16 queues. Once the packets are sent to the output CQ interface they arrive to the CBWFQ input interface. CBWFQ packet-classification mechanism, attached behind the custom queuing mechanism, arranges traffic into traffic classes defined by a class-based weighted fair queuing algorithm. Such classes are then ensured with fixed amounts of bandwidth. All the advantages of the CBWFQ are retained. With this method it is reduced the delays within the network, which is not the case with the ordinary CQ scheme.

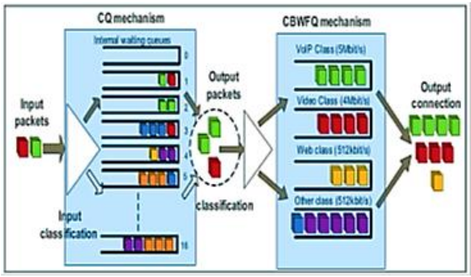

Figure 7: The hybrid queuing mechanism consisting of the CQ and the CBWFQ

PQ-CBWFQ hybrid waiting queue: This mechanism consists of two previously mentioned queuing mechanisms; the priority queuing mechanism and the admin class-defined queuing mechanism (CBWFQ). Since the properties of both mechanisms that construct the PQ-CBWFQ method have been already mentioned in previous sections, we should now take a look at the hybrid mode concept shown in Fig. 8.

In the first step the traffic is arranged into waiting queues according to the priorities set in individual packets' ToS fields. According to the ToS priorities the packets are arranged by their importance to four different internal priority queues. In the second step, the output interface algorithm first serves the highest-priority data stream (packets that are in the queue with the highest importance) and then all other lower ranking queues. Once the packets appear at the outgoing interface of the priority queuing mechanism, they are again scheduled into admin-defined classes of CBWFQ mechanism. Such defined classes already have the needed bandwidth pre-reserved as set by the network administrator. This way the packets at the CBWFQ mechanism output interface do not need to fight for bandwidth, as it is guaranteed in advance. This accelerates the transfer of high-priority flows, and such flows become independent of all other lower-priority flows.

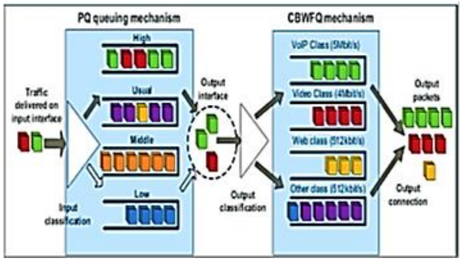

Figure 8: The hybrid queuing mechanism consisting of the PQ and the CBWFQ

WFQ-CBWFQ hybrid waiting queue: This mechanism consists of the weighted fair queuing (WFQ) and the class-based weighted fair queuing (CBWFQ). At the first step, we ensure undisturbed flow throughput for all active applications that appear at the WFQ mechanism's outgoing interface. In the next step the CBWFQ classification takes care of proper packets' assignment into admin-defined classes. This way every application at the first stage gets a fair treatment, and in the second phase high-priority applications get its own classes with the pre-reserved bandwidth. The rest of the bandwidth is left for all other active applications. The fairness and fluidity movement apply for all active applications (Fig. 9).

WFQ is suitable for operating with IP priority settings, such as Resource Reservation Protocol (RSVP) which is also capable of managing round-trip delay problems. Such queuing clearly improves algorithms such as SNA (Systems Network Architecture) - Cisco SNA (CSNA) which is an application that provides support for SNA protocols to the IBM mainframe. Using a Cisco 7000, 7200, or a 7500 Series router with a Channel Interface Processor (CIP) or Channel Port Adapter (CPA) and Cisco SNA (CSNA) support enabled, we can connect two mainframes (either locally or remotely), to a physical unit (PU) 2.0 or 2.1, or connect a mainframe to a front-end processor (FEP) in another Virtual Telecommunications Access Method (VTAM) domain logical link control (LCC) or, transmission control protocol (TCP). WFQ-CBWFQ is at the same time capable of accelerating slow features and removing congestions in the network.

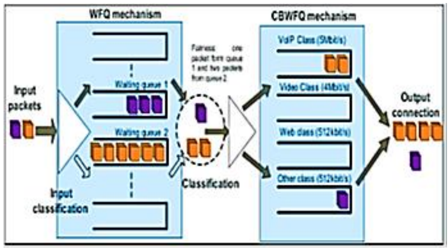

Figure 9: The hybrid queuing mechanism consisting of the WFQ and the CBWFQ

6. Experimental Description

The scenario of the simulation topology consists of remote servers, VoIP and Web clients (spread across specific geographic areas), switches, routers, etc. Some properties of the entire wide area network, such as delay, packet loss, etc. are described in "IP Cloud" element. The whole network structure (Fig, 10), public network, individual users, etc. is connected through an IP cloud to remote servers in the WAN network. Four external LANs (LAN1, LAN2, LAN3 and LAN4), where each of them contains of 50 VoIP users, establish connections to the VoIP users at the other end of the WAN network using a 10 Mbit/s wired broadband connection. In each of the local area networks, there are also World Wide Web (WWW) users, which exploit a part of the available bandwidth. These users causing affect the VoIP traffic delay, but only in the cases, when inappropriate QoS and queuing mechanisms are used. A fast connection allows exchange of large amounts of data between units, and at the same time ensures small time delays, which is crucial for the VoIP sessions. The wide area network (WAN) simulation structure is shown in Fig. 10. All active applications are designed in the OPNET Modeler simulation tool in the form of three different scenarios. The first scenario consists of the CQ queuing method, second only of the PQ queuing method; while the third scenario consists of the PQ-CBWFQ queuing regime, which belongs to the low latency queuing group. Through a comparison of all mentioned scenarios, the following results have been obtained.

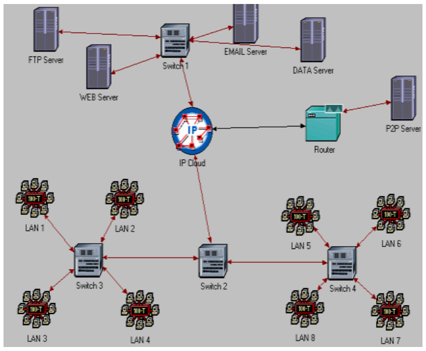

Figure 10: Simulation structure of the wide area network

7. Simulation Results

The scenario of the simulation topology consists of remote servers, VoIP and Web clients (spread In these network simulations different queuing methods have been used for IP traffic and have been measured the traffic delays corresponding to typical and hybrid queuing method. Results are presented in Fig. 11 Curves (1), (2) and (3) shows the average VoIP traffic delay for the used CQ, PQ and PQ-CBWFQ queuing mechanisms. Based on the simulation results shown in Fig. 11, the relationship factors have been calculated and evaluated. From this evaluation, it is found that how much the chosen method of classification for the specific observed network traffic better than the typical method.

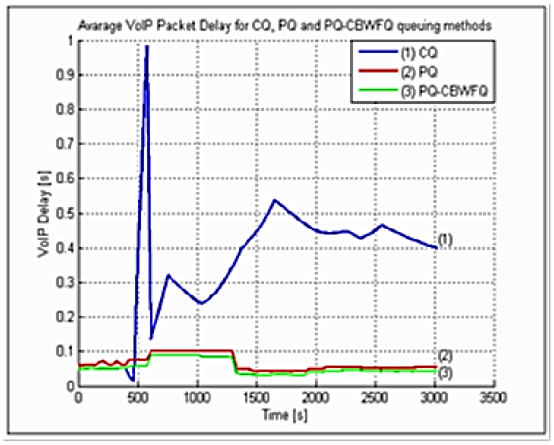

Figure 11: The average VoIP traffic delay for the used CQ, PQ and PQ-CBWFQ queuing mechanisms.

CQ and PQ are used as a basic reference queue method in comparisons. Relationships were calculated by averaging CQ delay and PQ delay of VoIP traffic and dividing it by averaged delays of the VoIP traffic with other methods (Table 1).

Table I: The calculated average delay for each of the individual waiting queue methods.

Method	**CQ**	**PQ**	**PQ-CBWFQ**
Avg. delay in second	0.346488	0.06444	0.052956

Table II: Calculated relationship factors, which describe the usefulness of an individual method in comparison to others.

Methods in comparison	$\frac{CQ}{PQ}$	$\frac{CQ}{PQ - CBWFQ}$	$\frac{PQ}{PQ - CBWFQ}$
Relationship factors	5.37	6.54	1.21

Table III: Calculated relationship factors, which describe the usefulness of an individual method in comparison to others.

Methods in comparison	$\frac{PQ}{CQ}$	$\frac{\mathrm{PQ - CBWFQ}}{\mathrm{CQ}}$	$\frac{\mathrm{PQ - CBWFQ}}{\mathrm{PQ}}$
Relationship factors	0.186	0.152	0.821

From the calculated factors, it is seen that the PQ and PQ-CBWFQ queuing mechanisms are most suitable for time-sensitive applications, especially for VoIP. From their comparison, it can be concluded that the PQ method is better for a factor 5.37 than the custom queuing method, and PQ-CBWFQ combination is for a factor 6.54 better than the basic CQ method. In simulation results, this can be observed in the form of the smallest delays for a specific application. In PQ and PQ-CBWFQ cases, the VoIP delay is lower than in the CQ case, and it does not exceed the critical delay (150ms), which represents the limit where the human ear can detect it. When both sophisticated methods are compared, the PQ-CBWFQ is for a factor 1.21 better than PQ queuing regime. Simulation results show how important the right choice and configuration of the queuing mechanisms are for time-sensitive traffic.

8. Conclusion

The scenario of the simulation topology consists of remote servers, VoIP and Web clients (spread In conclusion, queuing combinations express satisfactory results for one criterion as time sensitive application preferably in the case of VoIP application has been observed. Regarding VoIP delays the PQ and PQ-CBWFQ queuing schemes are most suitable. In such cases also the voice quality is on a higher level, compared to those where ordinary queuing schemes (CQ, for example) are used. In cases where it is needed to make a compromise between important traffic and traffic of lower importance, the PQ-CBWFQ hybrid method gives satisfying results. Based on the simulation results, it has been proved that PQ-CBWFQ is a low-latency queuing scheme and a proper solution for a VoIP time-sensitive application because it has the smallest average packet delay compared to any other queuing schemes.

9. References

1. Davide Astuti,“ Packet Handling”, http://marco.uminho.pt/disciplinas/ST/ packethandling.pdf
2. Influences of Classical and Hybrid Queuing Mechanisms on VoIP’s QoS PropertiesSasa Klampfer, Amor Chowdhury, Joze Mohorko2 and Zarko Cucej
3. Mansour J. Karam, Fouad A. Tobagi, “Analysis of the Delay and Jitter of Voice Traffic Over the Internet”, IEEE InfoCom 2001
4. Velmurugan, T.; Chandra, H.; Balaji, S.; , "Comparison of Queuing Disciplines for Differentiated Services Using OPNET," Advances in Recent Technologies in Communication and Computing, 2009. ARTCom '09., Vol., no., pp.744-746, 27-28 Oct. 2009
5. Fischer, M.J.; Bevilacqua Masi, D.M.; McGregor, P.V.; "Efficient Integrated Services in an Enterprise Network," IT Professional, vol.9, no.5, pp.28-35, Sept. Oct. 2007Cisco Systems, Understanding Jitter in Packet Voice Networks (Cisco IOS Platforms).
6. Impact of hybrid queuing disciplines on the VoIP traffic delay Saša Klampfer, Jože Mohorko, Žarko Čučej Faculty of Electrical Engineering and Computer Science, 2000 Maribor, Slovenia E-pošta: sasa.klampfer@uni-mb.si
7. Acomparative study of different queuing techniques in VoIP, video conferencing and file transferbyMohammad Mirza Golam Rashed and Mamun Kabir Department of ETE, Daffodil International University
8. VoIP performance measurement using QoS parameters A.H.Muhamad Amin IT/IS Department, Universiti Teknologi PETRONAS, 31750 Tronoh, Perak Darul Ridzuan, Malaysia
9. http://www.eng.tau.ac.il/~netlab/resources/booklet/lab9.pdf
10. http://www.cse.iitb.ac.in/~varsha/allpapers/packet-scheduling/ wfqJuniper.pdf
11. Internetworking Technology Handbook – Quality of Service (QoS), Cisco Systems.
12. G. 729 Data Sheet.
13. http://en.wikipedia.org/wiki/Type_of_Service.