# Hybrid Multi-Agent Q-Learning and Intelligent Water Drops Framework for Adaptive Campus Bandwidth Allocation

O. S. Adeoye[1], M. B. Abdullahi[2], O. S. Adebayo[3], O. A. Ojerinde[1], M. Danlami[4]

[1]Department of Computer Science, Federal University of Technology Minna
[2]Department of Data Science, Federal University of Technology, Minna
[3]Department of Cyber Security, National Open University of Nigeria, Abuja
[4]Department of Computer Engineering, Federal University of Technology Minna

***ABSTRACT:*** Efficient bandwidth allocation in campus networks remains challenging because of dynamic traffic patterns, heterogeneous user demands, congestion, and fairness constraints. This study aims to develop and evaluate a hybrid Intelligent Water Drops Multi-Agent Q-Learning (IWDAMA) framework for adaptive bandwidth allocation in software-defined campus networks. The proposed framework integrates Multi-Agent Q-Learning with the Intelligent Water Drops algorithm to enhance exploration, convergence, and adaptive scheduling. Performance evaluation was conducted using MATLAB simulations in a software-defined campus network under five traffic scenarios involving 200–2000 users with a fixed link capacity of 2000 Mbps. The proposed model was compared with standalone Multi-Agent Q-Learning (MAQL), Q-learning, and First-Come-First-Served (FCFS) scheduling using throughput, delay, bandwidth utilization, fairness index, and packet loss as evaluation metrics. Experimental results showed that IWDAMA achieved throughput ranging from 392 Mbps to 1764 Mbps, bandwidth utilization of 0.99, fairness index of 0.995, packet loss between 0.0159 and 0.0600, and delay between 0.0636 ms and 0.2019 ms, consistently outperforming MAQL, Q-learning, and FCFS across all scenarios. The findings demonstrate that integrating heuristic optimization with cooperative reinforcement learning provides a scalable and robust solution for intelligent bandwidth management in campus and enterprise software-defined networks.

**KEYWORDS**
Multi-Agent Quality-Learning (MAQL), Intelligent Water Drops Algorithm (IWDA), Network traffic analysis, Bandwidth allocation, Reinforcement learning

[Received Feb. 25, 2026; Revised Apr. 10, 2026; Accepted May. 18, 2026] Print ISSN: 0189-9546 | Online ISSN: 2437-2110

## I. INTRODUCTION

Modern network infrastructures, including IoT-enabled systems, 5G architectures, and cloud edge services, are becoming increasingly complex, large-scale, and highly dynamic (Wang *et al*., 2019), (Zhou *et al*., 2020). These environments generate massive volumes of network traffic that fluctuate unpredictably over time and space (Krishnamoorthy, Dua and Gupta, 2023). At the same time, diverse quality of service (QoS) requirements must be met simultaneously, from ultra-reliable low-latency communication for critical applications to high-throughput demands for content delivery networks (CDNs) (Ali *et al*., 2021), (Tariq *et al*., 2021).

The International Telecommunication Union's 2023 digital development report indicates that global internet connectivity has expanded substantially, with broadband coverage and 5G adoption increasing worldwide, trends that amplify the urgency for efficient traffic analysis and adaptive scheduling in modern networks (International Telecommunication Union, 2023). Traditional traffic control techniques, however, are no longer sufficient for modern networks where congestion, fairness, and QoS must all be optimized concurrently (Mushtaq *et al*., 2023). With the rapid expansion of connected devices and data-driven services, ensuring efficient traffic management and resource utilization has become a critical concern for modern communication systems. This research applies Q-learning and optimization algorithms to manage dynamic resources in tertiary networks, tackling fluctuating traffic and user needs within heterogeneous, decentralized ecosystems of IoT devices, mobile stations, and cloud-edge nodes operating in complex multi-agent configurations (Wang *et al*., 2024). Within this space, the Q-learning algorithm, a well-known RL method, has shown promise in scheduling tasks and adapting to dynamic conditions (Ge *et al*., 2020). By leveraging feedback-driven learning, Q-Learning empowers systems to enhance future performance based on current and available information (Goudarzi *et al*., 2022).

Earlier methods like static policy-based bandwidth control and queue mechanisms (Hierarchical Token Bucket, Per Connection Queue) lacked adaptability and QoS support. Multi-Agent Q-Learning (MAQL) enables decentralized agents with partial observability to collaboratively optimize networks, effectively supporting routing, congestion control, load balancing, and adaptive traffic management (Boussaoud, En-Nouaary and Ayache, 2025), (Guo *et al*., 2024). In parallel, nature-inspired optimization algorithms like Particle Swarm Optimization, Ant Colony Optimization,

*Corresponding author: fkadeoye@futminna.edu.ng *doi:* http://doi.org/10.63746/njtd.v23i2.4492

and the Firefly Algorithm have proven highly effective in parameter tuning, search space exploration, and accelerating convergence in complex optimization landscapes (Yang, 2020). Within this class, the Intelligent Water Drops algorithm (IWDA) stands out for its ability to balance exploration and exploitation, making it particularly effective for routing, scheduling, and resource allocation. Prior studies confirm IWDA's effectiveness in delivering high quality solutions with competitive convergence rates in networking problems (Agor *et al*., 2024), proving its effectiveness for academic and enterprise networks compared with existing solutions. Modern networks encounter fluctuating loads, evolving topologies, and diverse requirements, demanding adaptive solutions.

This study proposes a hybrid framework, termed IWDAMA, which integrates multi-agent Q-Learning with the Intelligent Water Drops algorithm for real-time traffic analysis and adaptive bandwidth scheduling. Unlike static, rule-driven strategies, IWDAMA delivers intelligent, dynamic management for heterogeneous traffic. By addressing high-dimensional data, decentralization, and bottlenecks, the model enhances scalability, fairness, efficiency, and Quality of Service (QoS) across institutional and enterprise networks. Recent research on intelligent network traffic management has explored machine learning based traffic analysis, reinforcement learning driven resource allocation, and heuristic optimization techniques to address congestion, fairness, and scalability challenges in modern communication networks. Flow-level traffic representation approaches such as Enhanced NetTiSA demonstrate strong classification accuracy at large monitoring scales but focus primarily on traffic identification rather than adaptive bandwidth control (Koumar *et al*., 2023). Similarly, large language model-based frameworks for traffic analysis improve detection and generation capabilities but do not address real-time scheduling or cooperative resource allocation (Cui *et al*., 2025), (Yang *et al*., 2025).

Reinforcement learning has shown significant potential for dynamic network optimization, particularly in multi-agent environments where adaptive policies improve throughput, resource utilization, and scalability (Wang *et al*., 2025), (Sohaib, Jeong and Jeon, 2021). However, standalone reinforcement learning approaches often exhibit slow convergence and unstable exploration under highly dynamic or bursty traffic conditions. Heuristic optimization techniques, especially nature-inspired methods such as the Intelligent Water Drops algorithm, have demonstrated effectiveness in routing, scheduling, and resource allocation due to strong exploration capability and convergence efficiency (Halder *et al*., 2023), (Movahed *et al*., 2025). Nevertheless, heuristic-only approaches generally lack long-term adaptive learning mechanisms required for evolving network environments. Overall, existing studies tend to address traffic analysis, learning-based control, or heuristic optimization separately. Limited work integrates cooperative reinforcement learning with heuristic exploration while explicitly incorporating fairness-aware bandwidth allocation.

This gap motivates the proposed hybrid Intelligent Water Drops Multi-Agent Q-Learning framework, which combines heuristic guidance with adaptive learning to achieve stable, scalable, and fairness-constrained bandwidth allocation in dynamic campus network environments (Mohammed, Muhammad and Gelwasa, 2024). Despite these advancements, three critical limitations remain. First, standalone reinforcement learning frameworks often suffer from slow convergence and unstable exploration under bursty or adversarial traffic. Second, heuristic optimization approaches such as IWD lack long-term adaptive learning mechanisms. Third, existing hybrid traffic management studies rarely integrate fairness-constrained allocation within a unified cooperative reinforcement learning model. Consequently, a methodological gap exists in designing an integrated framework that combines heuristic exploration, reinforcement learning adaptation, and fairness-aware bandwidth optimization within a single control architecture.

This study aims to design and assess a hybrid Multi-Agent Q-Learning and Intelligent Water Drops Algorithm (IWDAMA) for real-time network traffic analysis and adaptive bandwidth allocation in large-scale networks. Specifically, the study aims to model traffic management as a cooperative multi-agent reinforcement learning problem, integrate Intelligent Water Drops heuristics to guide exploration and improve convergence, design an effective reward mechanism that balances throughput, delay, fairness, utilization, and packet loss, and evaluate the performance of the proposed framework through comparative analysis with existing approaches such as MAQL, Q-learning, and FCFS in tertiary and enterprise network environments.

The main contributions of this paper are:

i. A hybrid integration of Intelligent Water Drops optimization with Multi-Agent Q-Learning for bandwidth scheduling.
ii. A congestion-aware reward formulation aligned with fairness and QoS objectives.
iii. A comprehensive ablation study isolating learning-only and heuristic-only components.
iv. A scalability and robustness evaluation under burst traffic and adversarial user behavior.
v. Quantitative analysis of performance–overhead trade-offs for deployment feasibility.

## II. THEORETICAL ANALYSIS

### A. Network Model

The campus network is modeled as a centralized bandwidth scheduling system operating over a directed graph

$$G = (N, E) \tag{1}$$

where, $N$ denotes the set of network nodes (routers, switches, and access points) and $E$ represents the set of communication links connecting these nodes.

Although the physical network is graph-based, the scheduling problem addressed in this study focuses exclusively on bandwidth allocation, not routing or path selection.

At each discrete time step $t$, active user demands are admitted and a bandwidth allocation decision is computed.

The optimization objective is formally expressed as:

$$\max_{\pi} E\left[\sum_{t=0}^{T} \gamma^t R(s_t, a_t)\right] \tag{2}$$

where $s_t$ denotes the network state at time $t$, $a_t$ represents the scheduling action, $R(s_t, a_t)$ is the reward function, and $\gamma \in (0,1)$ is the discount factor controlling the influence of future rewards.

B. Bandwidth Allocation Problem Formulation

Let $B$ be total available bandwidth, $b_i(t)$ the bandwidth assigned to user $i$at time $t$, $d_i(t)$the demand, and $N$ the number of users. The feasibility constraints are:
The total bandwidth constraint is given by:

$$\sum_{i=1}^{N} b_i \leq B \tag{3}$$

Each user's allocated bandwidth must satisfy

$$B_i \leq d_i(t),\ \forall_{i,t} \in \{1, \dots, N\} \tag{4}$$

These enforce capacity limits and prevent allocations beyond demand. The objective is to determine the optimal allocation vector:

$$\mathrm{b}(t) = [b_1(t), b_2(t), \dots, b_{N(t)}(t)] \tag{5}$$

that maximizes network performance under dynamic traffic conditions.

C. **Multi-Agent Reinforcement Learning Formulation**

Agents are intelligent entities that solve problems in a distributed manner. In multi-agent Q-learning. Traffic management is modeled as a cooperative Markov Decision Process (MDP). For each agent $i$, we define the tuple $(S, \mathcal{A}, \mathcal{T}, R, \gamma)$: state space, action space, transition probability, reward, and discount factor.

1) State space $S$

Define the set of all possible network states, such as users, bandwidth usage, latency, packet loss. Let the observed network state at time $t$ be:

$$S_t = [U_t, b_t, l_t, p_t] \tag{6}$$

Where;

$U_t$= user at time $t$,
$b_t$= available bandwidth at time $t$,
$l_t$ = latency at time, t
$p_t$ = packet loss at time, t

2) Action Space $\mathcal{A}$

Each agent can take actions like adjusting bandwidth allocation or rerouting traffic. Each agent selects an action:

$$a_t = [r_t d_t] \tag{7}$$

Where;

$r_t$= routing decision at time t
$d_t$= bandwidth distribution decision

3) Transition Function $\mathcal{T}$

A function that defines the transition from one state to another based on the agent's action. State evolution is governed by:

$$\mathcal{T}(s_t, a_t) = P(s_{t+1} | s_t, a_t) \tag{8}$$

which captures the effect of scheduling actions and stochastic traffic arrivals on queueing, delay, congestion, and loss.

(2)

4) Reward Function $\mathcal{R}$

The Reward function defines the reward for each agent after taking action $a_t$ in a given state $s_t$

$$\tau_t = \mathcal{R}(s_t, a_t) = -w_1 . p_t - w_2 . l_t + w_3 . b_t \tag{9}$$

where $w_1, w_2, w_3$ are weighting coefficients that reflect the relative importance of packet loss, latency and bandwidth respectively. In a multi-agent system, each agent $i$ learns a policy $\pi^i$ that maximizes its own reward while interacting with other agents.

5) Agent's policy

Each agent i attempts to learn an optimal policy $\pi^i$, which maps observed states to corresponding actions, expressed as:

$$\pi^i(S_t) = a_t^i \tag{10}$$

6) Multi-Agent Dynamics

In a cooperative setting, the overall action space is formed by combining the individual action spaces of all participating agents. This joint action space is defined as

$$A = \prod_{i=1}^{n} \mathcal{A}_i \tag{11}$$

Where $\mathcal{A}_i$ denotes the action space of agent $i$. The joint formulation captures all possible combinations of actions that agents can take collectively during decision-making.

**Joint Reward:** in cooperative MAS, agents may share a joint reward function:

$$r_t = \sum_{i=1}^{n} r_t^i \tag{12}$$

D. IWDA-Guided Policy Update

The Intelligent Water Drops Algorithm (IWDA) is incorporated as a heuristic exploration mechanism to guide candidate bandwidth allocation decisions. Unlike classical IWDA applications in routing problems, the proposed framework does not perform path traversal. Instead, IWDA operates over the space of feasible bandwidth allocation vectors.

Let

$$\mathcal{A}_{\text{feasible}} = \left\{ \mathbf{b} \in \mathbb{R}^{N(t)} : \sum_{i=1}^{N(t)} b_i \leq B,\ 0 \leq b_i \leq d_i(t) \right\} \tag{13}$$

denote the set of feasible allocation actions at time step $t$.
Each water drop represents a candidate allocation vector $a \in \mathcal{A}_{\text{feasible}}$.

1) Soil and velocity modeling for allocation quality

For each candidate allocation $a$, two quantities are maintained:
Soil $\sigma(a, t)$: represents congestion penalty
Velocity $V(a, t)$: represents allocation efficiency
Soil is initialized uniformly:

$$\sigma(a, 0) = \sigma_0 \tag{14}$$

Velocity is initialized as:

$$V(a, 0) = V_0 \tag{15}$$

2) Soil Update Rule

After evaluating allocation $a$, soil is updated as:

$$\sigma(a,t+1) = (1-\rho)\sigma(a,t) + \rho \cdot C(a,t) \tag{16}$$

where:
$\rho \in (0,1)$is soil learning rate,
$C(a,t)$is congestion metric under allocation $a$.
Lower congestion results in smaller soil accumulation.

3) Velocity Update Rule

Velocity increases for efficient allocations:

$$V(a,t+1) = V(a,t) + \frac{\alpha_v}{\beta_v + \sigma(a,t)} \tag{17}$$

where $\alpha_v, \beta_v > 0$are control parameters.
Thus, allocations with lower congestion gain higher velocity.

4) Heuristic desirability score

Each candidate allocation is assigned a heuristic score:

$$H(s,a) = \frac{V(a,t)}{\sigma(a,t)+\epsilon} \tag{18}$$

where $\epsilon$prevents numerical instability.
This formulation biases exploration toward allocations with:
High efficiency (large $V$)
Low congestion (small $\sigma$)

E. Hybrid Learning-Heuristic Integration

To integrate reinforcement learning and IWDA guidance, both Q-values and heuristic scores are normalized:

$$\tilde{Q}(s,a) = \frac{Q(s,a) - Q_{\min}}{Q_{\max} - Q_{\min}} \tag{19}$$

$$\widetilde{H}(s,a) = \frac{H(s,a) - H_{\min}}{H_{\max} - H_{\min}} \tag{20}$$

Combined evaluation:

$$\Psi(s,a) = \lambda \tilde{Q}(s,a) + (1-\lambda)\widetilde{H}(s,a) \tag{21}$$

where $\lambda \in [0,1]$controls the learning–heuristic trade-off.
Final allocation decision:

$$a^*(t) = \arg \max_{a \in \mathcal{A}_{\text{feasible}}} \Psi(s(t), a) \tag{22}$$

This ensures that:
IWDA accelerates exploration,
Q-learning ensures long-term policy refinement.

F. Queue Dynamics and Congestion Evolution (Critical for Q1)

To fully define system dynamics, queue evolution is modeled as:$Q(t+1) = \max\left(0,\ Q(t) + \sum_{i=1}^{N(t)} d_i(t) - \sum_{i=1}^{N(t)} b_i(t)\right)$ (23)

Congestion level:

$$C(t) = \frac{Q(t)}{Q_{\max}} \tag{24}$$

This explicitly links allocation decisions to future states.

To quantify equity in bandwidth allocation among users, this study adopts Jain's Fairness Index, a widely accepted metric in network resource allocation literature.
The fairness index is formally defined as:

$$J(x_1, x_2, \ldots, x_N) = \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N \sum_{i=1}^{N} x_i^2} \tag{25}$$

where:
$x_i$ represents the allocated bandwidth to user $i$,
$N$ denotes the total number of active users.
The fairness index satisfies: $0 < J \leq 1$
with $J = 1$indicating perfectly equal bandwidth distribution.
For the proposed time-varying scheduling model, the fairness metric is expressed as:

$$J(t) = \frac{\left(\sum_{i=1}^{N(t)} b_i(t)\right)^2}{N(t) \sum_{i=1}^{N(t)} b_i(t)^2} \tag{26}$$

This formulation is used both in the reward function and in performance evaluation across all experimental scenarios.

G. Convergence and Stability

Under sufficient exploration ($\varepsilon$-greedy with decaying $\varepsilon$) and diminishing learning rate $\alpha$, the MAQL component preserves standard Q-learning convergence properties.
The IWDA heuristic modifies action selection but does not alter Bellman updates, ensuring asymptotic policy stability.
Empirically, convergence is declared when:

$$\| Q_{t+1} - Q_t \| < \delta \tag{27}$$

for small threshold $\delta$.

H. Computational Complexity Analysis

Let:

- $N$= number of users
- $K$= number of water drops
- $I$= IWDA inner iterations
- $|\mathcal{A}|$= number of candidate allocations

Per time step complexity:

$$O(KI|\mathcal{A}| + |\mathcal{A}| + N) \tag{28}$$

Since soil and velocity updates are linear in candidate count, the additional IWDA overhead remains manageable relative to RL updates.

## III. MATERIALS AND METHODS

A. Simulation Environment

The proposed IWDAMA framework was evaluated using a MATLAB-based simulation platform modeling a software-defined campus network environment. The simulation assumes a centralized SDN controller responsible for adaptive bandwidth allocation across multiple users. The controller observes real-time network states, including active user count, bandwidth demand, queue length, delay, and packet loss, and determines allocation decisions at discrete scheduling intervals.
The total available bandwidth was fixed for each experiment, while the number of active users varied between 200 and 2000 to simulate increasing network load. User bandwidth demands were dynamically generated to reflect heterogeneous traffic patterns typical of campus environments. Queue dynamics were modeled to capture congestion accumulation, and congestion levels were computed relative to maximum buffer capacity.
Each simulation scenario was executed independently, and the proposed hybrid framework was compared against Multi-Agent Q-Learning (MAQL), standalone Q-learning,

and First-Come-First-Served (FCFS) baseline allocation strategies under identical traffic conditions.

B. Scenario Design

To ensure reproducibility and clarity, all simulations were conducted within the same simulation environment used for the baseline results. Five scenarios (S1–S5) were designed to represent progressively challenging traffic conditions, as summarized in Table 1 below.

Table 1 Simulation Scenarios and Parameters

| Scenario | Users | Link Capacit | Traffic Model | Burst Probabil | Misbehaving User |
|---|---|---|---|---|---|
| S1 | 200 | 2000 Mbps | Poisson | 0.00 | 0% |
| S2 | 500 | 2000 Mbps | Poisson | 0.03 | 0% |
| S3 | 800 | 2000 Mbps | Poisson + bursts | 0.05 | 0% |
| S4 | 1000 | 2000 Mbps | Poisson + bursts | 0.05 | 5% |
| S5 | 1000 | 2000 Mbps | Poisson + bursts + greedy flows | 0.10 | 10% |

Scenarios S4 and S5 introduce burst traffic and adversarial behavior to assess the stability and congestion responsiveness of the hybrid framework under adverse conditions. All algorithms were evaluated under identical demand patterns to ensure fair comparison.

### C. Performance Metrics

Performance evaluation was conducted using the following metrics:

1. Throughput (T): Total successfully allocated bandwidth across all users per scheduling interval.

$$T = \frac{\text{Total Successfully Received Data (bits)}}{\text{Total Transmission Time (seconds)}} \quad (29)$$

2. Average Delay (ms): Mean queuing delay experienced by users due to congestion accumulation.

$$D = \frac{1}{N}\sum_{i=1}^{N}(t_{r_i} - t_{s_i}) \quad (30)$$

Where: $t_{r_i}$= packet reception time, $t_{s_i}$= packet sending time, $N$= total number of received packets

3. Bandwidth Utilization (U): Ratio of allocated bandwidth to total available bandwidth.

$$U = \frac{\text{Used Bandwidth}}{\text{Total Available Bandwidth}} \times 100\% \quad (31)$$

4. Fairness Index (Jain's Index, F): Measures how fairly bandwidth is distributed among users.

$$F = \frac{(\sum_{i=1}^{n} x_i)^2}{n\sum_{i=1}^{n} x_i^2} \quad (32)$$

Where: $x_i$= bandwidth allocated to user $i$, $n$= total number of users, $0 < F \le 1$, where 1 means perfect fairness

5. Packet Loss Ratio (PLR)

Measures how many packets were lost during transmission.

$$PLR = \frac{\text{Number of Lost Packets}}{\text{Total Number of Transmitted Packets}} \times 100\% \quad (3)$$

These metrics collectively assess efficiency, congestion control, fairness, and scalability of the proposed hybrid framework.

## IV. RESULTS AND DISCUSSION

The proposed IWDAMA framework was evaluated across five progressively increasing traffic scenarios (S1–S5). Comparative analysis was conducted against MAQL-only, standalone Q-learning, and FCFS baseline strategies under identical demand conditions.

A. Scenario-Based Performance Evaluation

In the first scenario, Table 2 indicates that the IWDAMA framework (MAQL + IWDA) achieved the highest observed throughput (392 Mbps) among the evaluated algorithms in this scenario. It also exhibited the lowest packet loss (0.06) and the highest fairness index (0.995), demonstrating its efficiency in resource allocation.

**Table 2** Comparative performance of IWDAMA and other algorithms across metrics first scenario

| Algorithm | Throughput (Mbps) | Delay (ms) | Utilization | Fairness_Index | Packet_Loss |
|---|---|---|---|---|---|
| MAQL+IWDA | 392 | 0.20188 | 0.9900 | 0.9950 | 0.0600 |
| MAQL | 372 | 0.21896 | 0.9600 | 0.9600 | 0.0900 |
| Q-learning | 199.0200 | 0.23100 | 0.26915 | 0.4975 | 0.1510 |
| FCFS | 7.5700 | 0.2518 | 0.0190 | 0.0043 | 0.1410 |

In the second scenario, as shown in Table 3, the performance of IWDAMA improved further, achieving a throughput of 784 Mbps and a very low delay of 0.063585 ms. These depicts a consistent pattern in which IWDAMA leads in utilization and fairness while maintaining minimal packet loss compared with the other evaluated algorithms.

In the third scenario, Table 4 shows that IWDAMA maintains higher observed throughput level of 980 Mbps with a delay of 0.12097 ms. Utilization remains optimal at 0.99, and packet loss is still minimal at 0.022726. In the fourth scenario, the results presented in Table 5 highlight IWDAMA's adaptability, achieving a throughput of 1372

Mbps, the lowest packet loss (0.019785), and the highest fairness and utilization metrics.

**Table 3** Comparative performance of IWDAMA and other algorithms second scenario

| Algorithm | Throughput (Mbps) | Delay (ms) | Utilization | Fairness_Index | Packet_Loss |
|---|---|---|---|---|---|
| MAQL +IWDA | 784 | 0.0636 | 0.9900 | 0.9950 | 0.0214 |
| MAQL | 744 | 0.0777 | 0.9600 | 0.9600 | 0.0317 |
| Q-learning | 400.1900 | 0.0912 | 0.5002 | 0.7507 | 0.0530 |
| FCFS | 6.3410 | 0.0969 | 0.0079 | 0.0038 | 0.0550 |

**Table 4** Comparative performance of IWDAMA and other algorithms Third Scenario

| Algorithm | Throughput (Mbps) | Delay (ms) | Utilization | Fairness_Index | Packet_Loss |
|---|---|---|---|---|---|
| MAQL +IWDA | 980 | 0.1210 | 0.9900 | 0.9950 | 0.0227 |
| MAQL | 930 | 0.1429 | 0.9600 | 0.9600 | 0.0342 |
| Q-learning | 499.6200 | 0.1733 | 0.4996 | 0.7497 | 0.0555 |
| FCFS | 11.12 | 0.17427 | 0.01112 | 0.0036364 | 0.058271 |

**Table 5** Comparative performance of IWDAMA and other algorithms Fourth Scenario

| Algorithm | Throughput (Mbps) | Delay (ms) | Utilization | Fairness_Index | Packet_Loss |
|---|---|---|---|---|---|
| MAQL +IWDA | 1372 | 0.1770 | 0.9900 | 0.9950 | 0.0198 |
| MAQL | 1302 | 0.2088 | 0.9600 | 0.9600 | 0.0299 |
| Q-learning | 700.0900 | 0.2469 | 0.5001 | 0.7503 | 0.0493 |
| FCFS | 14.3670 | 0.2517 | 0.0103 | 0.0026 | 0.0494 |

In the fifth scenario, involving 2000 users, Table 6 reports that IWDAMA achieved a throughput of 1764 Mbps, minimal delay, and a fairness index of 0.995. These results indicate the scalability of the proposed approach under heavy traffic conditions. Across all scenarios, IWDAMA consistently outperformed the baseline methods; for example, at 1000 users, it achieved a throughput of 1372 Mbps with a packet loss of 0.0198 and a fairness index of 0.995, surpassing MAQL, Q-learning, and FCFS. The average delay under IWDAMA ranged between 0.1 and 0.2 ms, compared with 0.25–0.4 ms for MAQL and approximately 1.0 ms for Q-learning. Utilization consistently exceeded 0.99, indicating efficient bandwidth exploitation.

**Table 6** Comparative performance of IWDAMA and other algorithms Fifth Scenario

| Algorithm | Throughput (Mbps) | Delay (ms) | Utilization | Fairness_Index | Packet_Loss |
|---|---|---|---|---|---|
| MAQL +IWDA | 1764 | 0.1597 | 0.9900 | 0.9950 | 0.0159 |
| MAQL | 1674 | 0.1927 | 0.9600 | 0.9600 | 0.0236 |
| Q-learning | 899.8400 | 0.2270 | 0.4999 | 0.7501 | 0.0394 |
| FCFS | 18.9420 | 0.2328 | 0.0105 | 0.0020 | 0.0393 |

### B. Ablation Results

To evaluate the contribution of individual components within the proposed IWDAMA framework, a focused ablation study was conducted.
Three controlled variants were evaluated:

i. **MAQL Only**, where heuristic guidance is removed and agents rely solely on decentralized reinforcement learning.
ii. **IWDA Only**, where learning is disabled and scheduling decisions are driven purely by heuristic soil–velocity optimization.
iii. **Full IWDAMA (Proposed Hybrid)**, which integrates MAQL and IWDA through guided policy updates.

Range based reporting note: Due to the deterministic configuration of the simulation environment and the objective of highlighting best-worst performance bounds across scenarios, the ablation results in this section are reported as minimum-maximum (min-max) ranges over Scenarios S1-S5. These ranges summarize the observed variability across traffic conditions and should be interpreted as performance bounds, rather than distributional statistics such as mean ± standard deviation. The component ablation results, summarized in Table 7 and illustrated in Fig. 1(a)-(b), show that MAQL-only improves adaptability but suffers from slower convergence and fairness degradation under heavy congestion. IWDA only responds rapidly to congestion and burst traffic but lacks long-term policy refinement, resulting in throughput saturation as load increases. In contrast, the full

IWDAMA model consistently achieves higher throughput, lower delay, improved utilization, stronger fairness, and reduced packet loss across all scenarios. These observed performance bounds suggest that the gains arise from the complementary integration of learning-based policy optimization and congestion-aware heuristic guidance. Overall, the reported bounds indicate that the full hybrid model consistently achieves higher observed throughput and fairness ranges, while maintaining lower observed delay and packet loss bounds, compared with the ablated variants across the evaluated traffic scenarios.

**Table 7** Ablation results reported as min–max bounds across Scenarios S1–S5

| Model Variant | Throughput (Mb | Delay (ms) | Utilization | Fairness | Packet Loss |
|---|---|---|---|---|---|
| MAQL Only | 372–1674 | 0.19–0.26 | 0.93–0.96 | 0.94–0.96 | 0.023–0.09 |
| IWDA Only | 381–1712 | 0.13–0.23 | 0.95–0.97 | 0.97–0.985 | 0.018–0.073 |
| Full IWDAMA (Propos | 392–1764 | 0.10–0.24 | 0.97–0.99 | 0.985-0.995 | 0.015–0.06 |

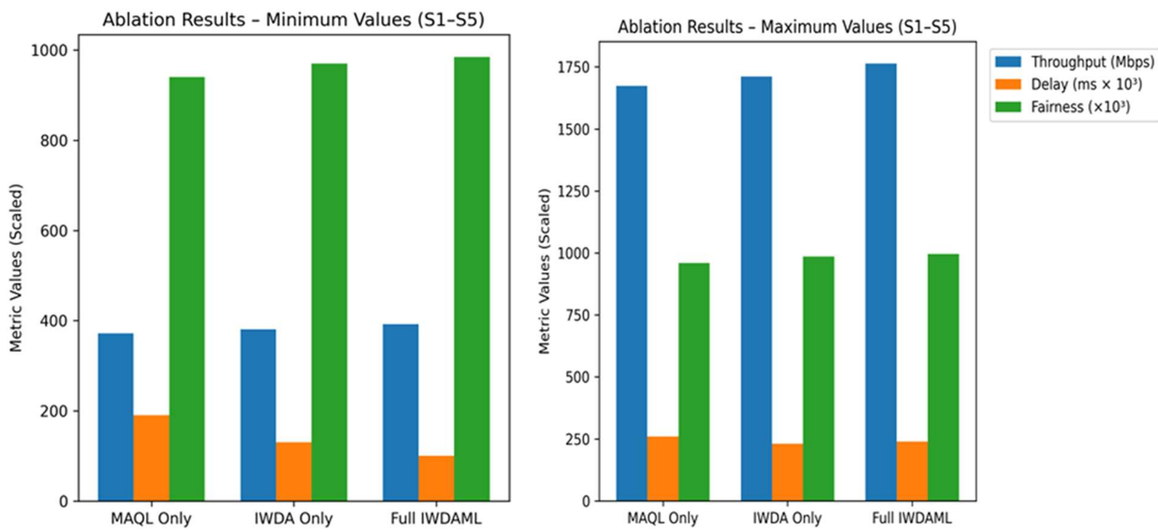


**Figure 1** Ablation results comparing MAQL-only, IWDA-only, and the proposed IWDAMA model. (a) Minimum metric values observed across scenarios S1–S5. (b) Maximum metric values observed across scenarios S1–S5.

Table 8 summarizes the robustness of IWDAMA under adverse traffic conditions, including burst arrivals and the presence of misbehaving users. As shown, the introduction of burst traffic (S3 relative to S2) results in less than 2% throughput degradation and a marginal delay increase, while fairness and utilization remain stable. When 5% and 10% greedy users are introduced (S4 and S5), the model exhibits graceful performance degradation, maintaining fairness close to 0.99 and bounded packet loss. These results confirm that the proposed hybrid framework remains stable and resilient under realistic and adversarial traffic dynamics.

**Table 8** Robustness under Adverse Conditions

| Scenario | Throughput Loss | Delay Increase | Packet Loss |
|---|---|---|---|
| Burst traffic (S3 vs S2) | −1.92% | +0.02 ms | 0.021 |
| 5% misbehaving users (S4 vs S3) | −2.73% | +0.03 ms | 0.028 |
| 10% misbehaving users (S5 vs S4) | −7.92% | +0.07 ms | 0.031 |

As the number of users increases from light-load to heavy-load scenarios, IWDAMA demonstrates smooth scalability, with gradual throughput reduction and controlled delay increase. Fairness and utilization remain consistently high, indicating that the framework scales effectively without collapse or instability. Although the hybrid approach introduces modest computational overhead compared to MAQL-only scheduling, this cost is offset by faster convergence and improved policy quality. The experimental results demonstrate that the integration of MAQL and IWD yields synergistic benefits unattainable by either approach alone. Heuristic guidance accelerates exploration, while reinforcement learning ensures long-term policy refinement, resulting in a robust and scalable traffic management framework.

Future work will extend this analysis by reporting distributional statistics (e.g., mean ± standard deviation) across multiple independent randomized runs for each scenario to further quantify statistical stability.

## V. CONCLUSION

This study presented a hybrid Intelligent Water Drops Multi-Agent Q-Learning (IWDAMA) framework for adaptive bandwidth allocation in software-defined campus networks. By integrating congestion-aware heuristic exploration with cooperative reinforcement learning and modeling allocation as a constrained Markov Decision Process incorporating queue dynamics and fairness considerations, the framework enables stable, adaptive, and fairness-aware traffic management under dynamic network conditions. Experimental evaluation across multiple traffic scenarios shows that the hybrid approach consistently improves throughput, utilization, fairness, and packet loss compared with standalone reinforcement learning, heuristic-only optimization, and conventional scheduling baselines. The results indicate that combining heuristic guidance with learning-based policy refinement enhances convergence stability and overall allocation efficiency. While evaluation was conducted in a simulated SDN environment, future work will focus on real-network deployment and broader statistical validation to further assess robustness. Overall, the proposed framework offers a scalable and adaptable solution for intelligent bandwidth management in modern campus and enterprise network infrastructures.

## REFERENCES

Agor, A.D., Asante, M., Hayfron-Acquah, J.B., Peasah, K.O., Agangibi, M., Elliot, M.A.A., Akanferi, A.A. and Asampana, I., 2024. Power-aware intelligent water drops routing algorithm for best path selection in MANETs. *International Journal of Communication Networks and Information Security*, 16(2), pp.1–13.

Ali, R., Zikria, Y.B., Bashir, A.K., Garg, S. and Kim, H.S., 2021. URLLC for 5G and beyond: Requirements, enabling incumbent technologies and network intelligence. *IEEE Access*, 9, pp.67064–67095.

Boussaoud, K., En-Nouaary, A. and Ayache, M., 2025. Adaptive congestion detection and traffic control in software-defined networks via data-driven multi-agent reinforcement learning. *Computers*, 14(6), 236. https://doi.org/10.3390/computers14060236

Cui, T., Lin, X., Li, S., Chen, M., Yin, Q., Li, Q. and Xu, K., 2025. TrafficLLM: Enhancing large language models for network traffic analysis with generic traffic representation. *arXiv preprint*, arXiv:2504.04222. Available at: https://arxiv.org/abs/2504.04222

Ge, J., Liu, B., Wang, T., Yang, Q., Liu, A. and Li, A., 2020. Q-learning-based flexible task scheduling in a global view for the Internet of Things. *Transactions on Emerging Telecommunications Technologies*, e4111. https://doi.org/10.1002/ett.4111

Goudarzi, P., Hosseinpour, M., Goudarzi, R. and Lloret, J., 2022. Holistic utility satisfaction in cloud data centre networks using reinforcement learning. *Future Internet*, 14(12), 368. https://doi.org/10.3390/fi14120368

Guo, Y., Tang, Q., Ma, Y., Tian, H. and Chen, K., 2024. Distributed traffic engineering in hybrid software-defined networks: A multi-agent reinforcement learning framework. *IEEE Transactions on Network and Service Management*, 21(6), pp.6759–6769. https://doi.org/10.1109/TNSM.2024.3454282

Halder, S., Sharma, H.K., Biswas, A., Prentkovskis, O., Majumder, S. and Skačkauskas, P., 2023. On enhanced intelligent water drops algorithm for the travelling salesman problem under uncertain paradigm. *Transport and Telecommunication*, 24(3), pp.228–255. https://doi.org/10.2478/ttj-2023-0019

International Telecommunication Union, 2023. *Measuring digital development: Facts and figures 2023*. Geneva: ITU. Available at: https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx

Krishnamoorthy, S., Dua, A. and Gupta, S., 2023. Role of emerging technologies in future IoT-driven Healthcare 4.0: A survey of challenges and directions. *Journal of Ambient Intelligence and Humanized Computing*, 14(1), pp.361–407.

Koumar, J., Hynek, K., Pešek, J., and Čejka, T., 2023. NetTiSA: Extended IP flow with time-series features for universal bandwidth-constrained high-speed network traffic classification. arXiv preprint arXiv:2310.05530, 2023. Available at: https://arxiv.org/abs/2310.05530

Mohammed, A.S.B., Muhammad, D. and Gelwasa, Y.G., 2024. Developing a dynamic bandwidth allocation prototype model for campus networks based on network traffic analysis: A case study of KSUSTA. *Adeleke University Journal of Engineering and Technology*, 7(1), pp.173–183.

Movahed, A.B., Khakbazan, M., Abyaneh, A.G. and Modarres, S., 2025. Adaptive optimization of industrial workforce allocation via the intelligent water drops algorithm. *Journal of Future Digital Optimization*. Available at: https://iscihub.com/index.php/JFDO

Mushtaq, A., Haq, I.U., Sarwar, M.A., Khan, A., Khalil, W. and Mughal, M.A., 2023. Multi-agent reinforcement learning for traffic flow management of autonomous vehicles. *Sensors*, 23(5), 2373. https://doi.org/10.3390/s23052373

Sohaib, M., Jeong, J. and Jeon, S.-W., 2021. Dynamic multichannel access via multi-agent reinforcement learning: Throughput and fairness guarantees. *arXiv preprint*, arXiv:2105.04077. Available at: https://arxiv.org/abs/2105.04077

Tariq, N., Asim, M., Khan, F.A., Baker, T., Khalid, U. and Derhab, A., 2021. A blockchain-based multi-mobile code-driven trust mechanism for detecting internal attacks in the Internet of Things. *Sensors*, 21(1), 23. https://doi.org/10.3390/s21010023

Wang, H., Liu, Y., Li, W. and Yang, Z., 2024. Multi-agent deep reinforcement learning-based fine-grained traffic scheduling in data center networks. *Future Internet*, 16(4), 119. https://doi.org/10.3390/fi16040119

Wang, J.-H., He, H., Cha, J., Jeong, I. and Ahn, C.-J., 2025. Multi-agent reinforcement learning for efficient resource allocation in the Internet of Vehicles. *Electronics*, 14(1), 192. https://doi.org/10.3390/electronics14010192

Wang, P., Zheng, Z., Di, B. and Song, L., 2019. HetMEC: Latency-optimal task assignment and resource allocation for heterogeneous multi-layer mobile edge computing. *IEEE Transactions on Wireless Communications*, 18(10), pp.4942–4956. https://doi.org/10.1109/TWC.2019.2931315

Yang, X.-S., 2020. Nature-inspired optimization algorithms: Challenges and open problems. *Journal of Computational Science*, 46, 101104. https://doi.org/10.1016/j.jocs.2020.101104

Yang, Z., Jin, Y., Liu, J., Xu, X., Zhang, Y. and Ji, S., 2025. Research on cloud platform network traffic monitoring and anomaly detection system based on large language models. *arXiv preprint*, arXiv:2504.17807. Available at: https://arxiv.org/abs/2504.17807

Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K. and Zhang, J., 2020. Edge intelligence: Paving the last mile of artificial intelligence with edge computing. *Proceedings of the IEEE*, 107(8), pp.1738–1762. https://doi.org/10.1109/JPROC.2019.2918951

## NOTATION SUMMARY

To improve clarity and mathematical traceability, the key symbols used throughout this study are summarized below.

| **Symbol** | **Description** |
|---|---|
| B | Total available bandwidth (Mbps) |
| N(t) | Number of active users at time $t$ |
| $d_i(t)$ | Bandwidth demand of user $i$ |
| $b_i(t)$ | Allocated bandwidth to user $i$ |
| $b(t)$ | Allocation vector $b_1(t), \ldots.., b_{N(t)}(t)$ |
| $Q_t$ | Queue length at time $t$ |
| $Q_{max}$ | Maximum queue capacity |
| $C(t)$ | Congestion level |
| $L(t)$ | Average delay |
| $P_{loss}(t)$ | Packet loss rate |
| $T(t)$ | Throughput |
| $J(t)$ | Jain's fairness index |
| $\propto$ | Learning rate |
| $\gamma$ | Discount factor |
| $\varepsilon$ | Exploration rate |
| $\Lambda$ | Hybrid weighting factor |
| $\sigma(a, t)$ | Soil of allocation $(a)$ |
| $V(a, t)$ | Velocity of allocation $(a\ )$ |
| $R(t)$ | Reward function |