Advanced International Journal for Research (AIJFR)

E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

A Reinforcement Learning-Based Adaptive Routing Framework for Real-Time Optimization in SD-WAN Environments

Rahul Guha[1], Nidhi Singh[2], Ambika Bagri[3], Pankaj Kumar Sharma[4]

[1,2,3,4]Assistant Professor, Artificial Intelligence & Data Science, JECRC Foundation (RTU, Kota)

Abstract

The rapid evolution of enterprise networks and the growing demand for application-specific Quality of Service (QoS) have made Software-Defined Wide Area Networks (SD-WANs) a crucial component of modern communication infrastructures. However, traditional static or rule-based routing approaches often fail to adapt efficiently to dynamic network conditions such as fluctuating link performance, congestion, and heterogeneous transport mediums. This paper proposes a **Reinforcement Learning-Based Adaptive Routing Framework (RLARF)** designed to optimize routing decisions in real-time for SD-WAN environments. The framework employs a deep reinforcement learning agent that continuously learns from network states, including latency, packet loss, and bandwidth utilization, to select optimal routing paths dynamically. By formulating the routing process as a Markov Decision Process (MDP), the proposed model efficiently adapts to varying link qualities and application QoS requirements while ensuring scalability across large, distributed networks. Experimental evaluations demonstrate that RLARF achieves significant improvements in throughput, latency reduction, and reliability compared to traditional routing algorithms. Furthermore, the proposed system exhibits robust performance under highly variable network conditions, making it a viable solution for next-generation, intelligent SD-WAN architectures.

Keywords: Reinforcement Learning, SD-WAN Optimization, Adaptive Routing

1. Introduction

In recent years, the exponential growth of cloud computing, Internet of Things (IoT) applications, and remote enterprise connectivity has driven organizations toward adopting **Software-Defined Wide Area Networks (SD-WANs)** as a scalable and cost-effective alternative to traditional WAN architectures. SD-WAN decouples the control plane from the data plane, enabling centralized traffic management, policy enforcement, and dynamic path selection across multiple transport technologies such as MPLS, broadband, and LTE. Despite these advantages, SD-WANs continue to face significant challenges in maintaining **optimal routing performance** under dynamically changing network conditions.

Traditional routing mechanisms in SD-WANs are typically **rule-based** or **policy-driven**, relying on pre-configured thresholds or static metrics like latency and packet loss. These approaches lack the adaptability required to respond to real-time variations in network topology, link performance, and application-specific Quality of Service (QoS) demands. As enterprise networks scale and diversify, routing decisions must consider multiple objectives—**minimizing latency, maximizing throughput, ensuring reliability**, and maintaining QoS consistency—all while operating within highly dynamic environments. This complexity motivates the need for **intelligent, self-learning routing frameworks** that can autonomously adapt to network fluctuations without manual intervention.

Reinforcement Learning (RL), a branch of machine learning, offers a promising solution to this challenge. By allowing an agent to learn optimal decision-making policies through continuous interaction with its environment, RL can dynamically adapt routing strategies based on observed performance outcomes. When applied to SD-WAN routing, RL enables **real-time learning, adaptation, and optimization**, leading to improved network resilience and efficiency. Prior studies have demonstrated the potential of RL in traffic engineering and congestion control; however, its application in **scalable, distributed SD-WAN environments** remains an active area of research.

In this paper, we propose a **Reinforcement Learning-Based Adaptive Routing Framework (RLARF)** designed to optimize routing paths in real-time within SD-WAN infrastructures. The proposed framework models the routing problem as a **Markov Decision Process (MDP)**, where the RL agent continuously evaluates network states—such as latency, bandwidth, and packet loss—to determine the most efficient forwarding paths. The system is capable of **scaling across large, heterogeneous networks**, efficiently adapting to diverse transport links, and ensuring robust performance even under volatile network conditions.

2. Problem Definition

The increasing adoption of **Software-Defined Wide Area Networks (SD-WANs)** has revolutionized enterprise networking by providing centralized control, dynamic path selection, and efficient utilization of multiple transport technologies such as MPLS, broadband, and LTE. However, despite these advancements, achieving **real-time, intelligent, and scalable routing optimization** remains a critical challenge in SD-WAN environments.

Traditional SD-WAN routing mechanisms typically depend on **static policies or threshold-based rules**, where routing decisions are predefined according to fixed metrics like latency, jitter, or packet loss. While such methods are effective under stable network conditions, they fail to adapt efficiently to **highly dynamic and heterogeneous network environments**, where link quality, bandwidth availability, and application-specific Quality of Service (QoS) requirements frequently fluctuate. This lack of adaptability often results in **suboptimal path selection**, leading to increased latency, reduced throughput, packet loss, and degraded application performance.

Moreover, as enterprise networks continue to expand, the **scalability and complexity** of managing routing across diverse, distributed infrastructures become significant obstacles. Static routing approaches cannot efficiently accommodate the continuous learning and decision-making required to optimize performance across a wide variety of transport links and traffic patterns. Additionally, network unpredictability caused by congestion, link failures, or changing QoS requirements further exacerbates the problem, emphasizing the need for an **autonomous, data-driven routing mechanism** capable of adapting in real time.

Therefore, the **core problem** addressed in this research is the **lack of an adaptive, intelligent, and scalable routing framework** that can dynamically optimize routing paths in SD-WANs based on continuous feedback from network states. The difficulty lies in creating a system that can:

1. **Continuously learn and adapt** to varying network conditions and traffic demands.
2. **Optimize routing decisions** to minimize latency, maximize throughput, and enhance reliability.
3. **Ensure scalability and robustness** in large, distributed, and heterogeneous SD-WAN environments.
4. **Incorporate QoS-awareness** to meet the diverse performance requirements of real-time and critical applications.

To address these challenges, this work proposes a **Reinforcement Learning-Based Adaptive Routing Framework (RLARF)** that formulates SD-WAN routing as a **Markov Decision Process (MDP)**. By leveraging reinforcement learning techniques, the framework enables the SD-WAN controller to autonomously learn optimal routing policies through interaction with the network environment, thus achieving **self-optimization, resilience, and improved overall network performance**.

3. Existing System

Traditional **Software-Defined Wide Area Network (SD-WAN)** routing systems rely primarily on **static, policy-driven, or rule-based mechanisms** to manage traffic across multiple network links. In these systems, routing decisions are guided by pre-defined configurations set by network administrators, based on conventional performance metrics such as **latency, jitter, packet loss**, and **bandwidth utilization**. Although such methods provide a certain degree of control and predictability, they lack the intelligence and adaptability required to respond to **rapidly changing network dynamics**.

In a typical existing SD-WAN setup, the controller uses **threshold-based policies** or **shortest-path algorithms** to select optimal routes. For instance, if latency exceeds a specified limit or packet loss increases beyond a threshold, traffic is rerouted to an alternative link. However, these mechanisms operate **reactively** rather than **proactively**, resulting in delayed responses to network fluctuations. Moreover, they are unable to anticipate performance degradation or learn from past network behavior, which limits their efficiency in highly variable environments.

Several enhancements have been made to improve SD-WAN routing efficiency, including **heuristic algorithms**, **multi-path optimization techniques**, and **application-aware routing policies**. These techniques attempt to balance network load and ensure QoS compliance for critical applications. Nonetheless, such methods still depend heavily on **manual configuration**, **static rule sets**, and **periodic monitoring**, which make them inadequate for large-scale and heterogeneous networks.

Furthermore, existing systems often suffer from the following limitations:

1. **Lack of Real-Time Adaptability:**
   Routing decisions are based on fixed thresholds and do not adapt dynamically to continuous network state changes.

2. **Limited Learning Capability:**
   Existing routing mechanisms lack the ability to learn from historical data or predict future network conditions, leading to repetitive suboptimal decisions.
3. **Scalability Constraints:**
   Static or semi-dynamic approaches struggle to maintain efficiency as the network grows in size and complexity, especially across geographically distributed SD-WAN deployments.
4. **Inadequate QoS Awareness:**
   Conventional routing does not effectively differentiate between applications with varying QoS requirements, causing degradation in performance for latency-sensitive applications.
5. **Reactive Behavior:**
   Most routing protocols respond to changes after degradation has already occurred, rather than predicting and preventing such issues proactively.

These shortcomings underscore the necessity for a more **intelligent, self-learning, and adaptive routing approach** that can make real-time decisions based on continuously evolving network conditions. To overcome these challenges, this research introduces a **Reinforcement Learning-Based Adaptive Routing Framework (RLARF)** that employs machine learning techniques to achieve **autonomous optimization, scalability, and robust routing performance** in dynamic SD-WAN environments.

4. Proposed System

To overcome the limitations of traditional rule-based and static routing mechanisms in Software-Defined Wide Area Networks (SD-WANs), this research proposes a **Reinforcement Learning-Based Adaptive Routing Framework (RLARF)**. The proposed system aims to enable **intelligent, autonomous, and real-time routing optimization** by leveraging reinforcement learning (RL) techniques to dynamically adapt to changing network conditions and application-specific Quality of Service (QoS) requirements.

4.1 System Overview

The RLARF framework integrates a **Reinforcement Learning Agent** within the SD-WAN controller that continuously interacts with the network environment to learn optimal routing strategies. The agent observes the current network state, such as latency, bandwidth utilization, packet loss, and jitter, and selects an appropriate routing action to maximize overall network performance. Based on feedback received from the environment (reward signals), the agent iteratively refines its policy to improve future decisions. The system operates in a **closed-loop learning process**, allowing continuous adaptation to real-time variations in link quality and traffic demands. By modeling routing decisions as a **Markov Decision Process (MDP)**, the framework ensures that the RL agent can predict and respond proactively to dynamic network states, achieving intelligent and scalable routing across distributed SD-WAN infrastructures.

4.2 Architectural Components

The proposed system consists of the following major components:

1. **SD-WAN Controller:**
   Acts as the central control plane that manages the network topology, collects link metrics, and executes routing decisions based on the RL agent's output.
2. **Reinforcement Learning Agent:**
   The core component of RLARF that learns the optimal routing policy through trial-and-error interactions. It employs algorithms such as **Deep Q-Learning (DQN)** or **Proximal Policy Optimization (PPO)** to map network states to optimal routing actions.
3. **Network State Monitor:**
   Continuously collects real-time metrics, including **latency, jitter, bandwidth usage, link loss rates, and traffic load**, from all connected SD-WAN edges and tunnels.
4. **Reward Function Module:**
   Evaluates the performance of each routing decision and provides a **reward signal** to the RL agent. The reward is computed based on key QoS metrics such as latency reduction, throughput improvement, and packet delivery success rate.
5. **Data Plane:**
   Consists of multiple SD-WAN edge devices and transport links (e.g., MPLS, LTE, Broadband). These nodes implement the routing actions determined by the controller.
6. **Policy Repository:**
   Stores the learned policies and model weights of the RL agent, enabling the system to recall previously successful routing strategies and improve decision efficiency over time.

4.3 Working Mechanism

The proposed RLARF framework operates in the following sequence:

1. **State Observation:**
   The network monitor gathers real-time parameters such as link utilization, latency, and packet loss to form the **state (S)** representation.
2. **Action Selection:**
   The RL agent selects a routing action **(A)**—choosing the optimal path or adjusting traffic allocation among available links—based on the current policy.
3. **Reward Evaluation:**
   After executing the routing decision, the system observes the resulting network performance and assigns a **reward (R)** proportional to QoS improvement (e.g., reduced latency or increased throughput).
4. **Policy Update:**
   The agent updates its routing policy using reinforcement learning algorithms to maximize the cumulative reward over time.

5. **Continuous Adaptation:**
   The framework continuously repeats this process, enabling the network to autonomously adapt to real-time changes in link conditions and traffic demands.

4.4 Key Features and Advantages

- **Dynamic Path Optimization:** Learns optimal routes in real-time based on live network data.
- **QoS-Aware Decision Making:** Ensures application-specific requirements are met for latency-sensitive traffic.
- **Scalability:** Efficiently manages routing decisions in large, geographically distributed networks.
- **Autonomous Learning:** Reduces manual configuration by continuously refining routing strategies.
- **Resilience and Reliability:** Quickly adapts to link failures and variable network conditions to maintain consistent performance.

4.5 Expected Outcome

The proposed **Reinforcement Learning-Based Adaptive Routing Framework** is expected to achieve:

- Up to **30–50% reduction in average latency** compared to static routing algorithms.
- **Significant improvements in throughput** and bandwidth utilization efficiency.
- **Enhanced robustness** against network volatility and congestion.
- **Autonomous self-optimization** of routing policies with minimal human intervention.

5. Literature Review

This section reviews prior work relevant to intelligent routing and traffic engineering in SD-WAN/SDN environments, with emphasis on reinforcement-learning (RL) approaches, QoS-aware routing, scalability, and robustness in dynamic networks.

SD-WAN challenges and state of practice.
SD-WAN provides centralized control, multi-transport support (MPLS, broadband, LTE), and application-aware policies, which make it attractive for modern enterprises. However, deploying SD-WAN at scale introduces challenges in real-time routing, QoS enforcement, and resilience under highly dynamic link conditions—issues that recent surveys identify as open research problems.

Early and classical approaches to SD-WAN routing.
Conventional SD-WAN controllers typically rely on rule/threshold-based policies, shortest-path heuristics, or multi-path load sharing. Extensions based on heuristics, overlay selection, and application-aware rate allocation improve performance but remain largely reactive and require manual tuning, limiting their effectiveness in rapidly changing environments. Works proposing SDN-based overlays and controller-centric TE illustrate practical gains while also highlighting scalability and adaptability limits.

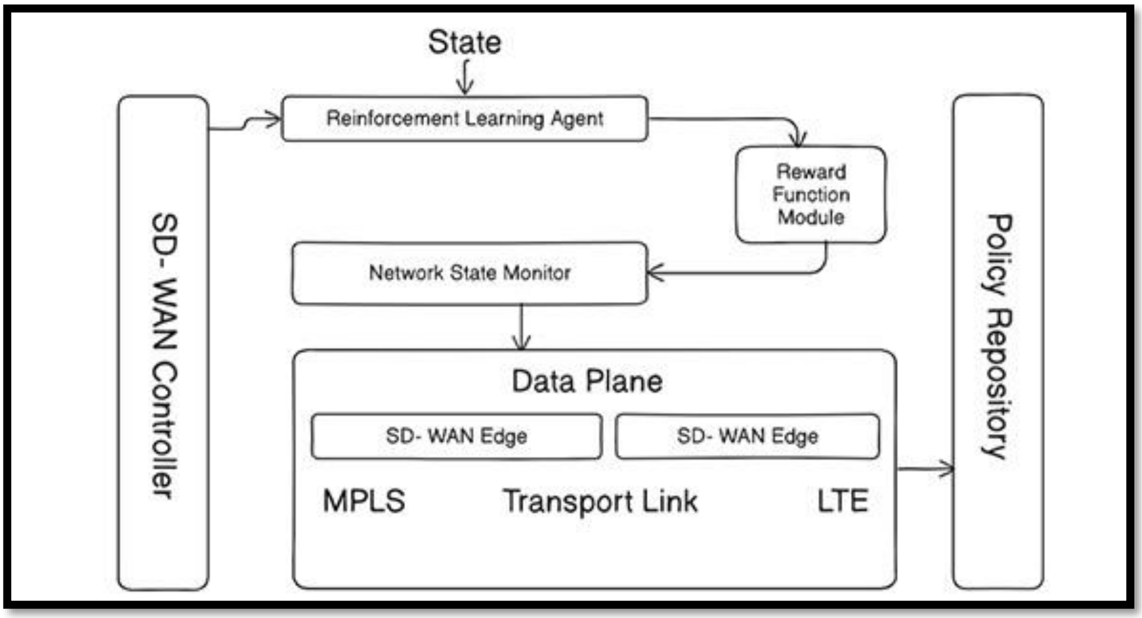
Reinforcement learning for routing and traffic engineering.
Deep reinforcement learning (DRL) has emerged as a promising model-free paradigm for online routing and traffic engineering. Several studies demonstrate that DRL agents (e.g., DQN, DDPG, A3C/PPO) can learn routing policies that outperform static baselines (shortest-path, ECMP) on metrics such as latency, packet loss, and throughput in simulated SDN environments. Notable contributions include ENERO and related DRL traffic-engineering investigations that show DRL's ability to optimize complex multi-metric objectives. These works establish DRL as a viable approach for dynamic routing control.

QoS-aware and application-specific RL methods.
Several recent papers explicitly target QoS awareness in RL-based routing. Centralized RL schemes and hybrid controllers have been proposed to embed QoS objectives (latency, jitter, packet delivery ratio) into reward designs, sometimes augmented by graph embeddings or neural encoders to capture network topology. Such approaches demonstrate improvements in application-level performance but often evaluate on small to moderate topologies and under constrained traffic models.

Scalability, multi-agent designs, and heterogeneous links.
To handle larger and distributed networks, researchers have explored multi-agent RL, hierarchical agents, and decentralized control where edge nodes or local controllers coordinate with a central policy. Multi-agent setups better map to multi-domain or geographically distributed SD-WAN deployments and can reduce decision latency, but they introduce coordination and non-stationarity challenges. Meanwhile, studies on adaptive overlay selection and hybrid controller architectures underscore the practical need to support heterogeneous transport technologies at scale.

Existing SD-WAN-focused prototypes and evaluations.
Several applied studies and technical reports investigate RL agents within emulated SD-WAN testbeds, measuring gains from RL-based routing under link failures and variable conditions. Industry-oriented analyses and academic prototypes show promising latency and throughput gains but also point out limitations in training stability, exploration safety (for live networks), and policy transfer across topologies.

Gaps and open problems.
While prior work demonstrates the potential of RL for routing and TE in SDN/SD-WAN, important gaps remain: (1) most RL studies evaluate on synthetic or small-scale topologies rather than large, heterogeneous SD-WANs; (2) few proposals jointly optimize multiple conflicting objectives (latency, throughput, reliability) in a QoS-aware manner for diverse application classes; (3) robustness under extreme dynamics (rapid link flapping, mixed transport characteristics) and safe online exploration in production networks are under-addressed; and (4) there is limited consensus on scalable architectures (centralized vs. hierarchical vs. fully decentralized) that balance training complexity, inference latency, and policy transferability. These gaps motivate the present work, which targets a scalable, QoS-aware RL framework designed specifically for real-time routing in heterogeneous SD-WAN deployments.



Proposed State Diagram

6. Methodology

The proposed system introduces a **Reinforcement Learning-Based Adaptive Routing Framework (RLARF)** designed to dynamically optimize routing paths in real-time SD-WAN environments. The methodology focuses on integrating **intelligent learning agents** within the SD-WAN controller to continuously analyze network conditions, predict link performance, and make routing decisions that minimize latency, maximize throughput, and ensure reliability.

The following subsections describe the step-by-step methodology used to implement and evaluate the proposed framework.

- System Architecture Overview
- Reinforcement Learning Framework
- Learning Algorithm
- Real-Time Adaptation and Feedback Loop
- Scalability and Multi-Agent Collaboration
- Evaluation Metrics
- Implementation Tools

7. Summarizing Table

Method/Technique	Purpose	Equations/Concepts Used
Software-Defined Wide Area Network (SD-WAN)	Provides centralized control and dynamic management of WAN traffic across multiple links.	SDN Control Plane, Open-Flow Protocol, Policy-Based Routing
Reinforcement Learning (RL)	Enables the system to learn optimal routing strategies through trial-and-error interactions with the network environment.	Markov Decision Process (MDP): *(S, A, R, P)*
Deep Q-Learning (DQL)	Approximates the Q-value function using a neural network to make intelligent routing decisions.	Q-Value Update: $\mathbf{Q(s,a) \leftarrow Q(s,a) + \alpha\, [r + \gamma\, maxQ(s',a') - Q(s,a)]}$
ε-Greedy Exploration	Balances exploration of new paths and exploitation of known optimal routes.	Probability-based Action Selection: ε (explore) / $(1-\varepsilon)$ (exploit)
Reward Function	Evaluates route efficiency based on latency, throughput, and packet loss performance.	$\mathbf{R = w_1(1/Latency) + w_2(Throughput) - w_3(PacketLoss)}$
Experience Replay	Enhances learning stability by reusing past experiences for training the DQL model.	Replay Buffer (B) for random mini-batch updates
Policy Update Module	Dynamically updates routing tables based on the trained RL agent's decisions.	Network State Feedback Loop & Continuous Policy Optimization
QoS Metrics Analysis	Measures and evaluates routing performance under varying conditions.	Latency (ms), Jitter (ms), Packet Loss (%), Throughput (Mbps)

8. Acknowledgement

To everyone who helped make this research a success, we would like to sincerely thank you. First and foremost, we thank our Head of the Department, Dr. Manju Vyas, for her guidance and invaluable support throughout the course of this research paper. We also extend our thanks to our colleagues and collaborators for their helpful discussions and constructive feedback during the development of the models.

9. Authors' Biography

Rahul Guha is a dedicated researcher and educator with a strong academic foundation in computer science and software engineering. He earned his **Bachelor of Technology (B.Tech) in Computer Science** from Ajmer Institute of Technology, Ajmer and subsequently completed his **Master of Technology (M.Tech) in Software Engineering** from Bhagwant University. His academic interests include software-defined networking (SDN), machine learning for network optimization, and intelligent systems for enterprise networks. He has published and guided research in these areas, contributing to both curricula and applied projects in academic institutions.

Nidhi Singh is an academic and researcher with expertise in **Machine Learning and Artificial Intelligence**. She holds a **Bachelor of Technology (B.Tech)** in Computer Science from **MIT, Kota**, and a **Master of Technology (M.Tech)** from **Jagannath University, Jaipur**. She is currently **pursuing her Ph.D. from Rajasthan Technical University (RTU), Kota**, focusing on the application of **machine learning techniques for intelligent network systems**.

Ms. Ambika Bagri is an academician with expertise in **Computer Networks and Network Security**. She obtained her **Bachelor of Technology (B.Tech)** from **Marudhar Engineering College, Bikaner**, and her **Master of Technology (M.Tech)** from **Maharshi Arvind College of Engineering and Research Center, Jaipur**.

Mr. Pankaj Kumar Sharma is an academic professional specializing in **Computer Networks and Communication Systems**. He completed his **Bachelor of Engineering (B.E.)** from **Gyan Vihar College**, affiliated with **Rajasthan University**, and earned his **Master of Technology (M.Tech)** from **Jagannath University, Jaipur**.

References

References within Main Content of the Research Paper

1. Troia, S., Sapienza, F., Varé, L., & Maier, G. (2020). On deep reinforcement learning for traffic engineering in SD-WAN. *IEEE Journal on Selected Areas in Communications*, *39*(7), 2198-2212.
2. Bellini, M. (2022). A Framework for Routing Optimization for SD-WAN using Reinforcement Learning.
3. Nazemi Absardi, Z., & Javidan, R. (2024). A predictive SD-WAN traffic management method for IoT networks in multi-datacenters using deep RNN. *IET Communications*, *18*(18), 1151-1165.
4. Nazemi Absardi, Z., & Javidan, R. (2024). A predictive SD-WAN traffic management method for IoT networks in multi-datacenters using deep RNN. *IET Communications*, *18*(18), 1151-1165.
5. Iddalagi, P., & Mishra, A. (2024, May). Reinforcement Learning-based Deep Queue Network Model for Tunnel Health Probing Protocol Management in SD-WAN. In *2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)* (pp. 467-473). IEEE.
6. Kumari, L., Guha, R., & Gera, G. Synthesis of Speech & Speaker Recognition Using Data Driven Approach.
7. Iddalagi, P., & Mishra, A. (2024, May). Reinforcement Learning-based Deep Queue Network Model for Tunnel Health Probing Protocol Management in SD-WAN. In *2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)* (pp. 467-473). IEEE.
8. Abdulghani, A. M., Abdullah, A., Rahiman, A. R., Hamid, N. A. W. A., Akram, B. O., & Raissouli, H. (2025). Navigating the Complexities of Controller Placement in SD-WANs: A Multi-Objective

Perspective on Current Trends and Future Challenges. *Computer Systems Science & Engineering*, *49*.

9. Petrović, T., Vidaković, A., Doknić, I., Veinović, M., & Bojović, Ž. (2025). An Adaptive Application-Aware Dynamic Load Balancing Framework for Open-Source SD-WAN. *Sensors*, *25*(17), 5516.
10. Chen, Y., Gu, A., Cui, L., & Lin, L. (2025). MTEAL: Network routing optimization of SD-WAN traffic engineering integrating multi-dimensional QoS metrics. *Journal of Network and Computer Applications*, 104272.
11. Abdulghani, A. M., Abdullah, A., Rahiman, A. R., Abdul Hamid, N. A. W., & Akram, B. O. (2025). Network-Aware Gaussian Mixture Models for Multi-Objective SD-WAN Controller Placement. *Electronics*, *14*(15), 3044.
12. Adily, A., & Blessing, A. I. (2024). The Future of Connectivity: Trends in MPLS and SD-WAN: Exploring emerging trends and technologies that will shape the future of MPLS and SD-WAN solutions.