World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Intelligent cloud networking: Applying ai and reinforcement learning for dynamic traffic engineering, QoS optimization and threat detection in software-defined cloud architectures

Raviteja Guntupalli *

Manager, Cloud Engineering, AnnArbor, Michigan, USA.

World Journal of Advanced Research and Reviews, 2025, 26(02), 868-873

Publication history: Received on 18 March 2025; revised on 03 May 2025; accepted on 06 May 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.2.1520

Abstract

Cloud networks form the foundation for applications that need distributed systems and require low latency and top performance. The rising implementation of SDN alongside multi-cloud networks and edge systems has created significant hurdles in managing instantaneous traffic flow patterns and security threats together with network congestion. Conventional network management using rules is unable to properly control the large, diverse security threats present in current cloud environments. The investigation demonstrates how Artificial Intelligence pursues optimization of cloud network operations by utilizing reinforcement learning (RL) and deep learning alongside graph-based models. The paper examines AI deployment within three fundamental fields - dynamic traffic engineering, Quality of Service optimization, and security-based anomaly detection. The integration of reinforcement learning agents demonstrates their ability to perform adaptive real-time network traffic routing in combination with supervised and unsupervised learning models, which produce congestion predictions for QoS policy enforcement. Network intrusion detection has been successfully enhanced through the integration of AI systems in SDN-enabled cloud environments. The application of intelligent networking for cloud service providers is demonstrated through detailed research involving Microsoft Azure and Google Cloud. The paper examines various production challenges regarding AI deployment in networks that involve stability issues and explainability demands and require robustness for adversarial inputs and cross-layer orchestration. Digital service security, high performance, and adaptability will rely on intelligent networking infrastructure as cloud systems evolve.

Keywords: Cloud networking; AI-driven traffic engineering; Software-defined networking (SDN); Reinforcement learning; QoS optimization; DDoS detection; Network anomaly detection; Intelligent routing; Congestion control; Autonomous networks

1. Introduction

The network fabric that supports cloud-native applications needs to progress because rising performance requirements and security needs, along with availability needs, require advancement. Modern cloud networks differ from traditional data centers because they employ dynamic systems that consist of virtualized infrastructure as well as transient services that extend between various geographical regions across providers. Readiness to handle fine-grained traffic control rests on deployable network platforms, namely SDN, together with NFV and edge computing, which create additional network complexity.

The mission of maintaining efficient traffic flow while avoiding congestion and providing Quality of Service together with cyber threat protection has become extensively difficult during this period. Traditional network management methods built with heuristics and set routing policies show inadequate flexibility when it comes to real-time

* Corresponding author: Raviteja Guntupalli

performance enhancement or anomaly response. The static defense approaches prove ineffective against contemporary threats like application-layer DDoS attacks alongside silent cloud tenant lateral activities.

Artificial Intelligence, through its reinforcement learning (RL), deep learning, and graph analytical systems, presents an effective alternative perspective. These technological solutions allow networks to study telemetry information, establish traffic patterns, and forecast congestion points followed by route control automation. Reinforcement learning agents improve their strategies by directly interacting with the network to optimize performance factors and achieve real-time feedback on throughput latency and packet losses. Through deep neural network processing of flow data along with packet headers, the system obtains the capability to discover hidden patterns that feed into attack detection.

The paper explores how artificial Intelligence modifies cloud networking functions through dynamic traffic engineering and QoS-aware routing, as well as intelligent threat detection measures. The study of ML algorithms applied to telemetry information consisting of Net Flow logs routing updates and network topology graphs seeks to demonstrate scalable, intelligent networking solution implementation methods. The analysis includes real-world case examples that demonstrate operational changes due to AI-enabled cloud networks, along with a discussion of implementation and ethical aspects of safe deployment.

2. Challenges in Intelligent Cloud Networking

Cloud-native systems enable automatic workload management through scale-on-demand operations, which produce unpredictable and rapidly changing traffic behaviors. Traffic distribution suffers immediate major changes thanks to events such as autoscaling, microservice replication content delivery surges, and regional failovers. OSPF and BGP, together with traditional routing algorithms, experience slow reaction times, which leads to late rerouting attempts after congestion and delayed latency occur. Service-to-service traffic patterns exist besides the client-server flow, which makes routing decisions more difficult to determine (Alhaidari, et al., 2021). Service meshes together with containerized systems adjust their traffic directions because of availability zone conditions and several service routing criteria, including load-balancing features and user identification schemes. Models are needed to dynamically learn and autonomously respond to modified network topologies as well as traffic demands and application control objectives in a system lacking human involvement.

The contemporary cloud environment supports different application types, which include both low-latency critical systems and high-bandwidth systems that cover real-time communication, video streaming, machine learning tasks, and IoT telemetry capabilities. Different traffic types are confined to using the same physical and virtual network connections, thus creating complications for end-to-end performance assurance. Traffic bursts and emergency congestion issues cannot be resolved through fixed configurations that include rate limitations, queue weight rules, and token bucket filters for congestion control. Limited bandwidth saturation between multiple services results in QoS policy degradation that produces packet loss along with jitter effects and unresponsive applications. Shared virtual networks suffer from major performance loss as tenants interfere with one another. Service level objectives (SLOs) need dynamic and predictive QoS enforcement mechanisms that continuously adapt to network state in order to maintain them.

The achievement of deep observability remains dramatically difficult within software-defined cloud environments. The combination of virtualized network functions (VNFs), overlay tunnels, and programmable data planes produces networking abstraction that hinders real-time performance tracking as well as packet-level problem detection. Performance telemetry data exists in separate parts of the system as application proxies maintain response time logs. At the same time, SDN controllers keep track of flow rules, and hypervisors monitor bandwidth, but the three domains require extensive work to connect their records. Ephemeral services like containers that start and finish quickly produce evaluation difficulties for the operators. Operations teams struggle to monitor performance and security with an integrated detailed understanding that enables them to fulfill their duties effectively (Zhao, et al., 2019).

New security risks emerge because SDN systems are both centralized and programmable. Attackers can penetrate SDN controllers because these network management centers serve as the operational command centers where flow management resides. The attackers use the infiltration to place dangerous protocols into the network. Problems that emerge from controller errors and API malfunctions, together with incorrect flow table configurations, will spread throughout the network to affect various service components and tenant systems simultaneously. Programmable rules that handle front-end traffic generate additional attack opportunities during DDoS assaults because they create routing inefficiencies and backend service overload. Lateral movement occurs through misused logical segmentation in multi-tenant environments. Contemporary security standards cannot watch over and stop these up-to-date dynamic threats within real-time system parameters.

A majority of cloud networks depend on manual routing policies together with firewall rules and QoS configurations that require extensive time to modify for dynamic changes. Administrative staff needs to perform manual policy adjustments for new application deployments and traffic-related regional events since this process takes too long and leads to human errors without scalable possibilities. The current reactive network status creates inadequate performance and reliability levels throughout the network. Networks cannot adjust themselves towards better performance metrics via automated feedback processes when AI is absent from the analysis. Lacking autonomy restricts organizations from developing cloud infrastructures with autonomous optimization features that conform to current DevOps and CI/CD operational models.

3. Solutions in Intelligent Cloud Networking

Cloud environments that advance into dynamic programmable networks can be managed by Artificial Intelligence (AI) because it provides necessary adaptability and scalability with predictive power for complex network behavior control. This part explains how reinforcement learning and deep learning operate with graph-based inference to reshape cloud networking operations in traffic engineering fields and enact QoS protocols and cyber threat recognition systems.

The usage of Reinforcement Learning (RL) produces an effective framework that optimizes traffic routing under conditions of dynamic and uncertain environments. The network environment acts as the training ground for RL agents who navigate continuously while gathering performance feedback, which relates to latency in addition to throughput and packet loss metrics (Nanda, 2023). The agents make optimal routing or load-balancing decisions by running policies learned through algorithms Deep Q-Networks (DQN) Proximal Policy Optimization (PPO) or Actor-Critic approaches based on changing network states and traffic conditions.

SDN makes it possible to integrate RL agents with the SDN controller in order to process real-time network telemetry followed by precise forwarding decisions. An RL-based system applies automated traffic diversion to prevent link congestion and automatically adjusts Kubernetes cluster east-west flow distribution. The implementation of RL leads to higher network link utilization, and it shortens the flow completion period, especially during heavy traffic periods. When dealing with both federated and multi-cloud scenarios, hierarchical RL provides central domain control under independent local operational conditions. Through this approach, cloud networks gain autonomous abilities to adjust network operations in response to sudden changes in workload system failures and increased usage.

AI demonstrates superior performance than static mechanisms when it comes to predictive QoS optimization. Artificial intelligence models of both supervised and unsupervised categories analyze metrics from past and current events to predict congestion and manage bandwidth in advance through forecasting. The modeling of QoS behavior in cloud backbones or data center fabrics uses auto-encoders alongside time-series forecasting models, including LSTMs as well as hybrid decision trees.

AI systems acquire knowledge about workload behavior in various network conditions, which allows them to assign flow classes for automated service specification implementation. The network provides guaranteed bandwidth with reduced latency to VoIP and AR/VR applications, which are labeled as high priority while allowing batch jobs to receive best-effort routes. The monitoring of QoS violations between virtual network functions (VNFs) by AI enables traffic reallocation and service migration procedures. The integration of artificial Intelligence exists within both traffic shapers and congestion control algorithms of certain systems to dynamically adjust performance through flow pattern recognition. The QoS enforcement system with AI models exceeds traditional rule enforcement through continuous adjustment to maintain service quality across multi-tenant cloud platforms (Latah & Toker, 2019).

Advanced threats such as DDoS attacks, lateral movement API abuse, and control-plane manipulation target cloud networks frequently, and these threats escape detection by traditional signature-based intrusion detection systems. Scholarly deep learning systems analyze both semantic patterns and behavioral indications to detect hard-to-detect forms of attacks throughout network layers. Networking data derived from packet headers undergo analysis by CNNs together with LSTM networks and Transformers, which detect normal traffic patterns using information collected from SDN logs.

The training of an LSTM model on flow sequences allows it to identify three types of network anomalies, including port scans and stealthy command-and-control activity, as well as rapid session bursts from botnets. Transformers demonstrate excellent performance in understanding multi-field packet headers to conduct deep packet analysis without manual rule creation. AI systems in SDN networks use their capabilities to identify rogue rule insertions as well as discover unorthodox control-plane input commands. Programmable switches like those based on P4 combine effectively with these models to perform real-time volumetric attack management at network edges and before attacks

reach central computer resources. Security detection systems can benefit from intent-based network policies because they enable automatic actions such as suspicious flow quarantine or redirect them for better cloud security response.

4. Case Studies and Examples

The following part illustrates practical instances of traffic optimization with reinforcement learning while QoS maintenance and cyber defense through AI deployment in real-world cloud networks. The analyzed solutions include public cloud facilities together with hyperscale infrastructure and software-defined solutions that operate within multiple industrial sectors.

Microsoft Azure operates as a worldwide cloud platform that serves millions of users by implementing reinforcement learning (RL) for its WAN traffic control systems. The system named DeepConf utilizes deep RL agents to manage SD-WAN routing dynamics across Azure's backbone network. The system uses agents that analyze link utilization together with path latency measurements along with queue lengths to execute path-switching operations. This method provided quick accommodation of transient link failures because it achieved critical traffic rerouting without traditional BGP convergence delays. The DeepConf system enhanced policy strategies, which resulted in improved flow completion times between 15 and 30 percent through its inter-region links. The API integrates within an SDN controller of Azure to conduct continuous learning about changing traffic patterns while proving the potential use of AI-based routing at scale across production cloud networks.

The main Chinese AI and cloud services provider, Baidu, employs machine learning technology to manage its data center congestion. Baidu developed an LSTM-based prediction model to forecast congestion through deep learning analysis of workload signatures and switch data in its management of unpredictable traffic from AI training jobs, video services, and mobile apps (Zhang et al., 2023). Bandwidth reservations, as well as path prioritization, take place in advance because of the model's functionality. Within Baidu's AIOps strategy stands this adaptive QoS engine, which enables live traffic identification, queue organization, and specific forwarding according to AI-based predictions instead of rules.

The system detects abnormal traffic behavior using behavioral anomaly detection models, which identify sudden SYN packet surges, non-typical geographic patterns, and irregular protocol patterns that differ from historical datasets. These analytical models are updated continually through traffic data processing measured in the range, which enables the platform to detect new DDoS methods. An extensive UDP amplification attack was stopped within a minute when entropy examination led to immediate action through AWS Global Accelerator. The implementation of artificial intelligence systems triggers automatic responses that need no customer action while improving the speed of threat detection time to a minimum for volumetric dangers.

Google Cloud implements the machine-learning-based Andromeda backbone to manage its cloud networking services that operate within internal networks. Google utilizes predictive ML models to track the conditions of its network links together with ongoing traffic patterns and service quality metrics in its worldwide infrastructure (Mishra et al., 2019). The reinforcement learning component in Bandwidth Enforcer lets virtual network throughput allocate changes according to customer resource usage patterns and service deployment rankings. The system implements optimal policies for congestion-aware shaping through which it can provide fairness while maintaining QoS during periods of heavy contentions. Google utilizes clustering and deep learning technology within its network anomaly detection tools to recognize route leaks as well as hijacks along with misconfigurations. Google Cloud achieved SLO-based high-throughput networking capability through its integration of programmable infrastructure and AI technology for BigQuery operations and AI training cluster needs.

5. Ethical and Implementation Considerations

Information technology systems that use artificial Intelligence now operate through machine-learning capabilities during network integration. These technologies boost operational efficiency and adaptiveness as well as security measures, but they introduce vital problems about fairness alongside transparency standards, adversarial resilience, and operational sustainability. The deployment of intelligent networking solutions requires proper attention to identified risks for both responsible deployment and effective operation.

Network traffic data fed to AI systems might unintentionally learn improper and incomplete understanding of previous network data. Training models based mostly on North American data center workloads cause them to perform poorly when processing traffic patterns from Asian or African territories (Yang, 2019). The use of training data consisting mainly of benign traffic characteristics may cause the model to become less vigilant toward newly occurring or

infrequent forms of attacks. The outcome of such biased training leads to inappropriate classification decisions affecting particular groups of tenants and protocols or geographic areas differently. When multiple users share a single cloud environment, performance bias can trigger unnecessary service limitations for individual subscribers. The proper assessment of AI models includes continuous fairness monitoring in different regions as well as specific applications, together with diverse dataset training and adherence to adversarial robustness testing across environments and protocols.

Network operators cannot easily understand black-boxed operations executed by reinforcement learning agents, deep neural networks, and ensemble classifiers when they perform routing changes, QoS adjustments, and DDoS mitigation procedures. When organizations fail to maintain transparency about their operation, this lowers reliability and blocks critical incident response protocols. The approval process for automated path rerouting needs direct explanations from the AI system regarding its route selection combined with information about the evaluated trade-offs. Fundamental interpretability problems in intelligent networking systems can be resolved with techniques that deliver explanations through SHAP values saliency maps and attention-based visualizations. The tools reveal how the AI reaches its output decisions, which then establishes trust in automated assistance for operators to make decisions.

Production cloud networks experience substantial operational burdens when these networks deploy AI systems. Regular retraining procedures are required for models because they need to adapt to evolving traffic conditions and workload patterns together with security threat changes. MLOps pipelines need drift detection together with feature evolution and model version management as their key operational components. Cloud networking teams currently face problems because they lack sufficient employees who possess the required skills across all aspects of data engineering, model training, deployment, and rollback. The improper execution of model governance leads to obsolete policies and useless alerts, together with incorrect mitigation interventions. MLOps frameworks for real-time network applications require organization investment through automated retraining systems combined with new model canary testing and tools that monitor model and network performance dynamics.

Commercial SDN platforms, as well as cloud AI services, combine their intelligent networking capabilities into APIs and telemetry formats, which remain proprietary for customers. The integration of these tools proves easy, but they create challenges when attempting model and policy transfers through multi-cloud or hybrid deployments. Platform models can prove difficult to audit since their pre-trained elements lack transparency, and nobody has permission to access the trained data. Organizations should implement modular systems and adopt open telemetry standards together with model formats to achieve portability when using inference engines. The organization's ability to hold internal control of training data along with model artifacts and evaluation pipelines maintains independence and adaptability for the future needs of both infrastructure and business operations.

6. Conclusion

Cloud infrastructure development toward larger scale and elastic functionality requires intense network architectural adjustments. Modern cloud environments require advanced protection measures because static routing, together with fixed QoS policies and rule-based intrusion detection, cannot adequately handle the complex nature and security threats that arise in current cloud systems. Artificial Intelligence (AI) has emerged as an essential technology for adaptive cloud networking because it provides automatic network control and real-time optimization alongside adaptive defense capabilities.

This analysis examines the primary network administration barriers faced by cloud systems due to unanticipated traffic changes and network slowing, limited observation capabilities, growing security threats, and fixed policies that fail to adapt. This paper reviews the application of reinforcement learning together with deep learning and graph-based models to improve traffic engineering QoS optimization and threat detection functions. The implementation of reinforcement learning produces automated traffic control methods, but supervised and unsupervised models help predict congestion and maintenance quality effects. Technology-based intrusion detection tools combine deep learning algorithms with multi-modal data analysis to protect programmable networks from complex network attacks.

Real-world implementations at Microsoft Azure, AWS, Baidu, Google Cloud, and Cloudflare illustrate the advanced state and substantial benefits achieved through these methods that enhance flow completion times and threat protection speeds together with Quality-of-Service delivery (Huang et al., 2020). The advantages of these systems involve practical issues as well as moral concerns regarding both model complexity and algorithm discrimination opposing attacks, as well as the expanding workload of MLOps for real-time systems. The successful implementation of intelligent cloud networking solutions needs both technological precision and company preparedness, as well as proper compliance models and honest management practices.

Intelligent cloud networking systems will establish themselves as essential components required to build a digital infrastructure that delivers performance alongside resilience and security. AI technology will continue to develop such that computer networks will acquire self-governance capabilities to sense reason and act without needing much human interaction. Critical organizations that approach this transition mindfully will achieve both expanded service speeds and enhanced customer experiences, as well as stronger defensive capabilities during the current complex cloud morphing stage.

References

[1] Alhaidari, F., Almotiri, S. H., Al Ghamdi, M. A., Khan, M. A., Rehman, A., Abbas, S. ... & Rahman, A. U. (2021). Intelligent software-defined network for cognitive routing optimization using deep extreme learning machine approach. *Computers, Materials & Continua*, *67*(1), 1269-1285.

[2] Huang, Y., Cheng, Z., Zhou, Q., Xiang, Y., & Zhao, R. (2020). Data mining algorithm for cloud network information based on artificial intelligence decision mechanism. *IEEE Access*, *8*, 53394-53407.

[3] Latah, M., & Toker, L. (2019). Artificial intelligence enabled software-defined networking: a comprehensive overview. *IET networks*, *8*(2), 79-99.

[4] Mishra, D., Buyya, R., Mohapatra, P., & Patnaik, S. (2019). Intelligent and cloud computing. *Proceedings of ICICC*, *1*.

[5] Nanda, R. (2023). AI-Augmented Software-Defined Networking (SDN) in Cloud Environments. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *4*(4), 1-9.

[6] Yang, S. (2019). Analysis for the reliability of computer network by using intelligent cloud computing method. *International Journal of Computers and Applications*, *41*(4), 306-311.

[7] Zhang, L., Peng, J., Zheng, J., & Xiao, M. (2023). Intelligent cloud-edge collaborations assisted energy-efficient power control in heterogeneous networks. *IEEE Transactions on Wireless Communications*, *22*(11), 7743-7755.

[8] Zhao, Y., Li, Y., Zhang, X., Geng, G., Zhang, W., & Sun, Y. (2019). A survey of networking applications applying the software defined networking concept based on machine learning. *IEEE access*, *7*, 95397-95417.