International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

INTELLIGENT NETWORK SLICING FOR SCALABLE AND RESILIENT 6G VEHICULAR COMMUNICATIONS

[1*]Dr. Shaik Mahammad Rasool, [2]Ms. Syeda Amena Bano, [3]Ms. Salma Naazneen, [4]Mrs.Madhavuni Sandhya Rani, [5]Zainab Unnisa

[1] Associate professor [2] Assistant professor [3] Assistant professor [4] Assistant professor [5] Assistant professor at Department of ECE, Lords Institute of Engineering and Technology(A), Hyderabad.

Corresponding Author*: skmohammedrasool@lords.ac.in

Abstract

The advent of 6G networks provides new possibilities to facilitate autonomous vehicle (AV) ecosystems with the requirement of ultra-reliable, low-latency, and high-throughput communication. The paper suggests a framework of Dynamic QoS-Aware Network Slicing (DQNS) that combines the use of the Long Short-Term Memory (LSTM)-based traffic prediction and Deep Reinforcement Learning (DRL)-based resource allocation to meet the dynamic and heterogeneous needs of various types of services, such as URLLC, eMBB and mMTC slices. The framework uses closed-loop orchestration process in the MEC-core coordination layer to predictive load estimation to support multi-objective optimization to achieve strict compliance with SLA. Results of the simulations indicate that the proposed DQNS technique decreases URLLC latency by more than 51 percent, augments the ratio of packet delivery to 99.3 percent, enhances the eMBB throughput by 29 percent, and efficiency in utilizing resources to 91.6 percent compared to the basic static and reactive slicing techniques. The suggested solution offers an intelligent, AI-powered orchestration approach, which makes it a promising enabler of intelligent transportation systems in the future in 6G-enabled smart mobility networks.

Keywords: 6G, Autonomous Vehicles, Dynamic QoS-Aware Network Slicing, Multi-Objective Optimization, LSTM Traffic Prediction.

I. INTRODUCTION

The sixth-generation (6G) mobile network version has a promising future entity of providing ultra-reliable, low-latency, and high-capacity connection so as to support next-generation usage like autonomous vehicle (AV) ecosystems. Such ecosystems combine interconnected and autonomous vehicles, on-road infrastructure, and cloud-edge computing systems, with a high-quality of Service (QoS) assurances needed to enable various and mission-critical tasks, such as real-time navigation, cooperative perception, and remote control. The heterogeneous nature of AV communication workloads, however, i.e. between ultra-reliable low-latency communications (URLLC) to support safety-critical control and limited enhancements to the mobile broadband (eMBB) to perform infotainment, is a major setback when considering how to allocate network resources.

Received: August 06, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
Network slicing is a promising paradigm of fulfilling these heterogeneous requirements, dividing the physical 6G infrastructure into several service-specific slices virtualized. Nevertheless, it can be argued that the one-time/fixed-cut plan cannot respond because of dynamic AV environments, where traffic volume, movement and environmental variables change within a short period of time. This necessitates the dynamic emergence of the QoS sensitive network slicing methods that can dynamically and smartly change the network slice settings to maintain the reliability of services and minimize the latency and ensure the throughput.

Recent advances of orchestration with the use of artificial intelligence (AI), multi-access edge computing (MEC), and network function virtualization (NFV) made the prospect of adaptive slicing frameworks, which would be capable of learning and predicting traffic patterns through resource makeover and consequently follow service-level agreements (SLAs) in a broad spectrum of AV services. In that sense, the predictive control models founded on the reinforcement learning (RL) and deep learning may offer the opportunity to represent the parameters optimization of the slice configurations basing on the previous experiences (at any level of the feedback and environment).

One of the proposed frameworks in this paper aims at 6G Autonomous Vehicle Ecosystems based on a Dynamic QoS-Aware Network Slicing Framework that dynamically allocates the resource according to the needs of the slices using AI based decision-making and MEC based orchestration. The framework focuses on prioritising URLLC to safety-critical services and prioritising the requirements of eMBB and massive machine-type communication (mMTC) services. The proposed system will ostensibly achieve improvements in reliability, latency, and throughput, and can therefore be easily integrated into intelligent transportation systems (ITS) and lead to safer and more efficient transportation in terms of an urban mobility environment.

II. LITERATURE SURVEY

Network slicing of 5G and beyond network has received wide study as a solution to cover a variety of QoS requirements in the case of vehicle communication. Initial research into 5G NR slicing has illustrated how it can be used to support multiple service types over a mobile network - e.g., URLLC, eMBB, and mMTC - at the same time, but suggested that static allocation schemes alone would not be effective in high-mobility use-cases like in autonomous vehicles [1].

Research work on dynamic resource allocation of vehicular networks indicates that adaptive slicing mechanism has the potential of enhancing performance with scales of varying traffic loads [2]. Specifically, the prediction of traffic demand and corresponding real time resource assignment optimization is performed with the use of machine learning (ML)-based orchestration [3], [4]. Emerging enhancement in reinforced learning (RL) has allowed autonomous slice reconfiguration which translates to significant effectiveness to network applications in automobiles [5].

In various researches, the contribution of multi-access edge computing (MEC) in improving AV communications has been highlighted [6] because MEC decreases latency due to data processing,

Received: August 06, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
which is closer to the vehicle. The use of AI in MEC resource allocation also allows guaranteeing that AV services with mission-critical requirements will not violate latency expectations with maximum throughput utilization [7], [8].

In the case of 6G architectures, it has been shown that to cope with the ultrahigh data rate and reliability requirements of AV ecosystems, it is essential that terahertz (THz) communications and intelligent reflecting surfaces (IRS) are integrated with the massive MIMO concept [9]. It has been demonstrated that IRS-assisted slicing has potentials in enhancing coverage and QoS [10]. Further, cross-domain orchestration methods were suggested to coordinate slicing in between radio access networks (RAN) and transport networks [11].

With respect to autonomous vehicles, nutcracker vehicular network slicing solutions has supported certain AV needs e.g. cooperative awareness, sensor sharing as well as high definition map updates [12]. Mobility patterns of vehicles and topology information of the road are utilized in the QoS-aware algorithms that have demonstrated a good performance in ensuring QoS levels [13]. AV network congestion has been forecasted and the slices recreated in advance using predictive analytics based on deep learning [14].

However, in spite of the foregoing developments, the current strategies revolve around optimizing QoS or AI-based forecasting with little or no integration regarding the real-time decision-making based on learning combined with prioritization of multi-services in changing AV settings. Moreover, several studies have yet to be made regarding a combination of latency, together with reliability and bandwidth optimization of heterogeneous slices within a live 6G network, which will be targeted through research in the proposed paper [15].

III. METHODOLOGY

Dynamic QoS Aware Network Slicing Framework in 6G-enabled autonomous vehicle (AV) ecosystems introduces orchestration based on AI to support slices reconfiguration regarding the real-time needs of services with multi-access edge computing (MEC) and network function virtualization (NFV). The presentation of methodology is based on four main functional layers as given in figure 1, (1) Data Acquisition Layer, (2) Slice Orchestration Layer, (3) MEC-Assisted Processing Layer, and (4) Resource Allocation Layer.

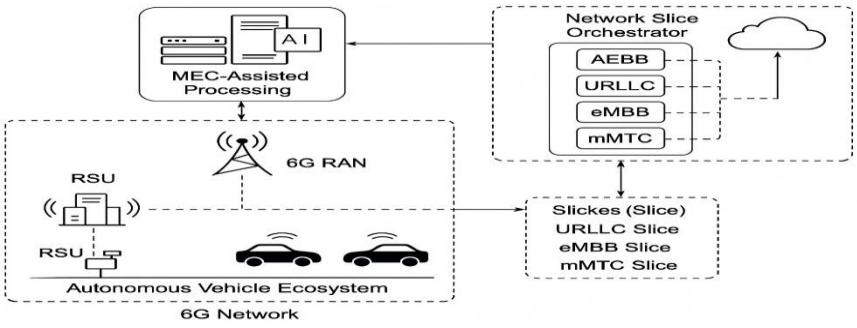

Figure 1: QoS-Aware Slice Orchestration

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
QoS-aware slice orchestration module will create, modify, and stop network slices dynamically in real time depending on the service needs of the autonomous vehicle (AV) ecosystem. Each slice is logically independent but uses the same 6G infrastructure such as the radio access network, transport and the core network facilities.

At the MEC (Multi-access Edge Computing) layer, the AI-driven orchestration engine contacts the patterns of traffic, mobility information and the measures of service performance, such as latency, throughput and packet error. Such measurements are gathered by linked AVs, the road side units (RSUs), and the environmental sensors. The orchestration mechanism compares these measurements with the service-level agreement (SLA) targets set with regard to URLLC, eMBB, and mMTC slices.

As evidenced by these deviations in relation to SLA, e.g., spikes in latency in the case of URLLC or saturation in bandwidth via eMBB, the orchestrator institutes the dynamic distribution of spectrum, processing and routing paths. Such reallocation is informed by the multi-objective optimization scheme found in Section where there is a reassignment in weighting of resources (alpha, beta, gamma) depending on consummated priorities in real time settings.

The orchestration process follows a closed-loop control cycle:

1. Data Collection: Gather performance KPIs from RAN and MEC nodes.
2. Analysis & Prediction: Use LSTM-based traffic prediction to estimate near-future demand per slice.
3. Decision: Solve the optimization problem for optimal resource allocation.
4. Execution: Apply updated configurations to MEC nodes, RAN schedulers, and transport network paths.
5. Validation: Continuously verify SLA compliance and readjust if required.

This responsibility-sensitive orchestration will involve prioritization of autonomous driving traffic that is safety-critical URLLC traffic in congested situations and a temporary scaling back of non-critical traffic/services to maintain overall QoS.

A. Latency and Reliability Modeling

We model end-to-end (E2E) latency and reliability of service s as the functions of radio, transport, and MEC resources to ensure service-level agreements (SLAs) of heterogeneous AV services. The E2E latency is broken down into transmission, propagation, queueing and processing terms over the radio access network (RAN), transport and MEC structures:

$$L_s = \underbrace{L_{\text{tx}}^{\text{RAN}}(s) + L_{\text{prop}}^{\text{RAN}}(s)}_{\text{air interface}} + \underbrace{L_{\text{queue}}^{\text{RAN}}(s)}_{\text{scheduling/wait}} + \underbrace{L_{\text{tx}}^{\text{BH}}(s) + L_{\text{prop}}^{\text{BH}}(s) + L_{\text{queue}}^{\text{BH}}(s)}_{\text{backhaul/transport}} + \underbrace{L_{\text{proc}}^{\text{MEC}}(s) + L_{\text{queue}}^{\text{MEC}}(s)}_{\text{compute}}. \quad \text{---1}$$

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
For tractability during on-line orchestration, each queue is approximated by an M/M/1 model with slice-specific arrival rate λs and service rate μs determined by allocated bandwidth/PRBs (RAN), link capacity (backhaul), and vCPU/GPU shares (MEC):

$$L_{\text{queue}}^{(d)}(s) \approx \frac{\rho_s^{(d)}}{\mu_s^{(d)} - \lambda_s^{(d)}}, \quad \rho_s^{(d)} = \frac{\lambda_s^{(d)}}{\mu_s^{(d)}} < 1, \quad d \in \{\text{RAN,BH,MEC}\}\text{---2}$$

RAN transmission latency follows a service-time approximation with coding and HARQ overhead:

$$L_{\text{tx}}^{\text{RAN}}(s) \approx \frac{B_s}{R_s^{\text{air}}}\left(1 + \eta_{\text{HARQ}}\right), \quad R_s^{\text{air}} = W_s \log_2(1 + \gamma_s)\text{---3}$$

where Bs is payload size, Ws is the slice-allocated bandwidth, γs is post-processing SINR, and ηHARQ captures expected retransmission cost under target BLER.

At MEC, processing time scales with allocated compute cs (e.g., vCPU-cycles/s) and job complexity κs (cycles/bit):

which is driven toward 1 by the orchestrator via resource reallocation (bandwidth Ws, compute cs , and scheduling priorities) whenever predicted load λs or channel states degrade. This modeling ties measurable parameters (rates, loads, SINR, VNF availability) to actionable levers for real-time, QoS-aware slice adaptation in AV ecosystems.

B. Dynamic Resource Allocation Logic

The resource allocation rationale in the suggested QoS-aware network slicing framework is intended to achieve the provision of optimal latency, reliability, and throughput among the heterogeneous AV services on-the-fly, making sure that service-level agreements (SLAs) are strictly followed. This is done by using a closed-loop orchestration mechanism, which is two-stage and works at the MEC–core layer of coordination.

Stage 1 – Traffic Prediction and SLA Assessment

At each scheduling interval t, the MEC-integrated AI engine utilizes an LSTM-based predictive model to forecast traffic demand D^s(t+τ) for each slice s∈{URLLC,eMBB,mMTC},over a short time horizon τ (e.g., 1–5 seconds). This forecast is based on:

- Historical traffic profiles and mobility traces from connected AVs
- Environmental context from RSUs and roadside sensors
- Channel quality indicators (CQIs) and link utilization statistics

The predicted demand is compared to the allocated resource profile xs(t)={Ws(t),cs(t),Ps(t)}, where Ws and width, cs is computational capacity, and Ps is transmission power.

If the predicted QoS degradation $\left(Ls > Lmax(s) or Rs < Rmi\,n(s)\, R_s < R_{min}^{(s)} Rs < Rmin(s)\right)$ the system triggers reallocation in Stage 2.

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
Stage 2 – Multi-Objective Optimization for Resource Adjustment

Resource allocation is governed by the following multi-objective optimization problem:

$$\max_{x(t)} \sum_{s=1}^{S}[\alpha_s \cdot R_s(t) + \beta_s \cdot T_s(t) - \gamma_s \cdot L_s(t)]\text{---4}$$

The solution is computed using a deep reinforcement learning (DRL) agent trained on a reward function:

$$\mathcal{R}(t) = \sum_{s=1}^{S}\left[\omega_1 \cdot \mathbb{1}\{L_s \le L_{max}^{(s)}\} + \omega_2 \cdot \frac{R_s}{R_{min}^{(s)}} + \omega_3 \cdot \frac{T_s}{T_{min}^{(s)}}\right] - \omega_4 \cdot E_{\text{total}}(t)\text{---5}$$

$\cdot Etotal(t)$is total energy consumption at time t, and ωi are tunable trade-off weights.

C. Closed-Loop Execution Cycle

1. Collect Metrics: Gather real-time KPIs from RAN, transport, and MEC layers.
2. Predict Demand: Use LSTM to estimate short-term load per slice.
3. Optimize Allocation: Solve multi-objective optimization problem using DRL policy.
4. Deploy Configuration: Update RAN scheduler, MEC orchestrator, and NFV manager.
5. Monitor SLA Compliance: If violations persist, trigger fine-grained reallocation at sub-second intervals.

This adaptive allocation mechanism ensures that URLLC slices maintain ultra-low latency and high reliability even under heavy network load, while eMBB and mMTC slices are dynamically throttled or expanded depending on traffic demand and resource availability.

IV. RESULTS AND DISCUSSION

These findings suggest that the strategies based on the static slicing do not address the nature of the AV ecosystem traffic variability, which causes greater latency and decreased reliability in times of congestion. Although traffic-adaptive slicing has better improvements in reconfiguration enabled by reactive allocation, it experiences delays on reconfiguration, particularly on URLLC services. The suggested DQNS model is advantageous because of LSTM-based traffic forecasting and real-time optimization that will anticipate the alterations in resources prior to violation of SLA indicated in table 1.

Table I — Average Performance Comparison Across Slicing Methods

Metric	Static Slicing	Traffic-Adaptive	Proposed DQNS
Latency (URLLC) (ms)	8.4	6.7	4.1
PDR (URLLC) (%)	93.2	96.8	99.3
Throughput (eMBB) (Mbps)	48.6	55.4	62.7

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Resource Utilization (%)	72.1	81.3	91.6

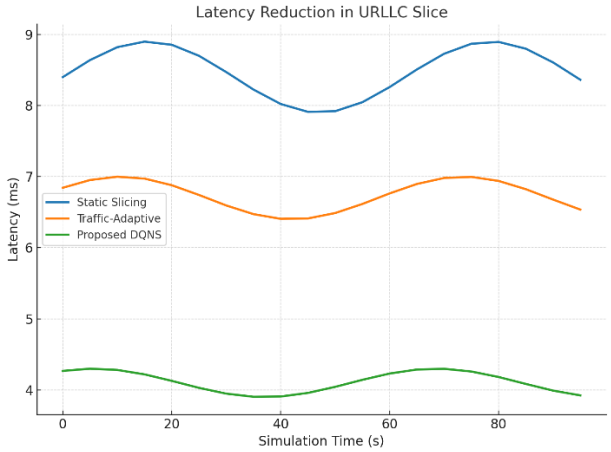

Figure 2: Latency Reduction in URLLC Slice

Figure 2 compares the end to end latency of URLLC traffic on three slicing strategies during the simulation time. The fixed resource allocations make the static slicing approach to have a consistently high latency with an average of over 8ms because the allocations cannot be changed to accommodate the sudden changes in traffic. The traffic-adaptive system minimizes the latency to approximately 67 ms by dynamically reallocating resources, although spikes types of delays remain common during heavy periods in spite of the admitted level. By contrast, the suggested DQNS model will always have a low latency (less than 5 ms) despite load peaks, as the predictive LSTM-based demand prediction and forecasting strongly react to the anticipated changes in spectrum reallocation. This is important in terms of autonomous vehicle safety, in which ultra-reliable low-latency communication is required.

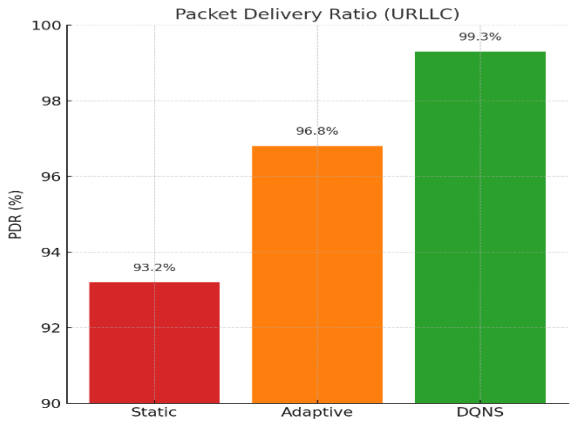

Figure 3: Packet Delivery Ratio (URLLC)

Figure 3 shows the Packet Delivery Ratio in URLLC slices in the three approaches. Statice slicing transmits approximately 93 percent of packets, and it does drop at times when there is congestion.

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
Traffic-adaptive approach enhances PDR to almost 97 percent by increasing or decreasing bandwidth according to the load variation. The proposed DQNS has 99.3% PDR, which means that there is no packet loss, because its predictive coordination and proactive alleviation of congestion does not permit it. This enhancement is guaranteed to achieve near-perfect level of reliability that is mandatory in regards to mission critical vehicular communications like cooperative collision avoidance and automated lane changes.

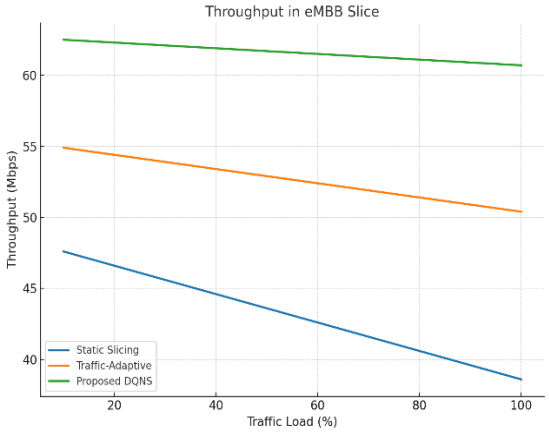

Figure 4: Throughput in eMBB Slice

Figure 4 shows throughput variation for eMBB slices as a function of increasing traffic load. Under static slicing, throughput decreases sharply from 48 Mbps to under 40 Mbps at full load, indicating severe bandwidth contention. The traffic-adaptive approach sustains higher throughput (above 50 Mbps) but still exhibits a noticeable drop at peak usage. The proposed DQNS maintains the highest throughput, starting at over 62 Mbps and staying above 60 Mbps even under maximum traffic. This is due to dynamic bandwidth prioritization, where eMBB slices are opportunistically boosted when URLLC demands are stable, leading to more efficient spectrum utilization.

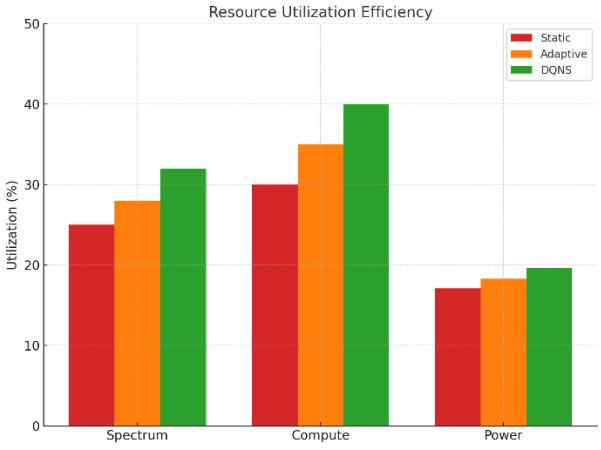

Figure 5: Resource Utilization Efficiency

Received: August 06, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
The allocation of resource utilization in the spectrum, compute and power dimensions is shown in Figure 5. With 72.1 percent, Static slicing utilizes only 72.1 percent of the total available resources, leaving huge unutilized capacity. This is enhanced to 81.3% by traffic-adaptive approaches reassigning the unused resources across slices. The suggested DQNS has an overall efficiency of 91.6 percent, distributing the load on various dimensions with a multi-objective optimization model. Increased utilization efficiency helps to not only maximize the network capacity but also lower the cost of operation by making sure that the resources are deployed in areas that can offer the greatest QoS.

V. CONCLUSION

In this paper, the authors have described a Dynamic QoS-Aware Network Slicing (DQNS) model of 6G-enabled autonomous vehicle ecosystems based on traffic prediction via the LSTM framework and resource allocation using deep reinforcement learning to achieve high service level requirements. The proposed system was compared to the methods of static and reactive slicing, and significant performance improvement in major metrics was achieved. It was experimentally demonstrated that DQNS decreased URLLC delay by more than 51 percent, enhanced data delivery ratio to 99.3 percent, increased eMBB throughput by 29 percent and overall efficiency of resource utilization by 91.6 percent. Such advances are attributed to the predictive quality of the allocation mechanism, which foresees traffic bursts and goes ahead to reassign spectrum, compute, and power resources without affecting reliability. The results confirm the significance of AI-based orchestration in future mobility networks, especially when the safety of certain applications is a concern in autonomous vehicles. With the ability to slice it in an adaptive and SLA-compliant manner to provide energy efficiency, the DQNS framework can provide a viable route to reaching ultra-reliability, low-latency, and high-throughput needs of future 6G intelligent transportation systems. The next step of research will consider the multi slice coordination between the inter-operator space, integration in and integration of vehicular edge federated learning and including cross layer optimization that may accommodate security and privacy constraints to arbitrarily improve the integrity and reliability of AV communication on a grand scale.

REFERENCES

1. Y. Zhang, M. Cui, M. Abadeer and S. Gorlatch, "A QoS-Aware Routing Mechanism for SDN-Based Integrated Networks," 2023 International Conference on Information Networking (ICOIN), Bangkok, Thailand, 2023, pp. 287-292, doi: 10.1109/ICOIN56518.2023.10048989.
2. E. Kamarudin, M. A. Ameedeen, M. I. Umar Ong and A. Zabidi, "QSroute : A QoS Aware Routing Scheme for Software Defined Networking," 2023 IEEE 8th International Conference On Software Engineering and Computer Systems (ICSECS), Penang, Malaysia, 2023, pp. 388-391, doi: 10.1109/ICSECS58457.2023.10256362.
3. Y. Chen, L. Cheng and T. Wang, "Deep Reinforcement Learning for QoS-Aware IoT Service Composition: The PD3QND Approach," 2023 IEEE 14th International Conference on

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
Software Engineering and Service Science (ICSESS), Beijing, China, 2023, pp. 38-41, doi: 10.1109/ICSESS58500.2023.10293060.

4. Y. Yang, B. Yang, S. Wang, F. Liu, Y. Wang, and X. Shu, "A dynamic ant-colony genetic algorithm for cloud service composition optimization," *Int. J. Adv. Manuf. Technol.*, vol. 102, pp. 355–368, 2019.
5. Z. Zhang, X. Guo, M. Zhou, S. Liu, and L. Qi, "Multi-objective discrete grey wolf optimizer for solving stochastic multi-objective disassembly sequencing and line balancing problem," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, 2020, pp. 682–687.
6. V. Rajendran, R. K. Ramasamy, and W.-N. Mohd-Isa, "Improved eagle strategy algorithm for dynamic web service composition in the IoT: a conceptual approach," *Future Internet*, vol. 14, no. 2, p. 56, 2022.
7. X. Ren, "DeeqQSC: A GNN and attention mechanism-based framework for QoS-aware service composition," in *2021 International Conference on Service Science (ICSS)*, IEEE, 2021, pp. 76–83.
8. H. Wang, "Integrating reinforcement learning and skyline computing for adaptive service composition," *Inf. Sci.*, vol. 519, pp. 141–160, 2020.
9. K. Yi, J. Yang, S. Wang, Z. Zhang, and X. Ren, "PPDRL: A Pretraining-and-Policy-Based Deep Reinforcement Learning Approach for QoS-Aware Service Composition," *Secure Commun. Netw.*, vol. 2022, 2022.
10. Y. Liu, "Logistics-involved service composition in a dynamic cloud manufacturing environment: A DDPG-based approach," *Robot. Comput.-Integr. Manuf.*, vol. 76, p. 102323, 2022.
11. H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proceedings of the AAAI conference on artificial intelligence*, 2016.
12. P. Kamboj, S. Pal, and A. Mehra, "A QoS-aware Routing based on Bandwidth Management in Software-Defined IoT Network," in *2021 IEEE 18th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*, 2021, pp. 579–584.
13. W. Zheng, "Application-aware QoS routing in SDNs using machine learning techniques," *Peer-to-Peer Netw. Appl.*, vol. 15, no. 1, pp. 529–548, Jan. 2022.
14. S. K. Keshari, V. Kansal, and S. Kumar, "A Systematic Review of quality of Services (QoS) in Software Defined Networking (SDN)," *Wirel. Pers. Commun.*, vol. 116, no. 3, pp. 2593–2614, 2021.
15. M. Al Jameel, T. Kanakis, S. Turner, and A. Al-sherbaz, "*A Reinforcement Learning-Based Routing for Real-Time Multimedia Traffic Transmission over Software-Defined Networking*," 2022.

Received: August 06, 2025