**International Journal of AI, Big Data, Computational and Management Studies**
Noble Scholar Research Group | Volume 7, Issue 2, 65-69, 2026
ISSN: 3050-9416 | https://doi.org/10.63282/3050-9416.IJAIBDCMS-V7I2P111

*Original Article*

# QoS-aware routing in Software Defined Networks

Shimul Shah
Independent Researcher Philadelphia, United States.

***Received On: 18/02/2026*** ***Revised On: 27/03/2026*** ***Accepted On: 05/04/2026*** ***Published On: 12/04/2026***

***Abstract:*** *The growing demand for bandwidth-intensive applications such as video streaming, multimedia services, and Internet of Things (IoT) deployments has intensified the need for efficient resource management that secures network operations without compromising Quality of Service (QoS). Addressing these challenges requires a comprehensive and centralized view of network resources. Software Defined Networking (SDN), as an emerging paradigm, offers centralized control and programmability, allowing administrators to dynamically manage network behavior and optimize routing. This paper introduces a QoS-aware routing scheme for SDN that considers parameters such as available bandwidth, packet delay, and packet loss to determine the most efficient routing paths. Optimal paths are selected based on predefined threshold criteria to meet users' QoS expectations. Furthermore, we explore the integration of SDN with the Service-Oriented Architecture (SOA) model, emphasizing a user-centric approach to traffic engineering and Quality of Experience (QoE) enhancement. QoS guarantees remain essential for optimizing performance metrics and ensuring reliability, especially under constrained network conditions. Controllers dynamically identify optimal routes to meet clients' core QoS requirements, highlighting QoS-aware delivery as a cornerstone for next-generation network design.*

***Keywords****: Qos Optimization, Communication, Traffic Engineering, Quality Of Experience (Qoe), Quality Of Service (Qos), Service Level Agreements (SLA), Software-Defined Network (SDN), Application-Aware Routing, Service Oriented Architecture (SOA), Qsroute (Qos Aware Routing).*

## 1. Introduction

Quality of Service (QoS) management and traffic engineering play a vital role in communication systems based on the Service-Oriented Architecture (SOA) paradigm, as they directly influence user satisfaction and service provider efficiency. Modern SOA systems, which increasingly rely on the Internet as the communication backbone, face the persistent challenge of guaranteeing QoS in accordance with service response time requirements. Service response time depends not only on application server processing but also on network delays introduced by switches and routers. Therefore, fulfilling QoS objectives demands effective coordination between computing and communication resource management. Achieving end-to-end QoS in complex, heterogeneous environments require adaptive, scalable approaches capable of integrating diverse QoS mechanisms while minimizing operational costs. A key challenge lies in developing elastic resource management and traffic engineering solutions that dynamically estimate network conditions and optimally allocate resources based on input load. Software Defined Networking (SDN) has emerged as a promising technology to address these challenges by separating the control and data planes, enabling centralized network management and dynamic configuration without modifying device hardware or system software. This flexibility has prompted many network operators to adopt SDN for infrastructure management, though traditional traffic engineering algorithms and routing practices largely remain unchanged.

As SDN continues to evolve, its potential for enhanced QoS provisioning becomes increasingly significant. By leveraging programmability and centralized control, SDN can monitor real-time network conditions and make intelligent routing decisions based on QoS parameters such as bandwidth, latency, jitter, throughput, queue length, and packet loss. Conventional Internet routing, constrained by best-effort service delivery, often fails to ensure consistent QoS for demanding applications such as multimedia streaming, online gaming, and IoT systems. SDN's architecture provides a foundation for implementing dynamic and QoS-aware traffic engineering that ensures reliable and responsive service delivery. This paper proposes a QoS-aware routing scheme for SDN that dynamically determines the optimal source-to-destination path by considering available bandwidth, packet delay, and packet loss. The proposed approach aims to improve end-user experience by maintaining predefined QoS thresholds while optimizing overall network resource utilization.

## 2. QoS in Real Time Traffic

Real-time data transmission typically experiences higher packet loss rates and longer delays compared to non-real-time communication. Common examples include Voice over IP (VoIP), online gaming, video conferencing, and high-rate data exchanges in remote learning environments. To support these real-time applications, various audio and video codecs have been standardized, such as ITU-T's H.261, H.263, and H.264 for video compression, and IEEE's MPEG-2, MPEG-4, G.711, GSM, and G.723 for audio encoding. In real-time

communication, audio streams generally employ fixed packet sizes and maintain a constant bit rate to ensure synchronized delivery and minimal distortion.

Quality of Service (QoS) refers to the capability of network components to provide specific performance guarantees for certain data flows, ensuring reliability and efficiency in data delivery. A network must therefore meet defined performance criteria associated with the services or applications it supports. QoS can be quantified by several key metrics, including average packet loss, latency, jitter (delay variation), and throughput. These parameters enable networks to prioritize and optimize the transmission of different traffic types through mechanisms such as priority queuing, application-specific routing, bandwidth allocation, and traffic shaping. QoS functionality can be classified across two primary layers:Application Layer QoS: Managed at the software level, this layer primarily handles parameters such as jitter and delivery timing, optimizing end-user experience.Network Layer QoS: Implemented within network infrastructure, this layer controls performance aspects like bandwidth allocation, delay management, and packet scheduling to maintain efficient transmission. Delay represents the time taken for data to travel from source to destination, typically measured in milliseconds. It results from factors including propagation time, switching and scheduling delays, and processing overheads. Jitter reflects the variation in packet arrival time at the destination, often termed delay difference or delay variation, and is likewise measured in milliseconds. Minimizing both delay and jitter is essential for maintaining consistent performance in real-time applications.

## 3. Models for QoS Implementation

Quality of Service (QoS) mechanisms for real-time traffic should be supported across all routers and switches in the network to ensure efficient end-to-end service delivery. In current networking environments, several QoS techniques are commonly used, including IEEE 802.1p/802.1q, Integrated Services (IntServ), and Differentiated Services (DiffServ). For maximum effectiveness, QoS must be implemented at both the sender and receiver sides, with intermediate routers and switches enforcing QoS policies for real-time traffic flows. QoS mechanisms are typically implemented at the data link and network layers of the protocol stack. At Layer 2, QoS supports congestion handling and traffic prioritization, while Layer 3 QoS is primarily enforced through routers. Given the limitations of available network resources, the objective of the proposed model is to reduce packet delay and packet loss for real-time multimedia traffic. Traffic may be classified as either variable bit rate or constant bit rate; for example, video traffic is generally variable bit rate, whereas many Voice over IP (VoIP) streams are treated as constant bit rate traffic. Accordingly, traffic classification is performed at the network edge router based on the type of incoming service. Three broad QoS models are commonly recognized: best-effort service delivery, IntServ, and DiffServ. The best-effort model provides no guarantees regarding delay, throughput, or reliability, and typically relies on first-in, first-out (FIFO) queuing. In contrast, IntServ was designed to provide per-flow QoS guarantees through resource reservation and signaling, commonly using the Resource Reservation Protocol (RSVP). However, its requirement for per-flow state at every router limits its scalability in large networks. DiffServ was introduced to address these scalability issues by aggregating flows into classes and applying per-hop behavior rather than maintaining state for each individual flow. It uses packet markings such as the Differentiated Services Code Point (DSCP) to classify traffic and apply appropriate forwarding treatment.

**Table 1: Comparison of Service Delivery Models**

| Factors of Quality | BE Model | IntServ QoS Decisions | DiffServ Decisions |
|---|---|---|---|
| Isolation of Data | There is no isolation | Isolation on a per-flow basis | Per aggregation isolation |
| Guarantee of QoS | No guarantee | Per-flow basis | Per aggregation (Traffic Class) |
| Scope of Service | End-to-end | End-to-end | Per domain |
| Complexity Setup | No setup | Per-flow basis setup | Long term setup |
| Scalability of Model | Highly scalable | Not scalable (each router maintains per-flow state) | Scalable (edge routers maintain per aggregate state; core routers per class state) |
| Acceptable for Real-Time Traffic | No | Yes, resource allocation | Yes, LLQ |
| Traffic Control Admission | No | Deterministic based on flows | Statistic based on Traffic Classes |
| The Applications | Internet Default | Scenarios involving small networks and flow aggregation | Any size of the network |
| Reservation of Resource | Not available | Each node in the source-destination route has one flow | On each node in the domain, per Traffic Class |
| Model Complexity | Low | High | Medium |

Within DiffServ, Assured Forwarding (AF) and Expedited Forwarding (EF) are two important service classes. AF provides differentiated dropping behavior and supports multiple priority levels, while EF is intended for low-delay, low-jitter, and low-loss traffic such as voice. For audio traffic, EF is generally preferred, whereas video

conferencing is often supported through AF-based scheduling and queue management. In practice, IntServ and DiffServ are not necessarily competing approaches; rather, they are often complementary, with IntServ providing fine-grained admission control at the edge and DiffServ offering scalable QoS enforcement across the core network. A significant aspect of QoS evaluation is the measurement of end-to-end latency and jitter. Latency refers to the time required for a packet to travel from source to destination, including propagation, processing, scheduling, and switching delays. Jitter, or delay variation, represents the variation in packet inter-arrival times and is particularly critical for real-time applications such as voice and video.

## 4. Network Architecture of QoS-aware routing

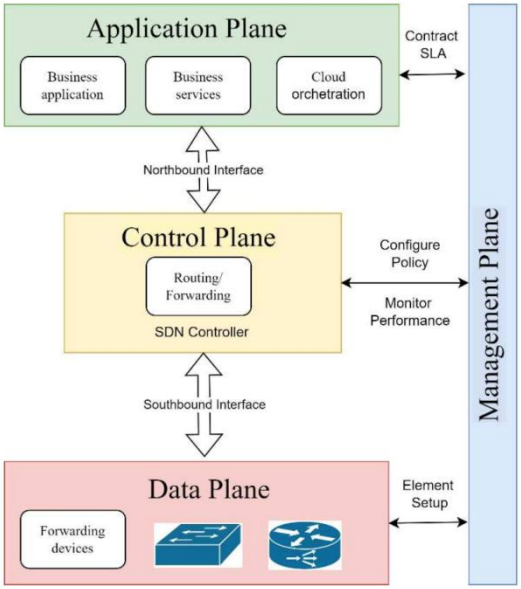


**Fig 1: Qos Models**

This section presents the overall network architecture of the proposed QSroute (QoS aware routing) framework. It begins with an overview of the SDN architecture, which forms the basis of QSroute, followed by a description of the QSroute architecture and its integration within SDN. The section concludes with a detailed discussion of the QSroute module.

### *4.1. SDN architecture*

The QSroute network architecture leverages the centralized control of Software Defined Networking (SDN) to deploy multiple programmable functional modules within the SDN controller. Unlike traditional networks—where switches and routers are aware only of their immediate neighbors—SDN maintains a global view, connecting the controller to every network device. This centralized design enhances flexibility, scalability, and overall manageability. As shown in Figure 1, SDN separates the control and data planes. The control plane, managed by a centralized controller such as RYU, ONOS, or Floodlight, determines how network traffic is forwarded and communicates with devices via the southbound interface, typically using the OpenFlow protocol. This arrangement provides network-wide visibility and intelligence, enabling administrators to define and enforce routing policies, Quality of Service (QoS), and security requirements. The data plane, in turn, handles the actual forwarding of packets based on instructions from the controller, allowing for dynamic and adaptive traffic management. The southbound and northbound interfaces are key components of SDN architecture. The southbound interface facilitates command and state exchange between the controller and network devices, ensuring real-time programmability. The northbound interface links the controller to management and application planes through APIs that translate high-level policies into enforceable network actions.

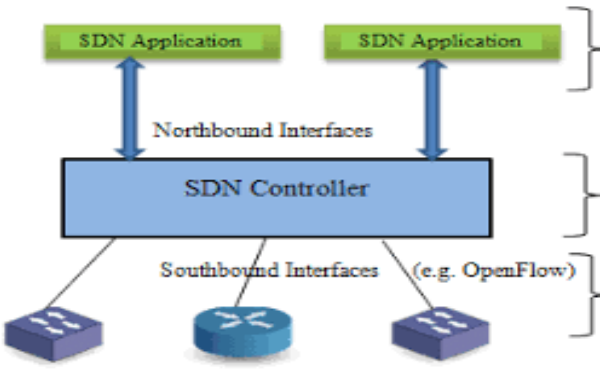


**Fig 1: SDN Architecture**

### *4.2. QSroute architecture in SDN*

The QSroute architecture enhances the capabilities of the SDN controller by integrating a Network Information Module composed of two sub-components: the Topology Statistics Module and the QSroute Module. The Topology Statistics Module is responsible for gathering real-time network data from devices in the data plane, such as link bandwidth, packet loss, delay, throughput, and utilization.

The QSroute Module, in turn, computes optimal routing paths based on three key Quality of Service (QoS) metrics. Figure 2 illustrates the overall QSroute network architecture and its operational process. First, the SDN controller periodically collects network status information—such as bandwidth availability, latency, and packet loss—from all switches through the southbound interface. This data is then stored within the Network Information Module, where the Topology Statistics Module maintains statistical information for each switch. The QSroute Module accesses this centralized dataset to perform optimal path calculations. Using the topology statistics, the QoS component determines the best route between each source–destination pair, as explained in the following section. Once the optimal paths are computed, the results are forwarded to the Flow Installation Module, which identifies the switches requiring updates. Finally, the controller creates and installs corresponding flow entries through the southbound interface, ensuring QoS-based path enforcement across the data plane.

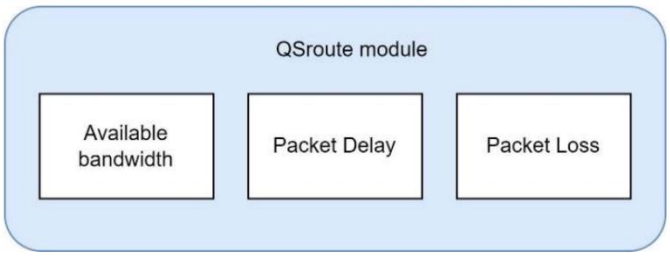


**Fig 2: Qos Aware Routing Architecture**

### *4.3. QSroute module*

The QSroute module determines optimal routing paths based on predefined Quality of Service (QoS) threshold parameters. Three QoS metrics are utilized: available bandwidth, packet delay, and packet loss. The available bandwidth threshold specifies the minimum acceptable bandwidth required to accommodate the anticipated traffic load without inducing congestion or performance degradation. Routing paths exceeding this threshold are prioritized, as they provide the necessary capacity for efficient data transmission. The packet delay threshold defines the maximum permissible end-to-end delay within the network, ensuring that latency remains within acceptable limits for time-sensitive and real-time applications. Paths meeting this criterion are preferred to maintain responsiveness and service quality. The packet loss threshold indicates the maximum tolerable loss rate of transmitted packets and serves as an indicator of network reliability. Routes maintaining packet loss within the specified threshold are considered stable and suitable for consistent data delivery, whereas paths exceeding the limit may cause degraded performance and reduced transmission integrity. The QSroute module aims at addressing the QoS-aware routing optimization to find a path between two nodes while guaranteeing multiple QoS requirements. Therefore, the objective function of the problem for routing is finding a route $p(v_{1,} v_n)$ the source node $v_1$ to destination node $v_n$ based on the threshold set for each of the QoS metrics: available bandwidth $B(p(v_{1,} v_n)$ , packet delay $D(p(v_{1,} v_n)$ and packet loss $L(p(v_{1,} v_n)$.

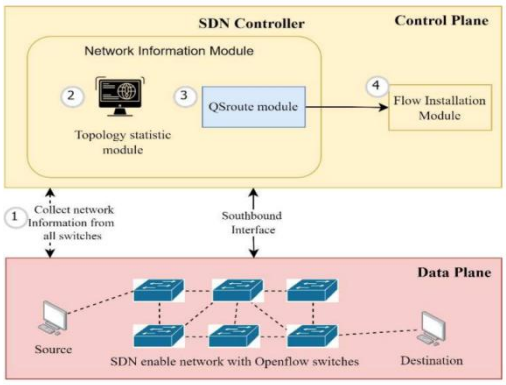


**Fig 3: QoS aware routing module components**

$$\begin{cases} B(p(v_1, v_n) \geq B_{min} \\ D(p(v_1, v_n) \leq D_{max} \\ L(p(v_1, v_n) \leq L_{max} \end{cases}$$

In the above formula, $B_{min}$ is the minimum available bandwidth require by applications in the route $p(v_{1,} v_n)$.$D_{max}$ and $L_{max}$ denote the maximum packet delay and packet loss rate which the applications can still tolerate before the degradation of QoS.

## 5. Conclusion

In conclusion, this study has presented the QSroute framework, a QoS-aware routing approach for Software Defined Networking (SDN) that utilizes three principal Quality of Service (QoS) metrics—available bandwidth, packet delay, and packet loss to determine optimal routing paths. The proposed scheme enhances the performance, reliability, and adaptability of SDN environments by enabling more efficient network resource allocation and improved traffic management for diverse application requirements. Despite its conceptual nature, several critical challenges have been identified for future investigation,

including scalability, performance optimization, and dynamic load balancing in large-scale SDN infrastructures. Furthermore, the inherent trade-offs among multiple QoS metrics introduce additional complexity in achieving balanced routing decisions. Addressing these limitations will be essential for the practical realization and refinement of the QSroute framework, ultimately contributing to the advancement of QoS provisioning and intelligent routing mechanisms within programmable network architectures.

## References


1. C. Lin, K. Wang, and G. Deng, "A QoS-aware routing in SDN hybrid networks," *Procedia Computer Science*, vol. 110, pp. 242–249, 2017.
2. Z. Li and P. Mohapatra, "QRON: QoS-aware routing in overlay networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 1, pp. 29–40, 2004.
3. L. Chen and W. B. Heinzelman, "QoS-aware routing based on bandwidth estimation for mobile ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 3, pp. 561–572, 2005.
4. M. Bagaa et al., "On SDN-driven network optimization and QoS-aware routing using multiple paths," *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4700–4714, 2020.
5. B. Nazir and H. Hasbullah, "Energy efficient and QoS aware routing protocol for clustered wireless sensor network," *Computers & Electrical Engineering*, vol. 39, no. 8, pp. 2425–2441, 2013.
6. K. Z. Ghafoor et al., "Quality of service aware routing protocol in software-defined internet of vehicles," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2817–2828, 2018.
7. K. Akkaya and M. Younis, "Energy and QoS aware routing in wireless sensor networks," *Cluster Computing*, vol. 8, no. 2, pp. 179–188, 2005.