

Review

Not peer-reviewed version

A Study of Bandwidth Management Techniques for the Implementation of Quality of Service in Internet Protocol Networks

[Joseph Kithinji](#)*

Posted Date: 10 June 2025

doi: 10.20944/preprints202506.0743.v1

Keywords: Bandwidth Management; micro segmentation; Full duplex Ethernet; Bandwidth Caps; IP network; Quality of Service (QoS); Load shedding



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

A study of Bandwidth Management Techniques for the implementation of Quality of Service in Internet Protocol Networks.

Joseph Kithinji

Department of computing and information technology of University of Embu, 6 Embu 60100, Kenya;
kithinji.joseph@embuni.ac.ke; ORCID iD: 0009-0002-8038-9652

Abstract: Bandwidth being a limited and expensive resource overprovisioning is not the viable option in this tough economic times. However management of the available bandwidth if implemented well can improve quality of experience among users in the network. In this paper a literature review of the various bandwidth management techniques available based on the TCP/IP model has been done. The results indicate those techniques based on network access layer are easy to implement while those in the network layer introduce additional overhead when used. Further the results shows that at the network layer and transport layer there is room for a lot of flexibility. At the network layer there could be policies implemented using routers to manage bandwidth dynamically. Even though there could be additional overhead for running the algorithm, a well-designed algorithm would require negligible computation time. At the transport layer varying window sizes could be used to dynamically adjust bandwidth. However the application layer uses bandwidth caps which are static and therefore not ideal for handling spikes in bandwidth demand. It is expected that the findings this paper will provide insights and guidelines to systems administrators when choosing appropriate bandwidth management technique to use.

Keywords: bandwidth management; micro segmentation; full duplex ethernet; bandwidth caps; IP network; Quality of Service (QoS); load shedding

1. Introduction

Bandwidth in computer networking is defined as the data rate supported by a network connection and is expressed in bits per second[1]. For the sake of this paper bandwidth management refers to the hardware and software techniques of regulating and assigning bandwidth. The term is derived from electrical engineering where it represents the distance between the highest and lowest frequencies of signals on the communication channel[18].

In a network there are a combination of different traffic that have different requirements based on the application they are using[3]. While we may have some applications or users in the network that are termed as being critical, others may not be. Although in the recent years the available bandwidth has increased, it is still one of the major causes of bottlenecks in communication networks if not managed especially with the huge spike in demand. Therefore making the utilization of bandwidth to be efficient and effective is still one of the key aspect that promote network performance. The management of bandwidth begins at the network planning and design stage and in the long run through a variety of techniques based on the various layers of the TCP/IP model. The TCP/IP protocol is the most popular suite used for network communications. A protocol suite is a layered architecture where each layer is associated by a certain functionality done by a protocol. The TCP/IP protocol stack consists of four layers namely network access layer, network layer, transport layer and application layer. Each of these layers includes a particular functionality that ensures end to end connectivity between devices. The popularity of TCP/IP is mainly due to its reliability which arises from its flexibility where connection options can be able to be created throughout the network

without any central administration. Furthermore TCP/IP is independent of the technology used in the network as well as the architecture of device node. This makes TCP/IP applicable for many purposes including bandwidth management.

Based on the above mention facts about TCP/IP this paper presents techniques for managing bandwidth based on the layers of TCP/IP model.

This paper provides a summary review of bandwidth management technique along with their advantages and disadvantages and when appropriate to use. The results and conclusions inform on the better method to use by a system administrator based on resources available.

2. Methodology

A systematic literature review was conducted for identification as well as assessing the various bandwidth management techniques. The aim of conducting the literature review was to identify the various bandwidth management techniques available as well as their advantages and disadvantages.

3. Network Access Layer

The network access layer of the TCP/IP model defines how data is physically transmitted over the network. The layer is responsible for the transmission of data from one node to the other. At the network layer bandwidth management can be achieved by installing the appropriate transmission media and networking equipment[9]. To achieve higher bandwidth additional hardware is installed.

3.1. Replacing Copper with Fiber Cable

Shielded twisted pair has been extensively used to connect computers in most organizations. To increase the bandwidth, the shielded twisted pair can be replaced by fiber optic cable for faster speeds[9].

3.2. Upgrading Ethernet Hubs and Networks Interface Cards

Replacing 100 Mbps/sec Ethernet network interface cards (NICs) with gigabit Ethernet increases bandwidth considerably[11]. This solution entails the purchase and installation of cabling. Besides the cost, other key areas of improvement include managerial overhead where external connections require managerial time and effort[15].

At the network access layer bandwidth management entails facilitating the flow of traffic along the cables and NICs[20]. It corresponds to the bigger pipe approach. The bigger pipe approach aim at providing high bandwidth for all users with undifferentiated services[4]. However the success of bigger pipe approach is dependent on the efficiency of implemented bandwidth management tool and the amount of bandwidth available. According to the proponents of bigger pipe approach, if enough bandwidth is made available to users then QOS won't be necessary. However the provisioning of more than enough bandwidth will depend on the price of bandwidth[3].

4. Network Layer Bandwidth Management

At network layer the main focus of bandwidth management is to eliminate congestion by use of routers which control rate at which packets are sent into the network. Congestion occurs when the traffic traversing the network exceeds the capacity of the network, and needs to be managed to prevent degradation of QOS[21]. If not managed congestion wastes whatever bandwidth is available and degrades the QOS in a network. The TCP/IP model network layer is responsible for assignment of IP address to hosts in the network. Bandwidth management at the network layer occurs in real time as the packets arrive[20]. And there after applying static or dynamic bandwidth allocations. The techniques used in routers to prevent congestion are discussed in the following sections[15].

4.1. Network Segmentation

Network segmentation is division of the network into smaller components. Network segmentation implemented using bridges, where the network is divided into many segments joined by bridges. This has the cost implications of buying new hubs and bridges[9]. Network segmentation can also be done using the network interface card. This is where each segment on the network is connected to a different NIC on the server. This means the servers will be required to route between NICs[17].

Network segmentation using switches is done by replacing hubs with switches. Since each port in a switch represents a collision domain, adding a switch breaks the network into small collision domains[20]. The creation of small collision using switches is known as micro segmentation. Micro segmentation comes with many benefits including low latency, support for virtual Local area networks (VLANs) and prioritization. In addition it is less costly as compared to layer one cost of installing cables[21].

4.2. Full Duplex Ethernet

Full duplex Ethernet technique is able to increase the bandwidth of a connection up to twice its capacity[18]. When a dedicated full duplex connection is used it further eliminates contention for the link. However to achieve full duplex Ethernet NICs must be able to support duplex mode. Therefore full duplex Ethernet requires NICs replacement and reconfiguration of the client and server software's[19].

4.3. Bandwidth Allocation, Sharing and Reservation

A network is made up of the users who need bandwidth to meet service demands[14]. However, in a realistic network, bandwidth is limited and some methods of allocating it are needed when total demand is greater than the resource limit. Bandwidth allocation is about efficiently allocating the network bandwidth among the sources[15].

Static bandwidth allocation technique assigns a maximum amount of bandwidth to each class and implements traffic shaping to control the data traffic. Classes are not restricted to use less than their bandwidth allocations however a class is limited to use more than its allocated bandwidth[17].

Dynamic bandwidth allocation techniques is an alternative that enables network resources to be adjusted in order to improve network utilization[7]. When a certain class of users require more bandwidth than that it is reserved for, a review is initiated to ask for more. If the allocated bandwidth to a class is more than enough, some of the bandwidth can be shared[11]. In this way bandwidth usage can be improved significantly. In dynamic bandwidth allocation various classes of algorithms are used. One such class of algorithm is based on parameter measures. In this case a parameter is calculated up to the current period and in future bandwidth is allocated based on the previous history of usage[14].

Due to the difficulty of establishing the relationship between the current parameter and the future bandwidth usage, these approach is not efficient in in enhancing QOS[15]. The other class of algorithm is based on prediction approaches. In this case bandwidth is assigned based on the available information and bandwidth is assigned based on the predictions. However proper utilization of network resources will significantly rely on accurate prediction of parameters. The major limitation associated with prediction approaches class of algorithms is the overhead created during negotiation of traffic flow at peak rate, sustained rate and burst length [19].

In a bandwidth sharing method, traffic is divided into classes and each class is allocated a percentage of the bandwidth [21]. When given classes reach its limit, no more data belonging to that class can be forwarded. However if all other classes are not utilizing their whole share, a class can borrow bandwidth for a short while and send its traffic[15].

The configuration of a class includes its fair share as well as the amount it is allowed to borrow. The spare capacity is mostly allocated to class having highest priority[16]. Dynamic bandwidth

sharing would result in proper utilization of the network bandwidth, due to the bursty nature of network traffic. Some users could be allocated extra bandwidth when others are not using it[19].

Bandwidth reservation is a technique where a certain data flow is allocated a specific amount of bandwidth. It ensures that some flows have been reserved bandwidth for guaranteed QOS. The reservation protocol enables one to reserve special QOS for their data[18]. When an application receives data packets for which it requires a certain QOS it sends a RSVP request back to the sending application[16]. As the data traverses the network. The QOS is negotiated with the routers and other network devices. Those network equipment's that do not contain the RSVP functionality simply ignores the RSVP traffic and do not participate in the negotiation[17].

4.4. Load Shedding and Buffer Allocation

Load shedding also referred to as packet dropping is a function performed by the router when it cannot handle all incoming packets[11]. In this case the packets may be dropped based on a certain priority. However priority schemes are challenging to implement owing to the fact negotiation from all users is required[18].

The dropping of excess packets has been proven not to be effective since it would require differentiation between data packets and acknowledgment packets. For example dropping TCP acknowledgment packets would delay the allocation of buffers and hence cause congestion.

In buffer allocation technique buffers are allocated in routers. By any chance congestion occurs, the packets can be temporarily stored in the pre allocated buffers to be transmitted later. In this way congestion is reduced because the stored packets are no longer in transit[12].

4.5. Flow Control Using Choke Packets and VLANs

Flow control using choke packets reduces congestion by ensuring that the router is not overwhelmed by many packets by transmitting a choke packet to the source. In response the source reduces the amount of packets sent[6]. To increase the speed of transmission for choke packets, they are sent to intermediate routers in which case the response becomes fast. The internet control message protocol is responsible for sending choke packets[9].

Flow control targets at shaping traffic from source to destination while in congestion control aims at regulating the traffic flow in the network[11]. Flow control reduces network throughput due to increased packet transit time which makes it an effective bandwidth management technique. Flow control is only effective if it manages to decrease the quantity of traffic in a link at critical points and time[15].

Virtual LANS are defined in IEEE standard 802.1q. The IEEE 802.1q is a standard that implements VLANs in an Ethernet network. IEEE 802.1q defines a system for tagging Ethernet frames and the mechanism to be used by networking devices to handle VLAN frames. VLANs allows for users to be grouped together irrespective of their physical locations. VLANs create a smaller broadcast domains which in turn reduces bandwidth consumed by broadcasts. In this way more bandwidth is availed to users[7].

5. Transport Layer Bandwidth Management

At the transport layer bandwidth is managed by not only controlling the amount of connections, but also regulating the packets flow amongst the hosts. In transport layer bandwidth management can be achieved through limiting the amount of end to end connections and controlling the flow of packets amongst two hosts[9]. In this case bandwidth management is achieved using Transmission Control Protocol Rate Control and rate control resource reservation protocol.

Sliding window is a TCP feature that is used to reduce congestion by regulating the amount to packets that can be transmitted by a given host. Acknowledgements are used to communicate to the transmitting host to continue transmitting otherwise the transmitting host stops transmitting[2]. TCP

sliding window can be adjusted dynamically in responses to any flow timeouts in the network by shaping traffic, TCP rate control makes it smoother and more predictable[1].

The resource reservation protocol (RSVP) method is used to optimize bandwidth by monitoring and controlling bandwidth along each link between sender and receiver[4]. The RSVP reserves bandwidth between a sender and receiver monitoring each connection and route[21]. Before information is sent, the receiver establishes whether each device along the route has spare bandwidth available, if not the transmitting device is informed[7].

6. Application Layer Bandwidth Management

Application layer bandwidth management is a composition of QOS tools used to provide priority to traffic based on application type to ensure bandwidth intensive applications do not crowd the network(Chang et al., 2017). This solution provides a predefined classification of protocols based on applications and an all-inclusive policies for traffic control such as rate shaping and priority marking(Yong et al., 2015).This makes it possible for network administrators to distinguish between desirable and undesirable traffic flows within the same protocol. Application layer bandwidth management is supported for all application matches, custom application rules and file transfer types[12].

The emergence of many web applications has led to the demand for application level QOS in many network setups. The high contention for bandwidth may lead to bandwidth sensitive applications not working properly[8].Application level bandwidth management is achieved by ensuring that network packets are associated with their process. Traffic generated by applications in a network is classified as either constant bit rate (CBR) or variable bit rate (VBR)[10].The CBR applications produce traffic flows at a constant rate and therefore can be allocated a given amount of bandwidth to achieve intended QOS. This implies that allocating more bandwidth does not improve user satisfaction [13].VBR applications produce varied traffic flows and are able to utilize all the available bandwidth. For VBR traffic the more the bandwidth the better the QOS. To achieve scalability both CBR and VBR servers need to regulate the data rate per user. This is done by putting upper bound on the bandwidth allocated to an individual user per unit time[15].

Bandwidth intensive applications such as peer to peer(P2P) traffic can be blocked to prevent performance degradation brought about by P2P file sharing[20]. Blocking can be done for only ports associated with P2P applications in an attempt to improve network performance. However blocking ports as a method of managing bandwidth has a number of drawbacks. First is that P2P applications such as bit torrents allow users to choose the port before the start of a download[21].

Using Bandwidth Caps technique is applied to discourage users from consuming huge amounts of bandwidth. If a certain user exceeds the amount he/she is allocated, his/her connection speed is reduced for a while[7]. Bandwidth caps is effective in bandwidth savings but cannot effectively manage congestion during peak hours. Additionally, this technique lacks the granularity to differentiate traffic. This technique can be improved by applying caps to certain applications at certain times of the day[3]. However bandwidth caps are difficult to distinguish from network failure and therefore should be used in consideration of the usage patterns of users to avoid punishing wrong users [5].

7. Results and Discussions

Table.1 highlights the various bandwidth management techniques and their suitability for use. Figure.1 represents the percentage of techniques that fall under each layer .From Table.1 it is evident that those techniques that fall under the network access layer are easy to implement, however they are costly. In addition a technique like big pipe approach is not sustainable. In the network layer we have techniques that can significantly improve bandwidth usage through policies implemented in routers. However there is overhead in implementing these policies which can be implemented using algorithms' poorly designed algorithm can cause a lot of overhead. However on the other hand these

techniques offer a lot of flexibility. At the transport layer TCP sliding window is used which has some dynamism hence ensuring optimal utilization of bandwidth. Techniques under the application layer are meant to save on bandwidth. A technique like use of bandwidth caps may not effectively manage congestion under peak hours.

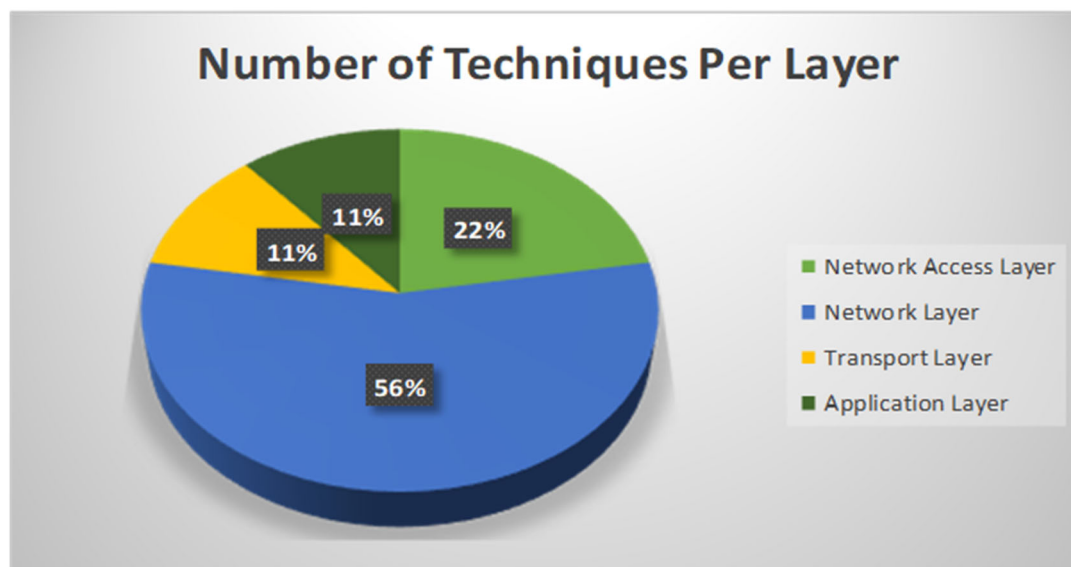


Figure 1. Percentage Bandwidth Management Techniques Distribution Per Layer.

Table 1. A summary of bandwidth management techniques based on the TCP/IP Model.

Technique	TCP/IP Layer	Strengths	Weaknesses
Replacing Copper with Fiber Cable	Network Access Layer	Easy to implement	Costly
Upgrading Ethernet Hubs and Networks Interface Cards	Network Access Layer	Easy to implement Big pipe approach ensure QOS for all users	Costly Managerial overhead due to external connections Big pipe approach is not sustainable
Network Segmentation	Network Layer	low latency, Support for prioritization. it is less costly as compared to network access layer on cost of installing cables	There are huge cost implications of buying new hubs and bridge
Full duplex Ethernet	Network Layer Bandwidth Management	Increases bandwidth capacity	Requires complete NICs replacement and reconfiguration of the client and server soft wares
Dynamic bandwidth allocation	Network Layer Bandwidth Management	Improves bandwidth usage significantly	Not efficient in in enhancing QOS overhead created during negotiation of traffic flow a peak rate, sustained rate and burst length

Load Shedding and Buffer Allocation	Network Layer Bandwidth Management	Reduces the amount of traffic in transit	Difficult to implement priority schemes
Flow Control Using Choke Packets and VLANs	Network Layer Bandwidth Management	Reduces bandwidth consumed by broadcast domains	Only effective if it manages to decrease the quantity of traffic in a link at critical points and time
TCP sliding window	Transport Layer Bandwidth Management	Dynamic based on situation Predictable	Dependent on TCP acknowledgements
Using Bandwidth Caps	Application Layer Bandwidth Management	Saves bandwidth	lacks the granularity to differentiate traffic bandwidth caps are difficult to distinguish from network failure cannot effectively manage congestion during peak hours

8. Conclusion and Recommendations

Despite the increase in speeds and capacity of various transmission media, bandwidth management is becoming more of a challenge due to the digital age. Different bandwidth management techniques have been reviewed with reference to the layers of the TCP/IP model. The review shows layer two as the layer with more techniques of managing bandwidth although it has its weaknesses. In TCP/IP high bandwidth may not necessarily reflect high throughput. This is because any devices on a TCP/IP network requires an acknowledgement to transmit data. If these acknowledgement delay others is bound to be reduced throughput even if bandwidth is well managed. Therefore this paper recommends on an approaches that would reduce roundtrip latency for time sensitive applications. This would ensure optimal utilization of the bandwidth available in the network which would translate to high throughput. It is therefore clear form the above discussion that system administrators will have to consider some tradeoffs when choosing a bandwidth management technique.in future we intend to explore the performance of integration of all the techniques under each layer.

References

1. Alkharasani, A. M., Othman, M., Abdullah, A., & Lun, K. Y. (2017). *An Improved Quality-of-Service Performance Using RED 's Active Queue Management Flow Control in Classifying Networks*. 5, 24467–24478.
2. By, P., Sovandara, T., April, M. U. M., & Penh, P. (2015). *Bandwidth Management and QOS*. (April).
3. Celik, A., Radaydeh, R. M., & Member, S. (2017). *Resource Allocation and Interference Management for D2D-Enabled DL / UL Decoupled Het-Nets*. 22735–22749.
4. Gémieux, M., Li, M., Savaria, Y., David, J., Zhu, G., & Member, S. (2018). *A Hybrid Architecture With Low Latency Interfaces Enabling Dynamic Cache Management*. 6. <https://doi.org/10.1109/ACCESS.2018.2876597>
5. He, L., Member, S., Wang, J., & Member, S. (2017). Bandwidth Efficiency Maximization for Single-Cell Massive Spatial Modulation MIMO : An Adaptive Power Allocation Perspective. *IEEE Access*, 5, 1482–1495. <https://doi.org/10.1109/ACCESS.2017.2668420>
6. Huang, X., Yuan, T., & Ma, M. (2018). Utility-Optimized Flow-Level Bandwidth Allocation in Hybrid SDNs. *IEEE Access*, 6, 20279–20290. <https://doi.org/10.1109/ACCESS.2018.2820682>
7. Ji, Y., & Member, S. (2018). Reconfigurable Optical OFDM Datacenter Datacenter Networks. *IEEE Photonics Journal*, 10(5), 1–16. <https://doi.org/10.1109/JPHOT.2018.2859275>

8. Kulkarni, S. (2015). *Implementation of iSCSI for Mobile Platforms and its Performance Optimization on Mobile Networks*. 3(5).
9. Lim, W., Kourtessis, P., Senior, J. M., Na, Y., Allawi, Y., Jeon, S., & Chung, H. A. E. (2017). *Dynamic Bandwidth Allocation for OFDMA-PONs Using Hidden Markov Model*. 5, 3–6.
10. Mahajan, N., & Mahajan, S. (2015). *AN EFFICIENT TOKEN BUCKET ALGORITHM INCREASING WIRELESS*. 3(3).
11. Mamman, M., & Hanapi, Z. M. (2017). An Adaptive Call Admission Control With Bandwidth Reservation for Downlink LTE Networks. *IEEE Access*, 5, 10986–10994. <https://doi.org/10.1109/ACCESS.2017.2713451>
12. Marir, N., Wang, H., Li, B., & Jia, M. (2018). Distributed Abnormal Behavior Detection Approach Based on Deep Belief Network and Ensemble SVM Using Spark. *IEEE Access*, 6, 59657–59671. <https://doi.org/10.1109/ACCESS.2018.2875045>
13. Nahrstedt, K., Arefin, A., & Rivas, R. (2011). *QoS and resource management in distributed interactive multimedia environments*. 99–132. <https://doi.org/10.1007/s11042-010-0627-7>
14. Paul, A. K., Tachibana, A., & Hasegawa, T. (2016). An Enhanced Available Bandwidth Estimation Technique for an End-to-End Network Path. *IEEE Transactions on Network and Service Management*, 13(4), 768–781. <https://doi.org/10.1109/TNSM.2016.2572212>
15. Randrianantenaina, I., & Member, S. (2017). *Interference Management in Full-Duplex Cellular Networks With Partial Spectrum Overlap*.
16. Sboui, L., Member, S., & Ghazzai, H. (2016). On Green Cognitive Radio Cellular Networks: Dynamic Spectrum and Operation Management. *IEEE Access*, 4, 4046–4057. <https://doi.org/10.1109/ACCESS.2016.2592178>
17. Song, M. (2018). Minimizing Power Consumption in Video Servers by the Combined Use of Solid-State Disks and Multi-Speed Disks. *IEEE Access*, 6, 25737–25746. <https://doi.org/10.1109/ACCESS.2018.2832221>
18. Wang, B., Sun, Y., & Cao, Q. I. (2018). Bandwidth Slicing for Socially-Aware D2D Caching in SDN-Enabled Networks. *IEEE Access*, 6, 50910–50926. <https://doi.org/10.1109/ACCESS.2018.2867542>
19. Wang, Z. (2018). DyCache : Dynamic Multi-Grain Cache Management for Irregular Memory Accesses on GPU. *IEEE Access*, 6, 38881–38891. <https://doi.org/10.1109/ACCESS.2018.2818193>
20. Wu, B., Wu, B., Yin, H., Liu, A., Liu, C., & Xing, F. (2017). Investigation and System Implementation of Flexible Bandwidth Switching for a Software-Defined Space Information Network. *IEEE Photonics Journal*, 9(3), 1–14. <https://doi.org/10.1109/JPHOT.2017.2705134>
21. Xu, W. (2018). A Novel Data Reuse Method to Reduce Demand on Memory Bandwidth and Power Consumption For True Motion Estimation. *IEEE Access*, 6, 10151–10159. <https://doi.org/10.1109/ACCESS.2018.2807406>

Authors' Profiles



Kithinji Joseph: Doctor of Philosophy (Computer Science), Lecturer, Department of computing and information Technology, University of Embu , Kenya. Areas of scientific interests: QOS in computer networks, database optimization, storage area networks, data science
Email: kithinji.joseph@embuni.ac.ke

How to cite this paper: Kithinji. Joseph, " A study of Bandwidth Management Techniques for the implementation of Quality of Service in Internet Protocol Networks", *International Journal of Computer Network and Information Security(IJCNIS)*,

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.