Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 6, 2025

OPTIMIZING CONGESTION CONTROL FOR QUALITY OF SERVICE (QOS) IN BANDWIDTH-CONSTRAINED WIRELESS NETWORKS

Abdulrehman Arif[1], Furqan Jamil[2], Syed Zohair Quain Haider[*3], Muhammad Zeeshan Haider Ali[4]

[1,*3,4]*Department Of Computer Science and Information Technology, University of Southern Punjab, Multan, Pakistan*
[2]*Department of Computer Science, National College of Business Administration and Economics Multan Sub Campus*

[*3]zohairhaider67@gmail.com

DOI: https://doi.org/10.5281/zenodo.15662476

Keywords
Quality of Service, Differentiated Services, Assured Forwarding, Expedited Forwarding, Best Effort, Asynchronous Transfer Mode.

Article History
Received on 06 May 2025
Accepted on 06 June 2025
Published on 14 June 2025

Corresponding Author: *
Syed Zohair Quain Haider

Abstract

Modern wireless networks typically operate on a best-effort service model, which, while able to support both real-time and non-real-time traffic, often falls short in ensuring the required Quality of Service (QoS) for real-time applications. Real-time applications, such as video streaming, voice over IP (VoIP), and online gaming, are highly sensitive to network conditions and require a predictable, low-latency environment to maintain performance. However, the best-effort model does not prioritize traffic effectively, leading to poor performance under high network load, with issues such as high jitter, excessive delay, and increased packet loss. QoS in wireless networks is traditionally assessed through performance metrics such as throughput, jitter, delay, and packet loss, all of which are crucial in determining the user experience in real-time applications. These metrics directly impact overall network efficiency and user satisfaction, with high delay or packet loss leading to degraded service quality, particularly for latency-sensitive applications. In this context, this study introduces a novel QoS framework tailored specifically for bandwidth-constrained networks, where managing limited resources is crucial. Instead of relying on the traditional approach of over-provisioning bandwidth, which can be inefficient and costly, the proposed model employs differentiated services combined with dynamic scheduling based on real-time measurements of incoming data rates and packet classification. By dynamically adapting the network's resource allocation to the changing traffic demands, the framework ensures that real-time applications receive the necessary priority, while non-real-time traffic is handled more flexibly. This results in a more efficient use of available resources, as bandwidth is allocated based on real-time traffic characteristics rather than fixed allocations. The framework incorporates an optimized queuing mechanism that prioritizes packets based on their type and current queue length, allowing for more accurate traffic management. This mechanism helps minimize delays for high-priority packets, such as those associated with real-time applications, while ensuring that lower-priority packets are processed appropriately without congesting the network. By reducing packet waiting times and minimizing the chances of packet loss, the approach significantly improves the QoS for real-time traffic, even in environments where bandwidth is limited. Furthermore, the model aims to minimize resource over-

provisioning, which is a common issue in traditional network designs that often result in underutilized resources or excessive costs for provisioning higher bandwidth than necessary.

INTRODUCTION

Wireless technology has become widely adopted, with its implementation defined under the IEEE 802.16 series of wireless standards. This series, endorsed by the Wireless Technology Forum, promotes interoperability across vendors offering static and mobile products. Originating from South Korea, wireless broadband technology supports various applications such as broadband internet access, cellular backhaul, and hotspot connectivity. While similar to Wi-Fi, wireless broadband covers longer distances.

The IEEE 802.16 standards aim to deliver long-range broadband wireless access (BWA) with guaranteed Quality of Service (QoS) for multiple classes of service (CoS), ensuring low latency, minimal jitter, low packet loss, and adequate bandwidth. The network architecture includes two main station types: Base Stations (BS), which are fixed, and Subscriber Stations (SS), which serve multiple users. Users connected via SS can be either stationary or mobile. The wireless network supports two communication modes: point-to-multipoint (PMP) and mesh mode.

Figure 1-1 illustrates the PMP configuration, where a central base station controls the transmission schedules of connected subscriber stations. Multiple base stations connect through an Access Service Network Gateway (ASN-GW), which in turn connects to a Connection Service Network (CSN) providing IP connectivity. The mesh mode supports direct communication between base stations and allows subscriber stations to communicate without intermediaries.

In PMP mode, the base station allocates bandwidth using either Grant Per Connection (GPC) or Grant Per Subscriber Station (GPSS) modes. The GPC mode assigns bandwidth individually per connection, while the GPSS mode treats all subscriber station connections as a single entity, distributing bandwidth equally across them.

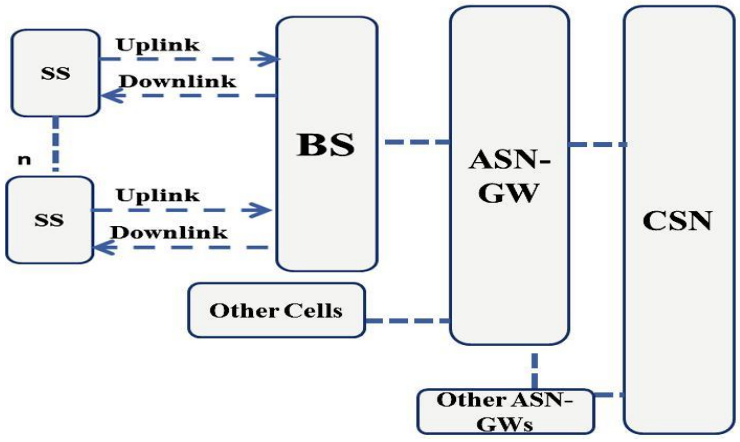

Figure 0-1: Basic Architecture [1]

Wireless technology is standardized under IEEE 802.16, providing a genuine broadband connection suitable for a variety of user scenarios. It operates on a static infrastructure that supports fixed, portable, and mobile access. Key features of this technology include:

1. **High Peak Data Rates:**

Wireless supports exceptionally high peak physical layer data rates, reaching up to 74 Mbps with a 20 MHz channel bandwidth and 25 Mbps with a 10 MHz bandwidth [7].

2. **Quality of Service (QoS) Support:**
QoS is a critical element in the wireless MAC layer design, implemented through a connection-oriented MAC framework. Both uplink and downlink transmissions are managed by the base station (BS) [7]. Important QoS parameters include service priority, maximum delay, jitter tolerance, ARQ (Automatic Repeat reQuest) mechanisms, and scheduling algorithms.

3. **Adaptive Modulation and Coding:**
Wireless networks support various modulation and coding schemes such as BPSK 1/2, QPSK 1/2, QPSK 3/4, 16-QAM 1/2, 16-QAM 3/4, 64-QAM 1/2, and 64-QAM 2/3. Here, the notation m/n represents the ratio between the number of source bits (m) and the total output bits (n), allowing flexible adaptation of modulation based on channel conditions. This adaptability maximizes throughput.

4. **Mobility Support:**
Wireless systems are designed for mobility, making them suitable for moving platforms. The architecture consists of base stations and mobile subscriber stations (SS). The system tracks SS as they move across different base stations with minimal handover delays [9].

5. **Robust Security:**
Wireless networks employ strong encryption standards such as Advanced Encryption Standard (AES) and Triple Data Encryption Standard (3DES) [9]. Additionally, coding techniques like Low-Density Parity-Check (LDPC) codes enhance both performance and security.

6. **OFDMA (Orthogonal Frequency Division Multiple Access):**
OFDMA enables multiple users to simultaneously access the wireless spectrum by assigning distinct OFDM subcarriers to each user. This method improves frequency diversity, extends the effective carrier, and enhances overall system capacity [9].

7. **Scalability:**
Mobile wireless technology is highly scalable, utilizing OFDM and Fast Fourier Transform (FFT) techniques. FFT allows flexible channel bandwidth allocation ranging from 1.25 MHz up to 20 MHz, facilitating easy deployment across different environments.

Major Benefits of Wireless Technology:
- Connects various devices across urban and rural areas to provide portable mobile broadband access.
- Offers a competitive alternative to DSL for network access using efficient wireless broadband.
- Supports multiple data services, including Voice over IP (VoIP) and Internet Protocol Television (IPTV).
- Provides fast internet connectivity solutions for business applications.
- Facilitates smart grid and metering implementations.

QoS Delivery in Wireless Networks
Wireless technology categorizes service delivery into five classes, also known as Classes of Service (CoS), to guarantee QoS:

1. **Unsolicited Grant Service (UGS):**
Designed for fixed-size, constant bit rate (CBR) applications like T1/E1 and VoIP. Key QoS parameters include:
- Maximum Sustained Traffic Rate (MSTR)
- Maximum Delay Tolerance
- Jitter Tolerance

2. **Extended Real-Time Polling Service (ertPS):**
Supports VoIP with activity detection, maintaining consistent bandwidth allocation during sessions.

3. **Real-Time Polling Service (rtPS):**
For applications generating variable-size data at regular intervals, such as streaming audio or video. QoS parameters include MSTR, Minimum Reserved Traffic Rate (MRTR), Maximum Delay Tolerance, and Jitter Tolerance.

4. **Non-Real-Time Polling Service (nrtPS):**
Suitable for variable-size, non-periodic data applications like file transfers, with no strict delay guarantees. QoS factors are MSTR, MRTR, and Traffic Priority.

5. **Best Effort (BE):**
Intended for web services without specific data rate or latency guarantees.

Despite these classes, wireless networks lack explicit congestion control mechanisms and detailed resource allocation strategies for each service category, making QoS a challenging area. Congestion results in buffer overflows and increased latency, reducing user experience. Researchers have proposed solutions such as threshold-based QoS enforcement [5], handover mechanisms [12], and scheduling algorithms [13] to improve QoS under load.

Problem Statement

Base stations can become overloaded during congestion, leading to QoS degradation. Existing wireless architectures do not provide effective base station overload avoidance mechanisms. Therefore, an intelligent congestion avoidance system is necessary to manage traffic scheduling and maintain base station buffers within target limits while preserving QoS.

DiffServ has been employed as a backbone for QoS and congestion control in wireless networks [32], but scalability and priority scheduling challenges remain [1].

Research Questions

- What are the primary challenges in maintaining QoS within data center networks?
- What features define an effective congestion control mechanism for optimizing QoS?
- How can DiffServ be utilized as a backbone in wireless networks to aid congestion control?
- How can data packets be classified via DiffServ to support QoS and congestion management?

Aims and Objectives

This research aims to leverage DiffServ to enhance QoS and congestion control in wireless networks by:

- Improving QoS to address limitations in current protocols
- Designing a model integrating DiffServ with wireless architectures for congestion management
- Evaluating the model's effectiveness through metrics such as packet delay and loss

LITERATURE REVIEW

Qos In Various Types of Real-Time Traffic

Real-time data transmission typically experiences higher delays and packet loss compared to non-real-time transmission. Common examples include applications such as remote classrooms, online gaming, video conferencing, and voice over ip (voip). Within real-time transmission, various audio and video codecs are employed, including ieee standards like mpeg-2, mpeg-4, and g.711, as well as gsm and g.723 for audio. For video, itu standards such as h.261, h.263, and h.264 are commonly used. Audio transmission generally involves fixed packet sizes and constant bit rates, whereas video transmission often uses variable packet sizes and bit rates.

Packet delay and loss in real-time transmission are primarily caused by factors like buffering in network devices. To improve real-time communication, an appropriate quality of service (qos) framework must be implemented. This section will discuss several qos techniques applicable to real-time traffic in subsequent sections.

What is quality of service (qos)?

Quality of service (qos) refers to the ability of a network to provide measurable performance guarantees. It is typically evaluated using metrics such as average packet loss, average delay, jitter (delay variation), and throughput. Qos mechanisms can be applied in different ways to enhance service for specific traffic types, including priority queuing, application-specific routing, bandwidth management, and traffic shaping.

Qos implementation is generally categorized into two layers:

- Application layer qos: manages jitter and ensures smooth media playback at the application level.
- Network layer qos: controls bandwidth allocation and latency at the network routing and switching level.

This research focuses on network layer qos techniques, which will be explored in more detail later. Qos ensures networks meet certain performance criteria to enhance reliability and data delivery, particularly for delay-sensitive applications.

Key qos parameters often used in evaluations include:

• Delay: the time taken for a packet to travel from source to destination, measured in milliseconds.
• Delay variation (jitter): the variation in delay between consecutive packets, also measured in milliseconds.

Applications requiring qos

The internet, being central to modern communications, often lacks guarantees for reliable and timely data delivery. For data-centric applications, transmission control protocol (tcp) over ip is widely used to ensure reliable packet delivery through retransmission of lost packets. However, real-time media streaming, such as live video and voice, demands timely delivery and cannot tolerate retransmission delays.

Udp, the transport protocol commonly used for real-time media, lacks retransmission capabilities and cannot ensure packet delivery, posing challenges for real-time applications. To address this, application-layer protocols like the real-time transport protocol (rtp) have been developed, which operate on top of udp to manage timing and synchronization. Rtp is defined by the ietf in rfc 1889.

Qos implementation in ip networks

For prioritizing real-time traffic, qos mechanisms must be implemented on routers and switches across the network. Common qos technologies include ieee 802.1p/q, differentiated services (diffserv), and integrated services (intserv). Effective qos requires support at both sender and receiver ends.

Qos operates primarily at network layers 2 and 3. Layer 3 qos focuses on bandwidth and delay management through routers, while layer 2 qos addresses congestion control on switches. Two primary models of qos at the network layer are:

• Best effort service model: provides no guarantees on delivery, delay, or throughput, typically using fifo queuing.
• Integrated services (intserv) model: uses resource reservation protocols like rsvp to allocate resources per flow, offering strong qos guarantees but limited scalability.
• Differentiated services (diffserv) model: assigns resources to classes of traffic rather than individual flows, offering scalable and flexible qos management.

Best effort model

In the best effort model, network resources are allocated without guarantees, and packets are processed on a first-come, first-served basis. Ip networks offer standard services with no prioritization. The differentiated services code point (dscp) field in the ip header contains bits to indicate the class of service but best effort treats all packets equally.

Integrated services (intserv) model

Intserv attempts to provide per-flow qos guarantees by reserving resources along the network path before transmission begins, using rsvp. It supports both unicast and multicast data streams and reserves cpu cycles, buffer space, and bandwidth as needed. However, due to the overhead of managing per-flow states on routers, intserv is not widely deployed on large-scale networks.

9.3.3 Differentiated Services (Diffserv) Model

Diffserv, described in rfc 2475, offers a scalable alternative by classifying traffic into aggregated groups rather than individual flows. Traffic classes are identified using dscp values, and packets are treated according to per-hop behavior (phb) at each router. Key diffserv classes include:

• Assured forwarding (af): provides different levels of forwarding assurance with priority classes and drop precedences.
• Expedited forwarding (ef): designed for low latency, low jitter, and low packet loss traffic, suitable for voice and video.

Tables in the original text compare dscp values for various classes and qos factors across different models.

Qos delivery in diffserv networks

Diffserv was developed to overcome the complexity of intserv and rsvp by offering class-based qos treatment. The dscp field in the ip header (6 bits) indicates the desired qos class. Edge routers perform diffserv was developed to overcome the complexity of intserv and rsvp by offering class-based qos treatment. the dscp field in the ip header (6 bits) indicates the desired qos class. edge routers perform complex classification, marking, and conditioning of

packets, while core routers implement simpler queuing and scheduling based on dscp values.

Traffic classification in diffserv can be based on:
Multi-field (mf): uses multiple header fields such as source/destination ip and ports.
Behavior aggregate (ba): uses only the dscp value for classification.

Traffic conditioning functions include:
. Metering: measuring traffic rates.
. Shaping: delaying packets to smooth traffic flows.
. Policing: dropping packets that exceed agreed rates.
• Marking: setting or modifying dscp values based on policies.

Diffserv provides three main service types:
• Expedited forwarding (ef): highest priority forwarding for critical traffic.
• Assured forwarding (af): priority classes with different drop precedences.
• Best effort (be): default, no qos guarantees.

The diffserv approach enables scalable and flexible qos management across ip networks.

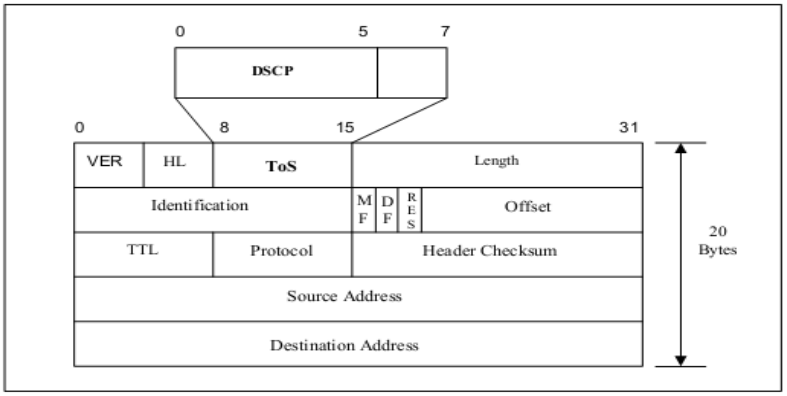

Figure 0-2: Type of Service (ToS) field in IP header [30]

The Internet Protocol (IP) mechanism originated from the broader generalization of internet protocols. It is considered straightforward because it prioritizes IP packets by categorizing them into low and high priority levels. This priority classification is determined by specific bits set within the IP header [21].
At the data link layer (Layer 2), quality of service can be provided using the IEEE 802.1p standard. This standard is often combined with IEEE 802.1q, which supports VLAN (Virtual Local Area Network) functionality. Both techniques utilize similar bits for identifying packet priority or VLAN membership. Each switch output port can have multiple queues, where traffic with higher priority is assigned to queues with greater bandwidth allocation.

QoS Using Multi-Protocol Label Switching (MPLS)
MPLS technology enhances network performance by enabling efficient forwarding, switching, and routing of traffic flows [11]. Positioned between Layer 2 and Layer 3.

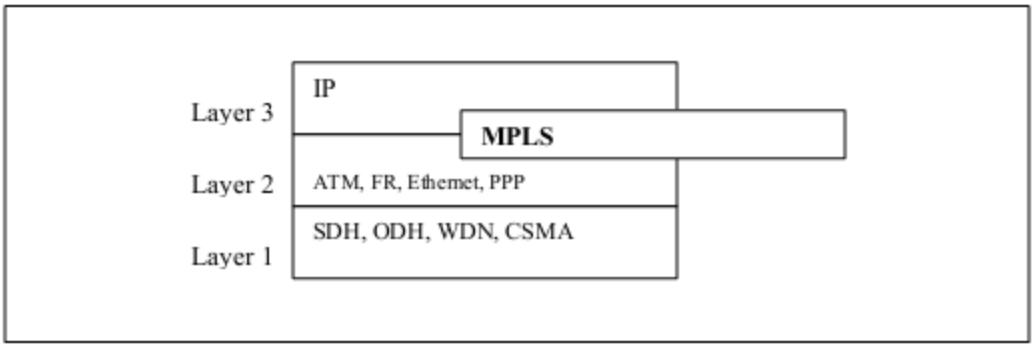

Figure 0-3: MPLS layer in between layers 2 and 3 [11].

As shown in Figure 2-2, MPLS provides a flexible framework with several key capabilities:

- It supports management of different types of traffic flows, whether between applications or across multiple devices.
- It operates independently of Layer 2 and Layer 3 protocols.
- It facilitates mapping IP addresses to fixed-length labels, which are used for packet classification and forwarding.
- It integrates with protocols like Open Shortest Path First (OSPF) and Resource Reservation Protocol (RSVP).
- MPLS can handle various Layer 2 protocols, including ATM, Frame Relay, and IP.

QoS Setup in IP Networks for Real-Time Communication

A service-oriented QoS approach is practical for handling real-time traffic. This method is typically implemented in two stages:

- Real-time packets, such as those carrying audio and video, are marked with specific DSCP (Differentiated Services Code Point) values at the switch level.
- Packets are then categorized into different service groups based on their DSCP markings, with tailored scheduling mechanisms applied to each group to meet desired QoS levels.

Switch operations are divided into multiple functional planes, primarily the Data Plane and the Control Plane. Packet classification and scheduling are managed within the Data Plane, whereas admission control and resource reservation requests are handled by the Control Plane. This division is depicted in Figure 2-3.

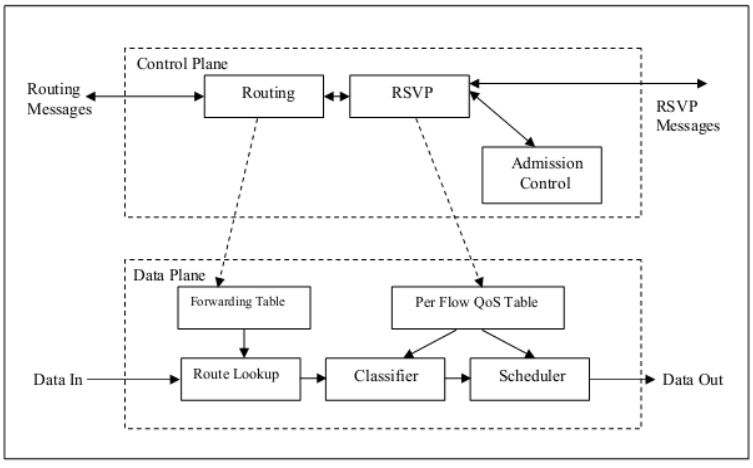

Figure 0-4: Different elements in modern router [10]

Packets are ordered relying on the DSCP estimation of parcel header and allotted to various lines (cradles) of sending classes as appeared in Figure 2-4.

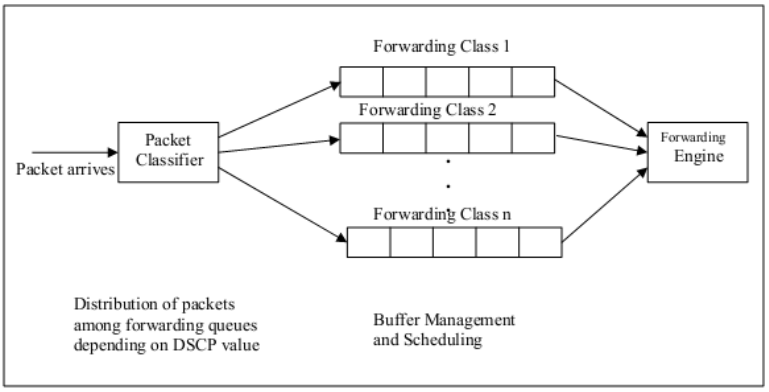

Figure 0-5: Classification of packets [13]

The primary reasons for packet loss and delay within a packet switching system are the queues. Queues are recognized as the main contributors to packet loss. Each output port has a queue where packets must wait before they can leave the system. If the queues are empty or have available space, incoming packets are immediately forwarded to the output link. However, when traffic volume is heavy, queues become full, causing packets to be delayed as they wait for all preceding packets in the queue to be transmitted. If the queues reach capacity and the traffic load remains high, the rate of packet loss increases significantly.

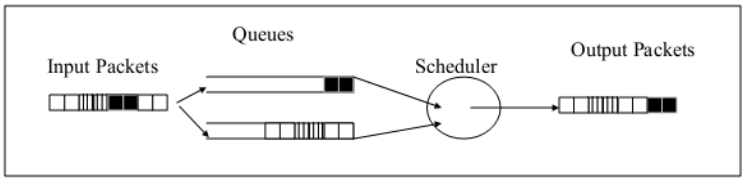

Figure 0-6: Packet queue diagram [14]

A priority queuing component adds additional queues at each shift and shift output port, dedicated to handling higher priority traffic. Figure 2-5 illustrates a two-level output queue. Within the lower queues, best-effort traffic is lined up, whereas higher priority traffic is placed in the upper-level queue. The queue management procedure determines how packets are dequeued. Packets in the higher priority queue are always served before those in the lower queue. As shown in Figure 2-5, the packets marked in dark are processed first, followed by those in the best-effort queue. If high-priority packets arrive while the lower queue is still being emptied, the system immediately switches to servicing the high-priority queue.

Several packet scheduling algorithms are employed here, such as:

- **FCFS (First Come First Serve):** Packets are transmitted in the order they arrive.
- **Priority Scheduling (PS):** Queues are assigned priority values, with higher priority queues served before lower ones.

- **Weighted Round Robin (WRR):** Service is distributed based on bandwidth allocation per queue. For example, if queue 1 is allocated 40% of bandwidth and queue 2 gets 60%, queue 2 will be served 1.5 times more frequently than queue 1.

Various buffer management strategies also exist, including.

- **Threshold method:** When buffer occupancy exceeds a threshold, low-priority packets are discarded.

• **Push-out method:** If a high-priority packet arrives when the buffer is full, it replaces a low-priority packet in the buffer.

There is a trade-off between queue size and packet loss. Small queues may lead to packet drops during bursts but have low waiting times, while large buffers reduce packet loss but increase waiting time. The Random Early Detection (RED) mechanism helps manage this by randomly dropping packets before the queue becomes full, avoiding congestion and packet loss during bursts.

MPLS traffic engineering reserves resources to create label-switched paths (LSPs) on links, ensuring bandwidth availability and reducing congestion. However, since LSPs are established only where resources exist, MPLS TE does not guarantee QoS for individual classes but operates on aggregate bandwidth. Traditional routing forwards packets individually based on metrics like shortest path, which is inefficient for real-time applications.

To address Quality of Service (QoS) challenges, the IETF introduced service models such as Integrated Services (IntServ) and Differentiated Services (DiffServ). IntServ allocates resources to meet strict delay requirements for real-time applications using Resource Reservation Protocol (RSVP), which supports end-to-end service guarantees over IP networks. However, IntServ requires maintaining extensive state information on routers, leading to scalability issues.

Simulations have shown that MPLS can better support real-time applications like VoIP than traditional IP routing due to faster processing and bandwidth efficiency. Combining DiffServ with MPLS has demonstrated improved end-to-end QoS for multiple traffic types in IP networks, though variable bit rate video traffic may still experience packet loss during bursts.

To evaluate QoS in DiffServ/MPLS networks, the Extended Quality of Service based Routing Simulator (EQRS) was developed. EQRS allows configuration of DiffServ/MPLS parameters and simulates constraint-based routing algorithms, confirming that QoS routing improves throughput and overall network performance compared to shortest-path routing.

Research has also explored integrating MPLS and DiffServ to leverage MPLS's fast packet forwarding and traffic engineering with DiffServ's scalable QoS framework, offering efficient solutions for backbone networks. Additionally, MPLS networks utilizing constraint-based Label Distribution Protocol (CR-LDP) and Label Distribution Protocol (LDP) focus on flow aggregation, distribution schemes, and tunnel management for enhanced performance.

Further studies have compared traditional IP networks, MPLS, and MPLS with traffic engineering, highlighting the benefits of MPLS TE in load balancing and traffic control to optimize service delivery and reduce costs. For voice traffic, Weighted Fair Queuing (WFQ) has been found effective in minimizing delay and jitter.

A new scheduler named WFQ-P was proposed to support DiffServ in MPLS core routers, improving bandwidth utilization during traffic bursts and simplifying connection and bandwidth management.

Recent Advances in QoS:

Researchers have proposed bandwidth-based QoS algorithms like QABAA with dynamic call admission control (CAC) mechanisms to optimize performance. For instance, QTBR uses threshold values to manage different service groups, triggering QABAA under certain conditions. However, delay requirements were not fully addressed.

Gateway relocation strategies like GRAC integrate admission control and gateway handover to reduce latency and prevent call blocking during network overload.

In mobile wireless networks, base station load balancing techniques initiate handovers when resource usage approaches thresholds, preventing prolonged overload. These schemes use hysteresis margins to avoid frequent handovers but may face limitations under dynamic conditions.

QoS frameworks combining admission control, load control, and scheduling have been proposed for HSDPA systems, adjusting parameters dynamically based on voice FER and prioritizing different service classes.

Other scheduling algorithms dynamically switch between policies to ensure QoS across multiple classes of service (CoS) in overloaded environments, using multi-level priority systems to meet varying resource requirements.

Analysis:
This research focuses on implementing a DiffServ-based QoS prototype within a network edge environment, requiring appropriate switches and traffic classification. Previous studies have provided mechanisms for load balancing and QoS through predefined parameters, typically set manually.

METHODOLOGY

This portion is divided into three key sections, as illustrated in Figure 3-1: Network Topology Design, Proposed QoS Model, and Evaluation of Results. The first section covers the design of the network topology, including details about the simulation tools and configuration settings used. The second section focuses on the proposed QoS model, providing an in-depth discussion of QoS scheduling parameters along with the presentation of the model's algorithm. The final section is dedicated to analyzing and evaluating the results obtained.

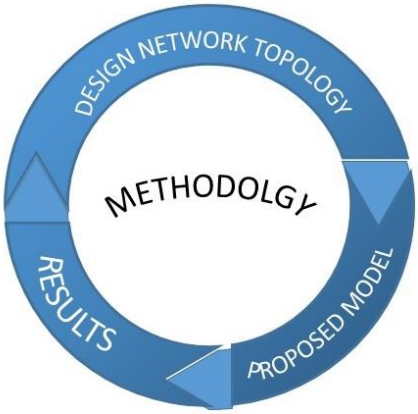

Figure 0-1: Sections of Methodology

This portion centers on the proposed model designed to implement Quality of Service (QoS) for real-time traffic, specifically targeting interactive services. A data traffic engineering model tailored for network environments is introduced, which relies on a differentiated services (DiffServ) based QoS approach. This model is applied at the source routers within the network. At the network edge, bandwidth allocation is shared among multiple clients and regulated according to traffic volume.

The primary goal of this study is to develop and deploy a data traffic engineering model that enhances QoS for real-time traffic by efficiently utilizing existing network resources. Popular QoS frameworks for real-time data include Multi-Level Switch Protocol, Differentiated Services, Resource Reservation Protocol (RSVP), and Integrated Services (IntServ). These mechanisms were briefly reviewed in the previous section. Differentiated Services is selected for this work due to its scalability advantages [22], making it well-suited for managing QoS for real-time constraints. This mechanism is implemented at the sender's network edge.

The proposed model incorporates dynamic buffer allocation alongside a priority-based scheme following the classification of incoming traffic. To evaluate its effectiveness, the model will be simulated and compared against existing approaches. The subsequent sections provide a detailed explanation of the proposed model.

9.1 Network Topology Design

The implementation involves simulating a network topology as illustrated in Figure 3.2. The sources r1, r2, and r3 represent video, data, and voice streams, respectively, at the transmitting side. These sources connect to the switch edge router r4 via 100 Mbps links. Router r4 is connected to the edge router r5 through a 1 Gbps link, and r5 connects to the network over a bottleneck link with bandwidth referred to as MAX. The receiving side mirrors the transmitting side, with router r6 linked to switch edge r7, which connects to data (r10), video (r8), and voice (r9) sources.

Packet formation occurs at switch edge r4 on the transmitting side, while packet scheduling is

managed at router edge r5. The video source r1 generates frames following a pattern of I, B, and P frames within fixed time intervals, with variable frame lengths. During simulation, three video sessions are assumed to run simultaneously at a data rate of 384 Kbps. The data source r2 simulates constant bit rate (CBR) traffic with fixed packet sizes, and voice traffic from r3 is modeled using a Poisson distribution at 64 Kbps. The total traffic load on the network's bottleneck link is the sum of the traffic loads from all sources (EF, BF, and AF traffic loads). Additional network specifications are provided in Table 3.1.

Component Type	Identifier	Description	Link Speed
Source	R1	Video Traffic Source	100 Mbps
	R2	Audio Traffic Source	100 Mbps
	R3	Text Traffic Source	100 Mbps
Sink	R8	Video Traffic Receiver	100 Mbps
	R9	Audio Traffic Receiver	100 Mbps
	R10	Text Traffic Receiver	100 Mbps
Switch	R4	Edge Switch for Network 1	1 Gbps
	R7	Edge Switch for Network 2	1 Gbps
Router	R5	Edge Router for Network 1 (Bottleneck Link)	2.1 Mbps
	R6	Edge Router for Network 2 (Bottleneck Link)	2.1 Mbps

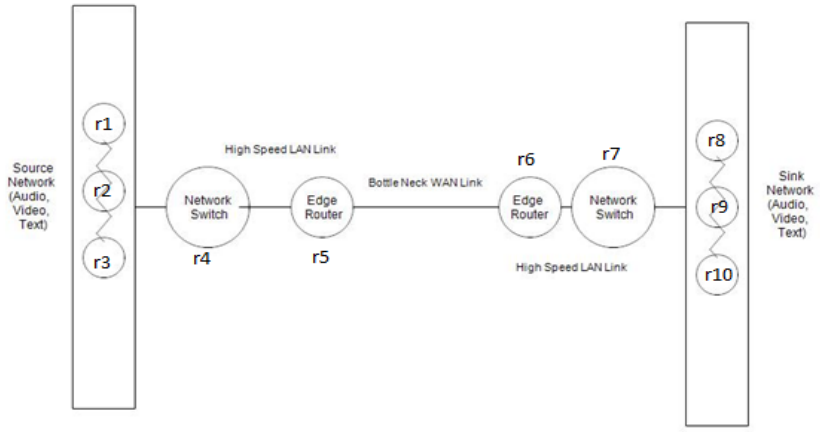

Figure 0-2: Network example for Simulation

Simulation Tool:

The network scenario is developed using the widely-used network simulator NS-2. The implementation of the proposed QoS model aims to enhance network performance by focusing on metrics such as packet loss and delay, particularly at the receiver end during traffic flow. The following sections present the real-time traffic monitoring data and corresponding analysis.

The model's performance is evaluated through simulation in NS-2 [34], demonstrating how QoS improvements affect actual-time data transmission. At the receiver, the effectiveness is measured by tracking packet loss and delay of live traffic [25]. Packet loss and delay are significant challenges in real-time traffic due to limited network resources. Reducing these issues, especially packet loss, is complex in constrained environments. This study targets minimizing both delay and packet loss.

Two traffic types are considered: Constant Bit Rate traffic (CBR-t) and Variable Bit Rate traffic (VBR-t). Video streams are modeled as VBR-t, while voice traffic is treated as CBR-t. Both belong to IP traffic categories. In a physical testbed, traffic would be

generated at the sender, but due to resource limitations, simulation is utilized.

Traffic flows through a QoS-enabled network to evaluate performance. Network devices, including routers and switches, are configured accordingly. Traffic classes adhere to a DiffServ (Differentiated Services) framework. The edge router and network switch implement QoS configurations to simulate bottleneck scenarios and assess performance impact. Real-world routers typically support QoS with four queues prioritized from 1 to 4. The proposed model excludes the highest priority queue and operates using the remaining three queues for traffic handling. Figure 3.2 illustrates the QoS configuration block model applied at the network edge.

Configuration Using Simulation Tool:

Given the limitations of available resources, the proposed model aims to reduce packet delay and loss in real-time collaborative data traffic. The model addresses both variable and constant bit rate traffic types. Video traffic is classified as variable bit rate, while voice over IP is considered constant bit rate. Traffic with real-time constraints is generated at the transmitter and sent through the QoS-enabled network, passing through edge devices.

Differentiated Services-based QoS is applied to various traffic groups. The mechanism is implemented on routers and switches at the network edge to manage bottleneck conditions at the transmitter. Typically, routers provide four queues for DiffServ. The first queue, known as the peak priority queue, is reserved for specific signaling traffic and is excluded from this model, which uses the other three queues.

Differentiated Services can be divided into three main stages:

- Marking: Incoming traffic from the local area network is labeled with different DSCP (Differentiated Services Code Point) priority values depending on the traffic type. For example, video traffic is tagged as AF (Assured Forwarding), voice traffic as EF (Expedited Forwarding), and best-effort traffic as BE (Best Effort). This marking occurs at the edge routers on the transmitter side.

- **Classifying: Traffic is categorized into four groups based on DSCP values:**
  - Assured Forwarding (AF) for video traffic
  - Best Effort (BE) for non-real-time traffic
  - Expedited Forwarding (EF) for voice traffic

  Classification is performed at central routers on the transmitting side.
- Scheduling: Routers allocate queues to traffic groups and distribute bandwidth accordingly. Queue management significantly impacts packet delay and loss. The proposed QoS scheduling model uses experimental queue management techniques to optimize performance.

QoS Scheduling Model:

This model dynamically allocates available bandwidth and buffer resources among different traffic classes. Buffer allocation adapts based on current queue lengths and load conditions. Bandwidth assignment uses a weighted round-robin approach, where weights determine priority bias. Each queue's allocated bandwidth is proportional to its weight. The model first defines scheduling factors and then calculates resource distribution for each queue.

Defining Nodes and Specifications in NS-2 Simulator:

In the initial step, six nodes are created in the NS-2 environment, including two edge routers and four nodes acting as sources and destinations, as depicted in Figure 3-3.

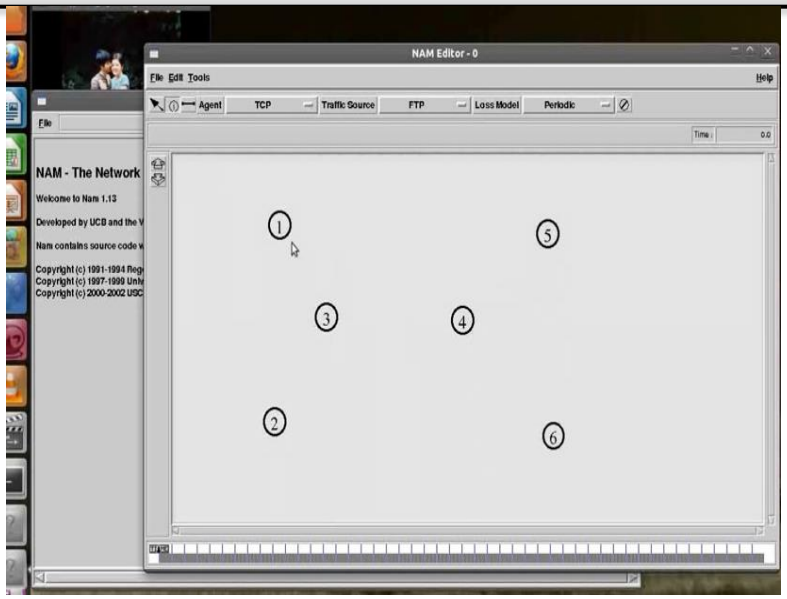

Figure 0-3: Step 1 Selection of 6 Nodes

In the second step, the source and destination nodes are identified using capital letters such as A, B, C, and D. Additionally, each node is assigned specific colors and sizes, as illustrated in Figure 3-4.

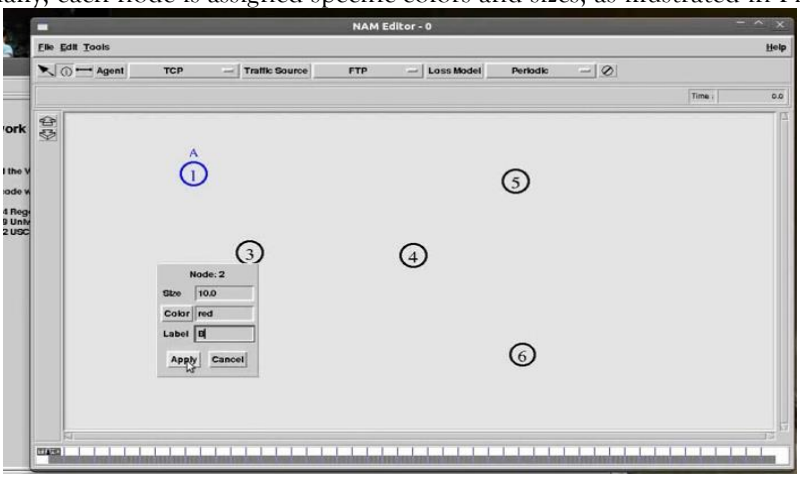

Figure 0-4: Step 2 labeling of all nodes

In step 3, all nodes were connected, and configurations for both TCP and UDP traffic were applied, as illustrated in Figure 3-5.

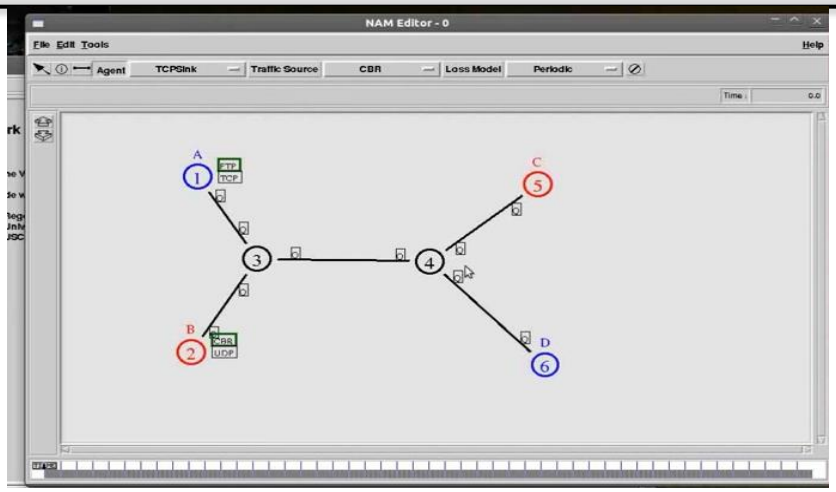

Figure 0-5: Step 3 Apply TCP and UDP Traffic

In step 4, the start and end times for UDP and TCP traffic were defined for each session, as illustrated in Figure 3-6.

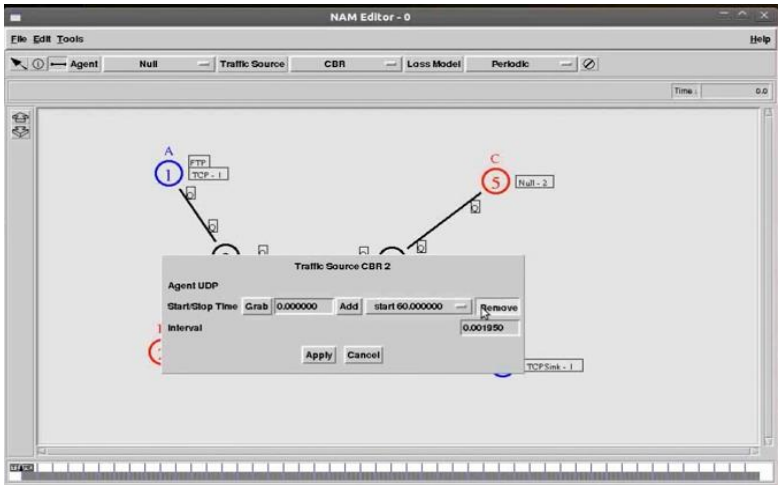

Figure 0-6: Step 4 Start and End time of UDP and TCP traffic

In step 5, we evaluated our setup and recorded the results for packet loss and packet delay for both UDP and TCP traffic during each iteration, as illustrated in Figures 3.7

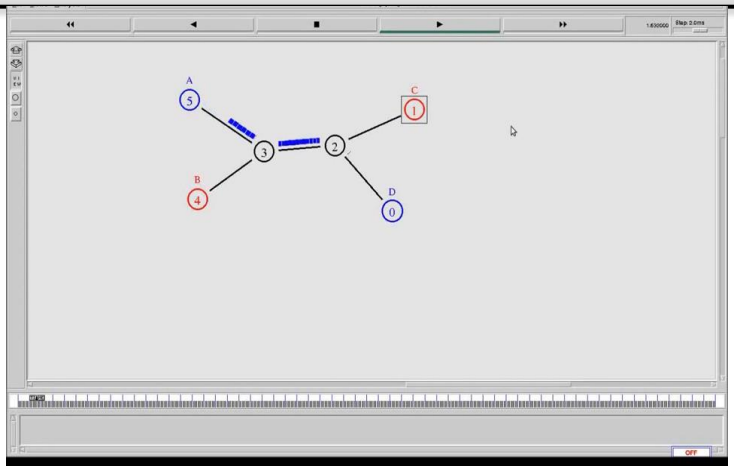

Figure 0-7: Step 5 Testing of TCP and UDP traffic

SIMULATION CODES

In this study, the NS2 event-driven network simulator is utilized. NS2 features a core engine written in C++ that is extended to perform simulations, while OTcl serves as the configuration and command interface. Essentially, the entire software framework is developed using C++, with OTcl functioning as the front-end scripting language. OTcl stands for Object-Oriented Tcl.

Topology Creation

A link connects two nodes (vertices). To create a node, the following command is used:

set n0 [$ns node]
set n1 [$ns node]

To establish a link between these nodes, the command format is:

$ns <link-type> $n0 $n1 <bandwidth> <delay>

For example:

$ns duplex-link $n0 $n1 1Mb 10ms dsRED

This creates a duplex link between nodes n0 and n1 with a bandwidth of 1 Mbps and a delay of 10 milliseconds, using the 'dsRED' queue management.

```
exec nam out.nam &
exit 0
}
#Create node
set n0 [ $ns node ]
set n1 [ $ns node ]
set n2 [ $ns node ]
set n3 [ $ns node ]
# Create links between the nodes
$ns duplex-link $n0 $n2 2Mb 10ms DropTail
$ns duplex-link $n1 $n2 2Mb 10ms DropTail
$ns duplex-link $n2 $n3 1.5Mb 20ms DropTail
# Set queue size for link
```

```
#Setup a CBR over UDP connection
set cbr [ new Application/Traffic/CBR ]
$cbr attach-agent $udp
$cbr set type_ CBR
$cbr set packet_size 1000
$cbr set rate_ 1mb
$cbr set random_ false

# schedule event for FTP agent
$ns at 0.5 "$cbr start"
$ns at 1.0 "$ftp start"
$ns at 4.0 "$ftp stop"
$ns at 5.0 "finish"
$ns run
```

Tcl ▾ Tab Width: 8 ▾ Ln 68, Col 23 INS

PROPOSED QOS MODEL

This section introduces and validates the proposed qos model through simulation. A detailed explanation of its components follows.

QOS SCHEDULING PARAMETERS

Two main groups of factors influence packet loss and delay in outgoing data traffic.

QUEUE LENGTH

The queue length determines the buffer size, which is crucial in handling incoming traffic efficiently. There are three distinct queues, each corresponding to a specific service category, reflecting the three types of data traffic: audio, video, and text. These are described as follows:

- qlbe: queue length or buffer size for best effort (be) traffic.
- qlef: queue length or buffer size for expedited forwarding (ef) traffic.
- qlaf: queue length or buffer size for assured forwarding (af) traffic.

LINE WEIGHT

Line weight represents how the queue length changes in response to variations in incoming traffic load. This factor is calculated separately for each traffic type using specific equations (provided at the end of this section). It serves as an input for a biased round-robin scheduler. The weights are defined as:

Qwbe: weight for be traffic.
Qwaf: weight for af traffic.
Qwef: weight for ef traffic.

QOS PARAMETERS

Qos is evaluated based on variables that impact performance, such as acceptable delay and packet size, specific to each traffic class within an active session. These metrics are measured independently for each class as follows:

Nef: number of sessions for ef traffic.
Dlvef: acceptable packet delay for ef traffic.
Pktszef: average packet size for ef traffic.
Idref: average data input rate for ef traffic.
Naf: number of sessions for af traffic.
Dlvaf: acceptable packet delay for af traffic.

DYNAMIC BUFFER ALLOCATION

Two main groups of factors influence packet loss and delay in outgoing data traffic. Buffer sizes are dynamically assigned based on traffic characteristics, including data rate, packet size, and acceptable delay per traffic class. Real-time traffic properties are analyzed to gather input data rates, while admission control mechanisms determine the number of sessions allowed. The following equations describe buffer allocation per traffic class:

$$Qlef = (idref * dlvef * nef) / pktszef \quad (1)$$

$$Qlaf = (idraf * dlvaf * naf) / pktszaf \quad (2)$$

$$Qlef = buff - (qlef + qlaf) \quad (3)$$

WEIGHTED ROUND ROBIN SCHEDULING

The model employs a weighted round robin (wrr) scheduling algorithm, which dynamically adjusts resource allocation based on traffic load. Line weights for each class influence the service rates:
Qwbe, qwaf, qwef: weights for be, af, and ef traffic respectively.
Queue sizes at a given time t are denoted as:
Qlbe, qlaf, qlef

Priority or superiority factors based on traffic type are assigned as:
Paf = superiority of af=3 (af traffic)
Pef = superiority of ef=2 (ef traffic)
Pbe = superiority of be=1(be traffic)
Weights are computed using the following formulas:
Qwbe= (qlbe * pbe)/(qlaf + qlbe + qlef) (4)
Qwaf= (qlaf * paf)/(qlaf + qlbe + qlef) (5)
Qwef= (qlef * pef)/(qlaf + qlbe + qlef) (6)
This approach ensures that service rates adapt to both queue lengths and priority levels. When queue lengths are similar, the scheduler prioritizes traffic classes according to their assigned superiority.

SCHEDULING PROCEDURE

The scheduling process using WAR is outlined as:
Step 1: determine the number of active sessions, as defined by admission control.
Step 2: calculate queue lengths for each traffic class using equations 1-3.
Step 3: compute the weighted service rates using equations 4-6.
Step 4: operate queues according to their weights:
If one queue is longer than others, serve that queue.
If queues are equal, prioritize based on class superiority.
The proposed scheduling model was simulated using ns2, and simulation results are presented subsequently.

PROPOSED ALGORITHM

The algorithm for qos measurement starts by initializing parameters such as input data rate, queue length, line weight, acceptable packet delay, and packet size for three traffic types: af (video), ef (audio), and be (text). At the router edge, it calculates these metrics based on buffer size and session count, assigns priority factors, and finally applies the weighted round robin scheduler to manage traffic efficiently.

Initialize idr_{AF} idr_{BE}, idr_{EF}, Ql_{AF}, Ql_{BE}, Ql_{EF},Qw_{AF}, Qw_{EF}, Qw_{BE}, dlv_{AF}, dlv_{EF}, $pktsz_{AF}$, $pktsz_{EF}$
F[]=incoming frame
Input frame-type
If frame-type=video then class=AF
Else if frame-type=audio then class=EF
Else if frame-type=text then class=BE
calculate idr from previous frame
if class=AF
input dlv_{AF}= maximum delay to observe
$pktsz_{AF=}$ Packet size from previous frame
N_{AF}= No of active sessions
$ql_{AF} = (idr_{AF} * dlv_{AF} * N_{AF}) / pktsz_{AF}$
P_{AF} =3
$qw_{AF}= (ql_{AF} * P_{AF})/(ql_{AF} + ql_{BE} + ql_{EF})$
if class=EF
input dlv_{EF}= maximum delay to observe
$pktsz_{EF=}$ Packet size from previous frame
N_{EF}= No of active sessions
$ql_{EF} = (idr_{EF} * dlv_{EF} * N_{EF}) / pktsz_{EF}$
P_{EF} =2
$qw_{EF}= (ql_{EF} * P_{EF})/(ql_{AF} + ql_{BE} + ql_{EF})$
if class=BE
ql_{EF} =BUFF – $(ql_{EF} + ql_{AF})$
P_{BE} =1
$qw_{BE}= (ql_{BE} * P_{BE})/(ql_{AF} + ql_{BE} + ql_{EF})$

Figure 0-8: Algorithm for Measuring QoS

RESULTS AND DISCUSSION

SIMULATION RESULTS AND ANALYSIS

A link connects two nodes (vertices). To create a node, the following command is used: To evaluate qos, key parameters such as packet delay and loss are examined under increased link loads, particularly focusing on multiple bottleneck links at the collector side. The analysis is conducted under two conditions:

- with the proposed qos method implemented.
- without applying the proposed qos model.

At the collector side, performance metrics are assessed to determine potential improvements offered by the proposed model in handling data traffic under real-time constraints.

TRAFFIC GENERATION:

Figures 4-1 (a) and 4-1 (b) illustrate sample screenshots from the traffic generation phase of the simulation performed using the ns2 simulator.

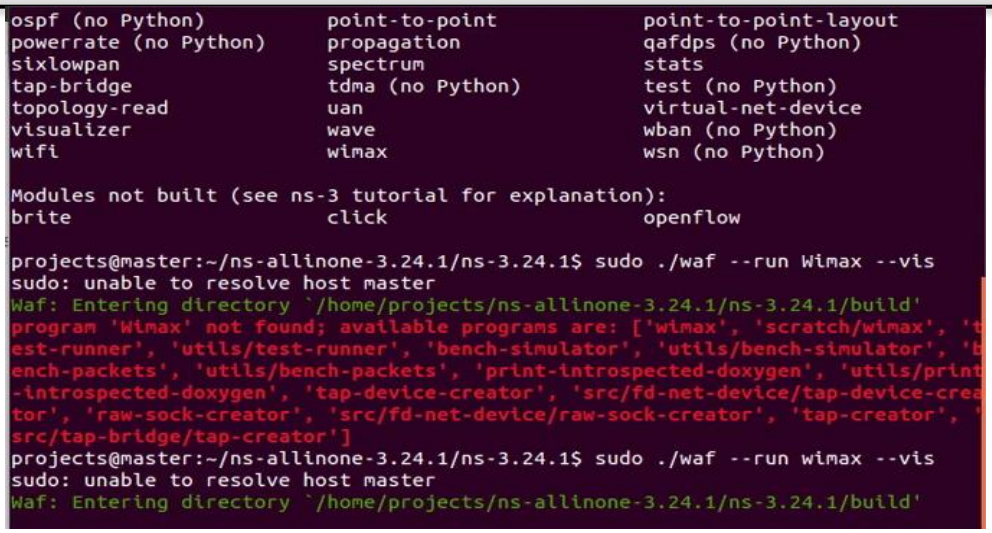

Figure 0-9(a) Traffic generation for NS2 Simulator

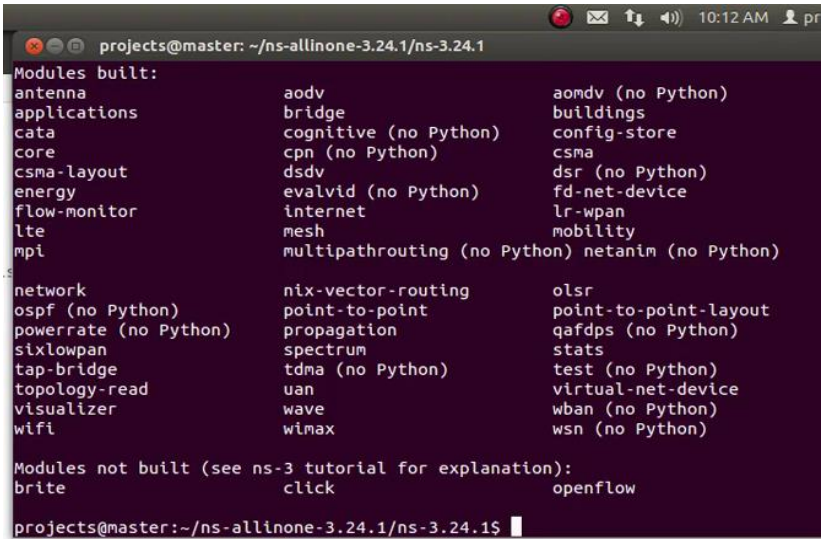

Figure 0-10(b) Traffic generation for NS2 Simulator

The results include an analysis of packet delay and packet loss across three types of traffic: AF, EF, and BE.

EFFECT OF QoS ON PACKET LOSS FOR VARIOUS TRAFFIC TYPES:

Packet loss was measured as a key indicator to evaluate QoS both before and after implementing the proposed dynamic buffer allocation and weight-based approach.

Packet Loss in AF Traffic:

At the receiver, packet loss was monitored to assess performance. It was found that the proposed QoS model significantly enhanced the handling of real-time traffic. The simulation results for different traffic classes are summarized below.

Table 4-1: Packet Loss Performance for AF Traffic under QoS

Bandwidth (Mbps)	Packet Loss Without QoS (%)	Packet Loss With QoS (%)
2.1	0.3	0.2
2.3	3	0.2
2.8	8	1
3.3	14	1
3.8	22	1

Table 4-1 illustrates that without applying the QoS model, packet loss increases proportionally with the network load, ranging from 0.3% at 2.1 Mbps to 22% at 3.8 Mbps. Conversely, when the QoS model is implemented, packet loss remains relatively stable, fluctuating between 0.2% and 1% across all tested loads. This improvement is attributed to the dynamic buffer allocation mechanism. Figure 4-2 presents a graphical comparison of AF traffic packet loss performance.

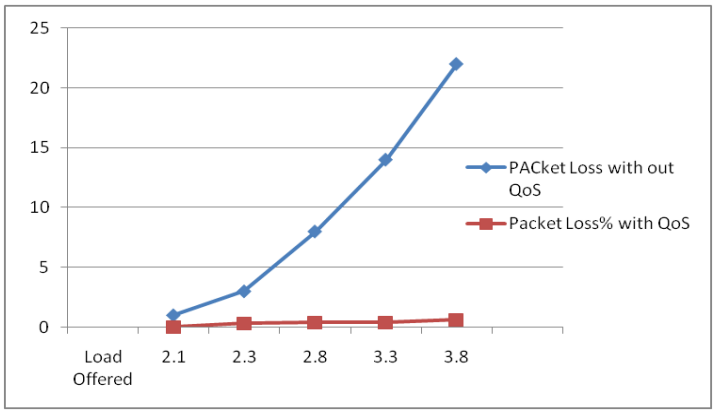

Figure 0-11: Performance for Packet Loss AF traffic

Packet Lost for EF traffic:

Table 4-2: QoS Performance for Packet Loss in EF Traffic

Bandwidth (Mbps)	Packet Loss Without QoS (%)	Packet Loss With QoS (%)
2.1	1	0.3
2.3	3.6	0.3
2.8	9	0.5
3.4	22	1
3.9	39	1.6

This table displays the packet loss ratio observed for EF traffic under varying network loads, which increase progressively from 2.1 Mbps up to 3.9 Mbps. It is evident that the packet loss ratio rises significantly once the network load surpasses the threshold of 2.1 Mbps. Without implementing a QoS model, the packet loss starts at 1% for 2.1 Mbps and escalates sharply, reaching 39% at the highest load of 3.9 Mbps.

Conversely, when the QoS mechanism is enabled—incorporating dynamic buffer allocation that adjusts buffer length based on load—the packet loss is maintained within a much lower range, between 0.3% and Conversely, when the QoS mechanism is enabled—incorporating dynamic buffer allocation that adjusts buffer length based on load—the packet loss is maintained within a much lower range, between 0.3% and 1.6%. The graphical representation of these results is illustrated in Figure 4-3, highlighting the improvement in packet loss performance for EF traffic with QoS applied.

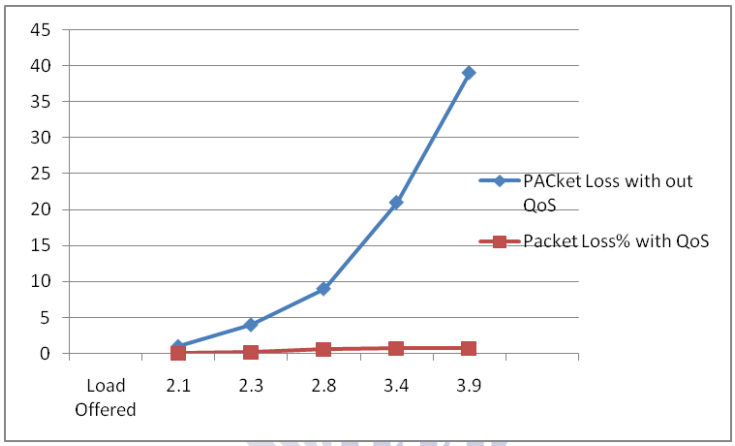

Figure 0-12: Performance for Packet Loss EF traffic

Packet Loss for BE Traffic:

Table 4-3: QoS Performance Regarding Packet Loss for BE Traffic

Bandwidth (Mbps)	Packet Loss Without QoS (%)	Packet Loss With QoS (%)
2.1	0.5	0.3
2.3	4	2
2.8	10	4
3.4	21	9
3.9	34	23

The performance of Best Effort (BE) traffic was analyzed under varying network loads both with and without implementing a Quality of Service (QoS) model. The bandwidth was gradually increased through values of 2.1, 2.3, 2.8, 3.4, and 3.9 Mbps, as presented in Table 4-3. Without QoS in place, the packet loss ratio rises sharply from 0.5% up to 34% once the traffic load surpasses the network's threshold.

Conversely, when the QoS model is applied, the packet loss shows some improvement. The loss increases from 0.3% to 23%, which is consistently lower than the losses recorded without QoS. However, the enhancement for BE traffic remains limited because BE packets are assigned the lowest priority in the queuing system. Priority is first given to Assured Forwarding (AF) traffic, followed by Expedited Forwarding (EF) traffic, which reduces the resources available for BE traffic. Despite this, there is still a noticeable improvement of about 9% in packet loss reduction for BE traffic with QoS.

Figure 4-4 illustrates the relationship between packet loss and network load for BE traffic in graphical form

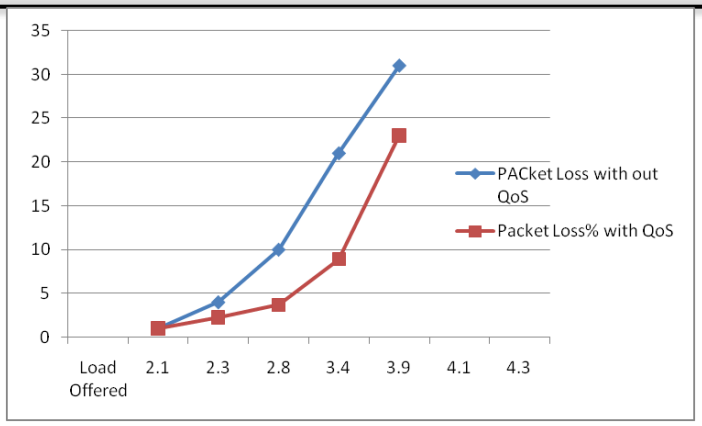

Figure 0-13: Ratio of lost packets with load for BE traffic

From the figures above, it is evident that the WRR scheduling algorithm allocates Quality of Service (QoS) to grantees by maintaining packet loss below a certain threshold for each service until the total offered load reaches the bandwidth limit of 2.1 Mbps. As a result, the EF (Expedited Forwarding) and AF (Assured Forwarding) streams, which have bandwidth constraints, experience lower packet loss percentages. Conversely, the BE (Best Effort) stream surpasses this threshold and suffers higher packet loss. When the total offered load exceeds the bottleneck or congested bandwidth, EF and AF streams consistently receive superior QoS guarantees, whereas the BE stream encounters significantly higher packet loss rates.

Therefore, the QoS mechanism ensures that real-time traffic maintains acceptable quality even when the total offered load exceeds the available bandwidth on a congested link. While it is important not to completely neglect the BE stream, fairness is preserved by allocating resources impartially.

9.1.1 Impact on Packet Delay for AF Streams

The following discussion analyzes how the QoS implementation influences packet delay across different traffic streams.

Bandwidth (Mbps)	Packet Delay Without QoS (ms)	Packet Delay With QoS (ms)
2.1	100	100
2.3	108	108
2.8	118	108
3.4	123	109
3.9	125	109
4.1	128	110
4.3	130	112
4.7	134	114

Table 4-5 illustrates that without applying the QoS model, packet delay increases as load rises from 2.1 Mbps to 4.7 Mbps, ranging from a minimum of 100 ms up to 134 ms. However, with the QoS scheme in place, the maximum packet delay is limited to 114 ms at the highest load of 4.7 Mbps. This represents a delay improvement of approximately 20 ms for AF traffic. The delay values remain relatively stable despite increasing load, which can be attributed to the WRR scheduler prioritizing this traffic and dynamically adjusting buffer allocation as the input data rate increases. Figure 4-4 provides a graphical comparison of packet delay versus load for AF traffic.

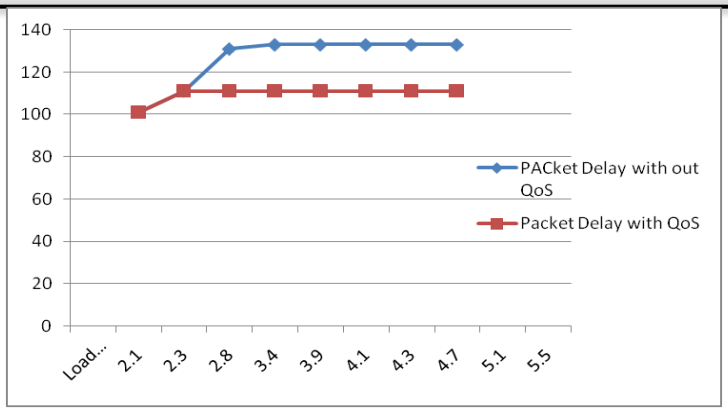

Figure 0-14: Effect on Packet delay for AF delay traffic

Effect of Packet Delay on EF Streams:

Table 4-5: QoS Performance for EF Delay Traffic

Bandwidth (Mbps)	Packet Delay Without QoS (ms)	Packet Delay With QoS (ms)
2.1	120	119
2.3	140	125
2.8	140	127
3.4	140	127
3.9	140	129
4.1	144	129
4.3	150	130
4.7	157	131

The data presented in Table 4-5 reflects the results observed during the evaluation of EF traffic performance. The bandwidth load was tested across various levels: 2.1, 2.3, 2.8, 3.4, 3.9, 4.1, 4.3, and 4.7 Mbps. Packet delay without QoS ranged from 120 ms up to 157 ms as the load increased.

In contrast, when QoS was implemented, packet delays remained more stable, varying between 119 ms and 131 ms across all load levels. The highest delay recorded with QoS was 131 ms at 4.7 Mbps, which is 26 ms lower compared to the delay experienced without QoS under the same load. Additionally, it was noted that delays plateau once the load surpasses a certain threshold. This behavior can be attributed to dynamic buffer management and prioritization mechanisms, especially when multiple sessions are active.

Figure 4-6 illustrates the relationship between packet delay and load for EF delay traffic, highlighting the beneficial impact of QoS.

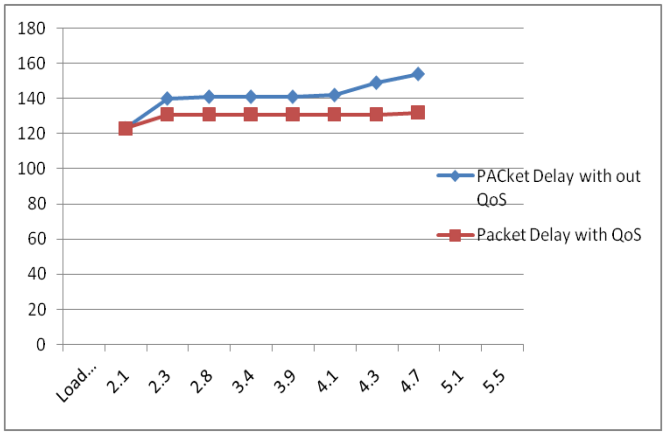

Figure 0-15: Effect on Packet delay for EF delay traffic

Effect of Packet Delay on Best Effort (BE) Traffic:

Table 4-6: QoS Performance for BE Delay Traffic

Bandwidth (Mbps)	Packet Delay Without QoS (ms)	Packet Delay With QoS (ms)
2.1	148	121
2.3	154	127
2.8	157	135
3.4	159	142
3.9	162	146
4.1	172	149
4.3	180	150
4.7	189	153

The network load applied to evaluate the QoS performance varies across bandwidths of 2.1, 2.3, 2.8, 3.4, 3.9, 4.1, 4.3, and 4.7 Mbps, as illustrated in Table 4-6. Without QoS, packet delay starts at a minimum of 148 ms and increases up to 189 ms as the load grows. When QoS is implemented, delays are notably reduced, ranging from 121 ms to 153 ms. The highest recorded delay under the QoS model is 153 ms at The network load applied to evaluate the QoS performance varies across bandwidths of 2.1, 2.3, 2.8, 3.4, 3.9, 4.1, 4.3, and 4.7 Mbps, as illustrated in Table 4-6. Without QoS, packet delay starts at a minimum of 148 ms and increases up to 189 ms as the load grows. When QoS is implemented, delays are notably reduced, ranging from 121 ms to 153 ms. The highest recorded delay under the QoS model is 153 ms at the 4.7 Mbps load, showing an improvement of 36 ms compared to the scenario without QoS. Figure 4-7 provides a graphical representation of how packet delay varies with increasing load for BE delay traffic.

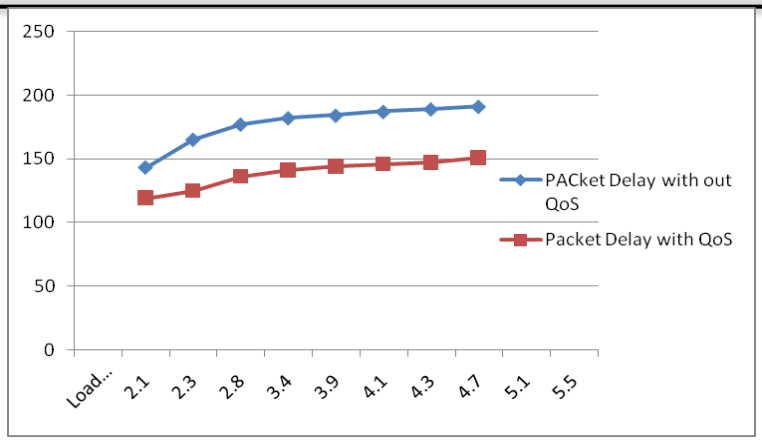

Figure 0-16: Effect on Packet delay for BE delay traffic

When the total incoming load remains below the available bandwidth of a congested link, applying a QoS model results in reduced delay.

The figures indicate that the QoS model guarantees quality of service with respect to packet delay for all service types once a certain load threshold is reached. At the upper limit of presented load, EF and AF packets experience a lower delay percentage, while BE traffic suffers more. When the total load surpasses the bottleneck bandwidth, the QoS mechanisms prioritize EF and AF traffic, consistently maintaining better service quality for these categories.

CONCLUSION & FUTURE WORK

CONCLUSION

This research focuses on designing an architecture that ensures Quality of Service (QoS) by implementing congestion control tailored for diverse data traffic types. Input traffic is routed through a bandwidth-limited bottleneck link. The proposed scheduler enhances traffic performance on this shared network segment. The approach introduces an adaptive algorithm integrated with the router, which monitors the router's current state and manages congestion based on available resources such as buffer capacity and input data rates.

The solution aims to guarantee QoS for real-time traffic, improving network performance despite limited bandwidth. It is built on the DiffServ model and deployed at the network edge. Unlike some approaches, it does not require over-provisioning of bandwidth. Validation is conducted through simulations that assess the scheduling algorithm's effectiveness. Due to practical constraints, real-world testing was not feasible; hence, simulation models were employed to analyze QoS parameters under various load conditions.

The method dynamically adapts to different traffic types and varying input rates across classes. Simulation results demonstrate that packet loss and delay increase with rising load, especially for AF and EF traffic classes. However, the model improves overall QoS for real-time services by reducing delay and packet loss in voice and video streams.

This traffic engineering strategy enhances QoS for limited-bandwidth real-time traffic by fine-tuning traffic management. Testing using Network Simulator 2 with constant and variable bit rate traffic shows that delay and packet loss remain within acceptable quality limits, even with multiple input sources competing for the same constrained bandwidth. A key parameter introduced controls the number of video sessions via a resource allocation index, optimizing performance during interactive video communication. Findings suggest that limiting active sessions frees bandwidth, improving best-effort traffic performance. Maintaining approximately half the resources as free capacity results in optimal traffic throughput, helping service providers avoid excessive resource over-provisioning for audio/video traffic.

The proposed process can be summarized as follows:

a) Incoming packets are marked with DSCP values based on their real-time source.

b)Traffic is classified and directed to appropriat queues.

c)Queues are scheduled according to available bandwidth and buffer capacity.
d)Input data rates are reported to the edge router.
e)The router schedules traffic based on these metrics.
f) Packets are forwarded through the constrained bandwidth link as output.
This method remains adaptive, adjusting to variable input rates detected before traffic reaches the router. Available bandwidth and buffer sizes are used as key parameters in scheduling decisions.

FUTURE WORK

While this work concentrates on QoS for real-time traffic, the approach can be extended to enhance best-effort TCP traffic delivery. Future research could explore:
a) Incorporating MPLS techniques to further improve QoS on shared network links.
b) Implementing the model with autonomous agents to enable learning and adaptive network management.

REFERENCE

Salahuddin, Syed Shahid Abbas, Prince Hamza Shafique, Abdul Manan Razzaq, & Mohsin Ikhlaq. (2024). Enhancing Reliability and Sustainability of Green Communication in Next-Generation Wireless Systems through Energy Harvesting. Journal of Computing & Biomedical Informatics.

Furqan, F., & Hoang, D. B. (2013, January). WFICC: A new mechanism for provision of QoS and Congestion Control in wireless. In Consumer Communications and Networking Conference (CCNC), 2013 IEEE (pp. 552-558). IEEE.

Ashraf, M., Jalil, A., Salahuddin & Jamil, F. (2024). DESIGN AND IMPLEMENTATION OF ERROR ISOLATION IN TECHNO METER. Kashf Journal of Multidisciplinary Research, 1(12), 49-66.

Tung, H. Y., Tsang, K. F., Lee, L. T., & Ko, K. T. (2008, January). QoS for mobile Wireless networks: call admission control and bandwidth allocation. In Consumer Communications and Networking Conference, 2008. CCNC 2008. 5th IEEE (pp. 576-580). IEEE.

Khan, S.U.R., Asif, S., Bilal, O. et al. Lead-cnn: lightweight enhanced dimension reduction convolutional neural network for brain tumor classification. Int. J. Mach. Learn. & Cyber. (2025). https://doi.org/10.1007/s13042-025-02637-6.

Casey, T., Veselinovic, N., & Jantti, R. (2008, September). Base station controlled load balancing with handovers in mobile Wireless. In Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on (pp. 1-5). IEEE.

Meeran, M. T., Raza, A., & Din, M. (2018). Advancement in GSM Network to Access Cloud Services. Pakistan Journal of Engineering, Technology & Science [ISSN: 2224-2333], 7(1).

Salahuddin, Hussain, M., & hamza Shafique, P. (2024). PERFORMANCE ANALYSIS OF MATCHED FILTER-BASED SECONDARY USER DETECTION IN COGNITIVE RADIO NETWORKS. Kashf Journal of Multidisciplinary Research, 1(10), 15-26.

Lucena, E. O., Lima, F. R. M., Freitas Jr, W. C., & Cavalcanti, F. R. P. (2010, December). Overload prediction based on delay in wireless OFDMA Systems. In Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE (pp. 1-5). IEEE.

Khan, S. U. R., Asim, M. N., Vollmer, S., & Dengel, A. (2025). Robust & Precise Knowledge Distillation-based Novel Context-Aware Predictor for Disease Detection in Brain and Gastrointestinal. arXiv preprint arXiv:2505.06381..

Hekmat, A., et al., Brain tumor diagnosis redefined: Leveraging image fusion for MRI enhancement classification. Biomedical Signal Processing and Control, 2025. 109: p. 108040.

Khan, Z., Hossain, M. Z., Mayumu, N., Yasmin, F., & Aziz, Y. (2024, November). Boosting the Prediction of Brain Tumor Using Two Stage BiGait Architecture. In 2024 International Conference on Digital Image Computing: Techniques and Applications (DICTA) (pp. 411-418). IEEE.

El-Shinnawy, A. H., Nassar, A. M., & Badawi, A. H. (2010, December). A switched scheduling algorithm for congestion relief in Wireless wireless networks. In Computer Engineering Conference (ICENCO), 2010 International (pp. 34-39). IEEE

Khan, S. U. R., Raza, A., Shahzad, I., & Ali, G. (2024). Enhancing concrete and pavement crack prediction through hierarchical feature integration with VGG16 and triple classifier ensemble. In 2024 Horizons of Information Technology and Engineering (HITE)(pp. 1-6). IEEE https://doi. org/10.1109/HITE63532.

Raza, A., Salahuddin, & Inzamam Shahzad. (2024). Residual Learning Model-Based Classification of COVID-19 Using Chest Radiographs. Spectrum of Engineering Sciences, 2(3), 367–396

Khan, S.U.R., Zhao, M. & Li, Y. Detection of MRI brain tumor using residual skip block based modified MobileNet model. Cluster Comput 28, 248 (2025). https://doi.org/10.1007/s10586-024-04940-3

Raza, A., Soomro, M. H., Shahzad, I., & Batool, S. (2024). Abstractive Text Summarization for Urdu Language. Journal of Computing & Biomedical Informatics, 7(02).

M. Wajid, M. K. Abid, A. Asif Raza, M. Haroon, and A. Q. Mudasar, "Flood Prediction System Using IOT & Artificial Neural Network", VFAST trans. softw. eng., vol. 12, no. 1, pp. 210–224, Mar. 2024.

Khan, U. S., & Khan, S. U. R. (2024). Boost diagnostic performance in retinal disease classification utilizing deep ensemble classifiers based on OCT. Multimedia Tools and Applications, 1-21.

Jaffar, J., Hashim, H., Abidin, H. Z., & Hamzah, M. K. (2009, October). Video quality of service in Diffserv-aware multiprotocol label switching network. In Industrial Electronics & Applications, 2009. ISIEA 2009. IEEE Symposium on (Vol. 2, pp. 963-967). IEEE.

Raza, A., & Meeran, M. T. (2019). Routine of encryption in cognitive radio network. Mehran University Research Journal of Engineering & Technology, 38(3), 609-618.

Salahuddin, Abdul Manan Razzaq, Syed Shahid Abbas, Mohsin Ikhlaq, Prince Hamza Shafique, & Inzimam Shahzad. (2024). Development of OWL Structure for Recommending Database Management Systems (DBMS). Journal of Computing & Biomedical Informatics, 7(02).

Al-Khasawneh, M. A., Raza, A., Khan, S. U. R., & Khan, Z. (2024). Stock Market Trend Prediction Using Deep Learning Approach. Computational Economics, 1-32.

M. Waqas, Z. Khan, S. U. Ahmed and A. Raza, "MIL-Mixer: A Robust Bag Encoding Strategy for Multiple Instance Learning (MIL) using MLP-Mixer," 2023 18th International Conference on Emerging Technologies (ICET), Peshawar, Pakistan, 2023, pp. 22-26.

Waqas, M., Ahmed, S. U., Tahir, M. A., Wu, J., & Qureshi, R. (2024). Exploring Multiple Instance Learning (MIL): A brief survey. Expert Systems with Applications, 123893.

Khan, U. S., Ishfaque, M., Khan, S. U. R., Xu, F., Chen, L., & Lei, Y. (2024). Comparative analysis of twelve transfer learning models for the prediction and crack detection in concrete dams, based on borehole images. Frontiers of Structural and Civil Engineering, 1-17.

Khan, S. U. R., & Asif, S. (2024). Oral cancer detection using feature-level fusion and novel self-attention mechanisms. Biomedical Signal Processing and Control, 95, 106437.

Waqas, M., Tahir, M. A., Al-Maadeed, S., Bouridane, A., & Wu, J. (2024). Simultaneous instance pooling and bag representation selection approach for multiple-instance learning (MIL) using vision transformer. Neural Computing and Applications, 36(12), 6659-6680.

Farooq, M. U., Khan, S. U. R., & Beg, M. O. (2019, November). Melta: A method level energy estimation technique for android development. In 2019 International Conference on Innovative Computing (ICIC) (pp. 1-10). IEEE.

Aslam, M., Salahuddin, Ali, G., & Batool, S. (2024). ASSESSING THE EFFECTS OF BIG DATA ANALYTICS AND AI ON TALENT ACQUISITION AND RETENTION. Kashf Journal of Multidisciplinary Research, 1(11), 73-84.

Mahmood, F., Abbas, K., Raza, A., Khan,M.A., & Khan, P.W. (2019). Three Dimensional Agricultural Land Modeling using Unmanned Aerial System (UAS). International Journal of Advanced Computer Science and Applications (IJACSA) [p-ISSN : 2158-107X, e-ISSN : 2156-5570], 10(1).

Raza, A.; Meeran, M.T.; Bilhaj, U. Enhancing Breast Cancer Detection through Thermal Imaging and Customized 2D CNN Classifiers. VFAST Trans. Softw. Eng. 2023, 11, 80–92.

Dai, Q., Ishfaque, M., Khan, S. U. R., Luo, Y. L., Lei, Y., Zhang, B., & Zhou, W. (2024). Image classification for sub-surface crack identification in concrete dam based on borehole CCTV images using deep dense hybrid model. Stochastic Environmental Research and Risk Assessment, 1-18.

Khan, S.U.R.; Asif, S.; Bilal, O.; Ali, S. Deep hybrid model for Mpox disease diagnosis from skin lesion images. Int. J. Imaging Syst. Technol. 2024, 34, e23044.

Waqas, M., Tahir, M. A., & Qureshi, R. (2023). Deep Gaussian mixture model based instance relevance estimation for multiple instance learning applications. Applied intelligence, 53(9), 10310-10325.

Lee, B., Kim, K., Kwon, T. G., & Lee, Y. (2010, April). Content classification of WAP traffic in Korean cellular networks. In Network Operations and Management Symposium Workshops (NOMS Wksps), 2010 IEEE/IFIP (pp. 22-27). IEEE.

Khan, S.U.R.; Zhao, M.; Asif, S.; Chen, X.; Zhu, Y. GLNET: Global–local CNN's-based informed model for detection of breast cancer categories from histopathological slides. J. Supercomput. 2023, 80, 7316–7348.

Hekmat, Arash, Zuping Zhang, Saif Ur Rehman Khan, Ifza Shad, and Omair Bilal. "An attention-fused architecture for brain tumor diagnosis." Biomedical Signal Processing and Control 101 (2025): 107221.

Waqas, M., Tahir, M. A., & Khan, S. A. (2023). Robust bag classification approach for multi-instance learning via subspace fuzzy clustering. Expert Systems with Applications, 214, 119113.

Khan, S.U.R.; Zhao, M.; Asif, S.; Chen, X. Hybrid-NET: A fusion of DenseNet169 and advanced machine learning classifiers for enhanced brain tumor diagnosis. Int. J. Imaging Syst. Technol. 2024, 34, e22975.

Khan, S.U.R.; Raza, A.;Waqas, M.; Zia, M.A.R. Efficient and Accurate Image Classification Via Spatial Pyramid Matching and SURF Sparse Coding. Lahore Garrison Univ. Res. J. Comput. Sci. Inf. Technol. 2023, 7, 10–23.

HUSSAIN, S., Raza, A., MEERAN, M. T., IJAZ, H. M., & JAMALI, S. (2020). Domain Ontology Based Similarity and Analysis in Higher Education. IEEEP New Horizons Journal, 102(1), 11-16.

Farooq, M.U.; Beg, M.O. Bigdata analysis of stack overflow for energy consumption of android framework. In Proceedings of the 2019 International Conference on Innovative Computing (ICIC), Lahore, Pakistan, 1–2 November 2019; pp. 1–9.

Shahzad, I., Khan, S. U. R., Waseem, A., Abideen, Z. U., & Liu, J. (2024). Enhancing ASD classification through hybrid attention-based learning of facial features. Signal, Image and Video Processing, 1-14.

Khan, S. R., Raza, A., Shahzad, I., & Ijaz, H. M. (2024). Deep transfer CNNs models performance evaluation using unbalanced histopathological breast cancer dataset. Lahore Garrison University Research Journal of Computer Science and Information Technology, 8(1).

Bilal, Omair, Asif Raza, and Ghazanfar Ali. "A Contemporary Secure Microservices Discovery Architecture with Service Tags for Smart City Infrastructures." VFAST Transactions on Software Engineering 12, no. 1 (2024): 79-92.

Waqas, M., & Khan, M. A. (2018). JSOPT: A framework for optimization of JavaScript on web browsers. Mehran University Research Journal of Engineering & Technology, 37(1), 95-104.

Khan, S. U. R., Asif, S., Zhao, M., Zou, W., Li, Y., & Li, X. (2025). Optimized deep learning model for comprehensive medical image analysis across multiple modalities. Neurocomputing, 619, 129182.

Khan, S. U. R., Asif, S., Zhao, M., Zou, W., & Li, Y. (2025). Optimize brain tumor multiclass classification with manta ray foraging and improved residual block techniques. Multimedia Systems, 31(1), 1-27.

Khan, S. U. R., Asim, M. N., Vollmer, S., & Dengel, A. (2025). AI-Driven Diabetic Retinopathy Diagnosis Enhancement through Image Processing and Salp Swarm Algorithm-Optimized Ensemble Network. arXiv preprint arXiv:2503.14209.

Khan, Z., Khan, S. U. R., Bilal, O., Raza, A., & Ali, G. (2025, February). Optimizing Cervical Lesion Detection Using Deep Learning with Particle Swarm Optimization. In 2025 6th International Conference on Advancements in Computational Sciences (ICACS) (pp. 1-7). IEEE.

Khan, S.U.R., Raza, A., Shahzad, I., Khan, S. (2025). Subcellular Structures Classification in Fluorescence Microscopic Images. In: Arif, M., Jaffar, A., Geman, O. (eds) Computing and Emerging Technologies. ICCET 2023. Communications in Computer and Information Science, vol 2056. Springer, Cham. https://doi.org/10.1007/978-3-031-77620-5_20

Asif Raza, Inzamam Shahzad, Ghazanfar Ali, and Muhammad Hanif Soomro. "Use Transfer Learning VGG16, Inception, and Reset50 to Classify IoT Challenge in Security Domain via Dataset Bench Mark." Journal of Innovative Computing and Emerging Technologies 5, no. 1 (2025).

Hekmat, A., Zuping, Z., Bilal, O., & Khan, S. U. R. (2025). Differential evolution-driven optimized ensemble network for brain tumor detection. International Journal of Machine Learning and Cybernetics, 1-26.

Shahzad, Inzamam, Asif Raza, and Muhammad Waqas. "Medical Image Retrieval using Hybrid Features and Advanced Computational Intelligence Techniques." Spectrum of engineering sciences 3, no. 1 (2025): 22-65.

Khan, S. U. R. (2025). Multi-level feature fusion network for kidney disease detection. Computers in Biology and Medicine, 191, 110214.

Khan, S. U. R., Asif, S., & Bilal, O. (2025). Ensemble Architecture of Vision Transformer and CNNs for Breast Cancer Tumor Detection From Mammograms. International Journal of Imaging Systems and Technology, 35(3), e70090.

Khan, S. U. R., & Khan, Z. (2025). Detection of Abnormal Cardiac Rhythms Using Feature Fusion Technique with Heart Sound Spectrograms. Journal of Bionic Engineering, 1-20.

Khan, M.A., Khan, S.U.R. & Lin, D. Shortening surgical time in high myopia treatment: a randomized controlled trial comparing non-OVD and OVD techniques in ICL implantation. BMC Ophthalmol 25, 303 (2025). https://doi.org/10.1186/s12886-025-04135-3.