Improving network performance with an integrated priority queue and weighted fair queue scheduling

Tsehay Admassu Assegie, Haymanot Derebe Bizuneh
Department of Computing Technology, College of Engineering and Technology, Aksum University, Ethiopia

Article Info

Article history:

Received May 28, 2019
Revised Dec 15, 2019
Accepted Jan 17, 2020

Keywords:

Congestion avoidance
Congestion management
Network congestion
Packet scheduling
Quality of service

ABSTRACT

Quality of service (QoS) is the measure of network service availability and transmission. There are many factors influencing QoS among which one is the increasing number of network service users. The increase in the number of network service users and communication traffic causes network congestion. And the traffic congestion results in delay or packet loss and jitter variation. As a result, an organization's network quality deteriorates and or even becomes unavailable. Therefore, to deliver a high quality network service to the users, a solution that avoids network traffic congestion is needed. In this study, the causes for network traffic congestion and the best solutions to eliminate traffic congestion in a network with congestion management and avoidance using an integrated priority queue (PQ) and weighted fair queue (WFQ) packet scheduling algorithms is proposed.

Corresponding Author:

Tsehay Admassu Assegie,
Department of Computing Technology,
1010 Aksum University, Ethiopia.
Email: tsehayadmassu2006@gmail.com

1. INTRODUCTION

In today's network communication, network service users' have an increasing demand for high quality of service. Because of a high demand for high quality network and the growing number of user's, ensuring quality of service is becoming an issue. When the network is congested, the communication quality of network services such as video, voice and video conferencing cannot be guaranteed. Therefore, a mechanism that improves end-to-end communication quality is required for avoiding the network congestion problem so that the quality of network can be improved. The factors affecting network quality are bandwidth, delay, jitter and packet loss. The bandwidth is the amount of data units transmitted per second [1].

The maximum bandwidth is the minimum link bandwidth on the transmission path. Hence, the minimum link bandwidth mainly affects the transmission rate of a link. By increasing network bandwidth better network quality can be achieved but, the increase in bandwidth incurs additional cost and time. Delay is another influencing factor affecting network performance. Delay refers the period of time during which packet is transmitted from source to destination [2].

The quality of real-time service such as voice and video conferencing depends on delay. In voice service, delay is the time during which a word is spoken and the word is heard by the receiver. The other factor that affects network performance is jitter which is the variation in end-to-end delay for each packet. The packet arrival time difference causes disruptions of voice and video conferencing services. In addition, the jitter affects processing of network protocols. Apart from bandwidth, delay and jitter, packet loss is another quality of service influencing factor that significantly affects the network performance. The packet loss occurs when a device drops packet because of limited buffer size. In a packet switched network, when a device such as switch or router receives packets that are destined to the same output interface, then some of the packets are stored in buffer until the packet leaving the output interface is forwarded to the destination. In cases where a large volume of packets are simultaneously received, packet loss will occur due to a limited buffer size.

Network congestion occurs when the network traffic increases and exceeds the inbound bandwidth of a device. When the network is congested, the quality of delay sensitive services, such as voice and video cannot be guaranteed. The major causes for network congestion are:

a) Rate mismatch: Packets are received by a device through a high-speed interface and are forwarded through a low-speed interface.
b) Traffic Aggregation: occurs when packets are received by a device from multiple links and are forwarded through a single link with the same bandwidth capacity.
c) Link bandwidth- the capacity of link transmission rate cannot accommodate the network traffic.

Network congestion is a condition in which an increased data transmissions in a network results in a proportionately smaller increase, or even a reduction, in throughput [3]. Network congestion has adverse effects in the network [4]. Packets with different quality of service requirements flow within the network and congestion causes loss of these packets. A congested network experiences the following adverse effects:

a) Increased packet transmission delay and jitter.
b) Packet retransmission due to packet loss and longer delays.
c) Lower network throughput.
d) Consumption of a large number of network resources, especially the storage resources.
e) Improper resource allocation causes resources to be locked and the system goes down.

2. RELATED WORKS

This section focuses on previous works on network performance and proposed solution to provide quality of service to delay sensitive services such as voice and video. In [5], there are two approaches proposed to solve the quality of service challenges and deliver a high quality communication in a network. The first one is, provisioning of unlimited resource, for example increasing the bandwidth of an interface or a link. The provisioning of unlimited resource is the simplest approach to improve quality of service but, provisioning of high bandwidth is costly and impractical hence, this approach is not the preferred approach for improving quality of service in network. The second approach is congestion management and avoidance with a limited resource and without incurring extra cost for link bandwidth. In this approach, congestion management techniques such as packet scheduling and traffic prioritization are employed to improve network performance.

Today’s network incorporates multimedia data such as video, voice and data which is transmitted over packet switched network [6]. The voice and video data are highly sensitive to quality of service. Therefore, delivering a high quality video and voice traffic service requires congestion management when the network is occasionally congested. The congestion management ensures that the key packets with higher priority are forwarded and packets with lower priority are placed in buffer until all higher priority packets are forwarded. A congestion management is queuing and queue scheduling technologies to determine the packet to be forwarded first and packet to be placed in buffer. In this approach if the queue is full of non-key packets, then all of the packets in the queue will be discarded. Hence, avoiding packet discarded by congestion management approach, congestion avoidance is applied to provide quality service.

As showcased by Anita Swain and Arun Kumar Ray in [7], priority queue (PQ) scheduling is designed for key service applications for congestion management. In this model, the key services such as voice and video which are sensitive to delay and jitter are scheduled favorably to reduce the response delay when congestion occurs in the network. In the PQ scheduling mechanism, four queues are required, namely high-priority, medium-priority, normal priority, and low-priority queues. When packets are forwarded out from queues, the device forwards packets in the higher-priority queue first. When all of the packets in the higher-priority queue are sent or forwarded, the device forwards packets in the medium-priority queue. Once, all of the packets in the medium-priority queue are sent or forwarded, the device forwards packets in the normal priority queue, and then packets in the lower priority queue are forwarded at the last. The packets of core services or delay sensitive data such as voice and video are placed in the higher-priority queues, and the packets of non-core services such as email and web traffic are placed in lower-priority queues. Hence, the packets of key services such as voice and video are processed first and non-core services such as email and web traffic are processed when the core services are processed. The open research question of applying this approach is that, if packets in the high-priority queue are forwarded continuously, then the packets in the lower-priority queue cannot be sent.

To improve the quality service, network traffic is classified into different categories [8]. This is because different packets have different quality of service requirement as summarized in Table 1.

For example, voice is delay sensitive but not bandwidth intensive and video is delay sensitive and bandwidth intensive.

Table 1. Packet types and quality of service requirement

Packet type	Behavior	Packet loss ratio
Voice	Delay sensitive	Low
Video	Bandwidth intensive	Low
Data	Loss sensitive	Low
FTP	Medium	High
Email, Web page browsing	low	Medium

To guarantee a quality of service in a network, different approaches are required to be implemented [9]. Among the approaches that improve the Quality of Service in a network queue scheduling algorithms such as priority queue and weighted fair priority queue are the most common approaches applied for congestion avoidance and congestion management. In this study, our focus is on the integration of priority queue scheduling (PQ) and weighted fair queuing (WFQ) scheduling algorithms to achieve a better quality of service to delay sensitive traffics such as voice and video.

The priority queue scheduling algorithm, as demonstrated in Figure 1, places packets into high priority, medium priority, normal and lower priority queues. When the packet is received, this algorithm places into the queues based on the traffic type or class. Likewise, the weighted fair queue (WFQ) scheduling algorithm, stores into different queues based on the traffic characteristics. Network packets with the same source Internet Protocol (IP) address, destination IP address, source port number, destination port number, protocol number, belong to the same flow. Each flow is assigned to a queue. Flows with different characteristics are sent to different queues. The number of queues allowed by WFQ is limited and configurable. When flows leave queues, WFQ allocates the egress bandwidth to each flow based on the precedence of each flow. A flow with the lowest priority gets the least bandwidth. In this approach, services of the same priority are treated in the same manner; services of different priorities are allocated with different weights. WFQ configuration is simple. Traffic is classified automatically, without manual intervention. Therefore, WFQ is inflexible. When multiple flows enter the same queue, WFQ limited by resources cannot provide accurate services and cannot ensure resources obtained by each service. WFQ balances the delay and jitter of each flow, hence, WFQ is not suitable for delay sensitive service applications.

In priority queue scheduling, the device first schedules traffic in higher priority queue, then medium priority queue, and lower priority queue. Important protocol packets or short-delay service packets are placed in queues using PQ scheduling so that they can be scheduled first. Other packets are placed in queues using WFQ scheduling. Using PQ and WFQ integrates advantages of PQ and WFQ. If only the PQ scheduling algorithm is applied to ensure quality of service, packets in queues with lower priorities may not get bandwidth for a longer period of time. When only the WFQ scheduling is applied, short-delay services such as voice service cannot be scheduled first. To avoid this issue, we proposed integrated PQ and WFQ scheduling algorithm to guarantee Quality of Service requirement for delay sensitive video and voice traffic.

In [10], the authors showcased that, the network traffic congestion can be avoided by implementing different congestion management and avoidance approaches at the data link and network layers. At data link layer, link rate limiting prevents an excess traffic beyond the capacity of the link bandwidth. At the network layer, routing policy like discard policy, traffic filtering policy, traffic shaping, traffic classification and packet queuing is applied to avoid congestions in a network. at the link layer bandwidth provisioning is applied to guarantee a higher quality of service in the network.

In another study [11] a flow control based network traffic congestion management and avoidance approach is proposed. In this approach, the volume of network traffic flow is controlled to ensure quality of service. The study showcased that the network performance improves when flow control is employed with token bucket approach for network traffic control.

In [12], the performance of the fair and delay adaptive scheduler and weighted fair queue scheduling algorithm are compared. The schedulers are evaluated against fairness on bandwidth allocation to different network traffic classes. The result shows that the fair and delay adaptive scheduler and weighted fair queue scheduling have identical bandwidth allocation.

In [13], the effect of mobility models on software defined wireless network is emulated and the result shows that different mobility models have different performance. This reveals that choosing a better performing mobility model improves the network performance in wireless network.

In [14], priority based dynamic quality of service management protocol is proposed to improve the quality of a network. The proposed model avoids network congestion and ensures on time delivery of packets to the intended destination.

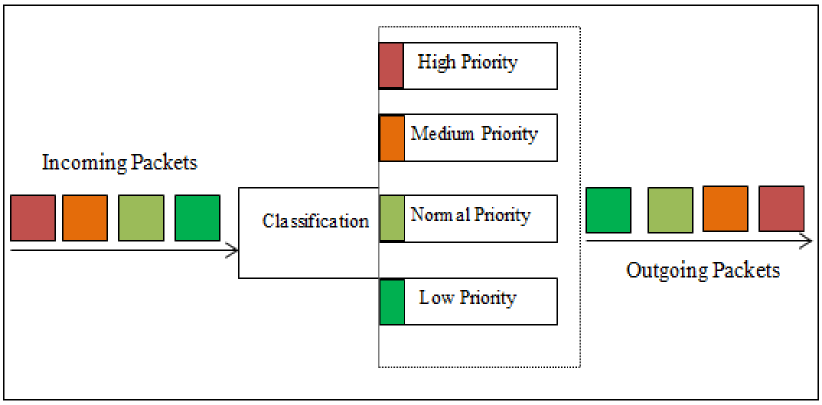

Figure 1. Priority queue scheduling algorithm

3. NETWORK CONGESTION MANAGEMENT

Quality of service ensures that the network applications and services are operating as expected [15-17]. The Quality of Service guarantees a high performance network. The performance of a network is the speed and reliability of information delivery in the network. The approaches to guarantee high network performance are:

a) Increasing interface bandwidth
b) Improving processing performance-using high end processors on routers and other network devices.
c) Compress and fragment packets-reducing the size of packet through data compression and fragmentation so that they can be transported over a link with limited bandwidth. This approach is also termed as link efficiency mechanism.
d) Implementing queue scheduling and congestion avoidance mechanism-ensure high priority packet not to be drooped and ensure that the lower priority packets are processed after the higher priority packets are processed.

Applying a better packet scheduling algorithm is critical to the provisioning of high network quality of service in communication [18-21]. But, packet queue scheduling algorithms have their own advantages and disadvantages. In situations where congestion cannot be managed through packet queue scheduling, the congestion is avoided by employing either of the following congestion avoidance approaches:

a) Trail drop-after a queue becomes full, the packet that is arriving will be dropped.
b) Random early detection (RED)-packets are dropped randomly after the queue has become full.
c) Weighted random early detection (WRED)-queue length and priority are taken into account to decide on the packet to be dropped, in this cases packets with low priority are discarded earlier as the drop probability for such packets is higher.

Traffic shaping and classification is implemented to manage network congestion and maximize bandwidth utilization [22-25]. The shaping classifies and places packets into different queues. When the network traffic exceeds the bandwidth limit, the incoming packets are placed in waiting queue until the data rate conforms to the bandwidth capacity. In traffic classification, packets are labeled with integer number and those labels are used to determine the queue in which an incoming packet is placed.

4. RESULTS AND ANALYSIS

To integrate PQ and WFQ scheduling algorithm and analyze the traffic behavior where these scheduling algorithms are applied, the simulation topology demonstrated in Figure 2 is used. In the experiment a Huawei AR2220 series router is used. The PQ and WFQ scheduling algorithms are implemented on the AR2220 series router. The delay variation for two scenarios, one a scenario in which PQ+WFQ is implemented and another scenario where PQ+WFQ is not implanted is analyzed. The delay time for each case on the network service is measured.

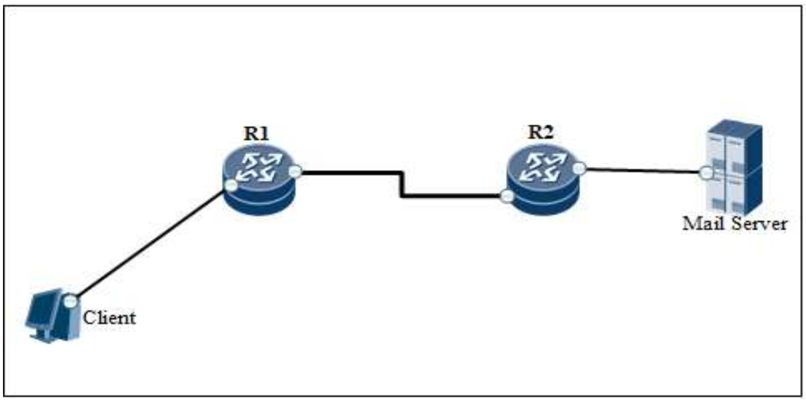

Figure 2. PQ+WFQ scheduling simulation topology

The PQ and WFQ configuration, a queue-based congestion management, puts data traffic into the WFQ and high priority and delay sensitive voice traffic to PQ queues as shown in Figure 3.

Queue	Schedule	Weight	Length(Bytes/Packets)	GTS(CIR/CBS)
1	WFQ	10	-/-	-/-
2	WFQ	10	-/-	-/-
3	WFQ	10	-/-	-/-
4	WFQ	10	-/-	-/-
5	PQ	-	-/-	-/-

Figure 3. Integrated PQ and WFQ scheduling

In the experiment, the data traffic is implemented so that the traffic enters WFQ queue and voice traffic enters the PQ queue. The WRED drop priority is based on DSCP priorities and set the upper drop threshold to 90, lower drop threshold to 50, and maximum drop probability to 30. The delay for network traffic in scenarios where no scheduling algorithms are implemented and the scheduling algorithms are implemented is shown in Figure 4. As shown in Figure 4, the delay for network traffic in the scenario where the integrated PQ and WFQ scheduling algorithms is implemented to control the congestion with PQ+WFQ scheduling algorithm against congested network is less than 500 millisecond and the delay for network traffic where the integrated scheduling algorithms are not implemented is 300 to 1000 milliseconds.

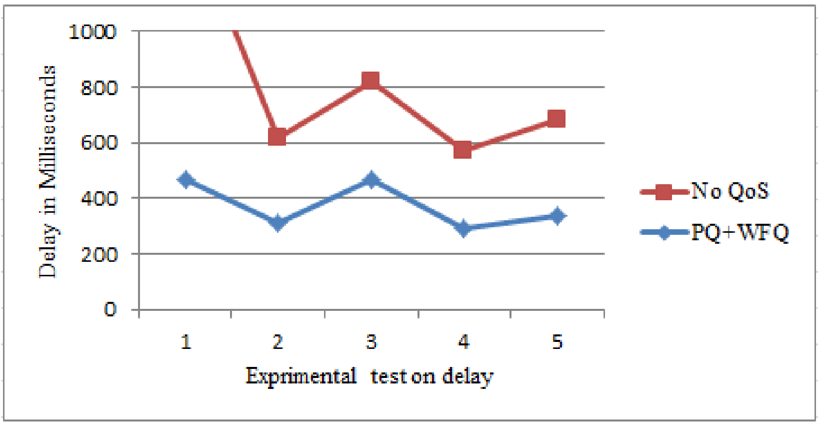

Figure 4. Delay for non-QoS vs PQ and WFQ traffic

5. CONCLUSION

In this study, we have proposed an integrated packet scheduling by employing the PQ and WFQ queue scheduling algorithms to avoid network congestion and improve network performance. The performance of the network is analyzed using the delay as performance metric and two scenarios, namely network with PQ and WFQ scheduling algorithm and the other scenario without packet scheduling. The performance analysis of the experimental result shows that, the integrated packet scheduling algorithm has significantly enhanced the network performance having lower delay compared to the test scenario where no packet scheduling algorithm is implemented to avoid network congestion.

REFERENCES

[1] Umeh Innocent Ikechukwu., "Network, Web & Security," *Global Journal of Computer Science and Technology*: Vol. 17, no. 4, 2017

[2] Abrar Noor Akramin, Wan Shahrum and Joseph William, Improving Security and Performance of IP Telephony (VoIP), August, 2015, [Online], Available : https://www.researchgate.net/publication/306099980_Improving_Security_and_Performance_of_IP_Telephony_VoIP

[3] Linfo, "Network Congestion Definition," 2005, [online], available, [http://www.linfo.org/congestion.html]

[4] Mohamed Nj, Sharin Sahib, Nanna Suryana, Burairah Hussin, "Understanding network congestion effects on performance," *Journal of Theoretical and Applied Information Technology,* October 2016

[5] Augustine C. Odinma, Lawrence Oborkhale, Quality of Service Mechanisms and Challenges for IP Networks, the Pacific Journal of Science and Technology, Vol. 7, No. 1, May 2006

[6] R.-T. Sheu, J.-L.C. Wu, "Performance analysis of rate control with scaling QoS parameters for multimedia transmissions," *IEE Proc.-Commun*, Vol. 150, No. 5, October 2003.

[7] Anita Swain, Arun Kumar Ray, QoS Management in WSN-MCN Convergence Network Using Priority Based Traffic Models. *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS),* vol. 12, no. 3, December 2018.

[8] Khurshid Lal Bhavan, Janpath, "Study paper on implementing quality of service in IP networks," R. Saji Kumar Director, J.M.Suri DDG, I Division, Telecom Engineering Centre, Department of Telecommunications, New Delhi.

[9] Swaroop R.Puranik, Vijayalakshmi M, Linganagouda Kulkarni, A Survey and Analysis on Scheduling Algorithms in IEEE 802.16e (WiMAX) Standard, *International Journal of Computer Applications,* vol. 79, no 12, October 2013.

[10] Y. Lan, Y. Sun, S. Liu and Z. Ma, "A real-time network traffic analysis and QoS management platform," 2017 *IEEE 9th International Conference on Communication Software and Networks (ICCSN*), pp. 266-270, Guangzhou, 2017.

[11] Ahmad Khafidin, Tatyantoro Andrasto, Suryono, "Implementation flow control to improve quality of service on computer networks," *Indonesian Journal of Electrical Engineering and Computer Science, (IJEECS,)* vol. 16, no. 3, December 2019.

[12] A. M. Elnaka, Q. H. Mahmoud and Xining Li, "Simulation based comparative performance analysis of QoS traffic scheduling using fair and delay adaptive scheduler (FDAS) versus WFQ and EDF," 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), pp. 916-923, Las Vegas, NV, 2016.

[13] Tsehay Admassu Assegie, Pramod Sekharan Nair, "The performance of Gauss Markov's mobility model in emulated software defined wireless mesh network," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 18, no. 1, April 2020.

[14] Madhumita Kathuria, Sapna Gambhir, Improvement of Quality of Service Parameters in Dynamic and Heterogeneous WBAN, *Indonesian Journal of Electrical Engineering and Informatics (IJEEI),* vol. 4, no. 4, December 2016.

[15] Winarno Sugeng, Jazi Eko Istiyanto, Khabib Mustofa, Ahmad Ashar, The Impact of QoS Changes towards Network Performance, *International Journal of Computer Networks and Communications Security,* vol. 3, no. 2, February 2015.

[16] Mohammed Dighriri, "Enhancing 5G Cellular network packets traffic by scheduling mechanisms," *Proceedings of the IRES International Conference*, Morocco, November 2011.

[17] W. Xie, X. Chen and D. Li, "Design and Implementation of Scalable QoS Policy Management System," 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), Beijing, 2015, pp. 1682-1685.

[18] Tsehay Admassu Assegie, Pramod Sekharan Nair, "The performance of Gauss Markov's mobility model in emulated software defined wireless mess network," *Indonesian Journal Of Electrical Engineering And Computer Scienc*e (IJEECS), 2020.

[19] Marius C. Breabăn, Adrian Graur, Alin D. Potorac, Doru G. Bălan. "New Approach of Traffic Limitation Management on Local Networks, *Conference: 2017 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM) & 2017 Intl Aegean Conference on Electrical Machines and Power Electronics (ACEMP), 2017.*

[20] Thaere Eido, Daniel Popa, Tulin Atmaca, Burst mode study and packet scheduling algorithm in optical packet switched network, *IEEE*, 2003.
[21] Changhee Joo and Saewoong Bahk, Weighted fair bandwidth allocation and active queue management for adaptive flows, *IEEE*, 2004.
[22] Mohamed Ashour, Tho Le-Ngoc, Performance Analysis of Weighted Fair Queues with Variable Service Rates, *IEEE*, 2006.
[23] Zhe Wang, Weisheng Hu, Weiqiang Sun, Hao He, Lilin Yi, An Efficient Aggregation Scheduling Algorithm for Unbalanced Traffic Distribution in Optical Packet Switch Network, *IEEE*, 2010.
[24] K. Kavitha, G. Suseendran, Priority based Adaptive Scheduling Algorithm for IoT Sensor Systems, *IEEE*, 2019.
[25] Laith Farhan, Laith Alzubaidi, Mohamed Abdulsalam, an Efficient Data Packet Scheduling Scheme for Internet of Things Networks, *IEEE*, 2018.